Technical Report

UCAM-CL-TR
ISSN 1476-2986

UNIVERSITY OF
CAMBRIDGE

Computer Laboratory

List of technical reports

July 2024

15 JJ Thomson Avenue
Cambridge CB3 0FD
United Kingdom
phone +44 1223 763500

https://www.cl.cam.ac.uk/

Technical reports published by the University of Cambridge
Computer Laboratory are freely available via the Internet:

https://www.cl.cam.ac.uk/techreports/

ISSN 1476-2986

UCAM-CL-TR-1

M.F. Challis:

The JACKDAW database package

October 1974, 15 pages, PDF

Abstract: This report describes a general database package which has been implemented in BCPL on an IBM 370/165 at the University of Cambridge. One current application is the provision of an administrative database for the Computing Service.

Entries within a database may include (in addition to primitive fields such as 'salary' and 'address') links to other entries: each link represents a relationship between two entries and is always two-way.

Generality is achieved by including within each database class definitions which define the structure of the entries within it; these definitions may be interrogated by program.

The major part of the package presents a procedural interface between an application program and an existing database, enabling entries and their fields to be created, interrogated, updated and deleted. The creation of a new database (or modification of an existing one) by specifying the class definitions is handled by a separate program.

The first part of the report describes the database structure and this is followed by an illustration of the procedural interface. Finally, some of the implementation techniques used to insure integrity of the database are described.

UCAM-CL-TR-2

J. Larmouth:

Scheduling for a share of the machine

October 1974, 29 pages, PDF

Abstract: This paper describes the mechanism used to schedule jobs and control machine use on the IBM 370/165 at Cambridge University, England. The same algorithm is currently being used in part at the University of Bradford and implementations are in progress or under study for a number of other British Universities.

The system provides computer management with a simple tool for controlling machine use. The managerial decision allocates a share of the total machine resources to each user of the system, either directly, or via a hierarchial allocation scheme. The system then undertakes to vary the turnaround of user jobs to ensure that those decisions are effective, no matter what sort of work the user is doing.

At the user end of the system we have great flexibility in the way in which he uses the resources he has received, allowing him to get a rapid turnaround for those (large or small) jobs which require it, and a slower turnaround for other jobs. Provided he does not work at a rate exceeding that appropriate to his share of the machine, he can request, for every job he submits, the 'deadline' by which he wants it running, and the system will usually succeed in running his job at about the requested time – rarely later, and only occasionally sooner.

Every job in the machine has its own 'deadline', and the machine is not underloaded. Within limits, each user can request his jobs back when he wants them, and the system keeps his use to within the share of the machine he has been given. The approach is believed to be an original one and to have a number of advantages over more conventional scheduling and controlling algorithms.

UCAM-CL-TR-3

A.J.M. Stoneley:

A replacement for the OS/360 disc space management routines

April 1975, 7 pages, PDF

Abstract: In the interest of efficiency, the IBM disc space management routines (Dadsm) have been completely replaced in the Cambridge 370/165.

A large reduction in the disc traffic has been achieved by keeping the lists of free tracks in a more compact form and by keeping lists of free VTOC blocks. The real time taken in a typical transaction has been reduced by a factor of twenty.

By writing the code in a more appropriate form than the original, the size has been decreased by a factor of five, thus making it more reasonable to keep it permanently resident. The cpu requirement has decreased from 5% to 0.5% of the total time during normal service.

The new system is very much safer than the old in the fact of total system crashes. The old system gave little attention to the consequences of being stopped in mid-flight, and it was common to discover an area of disc allocated to two files. This no longer happens.

UCAM-CL-TR-4

A.J.M. Stoneley:

The dynamic creation of I/O paths under OS/360-MVT

April 1975, 16 pages, PDF

Abstract: In a large computer it is often desirable and convenient for an ordinary program to be able to establish for itself a logical connection to a peripheral device. This ability is normally provided through a routine within the operating system which may be called by any user program at any time. OS/360 lacks such a routine. For the batch job, peripheral connections can only

be made through the job control language and this cannot be done dynamically at run-time. In the restricted context of TSO (IBM's terminal system) a routine for establishing peripheral connections does exist, but it is extremely inefficient and difficult to use.

This paper describes how a suitable routine was written and grafted into the operating system of the Cambridge 370/165.

UCAM-CL-TR-5

P. Hazel, A.J.M. Stoneley:

Parrot – A replacement for TCAM

April 1976, 25 pages, PDF

Abstract: The terminal driving software and hardware for the Cambridge TSO (Phoenix) system is described. TCAM and the IBM communications controller were replaced by a locally written software system and a PDP-11 complex. This provided greater flexibility, reliability, efficiency and a better "end-user" interface than was possible under a standard IBM system.

UCAM-CL-TR-6

Andrew D. Birrell:

System programming in a high level language

December 1977, 125 pages, PDF
PhD thesis (Trinity College, December 1977)

Abstract: This thesis is concerned with the construction of a high level language system suitable for the implementation of a general purpose operating system for a computer. There are three aspects to this task: firstly a suitable high level language must be chosen or designed; secondly, a suitable implementation of this language must be manufactured; thirdly, the operating system itself must be written. These three aspects inevitably overlap in time – experience in implementing the language may cause one to review decisions taken in the design of the language, and experience in constructing the operating system will bring to light inadequacies, inconveniences and inelegancies in both the implementation and design of the language.

Most previous work in this field has been concerned with the first of these aspects, and has adopted the approach of designing special 'System Programming Languages' (SPLs) or 'Machine Oriented Languages' (MOLs). Various such languages have been developed, although few have achieved the elegance and generality of general-purpose languages such as Pascal or Algol68. Little or no investigation has previously been made into the second of these aspects, the implementation of the language. This aspect can have a considerable effect on the practicability of using the resulting language for manufacturing an operating system. The implementation, however suitable the language makes the difference between the language being an aid or an impediment to the system programmer. It is with aspects of the implementation this this thesis is mainly concerned.

UCAM-CL-TR-7

Andrew Hopper:

Local area computer communication networks

April 1978, 192 pages, PDF
PhD thesis (Trinity Hall, April 1978)

Abstract: In this thesis a number of local area network architectures are studied and the feasibility of a LSI design for a universal local network chip is considered. The thesis begins with a survey of current network technologies and a discussion of some of the problems encountered in local network design. Existing implementations of local networks are then discussed, and their performance compared. Ultimately the design considerations for a general purpose, microprogrammed, LSI network chip is discussed. Such a circuit is able to handle a range of network architectures and can be reconfigured to suit various traffic patterns. Finally some of the protocol requirements of local networks are discussed, leading to a redesign of the Cambridge ring to provide hardware support for protocol implementation.

UCAM-CL-TR-9

Douglas John Cook:

Evaluation of a protection system

181 pages, PDF
PhD thesis (Gonville & Caius College, April 1978)

Abstract: The CAP research project was set up in 1969 to investigate memory protection by designing and building a computer with hardware support for a very detailed protection system based on the use of capabilities. The computer has been built and an operating system written which exploits its protection facilities. It is time, therefore, to assess how successful the project has been. A necessary component of such an assessment is an evaluation of the CAP's protection system and this thesis presents the results of the author's research in this area.

Protection in computer systems is first introduced with a brief description of various models of protection systems and mechanisms for the provision of protection. There follows a description in some detail of the CAP computer and the CAP operating system with particular attention paid to those aspects of the design which are relevant to the research reported. A brief introduction to performance evaluation techniques is

given followed by a discussion of performance evaluation on the CAP computer.

The need for measuring the benefits and costs of protection is discussed and there is a detailed critical description of the previous research in this area. A simple model of a protection system is presented as is a protection measure based on this model. There is then a discussion of how the services provided by modules in the system fit into the model and the protection measure. The application of the protection measure to the CAP operating system is described. The results led to suggestions for the improvement of the protection aspects of the operating system and these are discussed in detail. The implications of the results for operating system design in general are also discussed.

The experiements to investiage the cost of using the protection provided on the CAP are described next. Some performance evaluation work was done in connection with the protection cost experiments and this too is discussed.

UCAM-CL-TR-10

Mark Theodore Pezarro:

Prediction oriented description of database systems

190 pages, PDF
PhD thesis (Darwin College, October 1978)

Abstract: A descriptive model of database systems is presented. The model is intended to provide a general framework for the description of database systems which is not limited to any particular DBMS or even any of the three mainstream approaches to DBMS architecture. This generality is derived from a new analysis of file organisation methods on which the model is based. The model concentrates on the aspects of a database system relevant to first-order performance prediction. These include database structure, the hardware and software used in implementing the system, the size of the database at various points in its lifetime, and its known or anticipated usage. Particular attention has been devoted to arriving at a general treatment of the details of database systems at the physical level, including access paths and their encoding, storage devices and their operating characteristics, and the mapping of data representations to storage devices.

A formal language has been devised in which to write textual descriptions of a database system in terms of the model. In addition an experimental prediction program has been written which accepts a description of a database system expressed in the language and produces performance estimates for the described activity using computational methods based on expected value formulae. Some preliminary results obtained by comparing estimates given by the program with measurements of an operational database system are presented. Further experimentation that would allow a definitive evaluation of the prediction program is outlined and a review is made of the current limitations of the model and program with suggestions for further research.

UCAM-CL-TR-11

Branimir Konstatinov Boguraev:

Automatic resolution of linguistic ambiguities

222 pages, PDF
PhD thesis (Trinity College, August 1979)

Abstract: The thesis describes the design, implementation and testing of a natural language analysis system capable of performing the task of generating paraphrases in a highly ambiguous environment. The emphasis is on incorporating strong semantic judgement in an augmented transition network grammar: the system provides a framework for examining the relationship between syntax and semantics in the process of text analysis, especially while treating the related phenomena of lexical and structural ambiguity. Word-sense selection is based on global analysis of context within a semantically well-formed unit, with primary emphasis on the verb choice. In building structures representing text meaning, the analyser relies not on screening through many alternative structures – intermediate, syntactic or partial semantic – but on dynamically constructing only the valid ones. The two tasks of sense selection and structure building are procedurally linked by the application of semantic routines derived from Y. Wilks' preference semantics, which are invoked at certain well chosen points of the syntactic constituent analysis – this delimits the scope of their action and provides context for a particular disambiguation technique. The hierarchical process of sentence analysis is reflected in the hierarchical organisation of application of these semantic routines – this allows the efficient coordination of various disambiguation techniques, and the reduction of syntactic backtracking, non-determinism in the grammar, and semantic parallelism. The final result of the analysis process is a dependency structure providing a meaning representation of the input text with labelled components centred on the main verb element, each characterised in terms of semantic primitives and expressing both the meaning of a constituent and its function in the overall textual unit. The representation serves as an input to the generator, organised around the same underlying principle as the analyser – the verb is central to the clause. Currently the generator works in paraphrase mode, but is specifically designed so that with minimum effort and virtually no change in the program control structure and code it could be switched over to perform translation.

The thesis discusses the rationale for the approach adopted, comparing it with others, describes the system and its machine implementation, and presents experimental results.

UCAM-CL-TR-12

M.R.A. Oakley, P. Hazel:

HASP "IBM 1130" multileaving remote job entry protocol with extensions as used on the University of Cambridge IBM 370/165

September 1979, 28 pages, PDF

Abstract: This document brings together most of the information required to design, write and operate a HASP Remote Job Entry Terminal program. Most of the document describes facilities available using any host computer supporting the HASP protocols. The remainder of the document describes improvements to these facilities which have been made in order to enhance the reliability of the system, to make it easier to run, and to provide for a wider range of peripherals than the basic system.

UCAM-CL-TR-13

Philip Hazel:

Resource allocation and job scheduling

1980, 41 pages, PDF

Abstract: The mechanisms for sharing the resources of the Cambridge IBM 370/165 computer system among many individual users are described. File store is treated separately from other resources such as central processor and channel time. In both cases, flexible systems that provide incentives to thrifty behaviour are used. The method of allocating resources directly to users rather than in a hierarchical manner via faculties and departments is described, and its social acceptability is discussed.

UCAM-CL-TR-14

J.S. Powers:

Store to store swapping for TSO under OS/MVT

June 1980, 28 pages, PDF

Abstract: A system of store-to-store swapping incorporated into TSO on the Cambridge IBM 370/165 is described. Unoccupied store in the dynamic area is used as the first stage of a two-stage backing store for swapping time-sharing sessions; a fixed-head disc provides the second stage. The performance and costs of the system are evaluated.

UCAM-CL-TR-15

I.D. Wilson:

The implementation of BCPL on a Z80 based microcomputer

68 pages, PDF
BA dissertation (Downing College, May 1980)

Abstract: The main aim of this project was to achieve as full an implementation as possible of BCPL on a floppy disc based microcomputer, running CP/M or CDOS (the two being esentially compatible). On the face of it there seemed so many limiting factors, that, when the project was started, it was not at all clear which one (if any) would become a final stumbling block. As it happened, the major problems that cropped up could be programmed round, or altered in such a way as to make them soluble.

The main body of the work splits comfortably into three sections, and the writer hopes that, in covering each section separately, to be able to show how the whole project fits together into the finished implementation.

UCAM-CL-TR-16

Jeremy Dion:

Reliable storage in a local network

February 1981, 142 pages, PDF
PhD thesis (Darwin College, February 1981)

Abstract: A recent development in computer science has been the advent of local computer networks, collections of autonomous computers in a small geographical area connected by a high-speed communications medium. In such a situation it is natural to specialise some of the computers to provide useful services to others in the network. These server machines can be economically advantageous if they provide shared access to expensive mechanical devices such as discs.

This thesis discusses the problems involved in designing a file server to provide a storage service in a local network. It is based on experience gained from the design and implementation of a file server for the Cambridge ring.

An important aspect of the design of a file server is the choice of the service which is provided to client machines. The spectrum of choice ranges from providing a simple remote disc with operations such as read and write block, to a remote file system with directories and textual names. The interface chosen for the Cambridge file server is "universal" in that the services it provides are intended to allow easy implementation of both virtual memory systems and filing systems.

The second major aspect of the file server design concerns reliability. If the server is to store important

information for clients, then it is essential that it be resistant to transient errors such as communications or power failures. The general problems of reliability and crash resistance are discussed in terms of a model developed for this purpose. Different reliability strategies used in current data base and filing systems are related to the model, and a mechanism for providing atomic transactions in the Cambridge file server is described in detail. An improved mechanism which allows atomic transactions on multiple files is also described and contrasted with the first version. The revised design allows several file servers in a local network to cooperate in atomic updates to arbitrary collections of files.

UCAM-CL-TR-17

B.K. Boguraev, K. Spärck Jones, J.I. Tait:

Three papers on parsing

1982, 22 pages, PDF

Abstract: This collection of three papers examines current problems in the parsing of natural language. The first paper investigates the parsing of compound nouns, and suggests that the existing strategies are inadequate. Accepting that better approaches are needed, the paper then proceeds to examine the implications for natural language processing systems.

The second paper in the collection examines the task of recognising conjunctions within an ATN grammar. To do this only through the grammar specification is difficult and results in a bulky grammar. The paper therefore presents some ideas for extending the ATN mechanism to better deal with conjunctions.

The final paper considers ways in which semantic parsers can exploit syntactic constraints. Two specific semantic parsers are considered: those of Cater and Boguraev which are regarded as being representative of two styles of parsing. The main conclusion to be drawn is that there are significant disadvantages to semantic parsing without complete syntactic processing of the input.

UCAM-CL-TR-18

Burkard Wördenweber:

Automatic mesh generation of 2 & 3 dimensional curvilinear manifolds

November 1981, 128 pages, paper copy
PhD thesis (St John's College, November 1981)

UCAM-CL-TR-19

Arthur William Sebright Cater:

Analysis and inference for English

September 1981, 223 pages, PDF
PhD thesis (Queens' College, September 1981)

Abstract: AD-HAC is a computer program which understands stories. Its three principal components each deal with significant subareas of the overall language-processing task: it has a sentence analyser, which creates conceptual representations of the meanings of individual sentences; an inferencer, which assimilates these into the existing representation of a story, determining pronoun referents and answering questions as a byproduct of this activity; and a sentence generator, which produces english sentences conveying the meaning of conceptual representations. The research reported here has focussed on the analyser and the inferencer.

The analyser uses an ATN to identify low-level syntactic constituents, such as verb groups or prepositional phrases: 'requests' associated with words, particularly verbs, are then applied in a nondeterministic preference-directed framework, using the constituents as building blocks in the analysis of phrases, clauses and sentences: the requests fall into five distinct processing classes. The partial analyses which result from the application or non-application of particular requests are ordered by preference, and the most-preferred partial analysis is persued first, giving a predominantly left-to-right scan through the sentence. A surprising result is that the analyser performs better if it is permitted to keep only a small number of partial analyses.

The inferencer exploits the primitives of the conceptual representation language, using these as the main indicator of the appropriate set of inferences. The inferences are specified by means of inference networks associated with the conceptual primitives. Tests are applied to elementary propositions derived from input sentence analyses, and select paths through the networks where appropriate inferences are made. Inference networks are also associated with 'functions' of objects, permitting higher-level than can normally be made using the primitives alone: the resulting system offers a synthesis of low-level inference and script-like inference. The inferences made by the networks are also used to determine the referents of pronouns, and to provide the answers to questions: the program takes an identical approach to these two tasks.

The performance of the system is illustrated by reference to texts which have been successfully processed by AD-HAC.

UCAM-CL-TR-20

Avra Cohn, Robin Milner:

On using Edinburgh LCF to prove the correctness of a parsing algorithm

February 1982, 23 pages, PDF

Abstract: The methodology of Edinburgh LCF, a mechanized interactive proof system is illustrated through a problem suggested by Gloess – the proof of a simple parsing algorithm. The paper is self-contained, giving only the relevant details of the LCF proof system. It is shown how tactics may be composed in LCF to yield a strategy which is appropriate for the parser problem but which is also of a generally useful form. Also illustrated is a general mechanized method of deriving structural induction rules within the system.

UCAM-CL-TR-21

A. Cohn:

The correctness of a precedence parsing algorithm in LCF

April 1982, 38 pages, PDF

Abstract: This paper describes the proof in the LCF system of a correctness property of a precedence parsing algorithm. The work is an extension of a simpler parser and proof by Cohn and Milner (Cohn & Milner 1982). Relevant aspects of the LCF system are presented as needed. In this paper, we emphasize (i) that although the current proof is much more complex than the earlier one, mqany of the same metalanguage strategies and aids developed for the first proof are used in this proof, and (ii) that (in both cases) a general strategy for doing some limited forward search is incorporated neatly into the overall goal-oriented proof framework.

UCAM-CL-TR-22

M. Robson:

Constraints in CODD

18 pages, PDF

Abstract: The paper describes the implementation of the data structuring concepts of domains, intra-tuple constraints and referential constraints in the relational DBMS CODD. All of these constraints capture some of the semantics of the database's application.

Each class of constraint is described briefly and it is shown how each of them is specified. The constraints are stored in the database giving a centralised data model, which contains descriptions of procedures as well as of statistic structures. Some extensions to the notion of referential constraint are proposed and it is shown how generalisation hierarchies can be expressed as sets of referential constraints. It is shown how the stored data model is used in enforcement of the constraints.

UCAM-CL-TR-23

J.I. Tait:

Two papers about the scrabble summarising system

12 pages, PDF

Abstract: This report contains two papers which describe parts of the Scrabble English summarizing system. The first, "Topic identification techniques for predictive language analyzers" has been accepted as a short communication for the 9th International COnference on Computational Linguistics, in Prague. The second, "General summaries using a predictive language analyser" is an extended version of a discussion paper which will be presented at the European Conference on Artificial Intelligence in Paris. Both conferences will take place during July 1982.

The [second] paper describes a computer system capable of producing coherent summaries of English texts even when they contain sections which the system has not understood completely. The system employs an analysis phase which is not dissimilar to a script applier together with a rather more sophisticated summariser than previous systems. Some deficiencies of earlier systems are pointed out, and ways in which the current implementation overcomes them are discussed.

UCAM-CL-TR-24

B.K. Boguraev, K. Spärck Jones:

Steps towards natural language to data language translation using general semantic information

March 1982, 8 pages, PDF

Abstract: The aim of the work reported here is to maximise the use of general semantic information in an AI task processor, specifically in a system front end for converting natural language questions into formal database queries. The paper describes the translation component of such a front end, which is designed to work from the question meaning representation produced by a language analyser exploiting only general semantics and syntax, to a formal query relying on database-specific semantics and syntax. Translation is effected in three steps, and the paper suggests that the rich and explicit meaning representations using semantic primitives produced for input sentences by the analyser constitute a natural and effective base for further processing.

UCAM-CL-TR-25

Hiyan Alshawi:

A clustering technique for semantic network processing

May 1982, 9 pages, PDF

Abstract: This paper describes techniques for performing serial processing on the type of semantic network exemplified by NETL. They make use of an indexing scheme that can be based on semantic clustering. The basic algorithm is aimed at performing fast intersection operations. It is claimed that the scheme is suitable for its current application in text processing. The semantic criteria for clustering that have been tried are briefly described. Extensions of the scheme are suggested for use with large networks.

UCAM-CL-TR-26

Brian James Knight:

Portable system software for personal computers on a network

204 pages, PDF
PhD thesis (Churchill College, April 1982)

Abstract: This dissertation is concerned with the design of the portable operating system TRIPOS, and its use as the basis for an operating system to run in 'single connection' computers – that is, computers whose only peripheral is an interface to a local area network.

TRIPOS is a lightweight, yet powerful, multitasking operating system aimed at personal minicomputers. It is designed to be relatively straightforward to transport to new hardware, providing an almost identical user interface and program environment on each machine. Particular emphasis has been placed on avoiding unnecessary complexity, in order to make it simple to understand, explain, and adapt for special purposes. The majority of the system and utilities are written in the language BCPL, and can be moved without change to different computers. They run on a kernel and device drivers written in assembly language for each particular machine. The user's view of the system is presented first, with samples of console dialogue, and then its internal structure is described.

The main part of the work described concerns the building of a portable operating system presenting user and program interfaces as similar as possible to ordinary TRIPOS, but running in processors connected only to a local area network – the Cambridge Ring. The system makes use of 'server' computers on the ring in order to gain access to disc storage, terminals, and printers. Several methods are investigated for using the primitives provided by a universal file-server to construct a filing system which can be shared by machines of different types. Some conclusions are drawn on the effects of distributing operating system functions in this way.

UCAM-CL-TR-27

Martyn Alan Johnson:

Exception handling in domain based systems

September 1981, 129 pages, PDF
PhD thesis (Churchill College, September 1981)

Abstract: Modern operating systems allow the creation of protection domains; these enable subsystems to cooperate whilst being protected from each other. This creates a number of problems in the handling of exceptions such as the expiry of time limits or the receipt of console 'quit' signals. Particular problems arise when parts of the operating system are implemented as protection domains which cannot easily be distinguished from user programs by the underlying protection system.

The dissertation surveys some traditional methods of dealing with such problems, and explains why they are inadequate in a domain based system. In addition, work done on related topics in the operating system for the Cambridge CAP computer is described.

The major part of the research described is concerned with a class of exception not usually recognized by operating system designers. This arises from the observation that protection domains which implement subsystems can retain useful state information between invocations, and care needs to be taken to ensure that domains are given an opportunity to keep their private data structures in a consistent state. In particular, domains which fall into disuse need to be notified of the fact so that they can tidy up the data structures they manage before they are destroyed. An intuitively simple solution to the problem is discussed, and its limitations and implementation difficulties are noted. Refinements of the mechanism are proposed which provide an improved treatment of the problem; and it is suggested that the moderate run time overhead which these revisions impose can be minimized by providing hardware or microprogram support for the mechanism.

UCAM-CL-TR-28

D.C.J. Matthews:

Poly report

August 1982, 17 pages, PDF

Abstract: Poly was designed to provide a programming system with the same flexibility as a dynamically typed language but without the run-time oveheads. The type system, based on that of Russel allows polymorpphic operations to be used to manipulate abstract objects,

but with all the type checking being done at compile-time. Types may be passed explicitly or by inference as parameters to procedures, and may be returned from procedures. Overloading of names and generic types can be simulated by using the general procedure mechanism. Despite the generality of the language, or perhaps because of it, the type system is very simple, consisting of only three classes of object. There is an exception mechanism, similar to that of CLU, and the exceptions raised in a procedure are considered as part of its 'type'. The construction of abstract objects and hiding of internal details of the representation come naturally out of the type system.

UCAM-CL-TR-29

D.C.J. Matthews:

Introduction to Poly

May 1982, 24 pages, PDF

Abstract: This report is a tutorial introduction to the programming language Poly. It describes how to write and run programs in Poly using the VAX/UNIX implementation. Examples given include polymorphic list functions, a double precision integer package and a subrange type constructor.

UCAM-CL-TR-30

John Wilkes:

A portable BCPL library

October 1982, 31 pages, PDF

Abstract: Too often, programs written in BCPL are difficult to port from one system to another, not because of the language, but because of differences between 'standard' libraries. Almost without exception, the definitions of these libraries are loose, woolly and inaccurate – the proposed BCPL standards document being a prime example. The author has developed and implemented a new BCPL library which is explicitly designed to aid the portability of programs between systems. In addition to being largely portable itself, it has two other features of interest: it uses an exception handling system instead of return codes, and it makes no distinction between system and user defined stream handlers. This paper defines the interface to the package.

UCAM-CL-TR-31

J. Fairbairn:

Ponder and its type system

November 1982, 42 pages, PDF

Abstract: This note describes the programming language "Ponder", which is designed according to the principles of referencial transparency and "orthogonality" as in [vWijngaarden 75]. Ponder is designed to be simple, being functional with normal order semantics. It is intended for writing large programmes, and to be easily tailored to a particular application. It has a simple but powerful polymorphic type system.

The main objective of this note is to describe the type system of Ponder. As with the whole of the language design, the smallest possible number of primitives is built in to the type system. Hence for example, unions and pairs are not built in, but can be constructed from other primitives.

UCAM-CL-TR-32

B.K. Boguraev, K. Spärck Jones:

How to drive a database front end using general semantic information

November 1982, 20 pages, PDF

Abstract: This paper describes a front end for natural language access to databases making extensive use of general, i.e. domain-independent, semantic information for question interpretation. In the interests of portability, initial syntactic and semantic processing of a question is carried out without any reference to the database domain, and domain-dependent operations are confined to subsequent, comparatively straightforward, processing of the initial interpretation. The different modules of the front end are described, and the system's performance is illustrated by examples.

UCAM-CL-TR-33

John A. Carroll:

An island parsing interpreter for Augmented Transition Networks

October 1982, 50 pages, PDF

Abstract: This paper describes the implementation of an 'island parsing' interpreter for an Augmented Transition Network (ATN). The interpreter provides more complete coverage of Woods' original ATM formalism than his later island parsing implementation; it is written in LISP and has been modestly tested.

UCAM-CL-TR-34

Larry Paulson:

Recent developments in LCF: examples of structural induction

January 1983, 15 pages, PDF

Abstract: Manna and Waldinger have outlined a large proof that probably exceeds the power of current theorem-provers. The proof establishes the unification algorithm for terms composed of variables, constants, and other terms. Two theorems from this proof, involving structural induction, are performed in the LCF proof assistant. These theorems concern a function that searches for an occurrence of one term inside another, and a function that lists the variables in a term.

Formally, terms are regarded as abstract syntax trees. LCF automatically builds the first-order theory, with equality, of this recursive data structure.

The first theorem has a simple proof, induction followed by rewriting. The second theorem requires a cases split and substitution throughout the goal. Each theorem is proved by reducing the initial goal to simpler and simpler subgoals. LCF provides many standard proof strategies for attacking goals; the user can program additional ones in LCF's meta-language, ML. This felxibility allows users to take ideas from such diverse fields as denotational semantics and logic programming.

UCAM-CL-TR-35

Larry Paulson:

Rewriting in Cambridge LCF

February 1983, 32 pages, DVI

Abstract: Many automatic theorem-provers rely on rewriting. Using theorems as rewrite rules helps to simplify the subgoals that arise during a proof.

LCF is an interactive theorem-prover intended for reasoning about computation. Its implementation of rewriting is presented in detail. LCF provides a family of rewriting functions, and operators to combine them. A succession of functions is described, from pattern matching primitives to the rewriting tool that performs most inferences in LCF proofs.

The design is highly modular. Each function performs a basic, specific task, such as recognizing a certain form of tautology. Each operator implements one method of building a rewriting function from simpler ones. These pieces can be put together in numerous ways, yielding a variety of rewriting strategies.

The approach involves programming with higher-order functions. Rewriting functions are data values, produced by computation on other rewriting functions. The code is in daily use at Cambridge, demonstrating the practical use of functional programming.

UCAM-CL-TR-36

Lawrence Paulson:

The revised logic PPLAMBDA A reference manual

March 1983, 28 pages, PDF

Abstract: PPLAMBDA is the logic used in the Cambridge LCF proof assistant. It allows Natural Deduction proofs about computation, in Scott's theory of partial orderings. The logic's syntax, axioms, primitive inference rules, derived inference rules and standard lemmas are described as are the LCF functions for building and taking apart PPLAMBDA formulas.

PPLAMBDA's rule of fixed-point induction admits a wide class of inductions, particularly where flat or finite types are involved. The user can express and prove these type properties in PPLAMBDA. The induction rule accepts a list of theorems, stating type properties to consider when deciding to admit an induction.

UCAM-CL-TR-37

Christopher Gray Girling:

Representation and authentication on computer networks

154 pages, PDF
PhD thesis (Queens' College, April 1983)

Abstract: Controlling access to objects in a conventional operating system is a well understood problem for which solutions are currently in existence. Such solutions utilize mechanisms which accurately and trivially provide the identity of an accessing subject. In the context of a collection of computers communicating with each other over a network, provision of this mechanism is more complex. The design of such a mechanism and its implementation on the Cambridge Ring at Cambridge University is described.

The vehicle used to prove the identity of an object irrefutably is called a representation and the deduction of an object's identity is called authentication. Methods of authentication are given which show that the mechanism can cope with identification needs that arise in practice (even in a network where the function asigned to each computer is constantly changing). These generate representations for such important components of a computer network as people, services and addresses. The implementation of a representation system utilizing some of these methods is described, including the incorporation of its use into a real operating system. The place of representations within the communication protocols that must transport them is considered and some enhancements are proposed. in addition, some interesting variations and extensions of the system are explored.

UCAM-CL-TR-38

Mike Gray:

Views and imprecise information in databases

November 1982, 119 pages, PDF
PhD thesis (November 1982)

Abstract: Providing user views of a database is an important way of achieving data independence and ease of use of DBMSs. This dissertation discusses one aspect of the problem of supporting views. It is shown that a crucial factor in the support of views is the richness of the data model used, and in particular its ability to represent certain kinds of incomplete information. This dissertation discusses various ways of handling incomplete information, and the operations on views that can be supported. The implementation of an experimental system which supports views on a relational database is described.

The first chapter describes the problem of treating views as first-class objects, that is allowing all the usual database operations to be performed on data in views. It is shown how this is related to the problem of representing incomplete information in the conceptual schema. The second chapter proposes the use of lattices to represent incomplete information, and shows how this covers various particular kinds of imprecise information. The third chapter reviews other work relating to imprecise information in databases. The fourth chapter discusses certain further implications of representing imprecise information, and makes proposals regarding the interpretation of keys, constraints, and the open-world assumption in this environment. The fifth chapter discusses in detail the relational operations that are appropriate with imprecise data and proposes modified Join and Group-by operations. The implementation of a system with these features is discussed. Chapter six illustrates some of the points made by considering an example database, and finally chapter seven concludes this dissertation with a summary and examination of further possibilities.

UCAM-CL-TR-39

Lawrence Paulson:

Tactics and tacticals in Cambridge LCF

July 1983, 26 pages, PDF

Abstract: The tactics and tacticals of Cambridge LCF are described. Tactics reason about logical connectives, substitution and rewriting; tacticals combine tactics into more powerful tactics. LCF's package for managing an interactive proof is discussed. This manages the subgoal tree, presenting the user with unsolved goals and assembling the final proof.

While primarily a reference manual, the paper contains a brief introduction to goal-directed proof. An example shows typical use of the tactics and subgoal package.

UCAM-CL-TR-40

W. Stoye:

The SKIM microprogrammer's guide

October 1983, 33 pages, PDF

Abstract: This paper describes the design and implementation of the SKIM micorprocessor. The processor has a 24 bit ALU with 16 general purpose registers. The main unique feature is a large microcode store of up to 64K 40 bit words, with the intention that the microcode could be used like the machine code on a conventional processor, with operating system primitives being programmed in microcode.

The processor has been constructed from TTL logic, with a microcode assembler running on Phoenix. A debugger for both the hardware and microcode programs runs on the host machine, currently a BBC Microcomputer.

The processor architecture is discussed, with examples of microcode programming. comparisons with other processors are made, and some of the limitations of the present design are noted.

UCAM-CL-TR-41

Mike Gordon:

LCF_LSM, A system for specifying and verifying hardware

September 1983, 53 pages, PDF

Abstract: The LCF_LSM system is designed to show that it is practical to prove the correctness of real hardware. The system consists of a programming environment (LCF) and a specification language (LSM). The environment contains tools for manipulating and reasoning about the specifications. Verification consists in proving that a lov-level (usually structural) description is behaviourally equivalent to a high-level functional description. Specifications can be fully hierarchical, and at any level devices can be specified either functionally or structurally.

As a first case study a simple microcoded computer has been verified. This proof is described in a companion report. In this we also illustrate the use of the system for other kinds of manipulation besides verification. For example, we show how to derive an implementation of a hard-wired controller from a microprogram and its decoding and sequencing logic. The derivation is done using machine checked inference; this ensures that the hard-wired controller is equivalent to the microcoded one. We also show how to code a microassembler. These examples illustrate our belief that LCF is a good environment for implementing a wide range of tools for manipulating hardware specifications.

This report has two aims: first to give an overview of the ideas embodied in LCF_LSM, and second, to be a user manual for the system. No prior knowledge of LCF is assumed.

UCAM-CL-TR-42

Mike Gordon:

Proving a computer correct with the LCF_LSM hardware verification system

September 1983, 49 pages, PDF

Abstract: A machine generated correctness proof of a simple computer is described.

At the machine code level the computer has a memory and two registers: a 13 bit program counter and a 16-bit accumulator. There are 8 machine instructions: halt, unconditional jump, jump when the accumulator contains 0, add contents of a memory location to accumulator, subtract contents of a location from accumulator, load accumulator from memory, store contents of accumulator in memory, and skip. The machine can be interrupted by pushing a button on its front panel.

The implementation which we prove correct has 6 data registers, and ALU, a memory, and a microcode controller. The controller consists of a ROM holding 26 30-bit microinstructions, a microprogram counter, and some combinatorial microinstruction decode logic.

Formal specifications of the target and host machines are given, and we describe the main steps in proving that the host correctly fetches, decodes and executes machine instructions.

The utility of LCF_LSM for general manipulaton is illustrated in two appendices. In appendix 1 we show how to code a microassembler. In appendix 2 we use the LCF_LSM inference rules to design a hard-wired controller equivalent to the original microcoded one.

N.B. This report should be read in conjunction with LCF_LSM: A system for specifying and verifying hardware. University of Cambridge, Computer Laboratory technical report number 41.

UCAM-CL-TR-43

Ian Malcom Leslie:

Extending the local area network

February 1983, 71 pages, PDF
PhD thesis (Darwin College, February 1983)

Abstract: This dissertation is concerned with the development of a large computer network which has many properties associated with local area computer networks, including high bandwidth and lower error rates. The network is made up of component local area networks, specifically Cambridge rings, which are connected either through local ring-ring bridges or through a high capacity satellite link. In order to take advantage of the characteristics of the resulting network, the protocols used are the same simple protocols as those used on a single Cambridge ring. This in turn allows many applications, which might have been thought of as local area network applications, to run on the larger network.

Much of this work is concerned with an interconnection strategy which allows hosts of different component networks to communicate in a flexible manner without building an extra internetwork layer into protocol hierarchy. The strategy arrived at is neither a datagram approach nor a system of concatenated error and flow controlled virtual circuits. Rather, it is a lightweight virtual circuit approach which preserves the order of blocks sent on a circuit, but which makes no other guarantees about the delivery of these blocks. An extra internetwork protocol layer is avoided by modifying the system used on a single Cambridge ring which binds service names to addresses so that it now binds service names to routes across the network.

UCAM-CL-TR-44

Lawrence Paulson:

Structural induction in LCF

November 1983, 35 pages, PDF

Abstract: The fixed-point theory of computation can express a variety of recursive data types, including lazy types, conventional first-order (strict) types, mutually recursive types, and types with equational constraints. Lazy types contain infinite objects, regarded as the limit of a chain of finite objects. Structural induction for all these types follows from fixed-point induction, though induction for lazy types is only sound for a certain class of formulas.

The paper presents the derivation of structural induction for each type, and justifies the necessary axioms by furnishing models for them. It presents example type definitions of lazy lists, strict lists, syntax trees for expressions and finite sets. Strict data types are proved to be flat in their partial ordering. Primitive recursion operators are introduced for each type, providing theoretical insights as well as a concise notation for defining total functions.

The research was done using LCF, an interactive theorem-prover for the fixed-point theory. The paper documents the theory of LCF data types, and surveys several LCF proofs involving structural induction. In order to be self-contained, it makes little reference to LCF details and includes a summary of the fixed point theory.

UCAM-CL-TR-45

Karen Spärck Jones:

Compound noun interpretation problems

July 1983, 16 pages, PDF

Abstract: This paper discusses the problems of compound noun interpretation in the context of automatic language processing. Given that compound processing implies identifying the senses of the words involved, determining their bracketing, and establishing their underlying semantic relations, the paper illustrates the need, even in comparatively favourable cases, for inference using pragmatic information. This has consequences for language processor architectures and, even more, for speech processors.

UCAM-CL-TR-46

Nicholas Henry Garnett:

Intelligent network interfaces

May 1985, 140 pages, PDF
PhD thesis (Trinity College, May 1983)

Abstract: Local Area Networks are now an accepted part of computing research. The technology of the network itself and the hardware to interface it to a computer is standard and in the cases of networks like Ethernet and the Cambridge Ring is commercially available. The next level up from the hardware is the software interface between the host computer and the network. This dissertation is concerned with one specific type of interface where the host is not itself directly connected to the network, but must access it via a second Network Interface Processor (NIP).

The dissertation begins by describing the design and implementation of the two low level interfaces for the Cambridge Ring. The first of these, the type 2, is machine independent and although based on a simple processor offers some sophisticated facilities to its host. The second, Spectrum, is not so sophisticated, but is customized to interface to just one operating system. The difference between these two approaches is discussed.

We go on to introduce the High Level Interface, which removes all protocol and network related processing from the host machine. This can benefit both the protocol implementation, by reducing system overheads, and the host operating system, by freeing CPU time for other tasks. This is particularly true in the case of time-shared machines which rely on the network for terminal connections. The design and implementation of such an interface are described.

The dissertation concludes by considering the possible roles of the NIP in the areas of security, protection and reliability. Some thoughts are also given on the design of protocols which exploit the features of a NIP.

UCAM-CL-TR-47

John Irving Tait:

Automatic summarising of English texts

137 pages, PDF
PhD thesis (Wolfson College, December 1982)

Abstract: This thesis describes a computer program called Scrabble which can summarise short English texts. It uses large bodies of predictions about the likely contents of texts about particular topics to identify the commonplace material in an input text. Pre-specified summary templates, each associated with a different topic are used to condense the commonplace material in the input. Filled-in summary templates are then used to form a framework into which unexpected material in the input may be fitted, allowing unexpected material to appear in output summary texts in an essentially unreduced form. The system's summaries are in English.

The program is based on technology not dissimilar to a script applier. However, Scrabble represents a significant advance over previous script-based summarising systems. It is much less likely to produce misleading summaries of an input text than some previous systems and can operate with less information about the subject domain of the input than others.

These improvements are achieved by the use of three main novel ideas. First, the system incorporates a new method for identifying the idea or topics of an input text. Second, it allows a section of text to have more than one topic at a time, or at least a composite topic which may be dealt with by the computer program simultaneously applying the text predictions associated with more than one simple topic. Third, Scrabble incorporates new mechanisms for the incorporation of unexpected material in the input into its output summary texts. The incorporation of such material in the output summary is motivated by the view that it is precisely unexpected material which is likely to form the most salient matter in the input text.

The performance of the system is illustrated by means of a number of example input texts and their Scrabble summaries.

UCAM-CL-TR-48

Hiyan Alshawi:

A mechanism for the accumulation and application of context in text processing

November 1983, 17 pages, PDF

Abstract: The paper describes a mechanism for the representation and application of context information for automatic natural language processing systems. Context information is gathered gradually during the reading of the text, and the mechanism gives a way of combining the effect of several different types of context factors. Context factors can be managed independently, while still allowing efficient access to entities in focus. The mechanism is claimed to be more general than the global focus mechanism used by Grosz for discourse understanding. Context affects the interpretation process by choosing the results, and restricting the processing, of a number of important language interpretation operations, including lexical disambiguation and reference resolution. The types of context factors that have been implemented in an experimental system are described, and examples of the application of context are given.

UCAM-CL-TR-49

David Charles James Matthews:

Programming language design with polymorphism

143 pages, PDF
PhD thesis (Wolfson College, 1983)

Abstract: This dissertation describes the design and implementation of a programming language, Poly. By treating types as values, procedures can be written which can be applied to objects of many different types (polymorphism).

Poly was not designed specifically to investigate polymorphism, rather it was designed to provide a simple yet powerful alternative to large languages like Ada. The type system came out of a desire to treat several different programming language concepts by means of a single parameterisation mechanism, that of procedure application. For example, generic types are considered simply as procedures. Polymorphism as Poly can also be used to provide the effect of overloading without building resolution rules into the language. Unlike the language Russell, Poly does not require that procedures be "variable free". However, it is still possible to statically type-check a Poly program.

After an introduction to the principles behind modern languages, in particular types and their relation to abstraction, there is a survey of several languages. Adu, CLU, Russell, ML and the Cedar Mesa Kernel illustrate different aspects of language design. Poly is described by means of some examples and then the background to the design is discussed. The rationale behind the type system of Poly is considered and comparisons are made with two other polymorphic languages, ML and Russell. The remainder of the language is developed and some applications are discussed. There is a description of some problems encountered while implementing poly.

UCAM-CL-TR-50

Lawrence Paulson:

Verifying the unification algorithm in LCF

March 1984, 28 pages, PDF, DVI

Abstract: Manna and Waldinger's theory of substitutions and unification has been verified using the Cambridge LCF theorem prover. A proof of the monotonicity of substitution is presented in detail, as an example of interaction with LCF. Translating the theory into LCF's domain-theoretic logic is largely straightforward. Well-founded induction on a complex ordering is translated into nested structural inductions. Correctness of unification is expressed using predicates for such properties as idempotence and most-generality. The verification is presented as a series of lemmas. The LCF proofs are compared with the original ones, and with other approaches. It appears difficult to find a logic that is both simple and flexible, especially for proving termination.

UCAM-CL-TR-51

Glynn Winskel, Kim Guldstrand Larsen:

Using information systems to solve recursive domain equations effectively

July 1984, 41 pages, PDF

Abstract: This paper aims to make two main contributions. One is to show how to use the concrete nature of Scott's information systems to advantage in solving recursive domain equations. The method is based on the substructure relation between information systems. This essentially makes a complete partial order (cpo) of information systems. Standard domain constructions like function space can be made continuous on this cpo so the solution of recursive domain equations reduces to the more familiar construction of forming the least-fixed point of a continuous function. The second contribution again relies on the concrete nature of information systems, this time to develop a basic theory of effectively given information systems and through this present a simple treatment of effectively given domains.

UCAM-CL-TR-52

Steven Temple:

The design of a ring communication network

132 pages, PDF
PhD thesis (Corpus Christi College, January 1984)

Abstract: This dissertation describes the design of a high speed local area network. Local networks have been in use now for over a decade and there is a proliferation of different systems, experimental ones which are not widely used and commercial ones installed in hundreds of locations. For a new network design to be of interest from the research point of view it must have a feature or features which set it apart from existing networks and make it an improvement over existing systems. In the case of the network described, the research was started to produce a network which was considerably faster than current designs, but which retained a high degree of generality.

As the research progressed, other features were considered, such as ways to reduce the cost of the network and the ability to carry data traffic of many different types. The emphasis on high speed is still present but other aspects were considered and are discussed in the dissertation. The network has been named the Cambridge Fast Ring and and the network hardware is currently being implemented as an integrated circuit at the University of Cambridge Computer Laboratory.

The aim of the dissertation is to describe the background to the design and the decisions which were made during the design process, as well as the design itself. The dissertation starts with a survey of the uses of local area networks and examines some established networks in detail. It then proceeds by examining the characteristics of a current network installation to assess what is required of the network in that and similar applications. The major design considerations for a high speed network controller are then discussed and a design is presented. Finally, the design of computer interfaces and protocols for the network is discussed.

UCAM-CL-TR-53

Jon Fairbairn:

A new type-checker for a functional language

July 1984, 16 pages, PDF

Abstract: A polymorphic type checker for the functional language Ponder [Fairbairn 82] is described. The initial sections give an overview of the syntax of Ponder, and some of the motivation behind the design of the type system. This is followed by a definition of the relation of 'generality' between these types, and of the notion of type-validity of Ponder programs. An algorithm to determine whether a Ponder program is type-valid is then presented. The final sections give examples of useful types which may be constructed within the type system, and describe some of the areas in which it is thought to be inadequate.

UCAM-CL-TR-54

Lawrence Paulson:

Lessons learned from LCF

August 1984, 16 pages, PDF

Abstract: The history and future prospects of LCF are discussed. The introduction sketches basic concepts such as the language ML, the logic PPLAMBDA, and backwards proof. The history discusses LCF proofs about denotational semantics, functional programs, and digital circuits, and describes the evolution of ideas about structural induction, tactics, logics of computation, and the use of ML. The biography contains thirty-five references.

UCAM-CL-TR-55

Ben Moszkowski:

Executing temporal logic programs

August 1984, 27 pages, PDF

Abstract: Over the last few years, temporal logic has been investigated as a tool for reasoning about computer programs, digital circuits and message-passing systems. In the case of programs, the general feeling has been that temporal logic is an adjunct to existing languages. For example, one might use temporal logic to specify and prove properties about a program written in, say, CSP. This leads to the annoyance of having to simultaneously use two separate notations.

In earlier work we proposed that temporal logic itself directly serve as the basis for a programming language. Since then we have implemented an interpreter for such a language called Tempura. We are developing Tempura as a tool for directly executing suitable temporal logic specifications of digital circuits and other discrete time systems. Since every Tempura statement is also a temporal formula, we can use the entire temporal logic formalism for our assertion language and semantics. Tempura has the two seemingly contradictory properties of being a logic programming langauge and having imperative constructs such as assignment statements.

The presentation given here first describes the syntax of a first order temporal logic having the operators $\circ$ (next) and $\Box$ (always). This serves as the basis for the Tempura programming language. The lesser known temporal operator chop is subsequently introduced, resulting in Interval Temporal Logic. We then show how to incorporate chop and related constructs into Tempura.

UCAM-CL-TR-56

William Stoye:

A new scheme for writing functional operating systems

September 1984, 30 pages, PDF

Abstract: A scheme is described for writing nondeterministic programs in a functional language. The scheme is based on message passing between a number of expressions being evaluated in parallel. I suggest that it represents a significant improvement over previous methods employing a nondeterministic merge primitive, and overcomes numerous drawbacks in that approach. The scheme has been designed in a practical context, and is being used to write an operating system for SKIM, a functionally programmed machine. It is not yet well understood in a mathematical sense.

UCAM-CL-TR-57

Lawrence C. Paulson:

Constructing recursion operators in intuitionistic type theory

October 1984, 46 pages, PDF, DVI

Abstract: Martin-Löf's Intuitionistic Theory of Types is becoming popular for formal reasoning about computer programs. To handle recursion schemes other than primitive recursion, a theory of well-founded relations is presented. Using primitive recursion over higher types, induction and recursion are formally derived for a large class of well-founded relations. Included are $<$ on natural numbers, and relations formed by inverse images, addition, multiplication, and exponentiation of other relations. The constructions are given in full detail to allow their use in theorem provers for Type Theory, such as Nuprl. The theory is compared with work in the field of ordinal recursion over higher types.

UCAM-CL-TR-58

Glynn Winskel:

Categories of models for concurrency

October 1984, 35 pages, PDF

Abstract: It is shown how a variety of models for concurrent processes can be viewed as categories in which familiar constructions turn out to be significant categorically. Constructions to represent various parallel compositions are often based on a product construction, for instance. In many cases different models can be related by a pair of functors forming an adjunction between the two categories. Because of the way in which such pairs of functors preserve categorical constructions, the adjunction serves to translate between the different models, so it is seen how semantics expressed in terms of one model translate to semantics in terms of another.

UCAM-CL-TR-59

Glynn Winskel:

On the composition and decomposition of assertions

November 1984, 35 pages, PDF

Abstract: Recently there has been a great deal of interest in the problem of how to compose modal assertions, in order to deduce the truth of an assertion for a composition of processes in a parallel programming language, from the truth of certain assertions for its components.

This paper addresses that problem from a theoretical standpoint. The programming language used is Robin Milner's Synchronous Calculus of Communicating Systems (called SCCS), while the language of assertions is a fragment of dynamic logic which, despite its simplicity, is expressive enough to characterise observational equivalence. It is shown how, with respect to each operation 'op' in SCCS, every assertion has a decomposition which reduces the problem of proving the assertion holds of a compound process built up using 'op' to proving assertions about its components. These results provide the foundations of a proof system for SCCS with assertions.

UCAM-CL-TR-60

Hiyan Alshawi:

Memory and context mechanisms for automatic text processing

192 pages, PDF
PhD thesis (Trinity Hall, December 1983)

Abstract: The thesis describes memory and context mechanisms for natural language text processing. The mechanisms were implemented as part of a computer system that successfully processed a number of short descriptive English texts producing output that can be used to create a relational database. The memory mechanism is concerned with representing and retrieving various kinds of knowledge, while the context mechanism is concerned with accumulating and applying information specifying which fragments of knowledge are currently more salient.

The mechanisms are used in the implemented system by an interpretation component dealing with common language interpretation problems that cannot be handled by simple sentence-level analysis. These problems include resolving references, disambiguating word senses, and discovering implicit relationships. The mechanisms are also used by a task-specific component which carries out the database capture application using database descriptions stored in memory. The choice and handling of the particular application task, interpretation operations, and types of context information,

were designed to check that the computational techniques developed for memory and context provide appropriate apparatus for non-trivial text processing involving a wide range of phenomena of language interpretation in context.

The memory representation formalism is based on hierarchies for classifying entities and the associations between them. It has the advantage of simplicity and a well designed semantics. Retrieval from memory is performed by marker processing on a network structure. The context mechanism represents instances of various types of context information as "context factors" which can be combined to derive activation values for memory entities. Context activation is used to choose the results of memory operations and to restrict memory searches. Context factors are created and modified as a result of text processing operations, leading to a gradual alteration of the context representation. Both the memory and context mechanisms utilize an indexing scheme that uses semantic clustering criteria. This increases the efficiency of retrieval from memory and allows efficient access to entities with high activations derived from several factors while individual factors can be managed independently.

UCAM-CL-TR-61

Karen Spärck Jones:

User models and expert systems

December 1984, 44 pages, PDF

Abstract: This paper analyses user models in expert systems in terms of the many factors involved: user roles, user properties, model types, model functions in relation to different aspects of system performance, and sources, e.g. linguistic or non-linguistic, of modelling information. The aim of the detailed discussion, with extensive examples illustrating the complexity of modelling, is to clarify the issues involved in modelling, as a necessary preliminary to model building.

UCAM-CL-TR-62

Michael Robson:

Constraint enforcement in a relational database management system

106 pages, PDF
PhD thesis (St John's College, March 1984)

Abstract: The dissertation describes the implementation of the data structuring concept of domains, intra-tuple constraints and referential constraints in a relational database management system (DBMS). The need for constraints is discussed and it is shown how they can be used to capture some of the semantics of the database's application. The implementation described was done within the framework of the particular DBMS CODD, the main features of which are presented.

Each class of constraint is described and it is shown how each of them is specifed to the DBMS. The descriptions of the constraints are stored in the database giving a centralised data model, which is used in the enforcement of the constraints. This data model contains descriptions not only of static structures but also of procedures to be used to maintain constraints. A detailed account is given of how each constraint is maintained.

The main focus of the dissertation is on referential constraints since inter-relational structure is an area in which relational systems are particularly weak. Referential constraints impose a network structure on the database and it is shown how referential constraints can be maintained by interpreting this network, using the data-pipelining facilities provided by CODD. It is also shown how referential constraints can be used to construct generalisation hierarchies, themselves an important data modelling tool. Further, some extensions to referential constraints, which allow them to capture more semantics, are suggested. The usefulness of referential constraints is illustrated by presenting a real database example (that of the University Computing Service), on which the ideas described in the dissertation have been tested.

UCAM-CL-TR-63

David C.J. Matthews:

Poly manual

February 1985, 46 pages, PDF

Abstract: Poly is a general purpose, High-level programming language. It has a simple type system which is also very powerful. Higher order procedures, polymorphic operations, parameterised abstract types and modules are all supported by a single mechanism.

Poly is strongly typed. All objects have a specification which the compiler can use to check that operations applied to them are sensible. Type errors cannot cause run time faults. The language is safe, meaning that any faults occuring at run time will result in exceptions which can be caught. All veriables must be initialised before use, so faults due to undefined variables cannot occur. Poly allows higher order procedures to be declared and used; these take another procedure as a parameter, or return a procedure as the result. Since Poly is statically scoped, this may still refer to the arguments and local variables of the procedure which returned it.

Poly allows polymorphic operations. Thus, it is possible to write one program to perform an operation on data of any type, provided only that the operation is available for the data type. Abstract types may be created and manipulated. These can be specified in such a way that only the functions to manipulate these objects are available to the user. This has the advantage that the

implementation can easily be changed, provided that it has the same external properties. Abstract types can be parameterised so that a set of types can be defined in a single definition. Types in Poly are similar to modules in other languages. For example, types can be separately compiled. An abstract type which makes use of other types can be written as though it were polymorphic; it will work if it is given any type which has the required operations. Its operation may be to return a new type which may be used directly or as a parameter to other polymorphic abstract types.

UCAM-CL-TR-64

Branimir K. Boguraev, Karen Spärck Jones:

A framework for inference in natural language front ends to databases

February 1985, 73 pages, paper copy

UCAM-CL-TR-65

Mark Tillotson:

Introduction to the programming language "Ponder"

May 1985, 57 pages, paper copy

UCAM-CL-TR-66

M.J.C. Gordon, J. Herbert:

A formal hardware verification methodology and its application to a network interface chip

May 1985, 39 pages, PDF

Abstract: We describe how the functional correctness of a circuit design can be verified by machine checked formal proof. The proof system used is LCF_LSM [1], a version of Milner's LCF [2] with a different logical calculus called LSM. We give a tutorial introduction to LSM in the paper.

Our main example is the ECL chip of the Cambridge Fast Ring (CFR) [3]. Although the ECL chip is quite simple (about 360 gates) it is nevertheless real. Minor errors were discovered as we performed the formal proof, but when the corrected design was eventually fabricated it was functionally correct first time. The main steps in verification were: (1) Writing a high-level behavioural specification in the LSM notation. (2) Translating the circuit design from its Modula-2 representation in the Cambridge Design Automation System [4] to LSM. (3) Using the LCF_LSM theorem proving system to mechanically generate a proof that the behaviour determined by the design is equivalent to the specified behaviour.

In order to accomplish the second of these steps, an interface between the Cambridge Design Automation System and the LCF_LSM system was constructed.

UCAM-CL-TR-67

Lawrence C. Paulson:

Natural deduction theorem proving via higher-order resolution

May 1985, 22 pages, PDF

Abstract: An experimental theorem prover is described. Like LCF it is embedded in the metalanguage ML and supports backward proof using tactics and tacticals. The prover allows a wide class of logics to be introduced using Church's representation of quantifiers in the typed lambda-calculus. The inference rules are expressed as a set of generalized Horn clauses containing higher-order variables. Depth-first subgoaling along inference rules is essentially linear resolution, but using higher-order unification instead of first-order. This constitutes a higher-order Prolog interpreter.

The rules of Martin Löf's Constructive Type Theory have been entered into the Prover. Special tactics inspect a goal and decide which type theory rules may be appropriate, avoiding excessive backtracking. These tactics can automatically derive the types of many Type Theory expressions. Simple functions can be derived interactively.

UCAM-CL-TR-68

Mike Gordon:

HOL
A machine oriented formulation of higher order logic

July 1985, 52 pages, PDF

Abstract: In this paper we describe a formal language intended as a basis for hardware specification and verification. The language is not new; the only originality in what follows lies in the presentation of the details. Considerable effort has gone into making the formalism suitable for manipulation by computer.

The logic described here underlies an automated proof generator called HOL. The HOL logic is derived from Church's Simple Type Theory by: making the syntax more readable, allowing types to contain variables, and building in the Axiom of Choice via Hilbert's ϵ-operator.

The exact syntax of the logic is defined relative to a theory, which determines the types and constants that are available. Theories are developed incrementally

starting from the standard theories of truth-values or booleans, and of individuals. This paper describes the logic underlying the HOL system.

UCAM-CL-TR-69

Lawrence C. Paulson:

Proving termination of normalization functions for conditional expressions

June 1985, 16 pages, PDF, DVI

Abstract: Boyer and Moore have discussed a recursive function that puts conditional expressions into normal form. It is difficult to prove that this function terminates on all inputs. Three termination proofs are compared: (1) using a measure function, (2) in domain theory using LCF, (3) showing that its "recursion relation", defined by the pattern of recursive calls, is well-founded. The last two proofs are essentially the same though conducted in markedly different logical frameworks. An obviously total variant of the normalize function is presented as the 'computational meaning' of those two proofs.

A related function makes nested recursive calls. The three termination proofs become more complex: termination and correctness must be proved simultaneously. The recursion relation approach seems flexible enough to handle subtle termination proofs where previously domain theory seemed essential.

UCAM-CL-TR-70

Kenneth Graham Hamilton:

A remote procedure call system

December 1984, 109 pages, PDF
PhD thesis (Wolfson College, December 1984)

Abstract: The provision of a suitable means for communication between software modules on different machines is a recognized problem in distributed computing research. Recently the use of language-level Remote Procedure Call (RPC) has been advocated as a solution to this problem.

This thesis discusses the rationale, design, implementation and supporting environment of a flexible RPC system for an extended version of the CLU programming language. It is argued that earlier RPC systems have adopted an undesirably rigid stance by attempting to make remotre procedure calls look as similar as possible to local procedure calls. It is suggested instead that the inevitable differences in performance and failure properties between local and remote calls should be regarded as being essentially different from local calls. Following from this, it is proposed that RPC systems should offer at least two complementary call mechanisms. One of these should attempt to recover from network errors and should only report unrecoverable failures. The other should never attempt automatic recovery from network errors, thereby giving implementors the convenience of a language-level mechanism without losing sight of the underlying network.

Other specific areas that are discussed include binding issues, protocols, transmission mechanisms for standard data types, and the particular problems posed by abstract data types. A new transfer mechanism for abstract types is proposed which would permit software using new representations to communicate with software using earlier representations. The provision of special operating system support for the CLU RPC mechanism is also discussed.

UCAM-CL-TR-71

Ben Moszkowski:

Executing temporal logic programs

August 1985, 96 pages, paper copy

UCAM-CL-TR-72

W.F. Clocksin:

Logic programming and the specification of circuits

May 1985, 13 pages, PDF

Abstract: Logic programming (see Kowalski, 1979) can be used for specification and automatic reasoning about electrical circuits. Although propositional logic has long been used for describing the truth functions of combinational circuits, the more powerful Predicate Calculus on which logic programming is based has seen relatively little use in design automation. Previous researchers have introduced a number of techniques similar to logic programming, but many of the useful consequences of the logic programming methodology have not been exploited. This paper first reviews and compares three methods for representing circuits, which will be called here the functional method, the extensional method, and the definitional method. The latter method, which conveniently admits arbitrary sequential circuits, is then treated in detail. Some useful consequences of using this method for writing directly executable specifications of circuits are described. These include the use of quantified variables, verification of hypothetical states, and sequential simulation.

UCAM-CL-TR-73

Daniel Hammond Craft:

Resource management in a distributed computing system

116 pages, PDF
PhD thesis (St John's College, March 1985)

Abstract: The Cambridge Distributed System, based on the Cambridge Ring local area network, includes a heterogeneous collection of machines known as the processor bank. These machines may run network servers, or may be loaded with services and allocated to users dynamically. The machines and the variety of services they can support (eg. different operating systems, compilers, formatters) are viewed as resources available to other components of the distributed system.

By using a processor bank, two fundamental limitations of the personal computer approach to distributed computing can be overcome: responsiveness for computation-intensive tasks is not limited by the single, personal machine because tasks may expand into processor bank machines as necessary; and applications are not limited to the operating system or languages available on the personal computer because all of the systems or languages which run on processor bank machines are at the users disposal, both for implementing new applications and for importing applications from other systems. Resource management is seen as one of the four areas which must be addressed to realize these advantages.

The resource management system must match client requirements for resources to those resources which are available on the network. To do this it maintains two data bases: one contains information describing existing resources, and the other contains information indicating how to obtain resources from servers or have them constructed from existing subresources by fabricators. The resource management system accepts resource requirements from clients and picks from the alternatives in these data bases the "best" match (as defined by the resource management policy).

The resource management issues addressed include resource description, location and allocation, construction, monitoring and reclamation, authentication and protection, and policy. The design and implementation of two resource management servers is discussed.

UCAM-CL-TR-74

Mike Gordon:

Hardware verification by formal proof

August 1985, 6 pages, PDF

Abstract: The use of mathematical proof to verify hardware designs is explained and motivated. The hierarchical verification of a simple n-bit CMOS counter is used as an example. Some speculations are made about when and how formal proof will become used in industry.

UCAM-CL-TR-75

Jon Fairbairn:

Design and implementation of a simple typed language based on the lambda-calculus

May 1985, 107 pages, PDF
PhD thesis (Gonville & Caius College, December 1984)

Abstract: Despite the work of Landin and others as long ago as 1966, almost all recent programming languages are large and difficult to understand. This thesis is a re-examination of the possibility of designing and implementing a small but practical language based on very few primitive constructs.

The text records the syntax and informal semantics of a new language called Ponder. The most notable features of the work are a powerful type-system and an efficient implementation of normal order reduction.

In contrast to Landin's ISWIM, Ponder is statically typed, an expedient that increases the simplicity of the language by removing the requirement that operations must be defined for incorrect arguments. The type system is a powerful extension of Milner's polymorphic type system for ML in that it allows local quantification of types. This extension has the advantage that types that would otherwise need to be primitive may be defined.

The criteria for the well-typedness of Ponder programmes are presented in the form of a natural deduction system in terms of a relation of generality between types. A new type checking algorithm derived from these rules is proposed.

Ponder is built on the λ-calculus without the need for additional computation rules. In spite of this abstract foundation an efficient implementation based on Hughes' super-combinator approach is described. Some evidence of the speed of Ponder programmes is included.

The same strictures have been applied to the design of the syntax of Ponder, which, rather than having many pre-defined clauses, allows the addition of new constructs by the use of a simple extension mechanism.

UCAM-CL-TR-76

R.C.B. Cooper, K.G. Hamilton:

Preserving abstraction in concurrent programming

August 1985, 16 pages, PDF

Abstract: Recent programming languages have attempted to provide support for concurrency and for modular programming based on abstract interfaces. Building on our experience of adding monitors to CLU, a language orientated towards data abstraction, we explain how these two goals conflict. In particular we discuss the clash between conventional views on interface abstraction and the programming style required

for avoiding monitor deadlock. We argue that the best compromise between these goals is a combination of a fine grain locking mechanism together with a method for explicitly defining concurrency properties for selected interfaces.

UCAM-CL-TR-77

Mike Gordon:

Why higher-order logic is a good formalisation for specifying and verifying hardware

September 1985, 28 pages, PDF

Abstract: Higher order logic was originally developed as a foundation for mathematics. In this paper we show how it can be used as: 1. a hardware description language, and 2. a formalism for proving that designs meet their specifications.

Examples are given which illustrate various specification and verification techniques. These include a CMOS inverter, a CMOS full adder, an n-bit ripple-carry adder, a sequential multiplier and an edge-triggered D-type register.

UCAM-CL-TR-78

Glynn Winskel:

A complete proof system for SCCS with model assertions

September 1985, 23 pages, PDF

Abstract: This paper presents a proof system for Robin Milner's Synchronous Calculus of Communicating Systems (SCCS) with modal assertions. The language of assertions is a fragment of dynamic logic, sometimes called Hennessy-Milner logic after they brought it to attention; while rather weak from a practical point of view, its assertions are expressive enough to characterise observation equivalence, central to the work of Milner et al. on CCS and SCCS. The paper includes a completeness result and a proof of equivalence between an operational and denotational semantics for SCCS. Its emphasis is on the theoretical issues involved in the construction of proof systems for parallel programming langauges.

UCAM-CL-TR-79

Glynn Winskel:

Petri nets, algebras and morphisms

38 pages, PDF

Abstract: It is shown how a category of Petri nets can be viewed as a subcategory of two sorted algebras over multisets. This casts Petri nets in a familiar framework and provides a useful idea of morphism on nets different from the conventional definition – the morphisms here respect the behaviour of nets. The categorical constructions with result provide a useful way to synthesise nets and reason about nets in terms of their components; for example various forms of parallel composition of Petri nets arise naturally from the product in the category. This abstract setting makes plain a useful functor from the category of Petri nets to a category of spaces of invariants and provides insight into the generalisations of the basic definition of Petri nets – for instance the coloured and higher level nets of Kurt Jensen arise through a simple modificationof the sorts of the algebras underlying nets. Further it provides a smooth formal relation with other models of concurrency such as Milner's Calculus of Communicating Systems (CCS) and Hoare's Communicating Sequential Processes (CSP).

UCAM-CL-TR-80

Lawrence C. Paulson:

Interactive theorem proving with Cambridge LCF A user's manual

November 1985, 140 pages, paper copy

UCAM-CL-TR-81

William Robert Stoye:

The implementation of functional languages using custom hardware

December 1985, 151 pages, PDF
PhD thesis (Magdalene College, May 1985)

Abstract: In recent years functional programmers have produced a great many good ideas but few results. While the use of functional languages has been enthusiastically advocated, few real application areas have been tackled and so the functional programmer's views and ideas are met with suspicion.

The prime cause of this state of affairs is the lack of widely available, solid implementations of functional languages. This in turn stems from two major causes: (1) Our understanding of implementation techniques was very poor only a few years ago, and so any implementation that is "mature" is also likely to be unuseably slow. (2) While functional languages are excellent for expressing algorithms, there is still considerable debate in the functional programming community over the way in which input and output operations should

be represented to the programmer. Without clear guiding principles implementors have tended to produce adhoc, inadequate solutions.

My research is concerned with strengthening the case for functional programming. To this end I constructed a specialised processor, called SKIM, which could evaluate functional programs quickly. This allowed experimentation with various implementation methods, and provided a high performance implementation with which to experiment with writing large functional programs.

This thesis describes the resulting work and includes the following new results: (1) Details of a practical turner-style combinator reduction implementation featuring greatly improved storage use compared with previous methods. (2) An implementation of Kennaway's director string idea that further enhances performance and increases understanding of a variety of reduction strategies. (3) Comprehensive suggestions concerning the representation of input, output, and nondeterministic tasks using functional languages, and the writing of operating systems. Details of the implementation of these suggestions developed on SKIM. (4) A number of observations concerning fuctional programming in general based on considerable practical experience.

UCAM-CL-TR-82

Lawrence C. Paulson:

Natural deduction proof as higher-order resolution

December 1985, 25 pages, PDF, DVI

Abstract: An interactive theorem prover, Isabelle, is under development. In LCF, each inference rule is represented by one function for forwards proof and another (a tactic) for backwards proof. In Isabelle, each inference rule is represented by a Horn clause. Resolution gives both forwards and backwards proof, supporting a large class of logics. Isabelle has been used to prove theorems in Martin-Löf's Constructive Type Theory.

Quantifiers pose several difficulties: substitution, bound variables, Skolemization. Isabelle's representation of logical syntax is the typed lambda-calculus, requiring higher-order unification. It may have potential for logic programming. Depth-first search using inference rules constitutes a higher-order Prolog.

UCAM-CL-TR-83

Ian David Wilson:

Operating system design for large personal workstations

203 pages, PDF
PhD thesis (Darwin College, July 1985)

Abstract: With the advent of personal computers in the mid 1970s, the design of operating systems has had to change in order to take account of the new machines. Traditional problems such as accounting and protection are no longer relevant, but compactness, efficiency and pertability have all become important issues as the number of these small systems has grown.

Since that time, due to the reductions in the costs of computer components and manufacture, personal workstations have become more common with not only the number of machines having increased, but also their CPU power and memory capacity. The work on software for the new machines has not kept pace with the improvements in hardware design, and this is particularly true in the area of operating systems, where there is a tendency to treat the new machines as small, inferior mainframes.

This thesis investigates the possibility of enhancing work done on the original personal computer operating systems, so that better utilisation of the new machines can be obtained. The work concentrates on two main areas of improvement: the working environment as perceived by the user, and the underlying primitives and algorithms used by the operating system kernel.

The work is illustrated by two case studies, the user environment of the TRIPOS operating system is described, along with a new command line interpreter and command programming language, and a series of techniques to make better use of the available hardware facilities is discussed. The kernel of the TRIPOS operating system is examined critically, particularly with respect to the way that machine resources are used, and finally, a new set of kernel primitives and algorithms is suggested, with reference to an experimental kernel for the real time implementation of network protocol software.

UCAM-CL-TR-84

Martin Richards:

BSPL: a language for describing the behaviour of synchronous hardware

April 1986, 56 pages, paper copy

UCAM-CL-TR-85

Glynn Winskel:

Category theory and models for parallel computation

April 1986, 16 pages, PDF

Abstract: This report will illustrate two uses of category theory: Firstly the use of category theory to define semantics in a particular model. How semantic constructions can often be seen as categorical ones, and, in particular, how parallel compositions are derived from a categorical product and a nun-deterministic sum. These categorical notions can provide a basis for reasoning about computations and will be illustrated for the model of Petri nets.

Secondly, the use of category theory to relate different semantics will be examined; specifically, how the relations between various concrete models like Petri nets, event structures, trees and state machines are expressed as adjunctions. This will be illustrated by showing the coreflection between safe Petri nets and trees.

UCAM-CL-TR-86

Stephen Christopher Crawley:

The Entity System: an object based filing system

April 1986, 120 pages, PDF
PhD thesis (St John's College, December 1985)

Abstract: Developments in programming languages have provided increasingly powerful facilities for algorithmic and data abstraction. Concepts such as record declarations and formal type checking have been developed by languages such as Pascal and Algol 68, while languages such as Simula 67 and Smalltalk supported object based type systems. Until recently, however, very little work has been done on extending data typing concepts beyond a single program, where I/O is typically performed by reading and writing data as an untyped stream of characters.

By contrast, database systems have traditionally taken a data and file based approach to storing complex data, and address the problems of many programs using the same data, while handing changing data descriptions and access requirements. Recently attention has been focussed on extending data typing beyond the bounds of a single program. The DTL language [Hughes 83] models a program as a data transformer which converts one typed data stream into another, while PS-Algol extends a representational type system by allowing data in the heap to persist from one run of a program to the next. None of these, however, really address the issues of evolving programs and data requirements.

In order to build the desired functionality in programming environments, the file system needs to provide considerably more functionality, by joining together the components of a modular program, and supporting both small and large components efficiently. Finally a mechanism was needed for ensuring that files were treated consistently. The term entity is used to describe an object held in the file system, which are modelled as a collection of strongly typed attributes with abstract interfaces. This thesis describes the experience gained in constructing such a system and the requirements of an effective persistent storage system.

UCAM-CL-TR-87

Kathleen Anne Carter:

Computer-aided type face design

May 1986, 172 pages, PDF
PhD thesis (King's College, November 1985)

Abstract: This thesis tackles the problems encountered when trying to carry out a creative and intuitive task, such as type face design, on a computer. A brief history of printing and type design sets the scene for a discussion of digital type. Existing methods for generating and handling digital type are presented and their relative merits are discussed. Consideration is also given to the nature of designing, independent of the tools used. The importance of intuition and experience in such a task is brought out. Any new tools must allow the designer to exercise his skills of hand and eye, and to judge the results visually. The different abstractions that can be used to represent a typeface in a computer are discussed with respect to the manner of working that they force upon the designer.

In the light of this discussion some proposals are made for a new system for computer-aided type face design. This system must be highly interactive, providing rapid visual feedback in response to the designer's actions. Designing is a very unstructured task, frequently with a number of activities being pursued at once. Hence the system must also be able to support multiple activities, with the user free to move between them at any time.

The characteristics of various types of interactive graphical environment are then considered. This discussion leads on to proposals for an environment suitable for supporting type face design. The proposed anvironment is based on the provision of a number of windows on the screen, each supporting a different activity. A mouse, graphics tablet and keyboard are all continuously available for interection with the system. The rest of the thesis discusses the implementation of this graphical environment and the type face design system that makes use of it. The final chapter evaluates the success of both the underlying software and of the type face design system itself.

UCAM-CL-TR-88

David Maclean Carter:

A shallow processing approach to anaphor resolution

May 1986, 233 pages, PDF
PhD thesis (King's College, December 1985)

Abstract: The thesis describes an investigation of the faesibility of resolving anaphors in natural language texts by means of a "shallow processing" approach which exploits knowledge of syntax, semantics and local focussing as heavily as possible; it does not rely on the presence of large amounts of world or domain knowledge, which are notoriously hard to process accurately.

The ideas reported are implemented in a program called SPAR (Shallow Processing Anaphor Resolver), which resolves anaphoric and other linguistic ambiguities in simple English stories and generates sentence-by-sentence paraphrases that show what interpretations have been selected. Input to SPAR takes the form of semantic structures for single sentences constructed by Boguraev's English analyser. These structures are integrated into a network-style text representation as processing proceeds. To achieve anaphor resolution, SPAR combines and develops several existing techniques, most notably Sidner's theory of local focussing and Wilks' "preference semantics" theory of semantics and common sense inference.

Consideration of the need to resolve several anaphors in the same sentence results in Sidner's framework being modified and extended to allow focus-based processing to ineract more flexibly with processing based on other types of knowledge. Wilks' treatment of common sense inference is extended to incorporate a wider range of types of inference without jeopardizing its uniformity and simplicity. Further his primitive-based formalism for word sense meanings is developed in the interests of economy, accuracy and ease of use.

Although SPAR is geared mainly towards resolving anaphors, the design of the system allows many non-anaphoric (lexical and structural) ambiguities that cannot be resloved during sentence analysis to be resolved as a by-product of anaphor resolution.

UCAM-CL-TR-89

Jon Fairbairn:

Making form follow function An exercise in functional programming style

June 1986, 9 pages, PDF

Abstract: The combined use of user-defined infix operators and higher order functions allows the programmer to invent new control structures tailored to a particular problem area.

This paper is to suggest that such a combination has beneficial effects on the ease of both writing and reading programmes, and hence can increase programmer productivity. As an example, a parser for a simple language is presented in this style.

It is hoped that the presentation will be palatable to people unfamiliar with the concepts of functional programming.

UCAM-CL-TR-90

Andy Hopper, Roger M. Needham:

The Cambridge Fast Ring networking system (CFR)

June 1986, 25 pages, PDF

Abstract: Local area networks have developed from slow systems operating at below 1MBs to fast systems at 50MBs or more. We discuss the choices facing a designer as faster speeds for networks are contemplated. The 100MBs Cambridge Fast Ring is described. The ring protocol allows one of a number of fixed size slots to be used once or repeatedly. The network design allows sets of rings to be constructed by pushing the bridge function to the lowest hardware level. Low cost and ease of use is normally achieved by design of special chips and we describe a two-chip VLSI implementation. This VLSI hardware forms the basis of a kit-of-parts from which many different network components can be constructed.

UCAM-CL-TR-91

Albert Camilleri, Mike Gordon, Tom Melham:

Hardware verification using higher-order logic

September 1986, 25 pages, PDF

Abstract: The Hardware Verification Group at the University of Cambridge is investigating how various kinds of digital systems can be verified by mechanised formal proof. This paper explains our approach to representing behaviour and structure using higher order logic. Several examples are described including a ripple carry adder and a sequential device for computing the factorial function. The dangers of inaccurate models are illustrated with a CMOS exclusive-or gate.

UCAM-CL-TR-92

Stuart Charles Wray:

Implementation and programming techniques for functional languages

June 1986, 117 pages, PDF
PhD thesis (Christ's College, January 1986)

Abstract: In this thesis I describe a new method of strictness analysis for lazily evaluated functional languages, and a method of code generation making use of the information provided by this analysis. I also describe techniques for practical programming in lazily evaluated functional languages, based on my experience of writing substantial functional programs.

My new strictness analyser is both faster and more powerful than that of Mycroft. It can be used on any program expressed as super-combinator definitions and it uses the additional classifications absent and dangerous as well as strict and lazy. This analyser assumes that functional arguments to higher order functions are completely lazy.

I describe an extension of my analyser which discovers more strictness in the presence of higher order functions, and I compare this with higher order analysers based on Mycroft's work. I also describe an extension of my analyser to lazy pairs and discuss strictness analysers for lazy lists.

Strictness analysis brings useful performance improvements for programs running on conventional machines. I have implemented my analyser in a compiler for Ponder, a lazily evaluated functional language with polymorphic typing. Results are given, including the surprising result that higher order strictness analysis is no better than first order strictness analysis for speeding up real programs on conventional machines.

I have written substantial programs in Ponder and describe in some detail the largest of these which is about 2500 lines long. This program is an interactive spreadsheet using a mouse and bitmapped display. I discuss programming techniques and practical problems facing functional languages with illustrative examples from programs I have written.

UCAM-CL-TR-93

J.P. Bennett:

Automated design of an instruction set for BCPL

June 1986, 56 pages, paper copy

UCAM-CL-TR-94

Avra Cohn, Mike Gordon:

A mechanized proof of correctness of a simple counter

June 1986, 80 pages, paper copy

UCAM-CL-TR-95

Glynn Winskel:

Event structures Lecture notes for the Advanced Course on Petri Nets

July 1986, 69 pages, PDF

Abstract: Event structures are a model of computational processes. They represent a process as a set of event occurrences with relations to express how events causally depend on others. This paper introduces event structures, shows their relationship to Scott domains and Petri nets, and surveys their role in denotational semantics, both for modelling laguages like CCS and CSP and languages with higher types.

UCAM-CL-TR-96

Glynn Winskel:

Models and logic of MOS circuits Lectures for the Marktoberdorf Summerschool, August 1986

October 1986, 47 pages, PDF

Abstract: Various models of hardware have been proposed though virtually all of them do not model circuits adequately enough to support and provide a formal basis for many of the informal arguments used by designers of MOS circuits. Such arguments use rather crude discrete notions of strength – designers cannot be too finicky about precise resistances and capacitances when building a chip – as well as subtle derived notions of information flow between points in the circuit. One model, that of R.E. Bryant, tackles such issues in reasonable generality and has been used as the basis of several hardware simulators. However Bryant's model is not compositional. These lectures introduce Bryant's ideas and present a compositional model for the behaviour of MOS circuits when the input is steady, show how this leads to a logic, and indicate the difficulties in providing a full and accurate treatment for circuits with changing inputs.

UCAM-CL-TR-97

Alan Mycroft:

A study on abstract interpretation and "validating microcode algebraically"

October 1986, 22 pages, PDF

Abstract: This report attempts to perfrom two roles: the first part aims to give a state-of-the-art introduction to abstract interpretation with as little mathematics as possible. The question of the 'best' meta-language for abstract interpretation is, however, left open. The second part gives a tutorial introduction to an application of abstract interpretation based on the relational style of Mycroft and Jones (1985). This report does not claim to have introduced any new techniques, but rather aims to make the existing literature understandable to a wider audience.

UCAM-CL-TR-98

E. Robinson:

Power-domains, modalities and the Vietoris monad

October 1986, 16 pages, PDF

Abstract: It is possible to divide the syntax-directed approaches to programming language semantics into two classes, "denotational", and "proof-theoretic". This paper argues for a different approach which also has the effect of linking the two methods. Drawing on recent work on locales as formal spaces we show that this provides a way in which we can hope to use a proof-theoretical semantics to give us a denotational one. This paper reviews aspects of the general theory, before developing a modal construction on locales and discussing the view of power-domains as free non-deterministic algebras. Finally, the relationship between the present work and that of Winskel is examined.

UCAM-CL-TR-99

David C.J. Matthews:

An overview of the Poly programming language

August 1986, 11 pages, PDF

Abstract: Poly is a general purpose programming language based on the idea of treating types as first class values. It can support polymorphic operations by passing types as parameters to procedures, and abstract types and parameterised types by returning types as results.

Although Poly is not intended specifically as a database programming language it was convenient to implement it is a persistent storage system. This allows the user to retain data structures from one session to the next, and can support large programming systems such as the Poly compiler and a Standard ML system.

UCAM-CL-TR-100

Jeff Joyce, Graham Birtwistle, Mike Gordon:

Proving a computer correct in higher order logic

December 1986, 57 pages, PDF

Abstract: Technical report no. 42, 'Proving a computer correct using the LSF_LSM hardware verification system', describes the specification and verification of a register-transfer level implementation of a simple general purpose computer. The computer has a microcoded control unit implementing eight user level instructions. We have subsequently redone this example in higher order logic using the HOL hardware verification system.

This report presents the specification and verification of Gordon's computer as an example of hardware specification and verification in higher order logic. The report describes how the structure and behaviour of digital circuits may be specified using the formalism of higher order logic. The proof of correctness also shows how digital behaviour at different granularities of time may be related by means of a temporal abstraction.

This report should be read with Technical report no. 68, 'HOL, a machine oriented formulation of higher order logic', which describes the logic underlying the HOL hardware verification system.

UCAM-CL-TR-101

David Russel Milway:

Binary routing networks

December 1986, 131 pages, PDF
PhD thesis (Darwin College, December 1986)

Abstract: Binary Routing Networks combine ideas from Wide Area Networks and Interconnection Networks with the principles of Local Area Networks. This results in a high performance network for use in the local and wide area environment. Simulation of this form of network shows that for certain structures the performance of the network can approach or even exceed that obtained from a cross-bar switch. This dissertation describes how network structures based on Binary Routing Networks can be used in applications where a network capable of high rates of throughput with low delay is required.

Binary Routing Networks use a switching fabric constructed from simple routing nodes to route packets from a source to a destination. Some network topologies allow many packets to pass through the network simultanously, giving the network an aggregate throughput mugh greater than the basic bit rate. Routing nodes do not require knowledge of the topology and are thus simple to construct. They use routing information in the packet to direct the packet through the network. Packets pass through the nodes with little delay except

where contention for a link occurs when the packet needs to be buffered.

A design for a non-buffered routing node is described where contention is resolved by discarding one of the packets. Discarded packets are retried later by the sending station. This form of network removes the buffers from the routing nodes making them even simpler to construct. Simulations of a network of 512 stations show that for loads per station of up to 20% of the basic bit rate, a non-buffered network can outperform a buffered network. This design allows the construction of a fault tolerant network which can pass packets through any number of different paths, avoiding broken links or congensted areas in the network.

A prototype of a Binary Routing Network is discussed. This network makes use of the non-buffered routing nodes and measurements of its performance are compared with results obtained from the simulations. A proposal for using this form of network in an Integrated Service environment are also given.

Structures similar to Binary Routing Networks are fast becoming the backbone of multiprocessor systems. Local Area Networks also need to apply this technology to meet the requirements that they are being asked to support.

UCAM-CL-TR-102

David C.J. Matthews:

A persistent storage system for Poly and ML

January 1987, 16 pages, PDF

Abstract: The conventional strategy for implementing interactive languages has been based on the use of a "workspace" or "core-image" which is read in at the start of a session and written out at the end. While this is satisfactory for small systems it is inefficient for large programs. This report describes how an idea originally invented to simplify database programming, the persistent store, was adapted to support program development in an interactive language.

Poly and ML are both semi-functional languages in the sense that they allow functions as first class objects but they have variables (references) and use call-by-value semantics. Implementing such languages in a persistent store poses some problems but also allows optimisations which would not be possible if their type systems did not apply certain constraints.

The basic system is designed for single-users but the problems of sharing data between users is discussed and an experimental system for allowing this is described.

UCAM-CL-TR-103

Mike Gordon:

HOL

A proof generating system for higher-order logic

January 1987, 56 pages, PDF

Abstract: HOL is a version of Robin Milner's LCF theorem proving system for higher-order logic. It is currently being used to investigate: how various levels of hardware behaviour can be rigorously modelled; and how the resulting behavioural representations can be the basis for verification by mechanized formal proof.

This paper starts with a tutorial introduction to the meta-language ML. The version of higher-order logic implemented in the HOL system is then described. This is followed by an introduction to goal-directed proof with tactics and tacticals. Finally, there is a little example showing the system in action. This example illustrates how HOL can be used for hardware verification.

UCAM-CL-TR-104

Avra Cohn:

A proof of correctness of the Viper microprocessor: the first level

January 1987, 46 pages, PDF

Abstract: The Viper microprocessor designed at the Royal Signals and Radar Establishment (RSRE) is one of the first commercially produced computers to have been developed using modern formal methods. Viper is specified in a sequence of decreasingly abstract levels. In this paper a mechanical proof of the equivalence of the first two of these levels is described. The proof was generated using a version of Robin Milner's LCF system.

UCAM-CL-TR-105

Glynn Winskel:

A compositional model of MOS circuits

April 1987, 25 pages, PDF

Abstract: This paper describes a compositional model for MOS circuits. Like the model of Bryant (1984), it covers some of the effects of capacitance and resistance used frequently in designs. Although this has formed the basis of several hardware simulators, it suffers from the inadequacy that it is not compositional, making it difficult to reason in a structured way.

The present paper restricts its attention to the static behaviour of circuits, representing this as the set of possible steady states the circuit can settle into. A good understanding of such static behaviour is necessary to

treat sequential circuits. This paper further takes the view that it is useful to have a language to describe the construction of circuits, and to this end borrows ideas from Hoare's Communicating Sequential Processes, and Milner's Calculus of Communicating Systems.

UCAM-CL-TR-106

Thomas F. Melham:

Abstraction mechanisms for hardware verification

May 1987, 26 pages, PDF

Abstract: It is argued that techniques for proving the correctness of hardware designs must use abstraction mechanisms for relating formal descriptions at different levels of detail. Four such abstraction mechanisms and their formalisation in higher order logic are discussed.

UCAM-CL-TR-107

Thierry Coquand, Carl Gunter, Glynn Winskel:

DI-domains as a model of polymorphism

May 1987, 19 pages, PDF

Abstract: This paper investigates a model of the polymorphic lambda calculus recently described by Girard (1985). This model differs from earlier ones in that all the types are interpreted as domains rather than closures or finitary projections on a universal domain. The objective in this paper is to generalize Girard's construction to a larger category called dI-domains, and secondly to show how Girard's construction (and this generalization) can be done abstractly. It demonstrates that the generalized construction can be used to do denotational semantics in the ordinary way, but with the added feature of type polymorphism.

UCAM-CL-TR-108

Andrew John Wilkes:

Workstation design for distributed computing

June 1987, 179 pages, PDF
PhD thesis (Wolfson College, June 1984)

Abstract: This thesis discusses some aspects of the design of computer systems for local area networks (LANs), with particular emphasis on the way such systems present themselves to their users. Too little attention to this issue frequently results in computing environments that cannot be extended gracefully to accommodate new hardware or software and do not present consistent, uniform interfaces to either their human users or their programmatic clients. Before computer systems can become truly ubiquitous tools, these problems of extensibility and accessibility must be solved. This dissertation therefore seeks to examine one possible approach, emphasising support for program development on LAN based systems.

UCAM-CL-TR-109

Jeffrey Joyce:

Hardware verification of VLSI regular structures

July 1987, 20 pages, PDF

Abstract: Many examples of hardware specification focus on hierarchical specification as a means of controlling structural complexity in design. Another method is the use of iteration. This paper, however, presents a third method, namely the mapping of irregular combinational functions to regular structures.

Regular structures often result in solutions which are economical in terms of area and design time. The automatic generation of a regular structure such as a ROM or PLA from a functional specification usually accommodates minor changes to the functional specification.

The mapping of irregular combinational functions to a regular structure separates function from circuit design. This paper shows how this separation can be exploited to derive a behavioural specification of a regular structure parameterized by the functional specification.

UCAM-CL-TR-110

Glynn Winskel:

Relating two models of hardware

July 1987, 16 pages, PDF

Abstract: The idea of this note is to show how Winskel's static-configuration model of circuits is related formally to Gordon's relational model. Once so related, the simpler proofs in the relational model can, for instance, be used to justify results in terms of the static-configurations model. More importantly, we can exhibit general conditions on circuits which ensure that

assertions which hold of a circuit according to the simpler model are correct with respect to the more accurate model. The formal translation makes use of a simple adjunction between (partial order) categories associated with the two models, in a way reminiscient of abstract interpretation. Preliminary results suggest similar lines of approach may work for other kinds of abstraction such as temporal abstraction used in e.g. Melham's work to reason about hardware, and, more generally, make possible a formal algebraic treatment of the relationship between different models of hardware.

UCAM-CL-TR-111

K. Spärck Jones:

Realism about user modelling

June 1987, 32 pages, PDF

Abstract: This paper reformulates the framework for user modelling presented in an earlier technical report, 'User Models and Expert Systems', and considers the implications of the real limitations on the knowledge likely to be available to a system for the value and application of user models.

UCAM-CL-TR-112

D.A. Wolfram:

Reducing thrashing by adaptive backtracking

August 1987, 15 pages, PDF

Abstract: Adaptive backtracking dynamically reduces thrashing caused by blind backtracking and recurring failures, by locating early backtrack points and deleting choices which are not part of any solution. Search problems with hereditary bounding properties are soluble by this method. These problems include searches in theorem proving, logic programming, reason maintenance, and planning. The location of a backtrack point uses a particular minimal inconsistent subset, which is called the cause set. A rejection set is computed from the union of cause sets and rejection sets at a failure are used to locate subsequent backtrack points. A choice is deleted when a rejection set is a singleton. The worst case overhead is O(nf(n)) in time if the bounding property can be tested in O(f(n)) time, and $O(n^2)$ in space. An implementation confirms the expected exponential speed-ups for problems whose solution involves much thrashing.

UCAM-CL-TR-113

Lawrence C. Paulson:

The representation of logics in higher-order logic

August 1987, 29 pages, PDF

Abstract: Intuitionistic higher-order logic — the fragment comtaining implication, universal quantification, and equality — can serve as a meta-logic for formalizing various logics. As an example, axioms formalizing first-order logic are presented, and proved sound and complete by induction on proof trees.

Proofs in higher-order logic represent derivations of rules as well as proofs of theorems. A proof develops by deriving rules using higher-order resolutions. The discharge of assumptions involves derived meta-rules for 'lifting' a proposition.

Quantifiers require a similar lifting rule or else Hilbert's ϵ-operator. The alternatives are contrasted through several examples. Hilbert's ϵ underlies Isabelle's original treatment of quantifiers, but the lifting rule is logically simpler.

The meta-logic is used in the latest version of the theorem prover Isabelle. It extends the logic used in earlier versions. Compared with other meta-logics, higher-order logic has a weaker type system but seems easier to implement.

UCAM-CL-TR-114

Stephen Ades:

An architecture for integrated services on the local area network

September 1987, 166 pages, PDF
PhD thesis (Trinity College, January 1987)

Abstract: This dissertation concerns the provision of integrated services in a local area context, e.g. on business premises. The term integrated services can be understood at several levels. At the lowest, one network may be used to carry traffic of several media—voice, data, images etc. Above that, the telephone exchange may be replaced by a more versatile switching system, incorporating facilities such as stored voice messages. Its facilities may be accessible to the user through the interface of the workstation rather than a telephone. At a higher level still, new services such as multi-media document manipulation may be added to the capabilities of a workstation.

Most of the work to date has been at the lowest of these levels, under the auspices of the Integrated Services Digital Network (ISDN), which mainly concerns wide area communications systems. The thesis presented here is that all of the above levels are important in a local area context. In an office environment,

sophisticated data processing facilities in a workstation can usefully be combined with highly available telecommunications facilities such as the telephone, to offer the user new services which make the working day more pleasant and productive. That these facilities should be provided across one integrated network, rather than by several parallel single medium networks is an important organisational convenience to the system builder.

The work described in this dissertation is relevant principally in a local area context—in the wide area economics and traffic balance dictate that the emphasis will be on only the network level of integration for some time now. The work can be split into three parts:

i) the use of a packet network to carry mixed media. This has entailed design of packet voice protocols which produce delays low enough for the network to interwork with national telephone networks. The system has also been designed for minimal cost per telephone—packet-switched telephone systems have traditionally been more expensive than circuit-switched types. The network used as a foundation for this work has been the Cambridge Fast Ring.

ii) use of techniques well established in distributed computing systems to build an 'integrated services PABX (Private Automatic Branch Exchange)'. Current PABX designs have a very short life expectancy and an alarmingly high proportion of their costs is due to software. The ideas presented here can help with both of these problems, produce an extensible system and provide a basis for new multi-media services.

iii) development of new user level Integrated Services. Work has been done in three areas. The first is multi-media documents. A voice editing interface is described along with the system structure required to support it. Secondly a workstation display has been built to support a variety of services based upon image manipulation and transmission. Finally techniques have been demonstrated by which a better interface to telephony functions can be provided to the user, using methods of control typical of workstation interfaces.

UCAM-CL-TR-115

I.S. Dhingra:

Formal validation of an integrated circuit design style

August 1987, 29 pages, PDF

Abstract: In dynamic circuit design many rules must be followed which govern the correctness of the design. In this paper a dynamic CMOS design style using a two phase non-overlapping clock with its intricate design rules is presented together with formal means of showing that a circuit follows these rules.

UCAM-CL-TR-116

Thierry Coquand, Carl Gunter, Glynn Winskel:

Domain theoretic models of polymorphism

September 1987, 52 pages, PDF

Abstract: The main point of this paper is to give an illustration of a construction useful in producing and describing models of Girard and Reynolds' polymorphic λ-calculus. The key unifying ideas are that of a Grothendieck fibration and the category of continuous sections associated with it, constructions used in indexed category theory; the universal types of the calculus are interpreted as the category of continuous sections of the fibration. As a major example a new model for the polymorphic λ-calculus is presented. In it a type is interpreted as a Scott domain. The way of understanding universal types of the polymorphic λ-calculus as categories of continuous sections appears to be useful generally, and, as well as applying to the new model introduced here, also applies, for instance, to the retract models of McCracken and Scott, and a recent model of Girard. It is hoped that by pin-pointing a key construction this paper will help towards a deeper understanding of the models for the polymorphic λ-calculus and the relations between them.

UCAM-CL-TR-117

J.M. Bacon, K.G. Hamilton:

Distributed computing with RPC: the Cambridge approach

October 1987, 15 pages, PDF

Abstract: The Cambridge Distributed Computing System (CDCS) is described and its evolution outlined. The Mayflower project allowed CDCS infrastructure, services and applications to be programmed in a high level, object oriented, language, Concurrent CLU. The Concurrent CLU RPC facility is described in detail. It is a non-transparent, type checked, type safe system which employs dynamic binding and passes objects of arbitrary graph structure. Recent extensions accomodate a number of languages and transport protocols. A comparison with other RPC schemes is given.

UCAM-CL-TR-118

B.K. Boguraev, K. Spärck Jones:

Material concerning a study of cases

May 1987, 31 pages, PDF

Abstract: This note describes and illustrates a study of deep cases using a large sample of sentences. We have used a language analyser which builds meaning representations expressing semantic case roles; specifically Boguraev's (1979) analyser, which builds dependency trees with word senses defined by semantic category primitive formulae, and with case labels, i.e. semantic relation primitives. This note highlights the importance of the source material for those interested in case-based representations of sentence meaning, and indicates the potential utility of the study results.

UCAM-CL-TR-119

Robert Cooper:

Pilgrim: a debugger for distributed systems

July 1987, 19 pages, PDF

Abstract: Pilgrim is a source level debugger for concurrent CLU programs which execute in a distributed environment. It integrates conventional debugging facilities with features for debugging remote procedure calls and critical region based process interactions. Pilgrim is unusual in that it functions on programs in the target environment under conditions of actual use. This has caused a trade-off between providing rich and detailed information to the programmer and avoiding any unwanted alteration to the computation being debugged. Another complication is debugging one client of a network server while avoiding interference with the server's other clients. A successful methodology for this case requires assistance from the server itself.

UCAM-CL-TR-120

D. Wheeler:

Block encryption

November 1987, 4 pages, PDF

Abstract: A fast and simple way of encrypting computer data is needed. The UNIX crypt is a good way of doing this although the method is not cryptographically sound for text. The method suggested here is applied to larger blocks than the DES method which uses 64 bit blocks, so that the speed of encyphering is reasonable. The algorithm is designed for software rather than hardware. This forgoes two advantages of the crypt algorithm, namely that each character can be encoded and decoded independently of other characters and that the identical process is used both for encryption and decryption. However this method is better for coding blocks directly.

UCAM-CL-TR-121

Jonathan Billington:

A high-level petri net specification of the Cambridge Fast Ring M-access service

December 1987, 31 pages, PDF

Abstract: Numerical Petri Nets (a high level inhibitor net) are used to characterise the Cambridge Fast Ring Hardware at a high level of abstraction. The NPN model describes the service provided to users of the hardware (stations, monitors, bridges and ring transmission plant), known as the M-Access service definition in order to remove ambiguities and as a basis for the development and verification of the protocols using the M-Access service.

UCAM-CL-TR-122

John Herbert:

Temporal abstraction of digital designs

February 1988, 34 pages, PDF

Abstract: Formal techniques have been used to verify the function of reasonably large digital devices ([Hunt85], [Cohn87]), and also to describe and reason about digital signal behaviour at a detailed timing level [Hanna85], [Herbert86]. Different models are used: simple synchronous models of components are the basis for verifying high-level functional specifications; more detailed models which capture the behaviour of signals in real time are the basis of proofs about timing. The procedure called temporal abstraction is a technique for formally relating these two behavioural models.

The background to temporal abstraction is presented and the details of its implementation in HOL. The HOL language ([Gordon85a]) is a computerised version of higher-order logic which has an associated proof assistant also called HOL. In HOL one may specify behaviour at both the functional and timing levels. This work describes how the relationship between these levels may also be described in HOL and reasoned about using the HOL system.

The formal transformation of descriptions of behaviour at the timing level to behaviour at the functional level involves generating and verifying timing constraints. This process can be identified with the conventional design activity of timing analysis. This work shows that timing verification can be viewed, not as a separate phase of design, but as part of a single verification process which encompasses functional and timing verification. A single formal language, HOL, is used to describe all aspects of the behaviour and a single verification system provides all the proofs of correctness. The use of uniform, formal techniques is shown to have a number of advantages.

UCAM-CL-TR-123

John Herbert:

Case study of the Cambridge Fast Ring ECL chip using HOL

February 1988, 38 pages, PDF

Abstract: This article describes the formal specification and verification of an integrated circuit which is part of a local area network interface. A single formal language is used to describe the structure and behaviour at all levels in the design hierarchy, and an associated proof assistant is used to generate all formal proofs. The implementation of the circuit, described as a structure of gates and flip-flops, is verified via a number of levels with respect to a high-level formal specification of required behaviour. The high-level formal specification is shown to be close to precise natural language description of the circuit behaviour.

The specification language used, HOL [Gordon85a], has the advantage of permitting partial specifications. It turns out that partial specification has an important effect on the specification and verification methodology and this is presented. We have also evaluated aspects of conventional design, such as techniques for locating errors and the use of simulation, within the case study of formal methods. We assert that proof strategies must assist error location and that simulation has a role alongside formal verification.

UCAM-CL-TR-124

John Herbert:

Formal verification of basic memory devices

February 1988, 46 pages, PDF

Abstract: Formal methods have been used recently to verify high-level functional specifications of digital systems. Such formal proofs have used simple models of circuit components. In this article we describe complementary work which uses a more detailed model of components and demonstrates how hardware can be specified and verified at this level.

In this model all circuits can be described as structures of gates, each gate having an independent propagation delay. The behaviour of digital signals in real time is captured closely. The function and timing of asynchronous and synchronous memory elements implemented using gates is derived. Formal proofs of correctness show that, subject to certain constraints on gate delays and signal timing parameters, these devices act as memory elements and exhibit certain timing properties.

All the proofs have been mechanically generated using Gordon's HOL system.

UCAM-CL-TR-125

Juanito Camilleri:

An operational semantics for Occam

February 1988, 24 pages, PDF

Abstract: Occam is a programming language designed to support concurrent applications, especially those implemented on networks of communicating processors. The aim of this paper is to formulate the meaning of the language constructs of Occam by semantic definitions which are intended as a direct formalisation of the natural language descriptions usually found in programming language manuals [Inmos 3]. This is done by defining a syntax directed transition system where the transitions associated to a phrase are a function of the transitions associated to its components. This method is by no means novel. The concepts here were introduced in [Plotkin 8] and are applied in [Plotkin 9] where an operational semantics for CSP [Hoare 2] was presented. The operational semantics for a subset of Ada is defined in [Li 6], where tasking and exception handling are modelled. For simplicity only a subset of Occam is defined. Timing, priority, replicators and BYTE subscription are omitted. Other features of Occam which deal with the associated components of an Occam program with a set of physical resources (i.e. configurations) are also omitted since they do not effect the semantic interpretation of a program.

UCAM-CL-TR-126

M.E. Leeser:

Reasoning about the function and timing of integrated circuits with Prolog and temporal logic

February 1988, 50 pages, PDF

Abstract: This article describes the application of formal methods to transistor level descriptions of circuits. Formal hardware verification uses techniques based on mathematical logic to formally prove that a circuit correctly implements its behavioral specification. In the approach described here, the structure of circuits and their functional behavior are described with Interval Temporal Logic. These specifications are expressed in Prolog, and the logical manipulations of the proof process are achieved with a Prolog system. To demonstrate the approach, the bahavior of several example circuits is derived from the behavior of their components down to the transistor level. These examples include a dynamic latch which uses a 2-phase clocking scheme and exploits charge storage. Timing as well as functional aspects of behavior are derived, and constraints on the way a circuit interacts with its environment are reasoned about formally.

UCAM-CL-TR-127

John Carroll, Bran Boguraev, Claire Grover, Ted Briscoe:

A development environment for large natural language grammars

February 1988, 44 pages, PDF

Abstract: The Grammar Development Environment (GDE) is a powerful software tool designed to help a linguist or grammarian experiment with and develop large natural language grammars. (It is also, however, being used to help teach students on courses in computational linguistics). This report describes the grammatical formalism employed by the GDE, and contains detailed instructions on how to use the system.

Source code for a Common Lisp version of the software is available from the University of Edinburgh Artificial Intelligence Applications Institute.

UCAM-CL-TR-128

Robert Charles Beaumont Cooper:

Debugging concurrent and distributed programs

February 1988, 111 pages, PDF
PhD thesis (Churchill College, December 1987)

Abstract: This thesis aims to make one aspect of distributed programming easier: debugging. The principles for designing and implementing an interactive debugger for concurrent and distributed programs are presented. These programs are written in a high-level language with type-checked remote procedure calls. They execute on the nodes of a local computer network and interact with the other programs and services which exist on such a network.

The emphasis is on debugging programs in the environment in which they will eventually operate, rather than some simulated environment oriented specifically to the needs of debugging. Thus the debugging facilities impose a low overhead on the program and may be activated at any time.

Ideally the actions of the debugger should be transparent to the execution of the program being debugged. The difficult problem of avoiding any alteration to the relative ordering of inter-process events is examined in detail. A method of breakpointing a distributed computation is presented which achieves a high degree of transparency in the face of arbitary process interactions through shared memory.

The problems of debugging programs that interact with network services, which are shared concurrently with other users of the distributed environment, are examined. A range of debugging techniques, some of which are directly supported by the debugger, are discussed.

A set of facilities for debugging remote procedure calls is presented, and the functions required of the operating system kernel and runtime system to support debugging are also discussed. A distributed debugger is itself an example of a distributed program and so issues such as functional distribution and authentication are addressed.

These ideas have been implemented in Pilgrim, a debugger for Concurrent CLU programs running under the Mayflower supervisor within the Cambridge Distributed Computing System.

UCAM-CL-TR-129

Jeremy Peter Bennett:

A methodology for automated design of computer instruction sets

March 1988, 147 pages, PDF
PhD thesis (Emmanuel College, January 1987)

Abstract: With semiconductor technology providing scope for increasingly complex computer architectures, there is a need more than ever to rationalise the methodology behind computer design. In the 1970's, byte stream architectures offered a rationalisation of computer design well suited to microcoded hardware. In the 1980's, RISC technology has emerged to simplify computer design and permit full advantage to be taken of very large scale integration. However, such approaches achieve their aims by simplifying the problem to a level where it is within the comprehension of a simple human being. Such an effort is not sufficient. There is a need to provide a methodology that takes the burden of design detail away from the human designer, leaving him free to cope with the underlying principles involved.

In this dissertation I present a methodology for the design of computer instruction sets that is capable of automation in large part, removing the drudgery of individual instruction selection. The methodology does not remove the need for the designer's skill, but rather allows precise refinement of his ideas to obtain an optimal instruction set.

In developing this methodology a number of pieces of software have been designed and implemented. Compilers have been written to generate trial instruction sets. An instruction set generator program has been written and the instruction set it proposes evaluated. Finally a prototype language for instruction set design has been devised and implemented.

UCAM-CL-TR-130

Lawrence C Paulson:

The foundation of a generic theorem prover

March 1988, 44 pages, PDF, DVI
This paper is a revised version of UCAM-CL-TR-113.

Abstract: Isabelle is an interactive theorem prover that supports a variety of logics. It represents rules as propositions (not as functions) and builds proofs by combining rules. These operations constitute a meta-logic (or 'logical framework') in which the object-logics are formalized. Isabelle is now based on higher-order logic – a precise and well-understood foundation.

Examples illustrate use of this meta-logic to formalize logics and proofs. Axioms for first-order logic are shown sound and complete. Backwards proof is formalized by meta-reasoning about object-level entailment.

Higher-order logic has several practical advantages over other meta-logics. Many proof techniques are known, such as Huet's higher-order unification procedure.

UCAM-CL-TR-131

Karen Spärck Jones:

Architecture problems in the construction of expert systems for document retrieval

December 1986, 28 pages, PDF

Abstract: The idea of an expert system front end offering the user effective direct accessto a document retrieval system is an attractive one. The paper considers two specific approaches to the construction of such an expert interface, Belkin and Brooks and their colleagues' treatment of the functions of such a front end based on the analysis of human intermediaries, and Pollitt's experimental implementation of a query formulator for searching Cancerline. The distributed expert system model proposed by Belkin and Brooks is a plausable one, and Pollitt's system can be regarded as a first step towards it. But there are major problems about this type of architecture, and the paper argues in particular that in seeking to develop more powerful front ends of the kind envisaged there is one important issue, the nature of the language used for communication between the contributing experts, that requires for attention than it has hitherto received.

UCAM-CL-TR-132

Miriam Ellen Leeser:

Reasoning about the function and timing of integrated circuits with Prolog and temporal logic

April 1988, 151 pages, PDF
PhD thesis (Queens' College, December 1987)

Abstract: The structure of circuits is specified with Prolog; their function and timing behaviour is specified with interval temporal logic. These structural and behavioural specifications are used to formally verify the functionality of circut elements as well as their timing characteristics. A circuit is verified by deriving its behaviour from the behaviour of its components. The derived results can be abstracted to functional descriptions with timing constraints. The functional descriptions can then be used in proofs of more complex hardware circuits.

Verification is done hierarchically, with transistors as primitive elements. Transistors are modelled as switch-level devices with delay. In order to model delay, the direction of signal flow through the transistor must be assigned. This is done automatically by a set of Prolog routines which also determine the inputs and outputs of each circuit component.

Interval temporal logic descriptions are expressed in Prolog and manipulated using PALM: Prolog Assistant for Logic Manipulation. With PALM the user specifies rewrite rules and uses these rules to manipulate logical terms. In the case of reasoning about circuits, PALM is used to manipulate the temporal logic descriptions of the components to derive a temporal logic description of the circuit.

These techniques are demonstrated by applying them to several commonly used complementary metal oxide semiconductor (CMOS) structures. Examples include a fully complementary dynamic latch and a 1-bit adder. Both these circuits are implemented with transistors and exploit 2-phase clocking and charge sharing. The 1-bit adder is a sophisticated full adder implemented with a dynamic CMOS design style. The derived timing and functional behaviour of the 1-bit adder is abstracted to a purely functional behavior which can be used to derive the behaviour of an arbitrary n-bit adder.

UCAM-CL-TR-133

Lawrence C. Paulson:

A preliminary users manual for Isabelle

May 1988, 81 pages, PDF, DVI

Abstract: This is an early report on the theorem prover Isabelle and several of its object-logics. It describes Isabelle's operations, commands, data structures, and organization. This information is fairly low-level, but could benefit Isabelle users and implementors of other systems.

UCAM-CL-TR-134

Avra Cohn:

Correctness properties of the Viper black model: the second level

May 1988, 114 pages, PDF

Abstract: Viper [8,9,10,11,22] is a microprocessor designed by J. Cullyer, C. Pygott and J. Kershaw at the Royal Signals and Radar Establishment in Malvern (RSRE), and is now commericially available. Viper is intended for use is safety-critical applications such as aviation and nuclear power plant control. To this end, Viper has a particularly simple design about which it is relatively easy to reason using current techniques and models. The designers at RSRE, who deserve much credit for the promotion of formal methods, intended from the start that Viper be formally verified. This report describes the partially completed correctness proof, in the HOL system, of the Viper 'block model' with respect to Viper's top level functional specification. The (fully completed) correctness proof of the Viper 'major state' model has already been reported in [5]. This paper describes the analysis of the block model in some detail (in sections 6 to 9), so is necessarily rather long. A less detailed account is to appear in future [6]. Section 2 is a discussion of the scope and limits of the word 'verification', and cautions against careless use of the term. The paper includes a very brief introduction to HOL (section 4), but does not attempt a description or rationalization of Viper's design. The possible uses of the paper are as follows:

It includes enough detail to support an attempt to repeat the proof in HOL, or possibly in other theorem-proving systems.

It serves as a guide for future analyses of Viper;

It completes the existing Viper documentation;

It covers some general issues in hardware verification;

It illustrates the probelms in managing large HOL proofs.

UCAM-CL-TR-135

Thomas F. Melham:

Using recursive types to reason about hardware in higher order logic

May 1988, 30 pages, PDF

Abstract: The expressive power of higher order logic makes it possible to define a wide variety of data types within the logic and to prove theorems that state the properties of these types concisely and abstractly. This paper describes how such defined data types can be used to support formal reasoning in higher order logic about the behaviour of hardware designs.

UCAM-CL-TR-136

Jeffrey J. Joyce:

Formal specification and verification of asynchronous processes in higher-order logic

June 1988, 45 pages, PDF

Abstract: We model the interaction of a synchronous process with an asynchronous memory process using a four-phase "handshaking" protocol. This example demonstrates the use of higher-order logic to reason about the behaviour of synchronous systems such as microprocessors which communicate requests to asynchronous devices and then wait for unpredictably long periods until these requests are answered. We also describe how our model could be revised to include some of the detailed timing requirements found in real systems such as the M68000 microprocessor. One enhancement uses non-determinism to model minimum setup times for asynchronous inputs. Experience with this example suggests that higher-order logic may also be a suitable formalism for reasoning about more abstract forms of concurrency.

UCAM-CL-TR-137

F.V. Hasle:

Mass terms and plurals: from linguistic theory to natural language processing

June 1988, 171 pages, PDF

Abstract: Two linguistic theories within the tradition of formal semantics are investigated. One is concerned with mass terms, and the other with plurals.

Special attention is paid to the possibility of implementing the theories on a computer. With this goal in mind their basic ideas are examined, and the linguistic implications are discussed. In the process, various features of the theories are made formally precise. This leads to two formal systems, one for representing the meanings of sentences with mass terms, and another similar one for plurals. The systems are specified by machine-executable translation relations from fragments of natural language into logical representations.

The underlying model-theoretic semantics of each theory is partially axiomatised. From the axiomatisations all of the paradigmatic inferences of each theory can be proved in a purely deductive manner. This is demanstrated by a number of rigorous proofs of natural language inferences.

Finally some methodological issues are raised. Both theories recommend a particular approach within formal semantics for natural language. I explore the methodological views underlying the theories, and discuss whether the authors actually follow the methods which they recommend.

UCAM-CL-TR-138

Michael Burrows, Martín Abadi, Roger Needham:

Authentication: a practical study in belief and action

June 1988, 19 pages, PDF

Abstract: Questions of belief and action are essential in the analysis of protocols for the authentication of principals in distributed computing systems. In this paper we motivate, set out and exemplify a logic specifically designed for this analysis; we show how protocols differ subtly with respect to the required initial assumptions of the participants and their final beliefs. Our fomalism has enabled us to isolate and express these differences in a way that was not previously possible, and it has drawn attention to features of the protocols of which we were perviously unaware. The reasoning about particular protocols has been mechanically verified.

This paper starts with an informal account of the problem, goes on to explain the formalism to be used, and gives examples of its application to real protocols from the literature. The final sections deal with a formal semantics of the logic and conclusions.

UCAM-CL-TR-139

Paul R. Manson:

Petri net theory: a survey

June 1988, 77 pages, PDF

Abstract: The intense interest in concurrent (or "parallel") computation over the past decade has given rise to a large number of languages for concurrent programming, representing many conflicting views of concurrency.

The discovery that concurrent programming is significantly more difficult than sequential programming has prompted considerable research into determining a tractable and flexible theory of concurrency, with the aim of making concurrent processing more accessible, and indeed the wide variety of concurrent languages merely reflects the many different models of concurrency which have also been developed.

This report, therefore introduces Petri nets, discussing their behaviour, interpretation and relationship to other models of concurrency. It defines and discusses several restrictions and extensions of the Petri net model, showing how they relate to basic Petri nets, while explaining why they have been of historical importance. Finally it presents a survey of the analysis methods applied to Petri nets in general and for some of the net models introduced here.

UCAM-CL-TR-140

Albert John Camilleri:

Executing behavioural definitions in higher order logic

July 1988, 183 pages, PDF
PhD thesis (Darwin College, February 1988)

Abstract: Over the past few years, computer scientists have been using formal verification techniques to show the correctness of digital systems. The verification process, however, is complicated and expensive. Even proofs of simple circuits can involve thousands of logical steps. Often it can be extremely difficult to find correct device specifications and it is desirable that one sets off to prove a correct specification from the start, rather than repeatedly backtrack from the verification process to modify the original definitions after discovering they were incorrect or inadequate.

The main idea presented in the thesis is to amalgamate the techniques of simulation and verification, rather than have the latter replace the former. The result is that behavioural definitions can be simulated until one is reasonably sure that the specification is correct. Furthermore, proving the correctness with respect to these simulated specifications avoids the inadequacies of simulation where it may not be computationally feasible to demonstrate correctness by exhaustive testing. Simulation here has a different purpose: to get specifications correct as early as possible in the verification process. Its purpose is not to demonstrate the correctness of the implementation – this is done in the verification stage when the very same specifications that were simulated are proved correct.

The thesis discusses the implementation of an executable subset of the HOL logic, the version of Higher Order Logic embedded in the HOL theorem prover. It is shown that hardware can be effectively described using both relations and functions; relations being suitable for abstract specification and functions being suitable for execution. The difference between relational and functional specifications are discussed and illustrated by the verification of an n-bit adder. Techniques for executing functional specifications are presented and various optimisation strategies are shown which make the

execution of the logic efficient. It is further shown that the process of generating optimised functional definitions from relational definitions can be automated. Example simulations of three hardware devices (a factorial machine, a small computer and a communications chip) are presented.

UCAM-CL-TR-141

Roy Want:

Reliable management of voice in a distributed system

July 1988, 127 pages, PDF
PhD thesis (Churchill College, December 1987)

Abstract: The ubiquitous personal computer has found its way into most office environments. As a result, widespread use of the Local Area Network (LAN) for the purposes of sharing distributed computing resources has become common. Another technology, the Private Automatic Branch Exchange (PABX), has benefited from large research and development by the telephone companies. As a consequence, it is cost effective and has widely infiltrated the office world. Its primary purpose is to switch digitised voice but, with the growing need for communication between computers it is also being adapted to switch data. However, PABXs are generally designed around a centralised switch in which bandwidth is permanently divided between its subscribers. Computing requirements need much larger bandwidths and the ability to connect to several services at once, thus making the conventional PABX unsuitable for this application.

Some LAN technologies are suitable for switching voice and data. The additional requirement for voice is that point to point delay for network packets should have a low upper-bound. The 10 Mb/s Cambridge Ring is an example of this type of network, but is relatively low bandwidth gives it limited application in this area. Networks with larger bandwidths (up to 100 Mb/s) are now becoming available comercially and could support a realistic population of clients requiring voice and data communication.

Transporting voice and data in the same network has two main advantages. Firstly, from a practical point of view, wiring is minimised. Secondly, applications which integrate both media are made possible, and hence digitised voice may be controlled by client programs in new and interesting ways.

In addition to the new applications, the original telephony facilities must also be available. They should, at least by default, appear to work in an identical way to our tried and trusted impression of a telephone. However, the control and management of a network telephone is now in the domain of distributed computing. The voice connections between telephones are virtual circuits. Control and data information can be freely mixed with voice at a network interface. The new problems that result are the management issues related to the distributed control of real-time media.

This thesis describes the issues as a distributed computing problem and proposes solutions, many of which have been demonstrated in a real implementation. Particular attention has been paid to the quality of service provided by the solutions. This amounts to the design of helpful operator interfaces, flexible schemes for the control of voice from personal workstations and, in particular, a high reliability factor for the backbone telephony service. This work demonstrates the advantages and the practicality of integrating voice and data services within the Local Area Network.

UCAM-CL-TR-142

Peter Newman:

A fast packet switch for the integrated services backbone network

July 1988, 24 pages, PDF

Abstract: With the projected growth in demand for bandwidth and telecommunications services, will come the reguirement for a multi-service backbone network of far greater efficiency, capacity and flexibility than the ISDN is able to satisfy. This class of network has been termed the Broadband ISDN, and the design of the switching node of such a network is the subject of much current research. This paper investigates one possible solution. The design and performance, for multi-service traffic, is presented of a fast packet switch based upon a non-buffered, multi-stage interconnection network. It is shown that for an implementation in current CMOS technology, operating at 50 MHz, switches with a total traffic capacity of up to 150 Gbit/sec may be constructed. Furthermore, if the reserved service traffic load is limited on each input port to a maximum of 80% of switch port saturation, then a maximum delay across the switch of the order of 100 μsecs may be guaranteed, for 99% of the reserved service traffic, regardless of the unreserved service traffic load.

UCAM-CL-TR-143

Lawrence C. Paulson:

Experience with Isabelle A generic theorem prover

August 1988, 20 pages, PDF, DVI

Abstract: The theorem prover Isabelle is described briefly and informally. Its historical development is traced from Edinburgh LCF to the present day. The main issues are unification, quantifiers, and the representation of inference rules. The Edinburgh Logical

Framework is also described, for a comparison with Isabelle. An appendix presents several Isabelle logics, including set theory and Constructive Type Theory, with examples of theorems.

UCAM-CL-TR-144

Juanito Camilleri:

An operational semantics for occam

August 1988, 27 pages, PDF
This is an extended version of UCAM-CL-TR-125, in which we include the operational semantics of priority alternation.

Abstract: Occam is a programming language designed to support concurrent applications, especially those implemented on networks of communicating processors. The aim of this paper is to formulate the meaning of the language constructs of Occam by semantic definitions which are intended as a direct formalisation of the natural language descriptions usually found in programming language manuals [Inmos 5]. This is done by defining a syntax directed transition system where the transitions associated to a phrase are a function of the transitions associated to its components. This method is by no means novel. The concepts here were introduced in [Plotkin 10] and are applied in [Plotkin 11] where an operational semantics for CSP [Hoare 4] was presented. The operational semantics for a subset of Ada is defined in [Li 6], where tasking and exception handling are modelled. For simplicity only a subset of Occam is defined. Timing, replicators and BYTE subscription are omitted. Other features of Occam which deal with the association of components of an Occam program with a set of physical resources (i.e. configurations) are also omitted since they do not effect the semantic interpretation of a program.

UCAM-CL-TR-145

Michael J.C. Gordon:

Mechanizing programming logics in higher order logic

September 1988, 55 pages, PDF

Abstract: Formal reasoning about computer programs can be based directly on the semantics of the programming language, or done in a special purpose logic like Hoare logic. The advantage of the first approach is that it guarantees that the formal reasoning applies to the language being used (it is well known, for example, that Hoare's assignment axiom fails to hold for most programming languages). The advantage of the second approach is that the proofs can be more direct and natural.

In this paper, an attempt to get the advantages of both approaches is described. The rules of Hoare logic are mechanically derived from the semantics of a simple imperative programming language (using the HOL system). These rules form the basis for a simple program verifier in which verification conditions are generated by LCF-style tactics whose validations use the derived Hoare rules. Because Hoare logic is derived, rather than postulated, it is straightforward to mix semantic and axiomatic reasoning. It is also straightforward to combine the constructs of Hoare logic with other application-specific notations. This is briefly illustrated for various logical constructs, including termination statements, VDM-style 'relational' correctness specifications, weakest precondition statements and dynamic logic formulae.

The theory underlying the work presented here is well known. Our contribution is to propose a way of mechanizing this theory in a way that makes certain practical details work out smoothly.

UCAM-CL-TR-146

Thomas F. Melham:

Automating recursive type definitions in higher order logic

September 1988, 64 pages, PDF

Abstract: The expressive power of higher order logic makes it possible to define a wide variety of types within the logic and to prove theorems that state the properties of these types concisely and abstractly. This paper contains a tutorial introduction to the logical basis for such type definitions. Examples are given of the formal definitions in logic of several simple types. A method is then described for systematically defining any instance of a certain class of commonly-used recursive types. The automation of this method in HOL, an interactive system for generating proofs in higher order logic, is also discussed.

UCAM-CL-TR-147

Jeffrey Joyce:

Formal specification and verification of microprocessor systems

September 1988, 24 pages, PDF

Abstract: This paper describes the use of formal methods to verify a very simple microprocessor. The hierarchical structure of the microprocessor implementation is formally specified in higher-order logic. The behaviour of the microprocessor is then derived from a switch level model of MOS (Metal Oxide Semiconductor) behaviour using inference rules of higher-order

logic with assistance from a mechanical theorem proving system. The complexity of the formal proof is controlled by a multi-level approach based on increasingly abstract views of time and data. While traditional methods such as multi-level simulation may reveal errors or inconsistencies, formal verification can provide greater certainty about the correctness of a design. The main difference with formal verification, and its strength, is that behaviour at one level is formally dervied from lower levels with a precise statement of the conditions under which one level accurately models lower levels.

UCAM-CL-TR-148

Jonathan Billington:

Extending coloured petri nets

September 1988, 82 pages, PDF

Abstract: Jensen's Coloured Petri Nets (CP-nets) are taken as the starting point for the development of a specification technique for complex concurrent systems. To increase its expressive power CP-nets are extended by including capacity and inhibitor functions. A class of extended CP-nets, known as P-nets, is defined that includes the capacity function and the threshold inhibitor extension. The inhibitor extension is defined in a totally symmetrical way to that of the usual pre place map (or incidence function). Thus the inhibitor and pre place maps may be equated by allowing a marking to be purged by a single transition occurrence, useful when specifying the abortion of various procedures. A chapter is devoted to developing the theory and notation for the purging of a place's marking or part of its marking.

Two transformations from P-nets to CP-nets are presented and it is proved that they preserve interleaving behaviour. These are based on the notion of complementary places defined for PT-nets and involve the definition and proof of a new extended complementary place invariant for CP-nets

The graphical form of P-nets, known as a P-Graph, is presented formally and draws upon the theories developed for algebraic specification. Arc inscriptions are multiples of tuples of terms generated by a many-sorted signature. Transition conditions are Boolean expressions derived from the same signature. An interpretation of the P-Graph is given in terms of a corresponding P-net. The work is similar to that of Vautherin but includes the inhibitor and capacity extension and a number of significant differences. in the P-Graph concrete sets are associated with places, rather than sorts and likewise there are concrete initial marking and capacity functions. Vautherin associates equations with transitions rather than the more general Boolean expressions. P-Graphs are useful for specification at a concrete level. Classes of the P-Graph, known as Many-sorted Algebraic Nets and Many-sorted Predicate/Transition nets, are defined and illustrated by a number of examples. An extended place capacity notation is developed to allow for the convenient representation of resource bounds in the graphical form.

Some communications-oriented examples are presented including queues and the Demon Game of international standards fame.

The report concludes with a discussion of future work. In particular, an abstract P-Graph is defined that is very similar to Vautherin's Petri net-like schema, but including the capacity and inhibitor extensions and associating boolean expressions with transitions. This will be useful for more abstract specifications (eg classes of communications protocols) and for their analysis.

It is believed that this is the first coherent and formal presentation of these extensions in the literature.

UCAM-CL-TR-149

Paul Ashley Karger:

Improving security and performance for capability systems

October 1988, 273 pages, PDF, PostScript
PhD thesis (Wolfson College, March 1988)

Abstract: This dissertation examines two major limitations of capability systems: an inability to support security policies that enforce confinement and a reputation for relatively poor performance when compared with non-capability systems.

The dissertation examines why conventional capability systems cannot enforce confinement and proposes a new secure capability architecture, called SCAP, in which confinement can be enforced. SCAP is based on the earlier Cambridge Capability System, CAP. The dissertation shows how a non-discretionary security policy can be implemented on the new architecture, and how the new architecture can also be used to improve traceability of access and revocation of access.

The dissertation also examines how capability systems are vulnerable to discretionary Trojan horse attacks and proposes a defence based on rules built into the command-language interpreter. System-wide garbage collection, commonly used in most capability systems, is examined in the light of the non-discretionary security policies and found to be fundamentally insecure. The dissertation proposes alternative approaches to storage management to provide at least some of the benefits of system-wide garbage collection, but without the accompanying security problems.

Performance of capability systems is improved by two major techniques. First, the doctrine of programming generality is addressed as one major cause of poor performance. Protection domains should be allocated only for genuine security reasons, rather than at every subroutine boundary. Compilers can better enforce modularity and good programming style without adding the expense of security enforcement to every subroutine call. Second, the ideas of reduced instruction set computers (RISC) can be applied to capability

systems to simplify the operations required. The dissertation identifies a minimum set of hardware functions needed to obtain good performance for a capability system. This set is much smaller than previous research had indicated necessary.

A prototype implementation of some of the capability features is described. The prototype was implemented on a re-microprogrammed VAX-11/730 computer. The dissertation examines the performance and software compatibility implications of the new capability architecture, both in the context of conventional computers, such as the VAX, and in the context of RISC processors.

UCAM-CL-TR-150

Albert John Camilleri:

Simulation as an aid to verification using the HOL theorem prover

October 1988, 23 pages, PDF

Abstract: The HOL theorem proving system, developed by Mike Gordon at the University of Cambridge, is a mechanism of higher order logic, primarily intended for conducting formal proofs of digital system designs. In this paper we show that hardware specifications written in HOL logic can be executed to enable simulation as a means of supporting formal proof. Specifications of a small microprocessor are described, showing how HOL logic sentences can be transformed into executable code with minimum risk of introducing inconsistencies. A clean and effective optimisation strategy is recommended to make the executable specifications practical.

UCAM-CL-TR-151

Inderpreet-Singh Dhingra:

Formalising an integrated circuit design style in higher order logic

November 1988, 195 pages, PDF
PhD thesis (King's College, March 1988)

Abstract: If the activities of an integrated circuit designer are examined, we find that rather than keeping track of all the details, he uses simple rules of thumb which have been refined from experience. These rules of thumb are guidelines for deciding which blocks to use and how they are to be connected. This thesis gives a formal foundation, in higher order logic, to the design rules of a dynamic CMOS integrated circuit design style.

Correctness statements for the library of basic elements are fomulated. These statements are based on a small number of definitions which define the behaviour of transistors and capacitors and the necessary axiomisation of the four valued algebra for signals. The correctness statements of large and complex circuits are then derived from the library of previously proved correctness statements, using logical inference rules instead of rules of thumb. For example, one gate from the library can drive another only if its output constraints are satisfied by the input constraints of the gate that it drives. In formalising the design rules, these constraints are captured as predicates and are part of the correctness statements of these gates. So when two gates are to be connected, it is only necessary to check that the predicates match. These ideas are fairly general and widely applicable for formalising the rules of many systems.

A number of worked examples are presented based on these formal techniques. Proofs are presented at various stages of development to show how the correctness statement for a device evolves and how the proof is constructed. In particular it is demonstrated how such formal techniques can help improve and sharpen the final specifications.

As a major case study to test all these techniques, a new design for a gigital phase-locked loop is presented. This has been designed down to the gate level using the above dynamic design style, and has been described and simulated using ELLA. Some of the subcomponents have been formally verified down to the detailed circuit level while others have merely been specified without formal proofs of correctness. An informal proof of correctness of this device is also presented based on the formal specifications of the various submodules.

UCAM-CL-TR-152

Andrew Mark Pullen:

Motion development for computer animation

November 1988, 163 pages, PDF
PhD thesis (Churchill College, August 1987)

Abstract: This thesis examines the problems posed by the use of computers in the production of animated sequences, and suggests possible solutions to some of them. Over the years increasing emphasis has been placed on the quest for visual realism, with the result that other considerations—such as ease of motion specification—have been overshadowed. Many current techniques put animation in the realm of the scientist programmer rather than the animation artist. This thesis in the main ignores image rendering issues but considers instead the motion specification phase of generating an animated sequence.

The thesis begins by examining the traditional hand animation process and asks whether speed or quality benefits can be achieved by automating parts of it. My own experiences in this area are described based on

the design and implementation of KAS, a computer-assisted animation system, which was then used to produce parts of a short animated film for Channel 4 television.

In the light of this experience, other computer animation techniques are considered—beginning with a survey of animation systems developed by other researchers over the years and a discussion of their relative merits. This survey identifies the two techniques in most common use today as being (i) an extension of the keyframing technique used for traditional hand animation, and (ii) a scripting approach, which essentially involves producing a textual description of the desired motion in a specially developed animation language. Both of these methods suffer from serious drawbacks—with keyframing it is difficult to control complex motion, whereas scripting forces artists into a style of working which does not exploit their traditional intuition and experience and is more suited to computer programmers than animators.

In an attempt to overcome these shortcomings, proposals are made for a new style of motion development environment making heavy use of interactive graphics and using computer simulation to guide the motion of the objects being animated. Once suitable simulation rules have been set up, the software becomes capable of dealing with the majority of situations that an object will encounter. This means that the animator need only intervene occasionally in order to steer the animation to the desired conclusion. Two major experiments aimed at determining the usefulness of this idea were conducted: one uses fixed rules in a simple environment (the game of snooker); the other considers a more general realm (cellular automata) and gives the animator the freedom to change the simulation rules at any time.

The conclusion drawn from these experiments is that the proposed method is capable of development to the stage where a powerful tool can be provided for animators to use in a novel but intuitive way—one which exploits their capability as artists and makes only minor demands on them to understand the underlying technology.

UCAM-CL-TR-153

Michael Burrows:

Efficient data sharing

December 1988, 99 pages, PDF
PhD thesis (Churchill College, September 1988)

Abstract: As distributed computing systems become widespread, the sharing of data between people using a large number of computers becomes more important. One of the most popular ways to facilitate this sharing is to provide a common file system, accessible by all the machines on the network. This approach is simple and reasonably effective, but the performance of the system can degrade significantly if the number of machines is increased. By using a hierarchical network, and arranging that machines typically access files stored in the same section of the network it is possible to build very large systems. However, there is still a limit on the number of machines that can share a single file server and a single network effectively.

A good way to decrease network and server load is to cache file data on client machines, so that data need not be fetched from the centralized server each time it is accessed. This technique can improve the performance of a distributed file system and is used in a number of working systems. However, caching brings with it the overhead of maintaining consistency, or cache coherence. That is, each machine in the network must see the same data in its cache, even though one machine may be modifying the data as others are reading it. The problem is to maintain consistency without dramatically increasing the number of messages that must be passed between machines on the network.

Some existing file systems take a probabilistic approach to consistency, some explicitly prevent the activities that can cause inconsistency, while others provide consistency only at the some cost in functionality or performance. In this dissertation, I examine how distributed file systems are typically used, and the degree to which caching might be expected to improve performance. I then describe a new file system that attempts to cache significantly more data than other systems, provides strong consistency guarantees, yet requires few additional messages for cache management.

This new file-system provides fine-grain sharing of a file concurrently open on multiple machines on the network, at the granularity of a single byte. It uses a simple system of multiple-reader, single writer locks held in a centralized server to ensure cache consistency. The problem of maintaining client state in a centralized server are solved by using efficient data structures and crash recovery techniques.

UCAM-CL-TR-154

I.B. Crabtree, R.S. Crouch, D.C. Moffat, N.J. Pirie, S.G. Pulman, G.D. Ritchie, B.A. Tate:

A natural language interface to an intelligent planning system

January 1989, 14 pages, PDF

Abstract: An intelligent planning system is an example of a software aid which, although developed by specialists, is intended to be used by non-programmers for a wide variety of tasks. There is therefore a need for a communication medium which allows the application specialist, and the non-expert user to specify their needs without knowing the details of the system.

This kind of system is one where the 'mice and menus' approach is unlikely to be able to provide a very flexible interface since the range and type of potential queries is not predictable in advance. Clearly, therefore,

some kind of language is a necessity here. The aim of this project is to experiment with the use of English language as the medium of communication. The kind of system we would eventually be able to build would be one where the user could use the planner to organise some external activity, trying out alternative scenarios, and then interact with the system during the execution of the resulting plans, making adjustments where necessary.

UCAM-CL-TR-155

S.G. Pulman, G.J. Russell, G.D. Ritchie, A.W. Black:

Computational morphology of English

January 1989, 15 pages, PDF

Abstract: This paper describes an implemented computer program which uses various kinds of linguistic knowledge to analyse existing or novel word forms in terms of their components. Three main types of knowledge are required (for English): knowledge about spelling or phonological changes consequent upon affixation (notice we are only dealing with isolated word forms); knowledge about the syntactic or semantic properties of affixation (i.e. inflexional and derivational morphology), and knowledge about the properties of the stored base forms of words (which in our case are always themselves words, rather than more abstract entities). These three types of information are stored as data files, represented in exactly the form a linguist might employ. These data files are then compiled by the system to produce a run-time program which will analyse arbitrary word forms presented to it in a way consistent with the original linguistic description.

UCAM-CL-TR-156

Steve Pulman:

Events and VP modifiers

January 1989, 10 pages, PDF

Abstract: This paper concerns the anaysis of adverbial and PP modifiers of VP suggested by Davidson, where verbs are regarded as introducing reference to events, and such modifiers are predicates of these events. Several problems raised by it are described and a solution is presented. The paper then goes on to suggest some extensions of the theory in order to be able to cope with temporal and aspectual modification of VPs.

UCAM-CL-TR-157

Juanito Camilleri:

Introducing a priority operator to CCS

January 1989, 19 pages, PDF

Abstract: In this paper we augment the syntax of CCS by introducing a priority operator. We present a syntax directed operational semantics of the language as a labelled transition system. A new equivalence relation which is based on Milner's strong observational equivalence [11] is defined and proved to be a congruence. We also give some examples which illustrate the use of the operator and emphasise the novelty of the approach used to introduce the notion prior to process algebras.

UCAM-CL-TR-158

Karen Spärck Jones:

Tailoring output to the user: What does user modelling in generation mean?

August 1988, 21 pages, PDF

Abstract: This paper examines the implications for linguistic output generation tailored to the interactive system user, of earlier analyses of the components of user modelling and of the constraints realism imposes on modelling. Using a range of detailed examples it argues that tailoring based only on the actual dialogue and on the decision model required for the system task is quite adequate, and that more ambitious modelling is both dangerous and unnecessary.

UCAM-CL-TR-159

Andrew M. Pitts:

Non-trivial power types can't be subtypes of polymorphic types

January 1989, 12 pages, PostScript

Abstract: This paper establishes a new, limitative relation between the polymorphic lambda calculus and the kind of higher-order type theory which is embodied in the logic of toposes. It is shown that any embedding in a topos of the cartesian closed category of (closed) types of a model of the polymorphic lambda calculus must place the polymorphic types well away from the powertypes $\sigma \to \Omega$ of the topos, in the sense that $\sigma \to \Omega$ is a subtype of a polymorphic type only in the case that σ isempty (and hence $\sigma \to \Omega$ is terminal). As corollaries we obtain strengthenings of Reynold's result on the non-existence of set-theoretic models of polymorphism.

UCAM-CL-TR-160

Andrew Gordon:

PFL+: A Kernal Scheme for Functions I/O

February 1989, 26 pages, PDF

Abstract: In place of the common separation of functional I/O into continuation and stream based schemes, an alternative division between Data Driven and Strictness Driven mechanisms for I/O is proposed. The data driven mechanism determines I/O actions by the Weak Head Normal Form of programs, while strictness driven I/O is based on suspensions – I/O actions are triggered when demand arises for the value of a suspension during normal order reduction. The data driven and strictness driven I/O mechanisms are exemplified by the output list and input list, respectively, in Landin's stream based I/O scheme.

PFL+ is a functional I/O scheme, able to express arbitary I/O actions and both data driven and strictness driven constructs in terms of a small kernel of primitives. PFL+ could be added to any functional language. It is based on Holmström's PFL [5], a parallel functional language with embedded communication and concurrency operators from CCS. PFL+ adds non-strict communication, behaviours with results and primitives to make suspensions.

Examples are given of how PFL+ can dervive from these primitives both stream based I/O and the representation of the file system as a function.

UCAM-CL-TR-161

D.C.J. Matthews:

Papers on Poly/ML

February 1989, 150 pages, PDF

Abstract: Various papers and research documents have been written while the Poly/ML project was underway and now that the project is completed it seemed appropriate to combine them into a single document. Some of these papers have been published, while others were written simply to help clarify thoughts on a particular subject.

Although Poly and ML have different syntax and type-systems their execution models are remarkably similar. A new parser and type-checker had to be written, but the code-generator and optimiser could be shared between Poly and ML. The same run-time system was used. Poly turned out to be an excellent language for the project, and modules based around Poly abstract types were used extensively. The lack of low level operations of a systems programming language was not a problem as, in practice, compilers involve very few such operations.

The papers in this report have been grouped into sections according to their likely audience. The first section describes the Poly/ML system and the extensions for windows and processes. Section two contains various discussion papers about Poly and ML, although some of the ideas covered were never implemented. The third section contains two papers on the persistent storage system and its implementation. The final section covers the implementation of Poly and ML and the runtime system.

UCAM-CL-TR-162

Claire Grover, Ted Briscoe, John Carroll, Bran Boguraev:

The Alvey natural language tools grammar (2nd Release)

April 1989, 90 pages, PDF

Abstract: The ANLT grammar is a wide-coverage syntactic description of English expressed in a computationally-tractable unification based formalism. We describe the formalism and give a detailed account of the analyses adopted for different English syntactic constructions in the current version of the grammar. Appendices provide a complete listing of the grammar, sample lexical entries, and a corpus of parsable sentences. The grammar is fully compatible with the Grammar Development Environment (Technical Report 127) which provides an integrated software environment, supporting automated parsing, generation, and modification of grammars expressed in the formalism described here.

UCAM-CL-TR-163

Ann Copestake, Karen Spärck Jones:

Inference in a natural language front end for databases

February 1989, 87 pages, PDF

Abstract: This report describes the implementation and initial testing of knowledge representation and inference capabilities within a modular database front end designed for transportability.

UCAM-CL-TR-164

Li Gong, David J. Wheeler:

A matrix key distribution system

October 1988, 20 pages, PDF

Abstract: A new key distribution scheme is presented. It is based on the distinctive idea that lets each node have a set of keys of which it shares a distinct subset with every other node. This has the advantage that the numbers of keys that must be distributed and maintained are reduced by a square root factor; moreover, two nodes can start conversation with virtually no delay. Two versions of the scheme are given. Their performance and security analysis shows it is a practical solution to some key distribution problems.

UCAM-CL-TR-165

Peter Newman:

Fast packet switching for integrated services

March 1989, 145 pages, PDF
PhD thesis (Wolfson College, December 1988)

Abstract: As the communications industry continues to expand two current trends are becoming apparent: the desire to support an increasing diversity of communications services (voice, video, image, text, etc.) and the consequent requirement for increased network capacity to handle the expected growth in such multi-service traffic. This dissertation describes the design, performance and implementation of a high capacity switch which uses fast packet switching to offer the integrated support of multi-service traffic. Applications for this switch are considered within the public network, in the emerging metropolitan area network and within local area networks.

The Cambridge Fast Packet Switch is based upon a non-buffered, multi-path, switch fabric with packet buffers situated at the input ports of the switch. This results in a very simple implementation suitable for construction in current gate array technology. A simulation study of the throughput at saturation of the switch is first presented to select the most appropriate switch parameters. Then follows an investigation of the swith performance for multi-service traffic. It is shown, for example, that for an implementation in current CMOS technology, operating at 50 Mhz, switches with a total traffic capacity of up to 150 Gbits/sec may be constructed. Furthermore, if the high priority traffic load is limited on each input port to a maximum of 80% of switch port saturation, then a maximum delay across the switch of the order of 100 μsecs may be guaranteed, for 99% of the high priority traffic, regardless of the lower priority traffic load.

An investigation of the implementation of the switch by the construction of the two fundamental components of the design in 3 μm HCMOS gate arrays is presented and close agreement is demonstrated between performance of the hardware implementation and the simulation model. It is concluded that the most likely area of application of this design is as a high capacity multi-service local area network or in the interconnection of such networks.

UCAM-CL-TR-166

Jean Bacon:

Evolution of operating system structures

March 1989, 28 pages, PDF

Abstract: The development of structuring within operating systems is reviewed and related to the simultaneous evolution of concurrent programming languages. First traditional, multi-user systems are considered and their evolution from monolithic closed systems to general domain structured systems is traced. Hardware support for protected sharing is emphasised for this type of system.

The technology directed trend towards single user workstations requires a different emphasis in system design. The requirement for protection in such systems is less strong than in multi-user systems and, in a single language system, may to some extent be provided by software at compile time rather than hardware at run time. Distributed systems comprising single user workstations and dedicated server machines are considered and the special requirements for efficient implementation of servers are discussed.

The concepts of closed but structured and open system designs are helpful. It is argued that the open approach is most suited to the requirements of single user and distributed systems. Experiences of attempting to implement systems over a closed operating system base are presented.

Progress towards support for heterogeneity in distributed systems, so that interacting components written in a range of languages may interwork and may run on a variety of hardware, is presented.

The benefits of taking an object orientated view for system-level as well as language-level objects and for specification, generation and design of systems are discussed and work in this area is described.

An outline of formal approaches aimed at specification, verification and automatic generation of software is given.

Finally, design issues are summarised and conclusions drawn.

UCAM-CL-TR-167

Jeffrey J. Joyce:

A verified compiler for a verified microprocessor

March 1989, 67 pages, paper copy

UCAM-CL-TR-168

J.M. Bacon, I.M. Leslie, R.M. Needham:

Distributed computing with a processor bank

April 1989, 15 pages, PDF

Abstract: The Cambridge Distributed Computing System (CDCS) was designed some ten years ago and was in everyday use at the Computer Laboratory until December 1988. An overview of the basic design of CDCS is given, an outline of its evolution and a description of the distributed systems research projects that were based on it. Experience has shown that a design based on a processor bank leads to a flexible and extensible distributed system.

UCAM-CL-TR-169

Andrew Franklin Seaborne:

Filing in a heterogeneous network

April 1989, 131 pages, PDF
PhD thesis (Churchill College, July 1987)

Abstract: Heterogeneity is becoming a common feature in local area networks as the variety of equipment that is marketed increases. Each such system will have its own mix of hardware and software but it is still desirable to be able to bring in new applications and machines and integrate them with the existing equipment.

Careful design is required to be able to introduce new subsystems into the network in a manner that reduces the manpower needed. If binary compatibility for application programs is achieved then new software can be introduced without the need for source code alterations. If the design of the computing environment is correctly constructed then the introduction of new hardware will not require alterations or cause disruption to the rest of the system.

There is a reduction in the ability to share information through files being accessible to many people and from many locations when there are a number of filing systems present in the network. Ideally a single filing system spanning all operating systems that exist in the distributed computing environment would give maximum possible sharing.

Any existing file service will provide a set of facilities for the construction of a name space by its client or enforce a predefined naming structure which is not compatible with any other. This thesis describes a single network filing system that has been constructed by separating file storage from file naming. By introducing a directory service to manage the name space, and using file servers only for file storage, the need for each client to be extended to take account of every file service is avoided. A single network file transfer protocol allows the directory service to authenticate each request and allows for the introduction of new equipment with no disruption to the existing system.

UCAM-CL-TR-170

Ursula Martin, Tobias Nipkow:

Ordered rewriting and confluence

May 1989, 18 pages, PDF

Abstract: One of the major problems in term rewriting theory is what to do with an equation which cannot be ordered into a rule. Many solutions have been proposed, including the use of special unification algorithms or of unfailing completion procedures.

If an equation cannot be ordered we can still use any instances of it which can be ordered for rewriting. Thus for example $x * y = y * x$ cannot be ordered, but if a, b are constants with $b * a > a * b$ we may rewrite $b * a \rightarrow a * b$. This idea is used in unfailing completion, and also appears in the Boyer-Moore system. In this paper we define and investigate completeness with respect to this notion of rewriting and show that many familiar systems are complete rewriting systems in this sense. This allows us to decide equality without the use of special unification algorithms. We prove completeness by proving termination and local confluence. We describe a confluence test based on recursive properties of the ordering.

UCAM-CL-TR-171

Jon Fairbairn:

Some types with inclusion properties in $\forall$, $\rightarrow$, μ

June 1989, 10 pages, PDF

Abstract: This paper concerns the $\forall$, $\rightarrow$, μ type system used in the non-strict functional programming language Ponder. While the type system is akin to the types of Second Order Lambda-calculus, the absence of type application makes it possible to construct types with useful inclusion relationships between them.

To illustrate this, the paper contains definitions of a natural numbers type with many definable subtypes, and of a record type with inheritance.

UCAM-CL-TR-172

Julia Rose Galliers:

A theoretical framework for computer models of cooperative dialogue, acknowledging multi-agent conflict

July 1989, 226 pages, PDF

Abstract: This thesis describes a theoretical framework for modelling cooperative dialogue. The linguistic theory is a version of speech act theory adopted from Cohen and Levesque, in which dialogue utterances are generated and interpreted pragmatically in the context of a theory of rational interaction. The latter is expressed as explicitly and formally represented principles of rational agenthood and cooperative interaction. The focus is the development of strategic principles of multi-agent interaction as such a basis for cooperative dialogue. In contrast to the majority of existing work, these acknowledge the positive role of conflict to multi-agent cooperation, and make no assumptions regarding the benevolence and sincerity of agents. The result is a framework wherein agents can resolve conflicts by negotiation. It is a preliminary stage to the future building of computer models of cooperative dialogue for both HCI and DAI, which will therefore be more widely and generally applicable than those currently in existence.

The theory of conflict and cooperation is expressed in the different patterns of mental states which characterise multi-agent conflict, cooperation and indifference as three alternative postural relations. Agents can recognise and potentially create these. Dialogue actions are the strategic tools with which mental states can be manipulated, whilst acknowledging that agents are autonomous over their metal states; they have control over what they acquire and reveal in dialogue. Strategic principles of belief and goal adoption are described in terms of the relationships between autonomous agents' beliefs, goals, preference and interests, and the relation of these to action. Veracity, mendacity, concealing and revealing are defined as properties of acts. The role of all these elements in reasoning about dialogue action and conflict resolution, is tested in analysis of two example dialogues; a record of a real trade union negotiation and an extract from "Othello" by Shakespeare.

UCAM-CL-TR-173

Roger William Stephen Hale:

Programming in temporal logic

July 1989, 182 pages, PDF
PhD thesis (Trinity College, October 1988)

Abstract: The idea of writing computer programs in logic is an attractive one, for such programs may be designed, verified, and implemented using a single formal language. This brings a number of practical benefits:

1. There is no room for ambiguity in the relationship between specification and implementation, and no need to learn a different language for each.

2. It is easy to test out specifications from the earliest stages of development, which avoids attempting to implement or verify an inappropriate design.

3. Computerised tools can be applied directly to transform and verify programs, using the established machinery of mathematical logic.

4. Logic supports hierarchical design, so a large project can be divided into smaller tasks which may be designed and verified independently.

Similar benefits may be bestowed by any formal programming language, but the idea only works if the language suits the intended application. All too often the application is forced to fit the language.

In this dissertation I describe an approach that suits the development of parallel and real-time systems. The approach is based on Tempura, a deterministic programming language developed by Moszkowski from his work on hardware specification using Interval Temporal Logic (ITL). I present the formal semantics of ITL in higher-order logic, and show how programs can be transformed and verified using the HOL theorem prover. Then I show how to represent a number of familiar programming concepts in ITL. First, I show that the language of while-programs can be embedded in temporal logic; and that includes the destructive assignment statement with the usual inertial assumption. An interesting corollary is that a simple sequential program, written in Pascal, say, becomes a logic program in Tempura. More advanced concepts include parallel processes and message passing, as well as real-time phenomena such as timeouts, interrupts and traps. Each idea is experimentally tested on a suitable example, using an interpreter for Tempura. The examples range from matrix multiplication and parallel sorting, to a pipelined parser and a real-time lift-controller.

UCAM-CL-TR-174

James Thomas Woodchurch Clarke:

General theory relating to the implementation of concurrent symbolic computation

August 1989, 113 pages, PDF
PhD thesis (Trinity College, January 1989)

Abstract: The central result of this work is the discovery of a new class of architectures, which I call D-RISC, sharing some characteristics of both dataflow and Von Neumann RISC computers, for concurrent computation. This rests on an original and simple theory which relates the demands of concurrent computation on hardware resources to the fundamental performance constraints of technology. I show that dataflow and Von Neumann architectures have different fundamental hardware constraints to performance, and that therefore and D-RISC architecture, which balances these two constraints, is likely to be optimum for concurrent computation.

The work forms four related sections:

A study of the nature of concurrent symbolic computation and the demands which it makes from any implementation. Two new results emerge from this. A model of computation which will be used extensively

in subsequent sections, and a way of incorporating imperative updates in a functional language, similar but superior to non-deterministic merge, which captures locally sequential updates in a computation with minimum constraint on global concurrency.

The computational model is udes to contrast different policies for localising data near a CPU. A new type of cache is proposed which renames all of its cached addresses in order to reduce CPU word-length.

CPU design is examined and a new class of architectures for concurrent computation, called D-RISCs, are proposed.

The multiple-thread implementation problems encountered in the new architectures are examined. A new analysis of the relationship between scheduling and intermediate store use in a symbolic concurrent computation is presented.

UCAM-CL-TR-175

Lawrence C. Paulson:

A formulation of the simple theory of types (for Isabelle)

August 1989, 32 pages, PDF, DVI

Abstract: Simple type theory is formulated for use with the generic theorem prover Isabelle. This requires explicit type inference rules. There are function, product, and subset types, which may be empty. Descriptions (the eta-operator) introduce the Axiom of Choice. Higher-order logic is obtained through reflection between formulae and terms of type bool. Recursive types and functions can be formally constructed.

Isabelle proof procedures are described. The logic appears suitable for general mathematics as well as computational problems.

UCAM-CL-TR-176

T.J.W. Clarke:

Implementing aggregates in parallel functional languages

August 1989, 13 pages, PDF

Abstract: Many constructions which are difficult to write efficiently in pure functional languages have as underlying semantics an aggregate. An aggregate is a collection of individual elements whose order does not matter, it can thus be constructed functionally using a commutative associative combining operator. Equivalent and more efficient implementations for aggregates exist which are operational. A new construction, the A-thread, an aggregate specified operationally which introduces provably local data indeterminacy, is defined. Operational specification of an aggregate, in which each element is specified by a separate function call, does not necessarily destroy referential transparency in a functional language. Aggregates defined using joins on partial orders allow early termination if an operational implementation is used: Arvind's 'I-structures' and Burton's 'improving values' are examples of this.

UCAM-CL-TR-177

P.A.J. Noel:

Experimenting with Isabelle in ZF Set Theory

September 1989, 40 pages, PDF

Abstract: The theorem prover Isabelle has been used to axiomatise ZF set theory with natural deduction and to prove a number of theorems concerning functions. In particular the axioms and inference rules of four theories have been derived in the form of theorems of set theory. The four theories are:

$\lambda\_\beta\eta$, a form of typed lambda calculus with equality,

O_0, a form of simple type theory,

an intuitionistic first order theory with propositions interpreted as the type of their proofs,

PPλ, the underlying theory of LCF.

Most of the theorems have deen derived using backward proofs, with a small amount of automation.

UCAM-CL-TR-178

Jeffrey J. Joyce:

Totally verified systems: linking verified software to verified hardware

September 1989, 25 pages, PDF

Abstract: We describe exploratory efforts to design and verify a compiler for a formally verified microprocessor as one aspect of the eventual goal of building totally verified systems. Together with a formal proof of correctness for the microprocessor this yields a precise and rigorously established link between the semantics of the source language and the execution of compiled code by the fabricated microchip. We describe in particular: (1) how the limitations of real hardware influenced this proof; and (2) how the general framework provided by higher order logic was used to formalize the compiler correctness problem for a hierarchically structured language.

UCAM-CL-TR-179

Ursula Martin, Tobias Nipkow:

Automating Squiggol

September 1989, 16 pages, PDF

Abstract: The Squiggol style of program development is shown to be readily automated using LP, an equational reasoning theorem prover. Higher-order functions are handled by currying and the introduction of an application operator. We present an automated version of Bird's development of the maximum segment sum algorithm, and a similar treatment of a proof of the binomial theorem.

UCAM-CL-TR-180

Tobias Nipkow:

Formal verification of data type refinement: Theory and practice

September 1989, 31 pages, PDF

Abstract: This paper develops two theories of data abstraction and refinement: one for applicative types, as they are found in functional programming languages, and one for state-based types found in imperative languages. The former are modelled by algebraic structures, the latter by automata. The automaton theoretic model covers not just data types but distributed systems in general. Within each theory two examples of data refinement are presented and formally verified with the theorem prover Isabelle. The examples are an abstract specification and two implementations of a memory system, and a mutual exclusion algorithm.

UCAM-CL-TR-181

Tobias Nipkow:

Proof transformations for equational theories

September 1989, 17 pages, PDF

Abstract: This paper contrasts two kinds of proof systems for equational theories: the standard ones obtained by combining the axioms with the laws of equational logic, and alternative systems designed to yield decision procedures for equational problems.

Two new matching alogorithms for (among other theories) associativity, associativity + commutativity, and associativity + commutativity + identity are presented, the emphasis is not so much on individual theories but on the general method of proof transformation as a tool for showing the equivalence of different proof systems.

After studying proof translations defined by rewriting systems, equivalence tests based on the notion of resolvant theories are used to derive new matching and in some cases unification procedures for a number of equational theories. Finally the combination of resolvant systems is investigated.

UCAM-CL-TR-182

John M. Levine, Lee Fedder:

The theory and implementation of a bidirectional question answering system

October 1989, 27 pages, PDF

Abstract: This paper describes a question answering system which is a limited instance of the general bidirectional architecture suggested by Appelt (1987), The novel features of our approach include the use of a linguistically well-motivated set of functional features; a bidirectional grammar which encodes these features directly; a question answering program which uses the thematic organisation of the user's input to construct a cooperative reply; and a tactical generation component which can be used with Montague semantics.

UCAM-CL-TR-183

Rachel Cardell-Oliver:

The specification and verification of sliding window protocols in higher order logic

October 1989, 25 pages, PDF

Abstract: This report describes the formal specification and verification of a class of sliding window protocols using higher order logic. It is proved that a model for implementations of the protocol locically implies safety and liveness invariants, and that these invariants in turm imply an abstract specification of the protocol. The specification and implementation models used in the proof are based on techniques developed for hardware verification in HOL at Cambridge. This model and proof will eventually be used as the basis for a more general sliding window protocol model.

UCAM-CL-TR-184

David Lawrence Tennenhouse:

Site interconnection and the exchange architecture

October 1989, 225 pages, PDF
PhD thesis (Darwin College, September 1988)

Abstract: The users of a site's telecommunication facilities rely on a colection of devices, transducers and computers, to provide the primary communications interface. In the traditional approach to site interconnection, some of these devices are directly attached to specialised carrier networks. The remaining devices are attached to local networks that are tailored to support communication amongst compatible devices at remote sites. This arrangement does not reap the full benefits of network and service integration: each local network has its own common carrier interfaces; and there is no provision for device independent processing, storage, and forwarding elements.

This dissertation describes a layered approach to site interconnection. Communication between peer sites is supported by the lower layer carrier networks, and associations between upper layer clients are supported by the local network layer. The site interconnection layer, inserted between the lacal and carrier layers, facilitates communication between peer local networks. This layer is composed of independent subsystems that offer the site interconnection service (SI-service) to their upper layer clients. Each SI-subsystem is a funnel through which various device-dependent symbol sets are encoded into a common digital format. The symbol streams of concurrent upper layer associations are multiplexed together for presentation at the shared carrier interfaces. Service integration is achieved through the encoding of many different styles of communication (vioce, video, facsimilie, tile transfer, etc.) into a common symbol set.

The first part of this dissertation develops the connected argument sharing this layered architecture. The second part describes the experimental development and analysis of the exchange architecture, which provides an SI-service that supports Asynchronous Transfer Mode (ATM) communication. The ATM approach is characterized by the use of small packets in conjunction with switch fabrics that provide comprable performance to circuit switching, and permit much greater variability in traffic patterns. The switch fabric of the pilot implementation is based on the Cambridge Fast Ring: the CFR packet structure is the basis of the ATM encoding; and the VLSI ring technology has been used to construct the individual SI-subsystems. In this application, the CFR provides ATM-based switching and multiplexing facilities.

This work is distinguished by its emphasis on site independence and universal access to telecommunication services. The principal contributions to the thesis relate to site interconnection; ATM encodings; out-of-band and non-invasive network management; particular analysis methodologies; and the design of multi-service networks.

UCAM-CL-TR-185

Guo Qiang Zhang:

Logics of domains

December 1989, 250 pages, PDF
PhD thesis (Trinity College, May 1989)

Abstract: This dissertation studies the logical aspects of domains as used in the denotational semantics of programming languages. Frameworks of domain logics are introduced which serve as basic tools for the systematic derivation of proof systems from the denotational semantics of programming languages. The proof systems so derived are guaranteed to agree with the denotational semantics in the sense that the denotation of any program coincides with the set of assertions true of it.

The study focuses on two frameworks for denotational semantics: the SFP domains, and the less standard, but important, category of dI-domains with stable functions.

An extended form of Scott's information systems are introduced to represent SFP objects. They provide better understanding of the structure of finite elements and open sets of domains. These systems generalise to a logic of SFP which uses inequational formulae to axiomatise entailment and non-entailment of open-set assertions. Soundness, completeness, and expressiveness results of the logic are obtained, and possible applications are investigated. A μ-calculus of Scott domains is introduced to extend the expressive power of the assertion language.

Special kinds of open sets called stable neighbourhoods are introduced and shown to determine stable functions in a similar sense to that in which Scott-open sets determine continuous functions. Properties and constructions of the stable neighbourhoods on various categories of dI-domains are investigated. Logical frameworks for Girard's coherent spaces and Berry's dI-domains are given in which assertions are interpreted as stable neighbourhoods. Various soundness, completeness, and expressiveness results are provided.

UCAM-CL-TR-186

Derek Robert McAuley:

Protocol design for high speed networks

January 1990, 100 pages, PostScript
PhD thesis (Fitzwilliam College, September 1989)

Abstract: Improvements in fibre optic communication and in VLSI for network switching components have led to the consideration of building digital switched networks capable of providing point to point communication in the gigabit per second range. Provision of bandwidths of this magnitude allows the consideration of a whole new range of telecommunications services, integrating video, voice, image and text. These multi-service networks have a range of requirements not met by traditional network architectures designed for digital telephony or computer applications. This dissertation describes the design, and an implementation, of the Multi-Service Network architecture and protocol family, which is aimed at supporting these services.

Asynchronous transfer mode networks provide the basic support required for these integrated services, and the Multi-Service Network architecture is designed primarily for these types of networks. The aim of the Multi-Service protocol family is to provide a complete architecture which allows use of the full facilities of asynchronous transfer mode networks by multimedia applications. To maintain comparable performance with the underlying media, certain elements of the MSN protocol stack are designed with implementation in hardware in mind. The interconnection of heterogeneous networks, and networks belonging to different security and administrative domains, is considered vital, so the MSN architecture takes an internetworking approach.

UCAM-CL-TR-187

Ann Copestake, Karen Spärck Jones:

Natural language interfaces to databases

September 1989, 36 pages, PostScript

Abstract: This paper reviews the state of the art in natural language access to databases. This has been a long-standing area of work in natural language processing. But though some commercial systems are now available, providing front ends has proved much harder than was expected, and the necessary limitations on front ends have to be recognised. The paper discusses the issues, both general to language and task-specific, involved in front end design, and the way these have been addressed, concentrating on the work of the last decade. The focus is on the central process of translating a natural language question into a database query, but other supporting functions are also covered. The points are illustrated by the use of a single example application. The paper concludes with an evaluation of the current state, indicating that future progress will depend on the one hand on general advances in natural language processing, and on the other on expanding the capabilities of traditional databases.

UCAM-CL-TR-188

Timothy E. Leonard:

Specification of computer architectures: a survey and annotated bibliography

January 1990, 42 pages, PDF

Abstract: I first define computer architecture and architecture specification, explain how the conflict between clarity and ambiguity makes writing specifications difficult, and introduce and consider the advantages and problems of formal specifications. I then survey all the literature on architecture specification, and introduce the literature on technical writing and on formal specification in general. I close with an annotated bibliography.

UCAM-CL-TR-189

Lawrence C. Paulson, Tobias Nipkow:

Isabelle tutorial and user's manual

January 1990, 142 pages, PDF, DVI

Abstract: This (obsolete!) manual describes how to use the theorem prover Isabelle. For beginners, it explains how to perform simple single-step proofs in the built-in logics. These include first-order logic, a classical sequent calculus, ZF set theory, Constructie Type Theory, and higher-order logic. Each of these logics is described. The manual then explains how to develop advanced tactics and tacticals and how to derive rules. Finally, it describes how to define new logics within Isabelle.

UCAM-CL-TR-190

Ann Copestake:

Some notes on mass terms and plurals

January 1990, 65 pages, PostScript

Abstract: This report describes a short investigation into some possible treatments of mass nouns and plurals. It aims to provide a grammar and axiomatisation with a reasonable coverage of these phenomena, so that a range of sentences can be parsed, and inferences made automatically.

The previous work on the subject, mainly due to Hasle (1988) is reviewed, and the limitations of both the original theories and Hasle's implementation are demonstrated. Some more recent work, especially that relevant to Link's theory, is also discussed.

The present grammar and axiomatisation is described. Although it is not the implementation of any

particular theory, it draws on the work of Link, Krifka and Roberts. Some of the problems with the present approach are discussed, although possible solutions would need to be considered in a wider context. The aim is to show what types of phenomena can be treated by a relatively simple approach.

The implemented grammar covers everything that was treated by Hasle's implementation, and extends that coverage in a varietry of ways, while providing a better integration of the treatment of mass nouns and plurals than the earlier work. It was written in the CFG+ formalism, and some parts of the axiomatisation have been tested using the HOL system.

UCAM-CL-TR-191

Cosmos Nicolaou:

An architecture for real-time multimedia communications systems

February 1990, 30 pages, PDF

Abstract: An architecture for real-time multimedia communications systems is presented. A multimedia communication systems includes both the communication protocols used to transport the real-time data and also the Distributed Computing system (DCS) within which any applications using these protocols must execute. The architecture presented attempts to integrate these protocols with the DCS in a smooth fashion in order to ease the writing of multimedia applications. Two issues are identified as being essential to the success of this integration: namely the synchronisation of related real-time data streams, and the management of heterogeneous multimedia hardware. The synchronisation problem is tackled by defining explicit synchronisation properties at the presentation level and by providing control and synchronisation operations within the DCS which operate in terms of these properties. The heterogeneity problems are addressed by separating the data transport semantics (protocols themselves) from the control semantics (protocol interfaces). The control semantics are implemented using a distributed, typed interface, scheme within the DCS (i.e. above the presentation layer), whilst the protocols themselves are implemented within the communication subsystem. The interface between the DCS and communications subsystem is referred to as the orchestration interface and can be considered to lie in the presentation and session layers.

A conforming prototype implementation is currently under construction.

UCAM-CL-TR-192

Lawrence C. Paulson:

Designing a theorem prover

May 1990, 57 pages, PDF, DVI

Abstract: The methods and principles of theorem prover design are presented through an extended example. Starting with a sequent calculus for first-order logic, an automatic prover (called Folderol) is developed. Folderol can prove quite a few complicated theorems, although its search strategy is crude and limited. Folderol is coded in Standard ML and consists largely of pure functions. Its complete listing is included.

The report concludes with a survey of other research in theorem proving: the Boyer/Moore theorem prover, Automath, LCF, and Isabelle.

UCAM-CL-TR-193

Julia Rose Galliers:

Belief revision and a theory of communication

May 1990, 30 pages, PDF

Abstract: This report concerns choices about changing belief. It describes research to establish and model a principled theoretical basis by which rational agents autonomously choose whether, as well as how to revise their beliefs. Aspects of the various problems in belief revision are discussed, and solved in the context of an AI tool for reason maintenance extended to cover situations of new evidence as not assumed 'truth'. Primarily this results from the inclusion of a non numeric theory of strength of belief, which relates strength to persistence in the context of challenge. Such autonomous belief revision is presented as the basis of a theory of communication, as a special case of reasoning about change in an uncertain world with incomplete information, comprising others similarly constrained.

UCAM-CL-TR-194

Julia Rose Galliers:

Proceedings of the First Belief Representation and Agent Architectures Workshop

March 1990, 199 pages, PDF

Abstract: The first Belief Representation and Agent Architectures workshop was organised by Cambridge University Computer Laboratory, and held at SRI International in Cambridge on the 22nd and 23rd March 1990. It was designed as a closed meeting of 15 researchers, all currently working in and familiar with this subfield of AI. The purpose of the meeting was not so much to present completed work, as to exchange ideas and explore issues with others equally as aware of the relevant problems and background. Each presenter was given 90 minutes in which to lead a discussion on a topic related to their research interests. Generally these were oriented around the presenter's current research

projects, outlines of which had been distributed prior to the meeting.

These proceedings comprise eight sections, each including the discussion report followed by copies of the presenter's overheads, followed by the summaries of the presenter's and rapporteur's current research projects. The sections are as follows: General introduction, different styles of agent architectures, a minimalist approach to agent architectures, models of belief revision, the value of formal approaches, knowledge action chance and utility, different value systems, and channels for dialogue.

UCAM-CL-TR-195

Jeffrey J. Joyce:

Multi-level verification of microprocessor-based systems

May 1990, 163 pages, PDF
PhD thesis (Pembroke College, December 1989)

Abstract: The idea of using formal logic to reason about small fragments or single layers of a software/hardware system is well-established in computer science and computer engineering. Recently, formal logic has been used to establish correctness properties for several realistic systems including a commercially-available microprocessor designed by the British Ministry of Defence for life-critical applications. A challenging area of new research is to verify a complete system by linking correctness results for multiple layers of software and hardware into a chain of logical dependencies.

This dissertation focuses specifically on the use of formal proof and mechanical proof-generation techniques to verify microprocessor-based systems. We have designed and verified a complete system consisting of a simple compiler for a hierarchically structured programming language and a simple microprocessor which executes code generated by this compiler. The main emphasis of our discussion is on the formal verification of the microprocessor. The formal verification of the compiler is described in a separate paper included as an appendix to this dissertation.

Combining correctness results for the compiler with correctness results for the microprocessor yields a precise and rigorously established link between the formal semantics of the programming language and the execution of compiled code by a model of the hardware. The formal proof also links the hardware model to the behavioural specification of an asynchronous memory interface based on a four-phase handshaking protocol.

The main ideas of this research are (1) the use of generic specification to filter out non-essential detail, (2) embedding natural notations from special-purpose formalisations such as temporal logic and denotational description, and (3) the use of higher-order logic as a single unifying framework for reasoning about complete systems.

Generic specification, in addition to supporting fundamental principles of modularity, abstraction and reliable re-usability, provides a mechanism for enforcing a sharp distinction between what has and what has not been formally considered in a proof of corectness. Furthermore, it is possible to create generic specifications in a pure formalism with the expressive power of higher-order logic without inventing new constructs.

Natural notations from special-purpose formalisms offer the advantage of concise and meaningful specifications when applied to particular areas of formal description. Semantic gaps between different notations are avoided by embedding them in a single logic. Special-purpose rules based on these notations can be derived as theorems with the aim of implementing more efficient proof strategies.

Finally it is argued that the primary purpose of using mechanical proof generation techniques to reason about software and hardware is to support the intelligent participation of a human verifier in the rigorous analysis of a design at a level which supports clear thinking.

UCAM-CL-TR-196

John Peter Van Tassell:

The semantics of VHDL with Val and Hol: towards practical verification tools

June 1990, 77 pages, PDF

Abstract: The VHSIC Hardware Description Language (VHDL) is an emerging standard for the design of Application Specific Integrated Circuits. We examine the semantics of the language in the context of the VHDL Annotation Language (VAL) and the Higher Order Logic (HOL) system with the purpose of proposing methods by which VHDL designs may be converted into these two forms for further validation and verification. A translation program that utilizes these methods is described, and several comprehensive VHDL design examples are shown.

UCAM-CL-TR-197

Thomas Clarke:

The semantics and implementation of aggregates or how to express concurrency without destroying determinism

July 1990, 25 pages, PDF

Abstract: This paper investigates the relationship between declarative semantics and concurrent computation. A fundamental programming construction, the aggregate, is identified. Aggregates have a simple declarative semantics, yet cannot be written in pure functional languages. The addition of aggregates to a functional language increases expressiveness without destroying determinism or referential transparency. Specific aggregates can be used to implememnt concurrent graph marking, time deterministic merge of lazy lists, and write once locations.

UCAM-CL-TR-198

Andrew M. Pitts:

Evaluation Logic

August 1990, 31 pages, PostScript

Abstract: A new typed, higher-order logic is described which appears particularly well fitted to reasoning about forms of computation whose operational behaviour can be specified using the Natural Semantics style of structural operational semantics. The logic's underlying type system is Moggi's computational metalanguage, which enforces a distinction between computations and values via the categorical structure of a strong monad. This is extended to a (constructive) predicate logic with modal formulas about evaluation of computations to values, called evaluation modalities. The categorical structure corresponding to this kind of logic is explained and a couple of examples of categorical models given.

As a first example of the naturalness and applicability of this new logic to program semantics, we investigate the translation of a (tiny) fragment of Standard ML into a theory over the logic, which is proved computationally adequate for ML's Natural Semantics. Whilst it is tiny, the ML fragment does however contain both higher-order functional and imperative features, about which the logic allows us to reason without having to mention global states explicitly.

UCAM-CL-TR-199

Richard Boulton, Mike Gordon, John Herbert, John Van Tassel:

The HOL verification of ELLA designs

August 1990, 22 pages, PostScript

Abstract: HOL is a public domain system for generating proofs in higher order predicate calculus. It has been in experimental and commercial use in several countries for a number of years.

ELLA is a hardware design language developed at the Royal Signals and Radar Establishment (RSRE) and marketed by Computer General Electronic Design. It supports simulation models at a variety of different abstraction levels.

A preliminary methodology for reasoning about ELLA designs using HOL is described. Our approach is to semantically embed a subset of the ELLA language in higher order logic, and then to make this embedding convenient to use with parsers and pretty-printers. There are a number of semantic issues that may affect the ease of verification. We discuss some of these briefly. We also give a simple example to illustrate the methodology.

UCAM-CL-TR-200

Tobias Nipkow, Gregor Snelting:

Type classes and overloading resolution via order-sorted unification

August 1990, 16 pages, PDF

Abstract: We present a type inference algorithm for a haskell-like language based on order-sorted unification. The language features polymorphism, overloading, type classes and multiple inheritance. Class and instance declarations give rise to an order-sorted algebra of types. Type inference esentially reduces to the Hindley/Milner algorithm where unification takes place in this order-sorted algebra of types. The theory of order-sorted unification provides simple sufficient conditions which ensure the existence of principal types. The semantics of the language is given by a translation into ordinary λ-calculus. We prove the correctness of our type inference algorithm with respect to this semantics.

UCAM-CL-TR-201

Thomas Frederick Melham:

Formalizing abstraction mechanisms for hardware verification in higher order logic

August 1990, 233 pages, PDF
PhD thesis (Gonville & Caius College, August 1989)

Abstract: Recent advances in microelectronics have given designers of digital hardware the potential to build devices of remarkable size and complexity. Along with this however, it becomes increasingly difficult to ensure that such systems are free from design errors, where complete simulation of even moderately sized circuits is impossible. One solution to these problems is that of hardware verification, where the functional behaviour of the hardware is described mathematically and formal proof is used to show that the design meets rigorous specifications of the intended operation.

This dissertation therefore seeks to develop this, showing how reasoning about the correctness of hardware using formal proof can be achieved using fundamental abstraction mechanisms to relate specifications of hardware at different levels. Therefore a systematic method is described for defining any instance of a wide class of concrete data types in higher order logic. This process has been automated in the HOL theorem prover, and provides a firm logical basis for representing data in formal specifications.

Further, these abstractions have been developed into a new technique for modelling the behaviour of entire classes of hardware designs. This is based on a formal representation in logic for the structure of circuit designs using the recursive types defined by the above method. Two detailed examples are presented showing how this work can be applied in practice.

Finally, some techniques for temporal abstraction are explained, and the means for asserting the correctness of a model containing time-dependent behaviour is described. This work is then illustrated using a case study; the formal verification on HOL of a simple ring communication network.

[Abstract by Nicholas Cutler (librarian), as none was submitted with the report.]

UCAM-CL-TR-202

Andrew Charles Harter:

Three-dimensional integrated circuit layout

August 1990, 179 pages, PDF
PhD thesis (Corpus Christi College, April 1990)

Abstract: Some recent developments in semiconductor process technology have made possible the construction of three-dimensional integrated circuits. Unlike other technological developments in two dimensional integration, these circuits present a new and inherently richer connection topology. This offers potential for improved layout in terms of increased density and reduced interconnect length. These circuits will be difficult and expensive to manufacture, at least in the short term, and the scale of the improvement in layout is not apparent. This dissertation presents a discussion of layout and design for three-dimensional integrated circuits.

A number of materials and techniques can be used in the manufacture of such circuits. This choice has a profound bearing on the topology of circuit layout. A classification relating process technolgy to layout topology is developed and illustrated with the design of a number of circuits. A layout system is presented as the vehicle for a series of experiments in three-dimensional layout. It is shown that the system can be constrained to perform circuit layout in a number of topologies in the classification.

Finally, some attempt to quantify the benefits of three-dimensional layout is made. The layout model is calibrated by designing examples of basic circuit elements. This is done using a set of design rules corresponding to a proposed three-dimensional process technology. Circuit layouts produced by the system are compared with conventional two-dimensional layouts, and the variation in layout quality as a function of the three-dimensionality of a layout is explored.

UCAM-CL-TR-203

Valeria C.V. de Paiva:

Subtyping in Ponder (preliminary report)

August 1990, 35 pages, PDF

Abstract: This note starts the formal study of the type system of the functional language Ponder. Some of the problems of proving soundness and completeness are discussed and some preliminary results, about fragments of the type system, shown.

It consists of 6 sections. In section 1 we review briefly Ponder's syntax and describe its typing system. In section 2 we consider a very restricted fragment of the language for which we can prove soundness of the type inference mechanism, but not completeness. Section 3 describes possible models of this fragment and some related work. Section 4 describes the type-inference algorithm for a larger fragment of Ponder and in section 5 we come up against some problematic examples. Section 6 is a summary of further work.

UCAM-CL-TR-204

Roy L. Crole, Andrew M. Pitts:

New foundations for fixpoint computations: FIX-hyperdoctrines and the FIX-logic

August 1990, 37 pages, PostScript

Abstract: This paper introduces a new higher-order typed constructive predicate logic for fixpoint computations, which exploits the categorical semantics of computations introduced by Moggi and contains a strong version of Martin Löf's 'iteration type'. The type system enforces a separation of computations from values. The logic contains a novel form of fixpoint induction and can express partial and total correctness statements about evaluation of computations to values. The constructive nature of the logic is witnessed by strong metalogical properties which are proved using a category-theoretic version of the 'logical relations' method.

UCAM-CL-TR-205

Lawrence C. Paulson, Andrew W. Smith:

Logic programming, functional programming and inductive definitions

29 pages, PDF, DVI

Abstract: This paper reports an attempt to combine logic and functional programming. It also questions the traditional view that logic programming is a form of first-order logic, arguing instead that the essential nature of a logic program is an inductive definition. This revised view of logic programming suggests the design of a combined logic/functional language. A slow but working prototype is described.

UCAM-CL-TR-206

Rachel Cardell-Oliver:

Formal verification of real-time protocols using higher order logic

August 1990, 36 pages, PDF

Abstract: A protocol is a distributed program which controls communication between machines in a computer network. Two or more programs are executed on different computers which communicate only via the medium connecting them.Protocol implementations are difficult to understand and write correctly because the interaction between programs and their nondeterministic, real-time environment is complex. For this reason protocols are often specified using an abstract model. However few abstract specification techniques model the problems which occur in real implementations. In particular, the correctness of many protocols depends on real-time issues such as the correct setting of timers and fast responses to incoming messages.

This paper presents techniques for modelling real-time protocols at different levels of abstraction, from implementation behaviour to abstract requirements specifications. The language used for these models is higher order logic. The techniques are illustrated by the specification and verification of the class of sliding window protocols. The HOL system, a machine implementation of higher order logic [2], as used to both specify and verify this example and a full listing of the HOL theories for sliding window protocols is given in Appendix B.

UCAM-CL-TR-207

Stuart Philip Hawkins:

Video replay in computer animation

October 1990, 161 pages, PDF
PhD thesis (Queens' College, December 1989)

Abstract: This dissertation presents a design for an animation system that supports video-rate replay of frame sequences within a frame buffer based graphics architecture.

In recent years framebuffer architectures have become dominant, largely displacing other forms of graphics display system. But a framebuffer representation is not well suited to the support of animation. In particular, two main problems are faced: (1) the generation of each new frame within a single frame time (typically 40ms); and (2) the updating of the framebuffer with the new frame representation, also within one frame time. Both of these problems stem from the fact that the large amount of data required to represent each frame has to be processed within a strictly limited time. The difficulty with updating the frame buffer representation has been largely addressed by the development of powerful new display processor architectures, made possible by developments in semiconductor technology. The generation of frames at replay rates, however, represents a much greater challenge and there are numerous situations for which real time animation is simply inpracticable. In such cases an alternative approach is that of frame-by-frame animation in which the frame sequence is pre-calculated off-line and stored for later replay at the correct speed. This technique is commonly referred to as real-time playback.

In this dissertation the requirements of real-time playback are discussed and a number of distinct approaches to the design of such systems identified. For each approach examples of previous real-time playback systems are examined and their individual shortcomings noted. In light of these observations the design of a new hardware-based animation system is proposed and its implementation described. In this system frames are stored digitally and image compression is used to address the non-video-rate transfer rate and storage capacity limitations of the frame storage device employed (an unmodified 5 1/4 inch magnetic disc drive). Such an approach has previously received little attention. Frame sequences are stored on the disc in a compressed form and during replay are decompressed in real-time using a hardware implementation of the coding algorithm. A variety of image compression strategies are supported within a generalised coding framework. This introduces operational flexibility by allowing the system to be tailored according to the needs of a particular application.

UCAM-CL-TR-208

Eike Ritter:

Categorical combinators for the calculus of constructions

October 1990, 43 pages, PDF

Abstract: This report describes the derivation of a small and intuitive set of categorical combinators for the Calculus of Constructions. The choice of an appropriate categorical semantics is the crucial step. A modification of Ehrhard's higher-order closed summable fibrations, yielding so called CC-categories, turns out to be the appropriate underlying categorical structure. Standard techniques can then be used to derive the combinators. The combinators can be turned directly into the classifying category for the Calculus of Constructions. This establishes a precise connection between the calculus, the combinators and the CC-categories.

UCAM-CL-TR-209

Andrew William Moore:

Efficient memory-based learning for robot control

November 1990, 248 pages, PDF
PhD thesis (Trinity Hall, October 1990)

Abstract: This dissertation is about the application of machine learning to robot control. A system which has no initial model of the robot/world dynamics should be able to construct such a model using data received through its sensors—an approach which is formalized here as the SAB (State-Action-Behaviour) control cycle. A method of learning is presented in which all the experiences in the lifetime of the robot are explicitly remembered. The experiences are stored in a manner which permits fast recall of the closest previous experience to any new situation, thus permitting very quick predictions of the effects of proposed actions and, given a goal behaviour, permitting fast generation of a candidate action. The learning can take place in high-dimensional non-linear control spaces with real-valued ranges of variables. Furthermore, the method avoids a number of shortcomings of earlier learning methods in which the controller can become trapped in inadequate performance which does not improve. Also considered is how the system is made resistant to noisy inputs and how it adapts to environmental changes. A well founded mechanism for choosing actions is introduced which solves the experiment/perform dilemma for this domain with adequate computational efficiency, and with fast convergence to the goal behaviour. The dissertation explains in detail how the SAB control cycle can be integrated into both low and high complexity tasks. The methods and algorithms are evaluated with numerous experiments using both real and simulated robot domains. The final experiment also illustrates how a compound learning task can be structured into a hierarchy of simple learning tasks.

UCAM-CL-TR-210

Tobias Nipkow:

Higher-order unification, polymorphism, and subsorts

15 pages, PDF

Abstract: This paper analyses the problems that arise in extending Huet's higher-order unification algorithm from the simply typed λ-calculus to one with type variables. A simple, incomplete, but in practice very useful extension to Huet's algorithm is discussed. This extension takes an abstract view of types. As a particular instance we explore a type system with ML-style polymorphism enriched with a notion of sorts. Sorts are partially ordered and classify types, thus giving rise to an order-sorted algebra of types. Type classes in the functional language Haskell can be understood as sorts in this sense. Sufficient conditions on the sort structure to ensure the existence of principal types are discussed. Finally we suggest a new type system for the λ-calculus which may pave the way to a complete unification algorithm for polymorphic terms.

UCAM-CL-TR-211

Karen Spärck Jones:

The role of artificial intelligence in information retrieval

November 1990, 13 pages, PDF

Abstract: This paper reviews four potential roles for artificial intelligence in information retrieval, evaluating AI from a realistic point of view and within a wide information management context. The conclusion is that AI has limited potential, not just because AI is itself insufficiently developed, but because many information management tasks are properly shallow information processing ones. There is nevertheless an important place for specific applications of AI or AI-derived technology when particular constraints can be placed on the information management tasks involved.

UCAM-CL-TR-212

K.L. Wrench:

A distributed and-or parallel Prolog network

December 1990, 82 pages, PDF

Abstract: A model is proposed for the parallel execution of Prolog, exploiting both dependent and- and full or-parallelism. The model is implemented on a distributed network of loosely-coupled processors and has no need of shared memory nor multiprocessor hardware.

Known as APPNet, the model makes use of oracles to partition the search space dynamically, thereby enabling processing elements to be allocated a unique portion of the computation. No communication takes place between processing elements. In executing problems that do not exhibit any and-parallelism, all solutions found represent final answers to the query. When an and-parallel problem is executed, the solutions generated are only partial solutions. The sets of partial solution are then joined to produce consistent final solutions. Back-unification is the process whereby partial solutions are unified according to a template derived from the program.

Prolog source programs need not be modified by the user. Static analysis is, however, carried out automatically on all programs by a preprocessor before their execution in the APPNet to ensure that clauses are not distributed before it is feasible to do so. Side-effecting constructs are identified and the appropriate restrictions are placed on the parallel execution strategy.

UCAM-CL-TR-213

Valeria Correa Vaz de Paiva:

The Dialectica categories

January 1991, 82 pages, PDF
PhD thesis (Lucy Cavendish College, November 1988)

Abstract: This work consists of two main parts. The first one, which gives it its name, presents an internal categorical version of Gödel's "Dialectica interpretation" of higher-order arithmetic. The idea is to analyse the Dialectica interpretation using a cetegory DC where objects are relations on objects of a basic category C and maps are pairs of maps of C satisfying a pullback condition. If C is finitely complete, DC exists and has a very natural symmetric monoidal structure. If C is locally cartesian closed then DC is symmetric monoidal closed. If we assume C with stable and disjoint coproducts, DC has cartesian products and weak-coproducts and satisfies a weak form of distributivity. Using the structure above, DC is a categorical model for intuitionistic linear logic.

Moreover if C has free monoids then DC has cofree comonoids and the corresponding comonad "!" on DC, which has some special properties, can be used to model the exponential "of course!" in Intuitionistic Linear Logic. The category of "!"-coalgebras is isomorphic to the category of comonoids in DC and, if we assume commutative monoids in C, the "!"-Kleisli category, which is cartesian closed, corresponds to the Diller-Nahm variant of the Dialectica interpretation.

The second part introduces the categories GC. The objects of GC are the same objects of DC, but morphisms are easier to handle, since they are maps in C in opposite directions. If C is finitely complete, the category GC exists. If C is cartesian closed, we can define a symmetric monoidal structure and if C is locally cartesian closed as well, we can define inernal homs in GC that make it a symmetric monoidal closed category. Supposing C with stable and disjoint coproducts, we can define cartesian products and coproducts in GC and, more interesting, we can define a dual operation to the tensor product bifunctor, called "par". The operation "par" is a bifunctor and has a unit "$\perp$", which is a dualising object. Using the internal hom and $\perp$ we define a contravariant functor "$(-)\perp$" which behaves like negation and thus it is used to model linear negation. We show that the category GC, with all the structure above, is a categorical model for Linear Logic, but not exactly the classical one.

In the last chapter a comonad and a monad are defined to model the exponentials "!" and "?". To define these endofunctors, we use Beck's distributive laws in an interesting way. Finally, we show that the Kleisli category GC! is cartesian closed and that the categories DC and GC are related by a Kleisli construction.

UCAM-CL-TR-214

J.A. Bradshaw, R.M. Young:

Integrating knowledge of purpose and knowledge of structure for design evaluation

February 1991, 20 pages, PDF

Abstract: This paper describes a knowledge representation strategy, for mechanical devices, which combines Knowledge of Structure and Knowledge of Purpose. Knowledge of Purpose specifies how devices are expected to behave and Knowledge of Structure details how devices are connected. Knowing 'correct' behaviour (Knowledge of Purpose) it is possible to usefully comment on any generated behaviour, predicted or actual. Generation of behaviour is a bottom up process (from components to systems) whereas behaviour evaluation is top down (from systems to components). Common purpose is used to group devices into systems.

The core evaluation activity is the generation of an envisionment graph (similar to that described by deKleer and Brown [deK84]). The complete graph represents the full set of predicted behaviour states for the represented device. These behaviour states are compared with the Knowledge of Purpose behaviour descriptions; if conflicts are found then these are described and the structure and purpose descriptions of the device are scanned to establish the source of the conflict. The ideas discussed in this paper are implemented in the Doris system which is described.

UCAM-CL-TR-215

Paul Curzon:

A structured approach to the verification of low level microcode

265 pages, PDF
PhD thesis (Christ's College, May 1990)

Abstract: Errors in microprograms are especially serious since all higher level programs on the machine depend on the microcode. Formal verification presents one avenue which may be used to discover such errors. Previous systems which have been used for formally verifying microcode may be categorised by the form in which the microcode is supplied. Some demand that it be written in a high level microprogramming language. Conventional software verification techniques are then applied. Other methods allow the microcode to be supplied in the form of a memory image. It is treated as data to an interpreter modelling the behaviour of the microarchitecture. The proof is then performed by symbolic execution. A third solution is for the code to be supplied in an assembly language and modelled at that level. The assembler instructions are converted to commands in a modelling language. The resulting program is verified using traditional software verification techniques.

In this dissertation I present a new universal microprogram verification system. It achieves many of the advantages of the other kinds of systems by adopting a hybrid approach. The microcode is supplied as a memory image, but it is transformed by the system to a high level program which may be verified using standard software verification techniques. The structure of the high level program is obtained from user supplied documentation. I show that this allows microcode to be split into small, independently validatable portions even when it was not written in that way. I also demonstrate that the techniques allow the complexity of detail due to the underlying microarchitecture to be controlled at an early stage in the validation process. I suggest that the system described would combine well with other validation tools and provide help throughout the firmware development cycle. Two case studies are given. The first describes the verification of Gordon's computer. This example being fairly simple, provides a good illustration of the techniques used by the system. The second case study is concerned with the High Level Hardware Orion computer which is a commercially produced machine with a fairly complex microarchitecture. This example shows that the techniques scale well to production microarchitectures.

UCAM-CL-TR-216

Carole Susan Klein:

Exploiting OR-parallelism in Prolog using multiple sequential machines

250 pages, PDF
PhD thesis (Wolfson College, October 1989)

Abstract: If the branches at each node of a tree are labelled, paths through the tree can be represented by a sequence of labels called an oracle. If an oracle leading to a node is followed, all of the bindings and other state information associated with a node will be recreated. Thus oracles are both a specification for a path through the tree and a concide format for representing the environment at a particular node.

This dissertation investigates the use of oracles for the parallel execution of Prolog programs. The execution of a Prolog program can be represented pictorially by an AND/OR tree. The branches of OR nodes within this tree have no binding dependencies so their evaluation can be performed on separate processors. If one of more of these OR branches is explored in parallel, OR-parallelism is exploited in the Prolog program.

A distributed system called the Delphi Machine has been designed and implemented to exploit the OR-parallelism inherent in Prolog programs. In the implementation described in this dissertation, Delphi runs on a group of uniprocessors connected by Ethernet. Various control strategies using oracles to control the parallel search are investigated. The execution times for Prolog programs run on the Delphi Machine are compared with those of a compiled and an interpreted sequential Prolog system. The results show that a distributed system using oracles to control the parallel search can be an efficient way to exploit OR parallelism in nondeterministic programs.

Because of overheads imposed by the Delphi algorithm, a program executed on a single processor Delphi machine runs at approximately one half the speed as the same program executed on the unmodified prolog system. For a twenty processor configuration, the speed ups obtained vary from approximately two to nine times depending on the amount of OR-parallelism which can be exploited by Delphi. Problems with large amounts of OR-parallelism show a nearly linear speedup.

UCAM-CL-TR-217

Bhaskar Ramanathan Harita:

Dynamic bandwidth management

160 pages, PDF
PhD thesis (Wolfson College, October 1990)

Abstract: Recent advances in semiconductor and optical technologies have contributed greatly to the evolution of broadband integration of multi-service traffic. The asynchronous transfer mode (ATM) has been proposed as the target technique for broadband integrated services digital networks (BISDNs) based on fast packet switching and optical fibre transmission. A primary advantage of ATM is that variable bit rate services can be supported efficiently, which meets the basic needs

of flexibility and service independence required of integrated services networks. In order to fully exploit this flexibility and enchance network efficiency by statistical multiplexing it is important that there be effective methods of bandwidth management and congestion control.

This dissertation describes the use of dynamic bandwidth management to support an ATM overlay superimposed on a public, primary rate ISDN. The overlay architecture provides for the flexible aggregation of switched circuits into larger bandwidth channels. The channels are formatted into a common packet encoding and packets from different sources are statistically multiplexed onto them. In this work, different control schemes that dynamically vary the bandwidth of the channels in a transparent fashion, using out-of-band signalling, are contrasted. The bandwidth is adjusted by adding or deleting circuits in reaction to the traffic rates and the queue sizes at the channels. Performance models of simple bandwidth control schemes as queueing schemes are analysed by the use of moment generating functions

Packet transfer on the overlay is virtual circuit based and connection requests are accepted on the basis of their bandwidth requirements. Dynamic bandwidth management is used to supplement static bandwidth allocations in a congestion control framework presented for the overlay. The cost effectiveness of dynamic bandwidth control is examined for the tarrif structure implemented by the underlying public ISDN.

The contributions of this dissertation are the development of schemes for dynamic bandwidth management, their implementation on an ATM testbed and the analysis of performance models for bandwidth control validated by simulation and experiment.

UCAM-CL-TR-218

Tobias Nipkow:

Higher-order critical pairs

April 1991, 15 pages, PDF

Abstract: We consider rewrite systems over simply typed λ-terms with restricted left-hand sides. This gives rise to a one-step reduction relation whose transitive, reflexive and symmetric closure coincides with equality. The main result of this paper is a decidable confluence criterion which extends the well-known critical pairs to a higher-order setting. Several applications to typed λ-calculi and proof theory are shown.

UCAM-CL-TR-219

Ian M. Leslie, Derek M. McAuley, Mark Hayter, Richard Black, Reto Beller, Peter Newman, Matthew Doar:

Fairisle project working documents

Snapshot 1

March 1991, 56 pages, PDF

Abstract: This report contains the current versions of the documents associated with the fairisle project. These include both papers and draft documents. This collection of documents was made on March 21, 1991. Updated versions will be issued with later snapshot numbers which will replace earlier versions. The present collection includes the following documents:

Fairisle: Network architecture and components / Ian Leslie and Derek McAuley.

Fairisle port controller: design and ideas / Mark Hayter and Richard Black.

Fairisle VME interface (draft) / Reto Beeler.

A Slotted ring copy fabric for a multicast fast packet switch / Peter Newman and Matthew Doar.

Universal Fairisle connector (proposed)

UCAM-CL-TR-220

Cosmos Andrea Nicolaou:

A distributed architecture for multimedia communication systems

192 pages, PDF
PhD thesis (Christ's College, December 1990)

Abstract: Technological advances in digital communications and in personal computer workstations are beginning to allow the generation, communication and presentation of multiple information media simultaneously. In particular, the ability to support real-time voice and video makes a new range of advanced and highly interactive multimedia applications possible. These applications are not restricted to the computer industry, but extend to other technologically intensive industries which have some form of multimedia communication requirement. Such industries include medicine, conferencing, teaching, broadcasting, publishing and printing. Each of these application areas has its own particular set of requirements and makes corresponding demands on the computer systems used.

Such a wide range of application areas leads to a correspondingly large and diverse set of requirements of the systems used to implement them. In addition, the real-time nature of voice, and especially video, place heavy demands on the underlying systems. Many of these requirements and demands are not met by existing computer communication systems. This is due to the fact that the architectural models used to design and implement these systems were constructed before the technological advances making multimedia communication possible took place. As a result existing multimedia systems have tended to concentrate either on low level implementation issues (e.g. communication networks and protocols)or on a single restricted application area, without paying any regard to their respective problems and requirements. The inevitable consequence is that there is a mismatch between the functions

provided at the lower levels and those actually required by higher level applications.

This dissertation presents an attempt to overcome these problems by defining a new architecture for multimedia communication systems which recognises and supports a wide range of application requirements, in addition to satisfying the requirements made by the information media themselves. A thorough survey of existing multimedia systems was conducted in order to identify and understand the requirements made by both applications and information media led to the formulation of a set of design principles. In recognition of the fact that any multimedia communication system is inherently distributed in nature, the architecture is presented as an extension of existing distributed systems.

The resulting architecture is called the Integrated Multimedia Applications Communication architecture (IMAC) and a prototype implementation of IMAC has been constructed and used to evaluate the utility and feasibility of the architecture and to identify its strength and weaknesses.

UCAM-CL-TR-221

Robert Milne:

Transforming axioms for data types into sequential programs

44 pages, PDF

Abstract: A process is proposed for refining specifications of abstract data types into efficient sequential implementations. The process needs little manual intervention. It is split into three stages, not all of which need always be carried out. The three stages entail interpreting equalities as behavioural equivalences, converting functions into procedures and replacing axioms by programs. The stages can be performed as automatic transformations which are certain to produce results that meet the specifications, provided that simple conditions hold. These conditions describe the adequacy of the specifications, the freedom from interference between the procedures, and the mode of construction of the procedures. Sufficient versions of these conditions can be checked automatically. Varying the conditions could produce implementations for different classes of specification. Though the transformations could be automated, the intermediate results, in styles of specification which cover both functions and procedures, have interest in their own right and may be particularly appropriate to object-oriented design.

UCAM-CL-TR-222

Jonathan Billington:

Extensions to coloured petri nets and their application to protocols

190 pages, PDF
PhD thesis (Clare Hall, May 1990)

Abstract: This dissertation develops a net theoretic specification technique for an area known as protocol engineering that covers the life-cycle of protocols. After surveying the application of net theory to protocol engineering, the fundamentals of the specification technique are presented. The technique is based on Jensen's Coloured Petri Nets (CP-nets).

To increase their expressive power, CP-nets are extended by including place capacities and an inhibitor function, leading to the definition of a class of extended CP-nets, known as P-nets. To allow the analysis techniques developed for CP-nets to be applied to P-nets, a transformation from P-nets to CP-nets is formalised and it is proved that it preserves interleaving behaviour. The transformation is based on the notion of contemporary places (known from Place/Transition-nets) and involves the definition and proof of a new complementary place invariant for CP-nets. A class of P-nets is defined where true concurrency is preserved under the transformation.

A graphical form of P-nets, known as a P-graph, is formally defined, drawing upon the notions developed for algebraic specification of abstract data types. Arc inscriptions are multisets of terms generated from a many-sorted signature. Transition conditions are Boolean expressions derived from the same signature. An interpretation of the P-Graph is given in terms of a corresponding P-net. In the P-Graph, concrete sets are associated with places, and likewise there are concrete initial marking and capacity multisets. P-Graphs are useful for specification at a concrete level, and allow classes of nets, such as CP-Graphs, many-sorted Algebraic nets and many-sorted Predicate/Transition nets, to be defined as special cases. They also provide the basis for a comparison with other high-level nets such as Predicate/Transition nets and Algebraic nets. An extended place capacity notation is developed to allow for the convenient representation of resource bounds in the graphical form.

Abstract P-Graphs are defined in a similar way to P-Graphs, but this time sorts are associated with places, and markings and capacities are defined at the syntactic level. This is useful for more abstract specifications (such as classes of communication protocols) and for their analysis.

Part of the motivation for the extensions to CP-nets has been to develop convenient constructs for the purging of a place's marking (or part of the marking), by the occurrence of a single transition. This is achieved by equating the inscriptions of the inhibitor and normal arc. Some convenient notation is developed for the P-Graph for purging parts of a place's marking.

Some simple communications-oriented examples are presented including queues and the Demon Game developed by the International Organisation for Standardisation as a test case for formal description techniques. A major case study of the M-Access service of the Cambridge Fast Ring is specified with the P-Graph to illustrate the utility of a number of the extensions developed for P-nets.

UCAM-CL-TR-223

Philip Gladwin, Stephen Pulman, Karen Spärck Jones:

Shallow processing and automatic summarising: a first study

May 1991, 65 pages, PDF

Abstract: This report describes a study of ten simple texts, investigating various discourse phenomena to see how they might be exploited, in shallow text processing, for summarising purposes. The processing involved was a simulation of automatic analysis which is in principle within reach of the state of the art. Each text was treated by a version of Sidner's focusing algorithm. The products of this were fed into subsidiary stages of analysis to provide an assessment of the activity of the various discourse entities within each text. A concurrent process examined the occurrence of orthographically identical noun phrase forms. Appendices give the ten texts, a complete specification of the version of the focusing algorithm in use, and the full experimental results. These suggest, especially when the brevity of the test texts is taken into account, that the type of information given by focusing has potential but limited value for summarising.

UCAM-CL-TR-224

Ted Briscoe, John Carroll:

Generalised probabilistic LR parsing of natural language (corpora) with unification-based grammars

45 pages, PDF

Abstract: We describe work towards the construction of a very wide-coverage probabilistic parsing system for natural language (NL), based on LR parsing techniques. The system is intended to rank the large number of syntactic analyses produced by NL grammars according to the frequency of occurrence of the individual rules deployed in each analysis. We discuss a fully automatic procedure for constructing an LR parse table from a unification-based grammar formalism, and consider the suitability of alternative LALR(1) parse table construction methods for large grammars. The parse table is used as the basis for two parsers; a user-driven interactive system which provides a computationally tractable and labour-efficient method of supervised learning of the statistical information required to drive the probabilistic parser. The latter is constructed by associating probabilities with the LR parse table directly. This technique is superior to parsers based on probabilistic lexical tagging or probabilistic context-free grammar because it allows for a more context dependent probabilistic language model, as well as use of a more linguistically adequate grammar formalism. We compare the performance of an optimised variant of Tomita's (1987) generalised LR parsing algorithm to an (efficiently indexed and optimised) chart parser. We report promising results of a pilot study training on 151 noun definitions from the Longman Dictionary of Contemporary English (LDOCE) and retesting on these plus a further 54 definitions. Finally we discuss limitations of the current system and possible extensions to deal with lexical (syntactic and semantic) frequency of occurrence.

UCAM-CL-TR-225

Valeria de Paiva:

Categorical multirelations, linear logic and petri nets (draft)

May 1991, 29 pages, PDF

Abstract: This note presents a categorical treatment of multirelations, which is, in a loose sense a generalisation of both our previous work on the categories GC, and of Chu's construction A_NC [Barr'79]. The main motivation for writing this note was the utilisation of the category GC by Brown and Gurr [BG90] to model Petri nets. We wanted to extend their work to deal with multirelations, as Petri nets are usually modelled using multirelations pre and post. That proved easy enough and people interested mainly in concurrency theory should refer to our joint work [BGdP'91], this note deals with the mathematics underlying [BGdP'91]. The upshot of this work is that we build a model of Intuitionistic Linear Logic (without modalities) over any symmetric monoidal category C with a distinguished object (N, $\leq$, $\circ$, e $-\circ$) – a closed poset. Moreover, if the category C is cartesian closed with free monoids, we build a model of Intuitionistic Linear Logic with a non-trivial modality '!' over it.

UCAM-CL-TR-226

Kwok-yan Lam:

A new approach for improving system availability

June 1991, 108 pages, paper copy
PhD thesis (Churchill College, January 1991)

UCAM-CL-TR-227

Juanito Albert Camilleri:

Priority in process calculi

June 1991, 203 pages, paper copy
PhD thesis (Trinity College, October 1990)

UCAM-CL-TR-228

Mark Hayter, Derek McAuley:

The desk area network

May 1991, 11 pages, PostScript

Abstract: A novel architecture for use within an end computing system is described. This attempts to extend the concepts used in modern high speed networks into computer system design. A multimedia workstation is being built based on this concept to evaluate the approach.

UCAM-CL-TR-229

David J. Brown:

Abstraction of image and pixel The thistle display system

August 1991, 197 pages, paper copy
PhD thesis (St John's College, February 1991)

UCAM-CL-TR-230

J. Galliers:

Proceedings of the Second Belief Representation and Agent Architectures Workshop (BRAA '91)

August 1991, 255 pages, paper copy

UCAM-CL-TR-231

Raphael Yahalom:

Managing the order of transactions in widely-distributed data systems

August 1991, 133 pages, paper copy
PhD thesis (Jesus College, October 1990)

UCAM-CL-TR-232

Francisco Corella:

Mechanising set theory

July 1991, 217 pages, PDF
PhD thesis (Corpus Christi College, June 1989)

Abstract: Set theory is today the standard foundation of mathematics, but most proof development sysems (PDS) are based on type theory rather than set theory. This is due in part to the difficulty of reducing the rich mathematical vocabulary to the economical vocabulary of the set theory. It is known how to do this in principle, but traditional explanations of mathematical notations in set theoretic terms do not lead themselves easily to mechanical treatment.

We advocate the representation of mathematical notations in a formal system consisting of the axioms of any version of ordinary set theory, such as ZF, but within the framework of higher-order logic with λ-conversion (H.O.L.) rather than first-order logic (F.O.L.). In this system each notation can be represented by a constant, which has a higher-order type when the notation binds variables. The meaning of the notation is given by an axiom which defines the representing constant, and the correspondence between the ordinary syntax of the notation and its representation in the formal language is specified by a rewrite rule. The collection of rewrite rules comprises a rewriting system of a kind which is computationally well behaved.

The formal system is justified by the fact than set theory within H.O.L. is a conservative extension of set theory within F.O.L. Besides facilitating the representation of notations, the formal system is of interestbecause it permits the use of mathematical methods which do not seem to be available in set theory within F.O.L.

A PDS, called Watson, has been built to demonstrate this approach to the mechanization of mathematics. Watson embodies a methodology for interactive proof which provides both flexibility of use and a relative guarantee of correctness. Results and proofs can be saved, and can be perused and modified with an ordinary text editor. The user can specify his own notations as rewrite rules and adapt the mix of notations to suit the problem at hand; it is easy to switch from one set of notations to another. As a case study, Watson has been used to prove the correctness of a latch implemented as two cross-coupled nor-gates, with an approximation of time as a continuum.

UCAM-CL-TR-233

John Carroll, Ted Briscoe, Claire Grover:

A development environment for large natural language grammars

July 1991, 65 pages, paper copy

UCAM-CL-TR-234

Karen Spärck Jones:

Two tutorial papers: Information retrieval & Thesaurus

August 1991, 31 pages, PDF

Abstract: The first paper describes the characteristics of information retrieval from documents or texts, the development and status of automatic indexing and retrieval, and the actual and potential relations between information retrieval and artificial intelligence. The second paper discusses the properties, construction and actual and potential uses of thesauri, as semantic classifications or terminological knowledge bases, in information retrieval and natural language processing.

UCAM-CL-TR-235

Heng Wang:

Modelling and image generation

145 pages, PDF
PhD thesis (St John's College, July 1991)

Abstract: Three dimensional (3D) volume representation, processing and visualisation have gained growing attention during the last ten years due to the rapid decrease in computer memory cost and the enhancement of computation power. Recent developments in massively parallel computer architectures and special purpose graphics accelerators also facilitate the solution of 3D volume manuipulation problems which usually have large memory and computation requirements. Volumentric graphics is becoming practically possible and finding many applications such as medical image processing, computer aided design and scientific visualisation.

A volumetric object is usually represented in one of two forms: a large 3D uniform grid of voxels (volume elements), and a relatively compact non-uniform collection of volumes. Objects in the latter form are obtained by adaptive, recursive decompositions. An octree is a special case in which each non-terminal volume is subdivided into eight sub-volumes. The problems of current implementation of octrees concern the speed and complexity of memory management. This dissertation looks into a novel approach of designing octree-related volumetric graphics algorithms based on Content Addressable Memories (CAMs). A CAM is an architecture consisting of elements which have data storage capabilities and can be accessed simultaneously on the basis of data contents instead of addresses. It is demonstrated that the main features of CAMs, their parallel searching, pattern matching and masked parallel updating capabilities, are suitable for implementing octree related algorithms.

New CAM algorithms are presented for transforming octrees, evaluating set operations (union, intersection, difference), displaying volumetric objects, calculating volumes, constructing octrees from other representations, and so on. These algorithms are remarkably simple and conceptively intuitive. The simplicity plays an important role in constructing robust solid 3D modelling systems. In addition to their simplicity, many algorithms are more efficient than their conventional counterparts.

A new method has been developed to speed up the image synthesis algorithm of ray tracing using CAM octrees. It is aimed to reduce the number of ray-object intersection tests without significantly increasing the overheads of storage and computation which are related to octree data structures and their traversals. The simulation results confirm the expected improvements in speed and memory management. Ray tracing can be accelerated by applying parallelism. Preliminary analysis shows possibilities of implementing the above CAM octree ray tracer on general parallel machines such as MIMD (Multiple Instriction stream, Multiple Data stream).

UCAM-CL-TR-236

John Anthony Bradshaw:

Using knowledge of purpose and knowledge of structure as a basis for evaluating the behaviour of mechanical systems

153 pages, paper copy
PhD thesis (Gonville & Caius College, June 1991)

UCAM-CL-TR-237

Derek G. Bridge:

Computing presuppositions in an incremantal language processing system

212 pages, PDF
PhD thesis (Wolfson College, April 1991)

Abstract: This thesis describes the design and implementation of a natural language analysis system for the computation of presuppositions. The system is one in which syntactic, semantic and pragmatic processing are interleaved with feedback to syntactic analysis from semantic and pragmatic processing. The thesis begins by illustrating how the system processes definite noun phrases. The mechanisms used for this are then shown to be easily extensible to processing other parts of speech such as indefinite noun phrases and verb phrases.

Definite noun phrases have been said to be presupposition triggers. This means that traditionally they have been seen as licensing certain inferences — presuppositions. In the system described herein, presuppositions are treated as a special kind of inference: preconditions. This treatment for definite noun phrases can be extended to give a uniform account of all presupposition triggers (e.g. factive verbs). It is a view that makes it clear that presuppositions are not 'optional extras' that might or might not be derived once a semantic representation of an utterance has been produced. Rather, they play an essential role in driving the utterance analysis process: the failure of a presupposition, i.e. failure to satisfy a precondition, can direct the system to choose an alternative reading of an utterance of an ambiguous sentence.

As it processes an utterance, the system builds and regularly consults a representation of contextual knowledge referred to as a discourse model. Importantly, the system checks whether presuppositions are satisfied against the discourse model. Presupposition failure, i.e. a presupposition not being satisfied by the discourse model, is not necessarily the same as a presupposition being false in, e.g., the 'real' world. Checking presuppositions for satisfaction in a discourse model and not for truth in a possible world offers new ideas on the behaviour of presuppositions in utterances of negative and complex sentences.

In utterances of negative sentences, presuppositions must still be satisfied by the discourse model. Presuppositions cannot be cancelled as they can in other accounts. Rather, presupposition "cancellation" data is explained in terms of utterances that make metalinguistic statements about the model-theoretic interpretation of the discourse model. It is shown that computing presuppositions in an incremental system gives a simple account of most of the data relating to the behaviour of presuppositions in utterancesof compound sentences and longer stretches of text (the so-called "projection problem"). Presuppositions must again be satisfied by the discourse model, but they may be satisfied by virtue of changes made to the discourse model by earlier parts of the utterance or text.

UCAM-CL-TR-238

Ted Briscoe, Ann Copestake, Valeria de Paiva:

Proceedings of the ACQUILEX Workshop on Default Inheritance in the lexicon

October 1991, 180 pages, PDF

Abstract: The ACQUILEX Esprit BRA (Basic Research Action) research project is concerned with the acquisition and representation of lexical information from machine readable dictionaries for use in Natural Language Processing. The Cambridge group of the ACQUILEX project organised a Workshop on Default Inheritance in April 1991, the main purpose of which was to review approaches to default inheritance for lexical organisation and representation. The emphasis from ACQUILEX's point of view was in implementing a practical system capable of supporting substantial lexicons, based on existing proposals to incorporate (default) inheritance into a unification-based framework similar to DATR (Gazdar and Evans, 1989) and HPSG (e.g. Carpenter, 1990).

The workshop consisted of two days of talks, where theoretical and implementational issues on default inheritance were discussed, as well as a last day of demonstrations of implemented systems. Papers from several European colaborative projects on the topic of the workshop were presented – see enclosed list of titles and affiliations. The Cambridge ACQUILEX group presented and demonstrated the ACQUILEX lexical knowledge base (LKB) system and provided a tutorial on use of the software. The TFS system of the project POLYGLOSS and the system ELU of the group at ISSCO were also discussed and demonstrated.

Many thanks to all the participants for the lively discussions – exactly what workshops are supposed to be for.

UCAM-CL-TR-239

Mark Thomas Maybury:

Planning multisentential English text using communicative acts

December 1991, 329 pages, PDF
PhD thesis (Wolfson College, July 1991)

Abstract: The goal of this research is to develop explanation presentation mechanisms for knowledge based systems which enable them to define domain terminology and concepts, narrate events, elucidate plans, processes, or propositions and argue to support a claim or advocate action. This requires the development of devices which select, structure, order and then linguistically realize explanation content as coherent and cohesive English text.

With the goal of identifying generic explanation presentation strategies, a wide range of naturally occurring texts were analyzed with respect to their communicative structure, function, content and intended effects on the reader. This motivated an integrated theory of communicative acts which characterizes text at the level of rhetorical acts (e.g. describe, define, narrate), illocutionary acts (e.g. inform, request), and locutionary acts (ask, command). Taken as a whole, the identified communicative acts characterize the structure, content and intended effects of four types of text: description, narration, exposition, argument. These text types have distinct effects such as getting the reader to know about entities, to know about events, to understand plans, processes, or propositions, or to believe propositions or

want to perform actions. In addition to identifying the communicative function and effect of text at multiple levels of abstraction, this dissertation details a tripartite theory of focus of attention (discourse focus, temporal focus and spatial focus) which constrains the planning and linguistic realization of text.

To test the integrated theory of communicative acts and tripartite theory of focus of attention, a text generation system TEXPLAN (Textual EXplanation PLANner) was implemented that plans and linguistically realizes multisentential and multiparagraph explanations from knowledge based systems. The communicative acts identified during text analysis were formalized over sixty compositional and (in some cases) recursive plan operators in the library of a hierarchical planner. Discourse, temporal and spatial models were implemented to track and use attentional information to guide the organization and realization of text. Because the plan operators distinguish between the communicative function (e.g. argue for a proposition) and the expected effect (e.g. the reader believes the proposition) of communicative acts, the system is able to construct a discourse model of the structure and function of its textual responses as well as a user model of the expected effects of its responses on the reader's knowledge, beliefs, and desires. The system uses both the discourse model and user model to guide subsequent utterances. To test its generality, the system was interfaced to a variety of domain applications including a neuropsychological diagnosis system, a mission planning system, and a knowledge based mission simulator. The system produces descriptions, narratives, expositions and arguments from these applications, thus exhibiting a broader ranger of rhetorical coverage then previous text generation systems.

UCAM-CL-TR-240

Juanito Camilleri:

Symbolic compilation and execution of programs by proof: a case study in HOL

December 1991, 31 pages, PDF

Abstract: This paper illustrates the symbolic compilation and execution of programs by proof using the proof assistant HOL. We formalise the operational semantics of an Occam-like programming language oc and show how synchronous communication in oc compiles to an intermediate programming language Safe, whose compilation yields instructions intended to drive machines that communicate via shared memory. We show how the symbolic formal manipulation of terms of a programming language, subject to the definition of its semantics, can animate a desired effect — be it compilation or execution. Needless to say, such compilation and execution by proof is rather slow, but it is fast enough to give vital feedback about the compilation algorithm being used. Without such animation it is hard to anticipate whether the compilation algorithm is reasonable before attempting to verify it. This is particularly true when attempting to find a plausible handshaking protocol that implements synchronous communication.

UCAM-CL-TR-241

Thomas Ulrich Vogel:

Learning in large state spaces with an application to biped robot walking

December 1991, 204 pages, PDF
PhD thesis (Wolfson College, November 1991)

Abstract: Autonomous robots must be able to operate in complex, obstacle cluttered environments. To do this the robots must be able to focus on the important aspects of their environment, create basic strategies to carry out their operations, generalise these strategies and finally learn from successful experiences.

Based on simulated dynamic biped robot walking, this thesis investigates these issues. An algorithm is given which analyses the state space of the robot and orders the dimensions of the state space by their importance relative to the task of the robot. Using this analysis of its state space, the robot is able to generate a set of macros (gaits) which enable it to operate in its immediate environment. We then present a control algorithm which allows the robot to control the execution of its gaits

Once the robot has learned to walk on an obstacle-free horizontal surface, it uses its knowledge about gaits in order to derive obstacle crossing gaits from existing gaits. A strategy based on the qualitative equivalence between two behaviours is introduces in order to derive new behavioural patterns from previous ones. This enables the robot to reason about its actions at a higher level of abstraction. This facilitates the transfer and adaptation of existing knowledge to new situations. As a result, the robot is able to derive stepping over an obstacle from stepping on a horizontal surface.

Finally, the robot analyses its successful obstacle crossings in order to generate a generic obstacle crossing strategy. The concept of a virtual evaluation function is introduced in order to describe how the robot has to change its search strategy in order to search successfully for obstacle crossing behaviours. This is done by comparing how the successful obstacle crossing of the robot differs from its normal behaviour. By analysing and operationalising these differences, the robot acquires the capability to overcome previously unencountered obstacles. The robot's obstacle crossing capabilities are demonstrated by letting the robot walk across randomly generated obstacle combinations

UCAM-CL-TR-242

Glenford Ezra Mapp:

An object-oriented approach to virtual memory management

January 1992, 150 pages, PDF
PhD thesis (Clare Hall, September 1991)

Abstract: Advances in computer technology are being pooled together to form a new computing environment which is characterised by powerful workstations with vast amounts of memory connected to high speed networks. This environment will provide a large number of diverse services such as multimedia communications, expert systems and object-oriented databases. In order to develop these complex applications in an efficient manner, new interfaces are required which are simple, fast and flexible and allow the programmer to use an object-oriented approach throughout the design and implementation of an application. Virtual memory techniques are increasingly being used to build these new facilities.

In addition since CPU speeds continue to increase faster than disk speeds, an I/O bottleneck may develop in which the CPU may be idle for long periods waiting for paging requests to be satisfied. To overcome this problem it is necessary to develop new paging algorithms that better reflect how different objects are used. Thus a facility to page objects on a per-object basis is required and a testbed is also needed to obtain experimental data on the paging activity of different objects.

Virtual memory techniques, previously only used in mainframe and minicomputer architectures, are being employed in the memory management units of modern microprocessors. With very large address spaces becoming a standard feature of most systems, the use of memory mapping is seen as an effective way of providing greater flexibility as well as improved system efficiency.

This thesis presents an object-oriented interface for memory mapped objects. Each object has a designated object type. Handles are associated with different object types and the interface allows users to define and manage new object types. Moving data between the object and its backing store is done by user-level processes called object managers. Object managers interact with the kernel via a specified interface thus allowing users to build their own object managers. A framework to compare different algorithms was also developed and an experimental testbed was designed to gather and analyse data on the paging activity of various programs. Using the testbed, conventional paging algorithms were applied to different types of objects and the results were compared. New paging algorithms were designed and implemented for objects that are accessed in a highly sequential manner.

UCAM-CL-TR-243

Alison Cawsey, Julia Galliers, Steven Reece, Karen Spärck Jones:

Automating the librarian: a fundamental approach using belief revision

January 1992, 39 pages, PDF

Abstract: This paper describes a current research project investigating belief revision in intelligent systems by modelling the librarian in interaction with a literature-seeking user. The work is designed to both test a theory of agent behaviour based on belief revision proposed by Galliers, and to evaluate a model of the librarian developed by Belkin, Brooks and Daniels, through computational implementation. Agent communication is seen as motivated by and motivating belief changes, where belief revision is determined by coherence, combining endorsement, connectivity and conservatism. The librarian is viewed as a distributed expert system with many individual specialised functions operating in particular belief domains. The paper describes our first implementation of the belief revision mechanism and of a very primative librarian, designed to test the basic viability of our ideas and to allow us to explore different forms of the distributed system architecture.

UCAM-CL-TR-244

T.F. Melham:

A mechanized theory of the π-calculus in HOL

January 1992, 31 pages, PDF

Abstract: The π-calculus is a process algebra developed at Edinburgh by Milner, Parrow and Walker for modelling concurrent systems in which the pattern of communication between processes may change over time. This paper describes the results of preliminary work on a mechanized formal theory of the π-calculus in higher order logic using the HOL theorem prover.

UCAM-CL-TR-245

Michael J. Dixon:

System support for multi-service traffic

January 1992, 108 pages, PDF
PhD thesis (Fitzwilliam College, September 1991)

Abstract: Digital network technology is now capable of supporting the bandwidth requirements of diverse applications such as voice, video and data (so called multi-service traffic). Some media, for example voice, have specific transmission requirements regarding the maximum packet delay and loss which they can tolerate. Problems arise when attempting to multiplex such traffic over a single channel. Traditional digital networks based on the Packet- (PTM) and Synchronous- (STM) Transfer Modes prove unsuitable due to their media access contention and inflexible bandwidth allocation properties respectively. The Asynchronous Transfer Mode (STM) has been proposed as a compromise between the PTM and STM techniques. The current state of multimedia research suggests that a significant amount of multi-service traffic will be handled by computer operating systems. Unfortunately conventional operating systems are largely unsuited to such a task. This dissertation is concerned with the system organisation necessary in order to extend the benefits of ATM networking through the endpoint operating system and up to the application level. A locally developed microkernel, with ATM network protocol support, has been used as a testbed for the ideas presented. Practical results over prototype ATM networks, including the 512 MHz Cambridge Backbone Network, are presented.

UCAM-CL-TR-246

Victor Poznański:

A relevance-based utterance processing system

February 1992, 295 pages, PDF
PhD thesis (Girton College, December 1990)

Abstract: This thesis presents a computational interpretation of Sperber and Wilson's relevance theory, based on the use of non-monotonic logic supported by a reason maintenance system, and shows how the theory, when given a specific form in this way, can provide a unique and interesting account of discourse processing.

Relevance theory is a radical theory of natural language pragmatics which attempts to explain the whole of human cognition using a single maxim: the Principle of Optimal Relevance. The theory is seen by its originators as a computationally more adequate alternative to Gricean pragmatics. Much as it claims to offer the advantage of a unified approach to utterance comprehension, Relevance Theory is hard to evaluate because Sperber and Wilson only provide vague, high-level descriptions of vital aspects of their theory. For example, the fundamental idea behind the whole theory is that, in trying to understand an utterance, we attempt to maximise significant new information obtained from the utterance whilst consuming as little cognitive effort as possible. However, Sperber and Wilson do not make the nature of information and effort sufficiently clear.

Relevance theory is attractive as a general theory of human language communication and as a potential framework for computational language processing systems. The thesis seeks to clarify and flesh out the problem areas in order to develop a computational implementation which is used to evaluate the theory.

The early chapters examine and criticise the important aspects of the theory, emerging with a schema for an ideal relevance-based system. Crystal, a computational implementation of an utterance processing system based on this schema is then described. Crystal performs certain types of utterance disambiguation and reference resolution, and computes implicatures according to relevance theory.

An adequate reasoning apparatus is a key component of a relevance based discourse processor, so a suitable knowledge representation and inference engine are required. Various candidate formalisms are considered, and a knowledge representation and inference engine based on autoepistemic logic is found to be the most suitable. It is then shown how this representation can be used to meet particular discourse processing requirements, and how it provides a convenient interface to a separate abduction system that supplies not demonstrative inferences according to relevence theory. Crystal's powers are illustrated with examples, and the thesis shows how the design not only implements the less precise areas of Sperber and Wilson's theory, but overcomes problems with the theory itself.

Crystal uses rather crude heuristics to model notions such as salience and degrees of belief. The thesis thefore presents a proposal and outline for a new kind of reason maintenance system that supports non-monotonic logic whose formulae re labelled with upper/lower probability ranges intended to represent strength of belief. This system should facilitate measurements of change in semantic information and shed some light on notions such as expected utility and salience.

The thesis concludes that the design and implementation of crystal provide evidence that relevance theory, as a generic theory of language processing, is a viable alternative theory of pragmatics. It therefore merits a greater level of investigation than has been applied to it to date.

UCAM-CL-TR-247

Roy Luis Crole:

Programming metalogics with a fixpoint type

February 1992, 164 pages, PDF
PhD thesis (Churchill College, January 1992)

Abstract: A programming metalogic is a formal system into which programming languages can be translated and given meaning. The translation should reflect both the structure of the language and make it easy to prove properties of programs. This thesis develops certain metalogics using techniques of category theory and treats recursion in a new way.

The notion of a category with a fixpoint logic is defined. Corresponding to this categorical structure there are type theoretic equational rules which will be present in all of the metalogics considered. These rules define the fixpoint type which will allow the interpretation of recursive declarations. With these core notions FIX categories are defined. These are the categorical equivalent of an equational logic which can be viewed as a very basic programming metalogic. Recursion is treated both syntactically and categorically.

The expressive power of the equational logic is increased by embedding it in an intuitionistic predicate calculus, giving rise to the FIX logic. This contains propositions about the evaluation of computations to values and an induction principle which is derived from the definition of a fixpoint object as an initial algebra. The categorical structure which accompanies the FIX logic is defined, called a FIX hyperdoctrine, and certain existence and disjunction properties of FIX are stated. A particular FIX hyperdoctrine is constructed and used in the proof of the above properties.

PCF-style languages are translated into the FIX logic and computational adequacy results are proved. Two languages are studied: both are similar to PCF except one has call by value recursive function declarations and the other higher order conditionals.

A dependently typed equational logic containing a fixpoint type and a universal type is given together with its related categorical structure, namely a FIX category with attributes. A representation theorem for Scott predomains is proved, which gives rise to a concrete example of such a FIX category with attributes. Recursive domain equations give rise to endofunctions on the universal type; using the fixpoint type we may solve for fixpoints of such endofunctions and thus obtain a solution the original domain as the type coded by the fixpoint.

UCAM-CL-TR-248

Richard J. Boulton:

On efficiency in theorem provers which fully expand proofs into primitive inferences

February 1992, 23 pages, DVI

Abstract: Theorem Provers which fully expand proofs into applications of primitive inference rules can be made highly secure, but have been criticized for being orders of magnitude slower than many other theorem provers. We argue that much of this relative inefficiency is due to the way proof procedures are typically written and not all is inherent in the way the systems work. We support this claim by considering a proof procedure for linear arithmetic. We show that straightforward techniques can be used to significantly cut down the computation required. An order of magnitude improvement in the performance is shown by an implementation of these techniques.

UCAM-CL-TR-249

John P. Van Tassel:

A formalisation of the VHDL simulation cycle

March 1992, 24 pages, PDF

Abstract: The VHSIC Hardware Description Language (VHDL) has been gaining wide acceptance as a unifying HDL. It is, however, still a language in which the only way of validating a design is by careful simulation. With the aim of better understanding VHDL's particular simulation process and eventually reasoning about it, we have developed a formalisation of VHDL's simulation cycle for a subset of the language. It has also been possible to embed our semantics in the Cambridge Higher-Order Logic (HOL) system and derive interesting properties about specific VHDL programs.

UCAM-CL-TR-250

Innes A. Ferguson:

TouringMachines: autonomous agents with attitudes

April 1992, 19 pages, PostScript

Abstract: It is becoming widely accepted that neither purely reactive nor purely deliberative control techniques are capable of producing the range of behaviours required of intelligent robotic agents in dynamic, unpredictable, multi-agent worlds. We present a new architecture for controlling autonomous, mobile agents – building on previous work addressing reactive and deliberative control methods. The proposed multi-layered control architecture allows a resource-bounded, goal-directed agent to react promptlyto unexpected changes in its environment; at the same time it allows the agent to reason predictively about potential conflicts by contrasting and projecting theories which hypothesise other agents' goals and intentions.

The line of research adopted is very much a pragmatic one. A single common architecture has been implemented which, being extensively parametrized allows an experimenter to study functionally- and behaviourally-diverse agent configurations. A principal aim of this research is to understand the role different functional capabilities play in constraining an agent's behaviour under varying environmental conditions. To this end, we have constructed an experimental testbed comprising a simulated multi-agent world in which a variety of agent configurations and bahaviours have been investigated. Some experience with the new control architecture is described.

UCAM-CL-TR-251

Xiaofeng Jiang:

Multipoint digital video communications

April 1992, 124 pages, PDF
PhD thesis (Wolfson College, December 1991)

Abstract: Ever since the emergence of high-speed communication networks and fast signal processing technology, digital video has been attracting increased research interest. However, problems associated with its use in a multipoint communication environment have not been thouroughly investigated. In particular, these include the avoidance of congestion on multicast paths when multiple wideband sources are transmitting simultaneously, and the ability to interchange different format signals properly and efficiently. This dissertation addresses these issues with a two-level communications architecture.

The congestion issue at the network level is dealt with by several stream multicast path finding algorithms which are either centralised or distributed to suit various application environments. Different ways of integrating communication link capacities are investigated for supporting simultaneous transmission of broadband signals with minimum effect on network traffic and maximum success in path finding. Simulation results demonstrate performance improvements over conventional multicast path finding algorithms.

The format issue at the presentation level is dealt with by an intermediate format or general representation of digital video streams. Signals under this scheme are organised in a form to facilitate their interchange and scalable receiving in multipoint communication applications. Issues including frame segmentation and coding description are investigated. An experimental system implementing a simple version of the scheme is presented along with test results on picture quality degredation from conversion of various types and related timing characteristics.

UCAM-CL-TR-252

Andrew M. Pitts:

A co-induction principle for recursively defined domains

25 pages, PostScript

Abstract: This paper establishes a new property of predomains recursively defined using the cartesian product, disjoint union, partial function space and convex powerdomain constructors. We prove that the partial order on such a recuirsive predomain D is the greatest fixed point of a certain monotone operator associated to D. This provides a structurally defined family of proof principles for these recursive predomains: to show that one element of D approximates another, it suffices to find a binary relation containing the two elements that is a post-fixed point for the associated monotone operator. The statement of the proof principles is independent of any of the various methods available for explicit construction of recursive predomains. Following Milner and Tofte, the method of proof is called co-induction. It closely resembles the way bisimulations are used in concurrent process calculi.

Two specific instances of the co-induction principle already occur in the work of Abramsky in the form of 'internal full abstraction' theorems for denotational semantics of SCCS and the lazy lambda calculus. In the first case post-fixed binary relations are precisely Abramsky's partial bisimulations, whereas in the second case they are his applicative bisimulations. The coinduction principle also provides an apparently useful tool for reasoning about the equality of elements of recursively defined datatypes in (strict or lazy) higher order functional programming languages.

UCAM-CL-TR-253

Antonio Sanfilippo:

The (other) Cambridge ACQUILEX papers

141 pages, paper copy

UCAM-CL-TR-254

Richard J. Boulton:

A HOL semantics for a subset of ELLA

April 1992, 104 pages, DVI

Abstract: Formal verification is an important tool in the design of computer systems, especially when the systems are safety or security critical. However, the formal techniques currently available are not well integrated into the set of tools more traditionally used by designers. This work is aimed at improving the integration by providing a formal semantics for a subset of the hardware description language ELLA, and by supporting this semantics in the HOL theorem proving system, which has been used extensively for hardware verification.

A semantics for a subset of ELLA is described, and an outline of a proof of the equivalence of parallel and recursive implementations of an n-bit adder is given as an illustration of the semantics. The proof has been performed in an extension of the HOL system. Some proof tools written to support the verification are also described.

UCAM-CL-TR-255

Rachel Mary Cardell-Oliver:

The formal verification of hard real-time systems

1992, 151 pages, paper copy
PhD thesis (Queens' College, January 1992)

UCAM-CL-TR-256

Martin Richards:

MCPL programming manual

May 1992, 32 pages, PDF

Abstract: MCPL is a systems programming language having much in common with BCPL but augmented by the pattern matching ideas of both ML and Prolog. Unlike ML, MCPL is typeless, runs using a contiguous runtime stack and has no built in garbage collector, but it does make extensive use of ML-like pattern matching. The low level aspects of the language resemble those of BCPL and C. For efficiency, MCPL uses its own function calling sequence, however a convenient mechanism for mixing MCPL and C programs is provided.

Notable features of MCPL are its pattern matching facilities and the simple way in which data structures are handled.

This document gives a complete definition of the language and includes, at the end, several example programs to demonstrate its capabilities.

UCAM-CL-TR-257

Rajeev Prakhakar Goré:

Cut-free sequent and tableau systems for propositional normal modal logics

May 1992, 160 pages, PDF

Abstract: We present a unified treatment of tableau, sequent and axiomatic formulations for many propositional normal modal logics, thus unifying and extending the work of Hanson, Segerberg, Zeman, Mints, Fitting, Rautenberg and Shvarts. The primary emphasis is on tableau systems as the completeness proofs are easier in this setting. Each tableau system has a natural sequent analogue defining a finitary provability relation for each axiomatically formulated logic L. Consequently, any tableau proof can be converted into a sequent proof which can be read downwards to obtain an axiomatic proof. In particular, we present cut-free sequent systems for the logics S4.3, S4.3.1 and S4.14. These three logics have important temporal interpretations and the sequent systems appear to be new.

All systems are sound and (weakly) complete with respect to their known finite frame Kripke semantics. By concentrating almost exclusively on finite tree frames we obtain finer characterisation results, particularly for the logics with natural temporal interpretations. In particular, all proofs of tableau completeness are constructive and yield the finite model property and decidability for each logic.

Most of these systems are cut-free giving a Gentzen cut-elimination theorem for the logic in question. But even when the cut rule is required, all uses of it remain analytic. Some systems do not possess the subformula property. But in all such cases the class of "superformulae" remains bounded, giving an analytic superformula property. Thus all systems remain totally amenable to computer implementation and immediately serve as nondeterministic decision procedures for the logics they formulate. Furthermore, the constructive completeness proofs yield deterministic decision procedures for all the logics concerned.

In obtaining these systems we domonstrate that the subformula property can be broken in a systematic and analytic way while still retaining decidability. This should not be surprising since it is known that modal logic is a form of second order logic and that the subformula property does not hold for higher order logics.

UCAM-CL-TR-258

David J. Greaves, Derek McAuley, Leslie J. French:

Two papers on ATM networks

May 1992, 22 pages, PDF

Abstract: Private ATM networks / by David J. Greaves and Derek McAuley.

This paper advocates the use of local area networks which use 48 byte ATM cells. Hosts connected to the network are fitted with ATM interfaces and run a new protocol stack up to the network level, which avoids multiplexing and efficiently handles the out-of-band signalling used by ATM.

The private network may be of WAN, MAN or LAN dimensions and contain several different network technologies, provided each is able to perform the basic function of carrying ATM cells from one point to another. The private network may be connected to the B-ISDN at one or more points.

Protocol and interface for ATM LANs / by David J. Greaves, Derek McAuley and Leslie J. French.

This paper advocates the use of local area networks using the Asynchronous Transfer Mode, where data is carried in the payloads of 48-byte cells. We describe the design and performance of a simple ATM host interface for the DEC Turbochannel together with the MSNA protocol architecture. We describe how MSNA creates a homogeneous internet for ATM hosts and devices.

We discuss the implementation of an adaption layer for computer data which is able to take full advantage of MSNA semantics, and which makes use of the end-to-end ATM layer header bit which has recently been accepted.

UCAM-CL-TR-259

Samson Abramsky, C.-H. Luke Ong:

Full abstraction in the Lazy Lambda Calculus

104 pages, paper copy

UCAM-CL-TR-260

Henrik Reif Anderson:

Local computation of alternating fixed-points

June 1992, 21 pages, PDF

Abstract: In this paper we consider the problem of alternating fixed-points of monotone functions on finite boolean lattices. We describe a local (demand-driven, lazy) algorithm for computing a boolean expression with two alternating fixed-points, i.e. with a minimal and a maximal fixed-point intertwined. Such expressions arise naturally in the modal μ-calculus and are the main source of its expressive power – and its difficult model checking problem. By a translation of the model checking problem of the modal μ-calculus into a problem of finding fixed-points on boolean lattices, we get a local model checker for two alternating fixed-points which runs in time $O(|A|(|T|^2)\log(|A||T|))$, where $|A|$ is the size of the assertion and $|T|$ the size of the model, a labelled transition system. This extends earlier results by the author and improves on earlier published local algorithms. We also sketch how the algorithm can be extended to arbitrary alternations.

Due to the generality of the algorithm it can be applied to other (alternating or non-alternating) fixed-point problems.

UCAM-CL-TR-261

Neil Anthony Dodgson:

Image resampling

August 1992, 264 pages, PDF
PhD thesis (Wolfson College)

Abstract: Image resampling is the process of geometrically transforming digital images. This report considers several aspects of the process.

We begin by decomposing the resampling process into three simpler sub-processes: reconstruction of a continuous intensity surface from a discrete image, transformation of that continuous surface, and sampling of the transformed surface to produce a new discrete image. We then consider the sampling process, and the subsidiary problem of intensity quantisation. Both these are well understood, and we present a summary of existing work, laying a foundation for the central body of the report where the sub-process of reconstruction is studied.

The work on reconstruction divides into four parts, two general and two specific:

1. Piecewise local polynomials: the most studied group of reconstructors. We examine these, and the criteria used in their design. One new derivation is of two piecewise local quadratic reconstructors.

2. Infinite extent reconstructors: we consider these and their local approximations, the problem of finite image size, the resulting edge effects, and the solutions to these problems. Amongst the reconstructors discussed are the interpolating cubic B-spline and the interpolating Bezier cubic. We derive the filter kernels for both of these, and prove that they are the same. Given this kernel we demonstrate how the interpolating cubic B-spline can be extended from a one-dimensional to a two-dimensional reconstructor, providing a considerable speed improvement over the existing method of extension.

3. Fast Fourier transform reconstruction: it has long been known that the fast Fourier transform (FFT) can be used to generate an approximation to perfect scaling of a sample set. Donald Fraser (in 1987) took this result and generated a hybrid FFT reconstructor which can be used for general transformations, not just scaling. We modify Fraser's method to tackle two major problems: its large time and storage requirements, and the edge effects it causes in the reconstructed intensity surface.

4. A priori knowledge reconstruction: first considering what can be done if we know how the original image was sampled, and then considering what can be done with one particular class of image coupled with one particular type of sampling. In this latter case we find that exact reconstruction of the image is possible. This is a surprising result as this class of images cannot be exactly reconstructed using classical sampling theory.

The final section of the report draws all of the strands together to discuss transformations and the resampling process as a whole. Of particular note here is work on how the quality of different reconstruction and resampling methods can be assessed.

UCAM-CL-TR-262

Nick Benton, Gavin Bierman,
Valeria de Paiva:

Term assignment for intuitionistic linear logic (preliminary report)

August 1992, 57 pages, PDF

Abstract: In this paper we consider the problem of deriving a term assignment system for Girard's Intuitionistic Linear Logic for both the sequent calculus and natural deduction proof systems. Our system differs from previous calculi (e.g. that of Abramsky) and has two important properties which they lack. These are the substitution property (the set of valid deductions is closed under substitution) and subject reduction (reduction on terms is well typed).

We define a simple (but more general than previous proposals) categorical model for Intuitionistic Linear Logic and show how this can be used to derive the term assignment system.

We also consider term reduction arising from cut-elimination in the sequent calculus and normalisation in natural deduction. We explore the relationship between these, as well as with the equations which follow from our categorical model.

UCAM-CL-TR-263

C.-H. Luke Ong:

The Lazy Lambda Calculus: an investigation into the foundations of functional programming

August 1992, 256 pages, paper copy
PhD thesis (Imperial College London, May 1988)

UCAM-CL-TR-264

Juanito Camilleri:

CCS with environmental guards

August 1992, 19 pages, PDF

Abstract: This paper investigates an extension of Milner's CCS with agents guarded by propositions on the environment. The agent g $\gg$ E, pronounced E in an environment of which g holds, depends on the set of actions the environment in ready to perform. This dependency is realised by an operational semantics in which transitions carry ready-sets (of the environment) as well as the normal action symbols from CCS. A notion of strong bisimulation is defined on guarded agents via this semantics. It is a congruence and satisfies new equational laws (including a new expansion law) whicyh are shown to be complete for finite guarded agents. The laws are conservative over agents of traditional CCS. The guarding operator $\gg$ provides a dynamic, local, and clean syntactic means of expressing the behaviour of an agent depending on circumstance; it is more expressive than the unless operator presented in [Cam91] and the priority choice operator presented in [Cam90] and [CaW91], and yields a much simpler expansion theorem.

UCAM-CL-TR-265

Juanito Camilleri, Tom Melham:

Reasoning with inductively defined relations in the HOL theorem prover

August 1992, 49 pages, paper copy

UCAM-CL-TR-266

Carole Klein:

Automatic exploitation of OR-parallelism in Prolog

September 1992, 18 pages, PDF

Abstract: A path through a search space can be defined by a sequence of integers called an oracle. The Delphi machine consists of a network of individual workstations co-operating to execute a Prolog program. Using oracles, these machines automatically partition the search space between them, thereby exploiting OR-parallelism. This report provides a brief description of the tree-searching algorithms (control strategies) implemented in the Delphi machine.

UCAM-CL-TR-267

Christine Ernoult, Alan Mycroft:

Untyped strictness analysis

October 1992, 13 pages, PDF

Abstract: We re-express Hudak and Young's higher-order strictness analysis for the untyped λ-calculus in a conceptually simpler and more semantically-based manner. We show our analysis to be a sound abstraction of Hudak and Young's whichis also complete in a sense we make precise.

UCAM-CL-TR-268

Paul W. Jardetzky:

Network file server design for continuous media

October 1992, 101 pages, PostScript
PhD thesis (Darwin College, August 1992)

Abstract: This dissertation concentrates on issues related to the provision of a network based storage facility for digital audio and video data. The goal is to demonstrate that a distributed file service in support of these media may be built without special purpose hardware. The main objective is to identify those parameters that affect file system performance and provide the criteria for making desirable design decisions.

UCAM-CL-TR-269

Alan Mycroft, Arthur Norman:

Optimising compilation

October 1992, 23 pages, PDF

Abstract: This report consists of pre-prints of two tutorial lectures on optimising compilation to be presented at the Czechoslovak 'SOFSEM 92' conference. The first discusses optimising compilers based on dataflow analysis for classical imperative languages like 'C'. The second turns attention to optimisation of lazy functional languages by 'strictness analysis'.

Part 1: Classical imperative languages

This tutorial considers the design of modern machine-independent optimising compilers for classical (C-like) languages. We draw from two sources (1) the literature and lectures by the authors at Cambridge and (2) the 'Norcroft' compiler suite jointly constructed by the authors.

Part 2: Lazy functional languages

This lecture considers the optimisation of functional programming languages (particularly 'lazy' languages) based on 'strictness analysis'. Such optimisations alter evaluation order to allow more efficient translation into von Neumann architecture or to increase the parallelism in a sequentially specified system (such as that implied by lazy-semantics).

UCAM-CL-TR-270

Chaoying Ma:

Designing a universal name service

133 pages, PDF
PhD thesis (Newnham College, October 1992)

Abstract: Generally speaking, naming in computing systems deals with the creation of object identifiers at all levels of system architecture and the mapping among them. Two of the main purposes of having names in computer systems are (a) to identify objects; (b) to accomplish sharing. Without naming no computer system design can be done.

The rapid development in the technology of personal workstations and computer communication networks has placed a great number of demands on designing large computer naming systems. In this dissertation, issues of naming in large distributed computing systems are addressed. Technical aspects as well as system architecture are examined. A design of a Universal Name Service (UNS) is proposed and its prototype implementation is described. Three major issues on designing a global naming system are studied. Firstly, it is observed that none of the existing name services provides enough flexibility in restructuring name spaces, more research has to be done. Secondly it is observed that although using stale naming data (hints) at the application level is acceptable in most cases as long as it is detectable and recoverable, stronger naming data integrity should be maintained to provide a better guarantee of finding objects, especially when a high degree of availability is required. Finally, configuring the name service is usually done in an ad hoc manner, leading to unexpected interruptions or a great deal of human intervention when the system is reconfigured. It is necessary to make a systematic study of automatic configuration and reconfiguration of name services.

This research is based on a distributed computing model, in which a number of computers work cooperatively to provide the service. The contributions include: (a) the construction of a Globally Unique Directory Identifier (GUDI) name space. Flexible name space restructuring is supported by allowing directories to be added to or removed from the GUDI name space. (b) The definition of a two class name service infrastructure which exploits the semantics of naming. It makes the UNS replication control more robust, reliable as well as highly available. (c) The identification of two aspects in the name service configuration: one is concerned with the replication configuration, and the other is concerned with the server configuration. It is notable that previous work only studied these two aspects individually but not in combination. A distinguishing feature of the UNS is that both issues are considered at the design stage and novel methods are used to allow dynamic service configuration to be done automatically and safely.

UCAM-CL-TR-271

Lawrence C. Paulson:

Set theory as a computational logic: I. from foundations to functions

November 1992, 28 pages, PDF, DVI

Abstract: A logic for specification and verification is derived from the axioms of Zermelo-Fraenkel set theory. The proofs are performed using the proof assistant Isabelle. Isabelle is generic, supporting several different logics. Isabelle has the flexibility to adapt to variants of set theory. Its higher-order syntax supports the definition of new binding operators. Unknowns in subgoals can be instantiated incrementally. The paper describes the derivation of rules for descriptions, relations and functions, and discusses interactive proofs of Cantor's

Theorem, the Composition of Homomorphisms challenge, and Ramsey's Theorem. A generic proof assistant can stand up against provers dedicated to particular logics.

UCAM-CL-TR-272

Martin David Coen:

Interactive program derivation

November 1992, 100 pages, PDF, DVI
PhD thesis (St John's College, March 1992)

Abstract: As computer programs are increasingly used in safety critical applications, program correctness is becoming more important; as the size and complexity of programs increases, the traditional approach of testing is becoming inadequate. Proving the correctness of programs written in imperative languages is awkward; functional programming languages, however, offer more hope. Their logical structure is cleaner, and it is practical to reason about terminating functional programs in an internal logic.

This dissertation describes the development of a logical theory called TPT for reasoning about the correctness of terminating functional programs, its implementation using the theorem prover Isabelle, and its use in proving formal correctness. The theory draws both from Martin-Löf's work in type theory and Manna and Waldinger's work in program synthesis. It is based on classical first-order logic, and it contains terms that represent classes of behaviourally equivalent programs, types that denote sets of terminating programs and well-founded orderings. Well-founded induction is used to reason about general recursion in a natural way and to separate conditions for termination from those for correctness.

The theory is implemented using the generic theorem prover Isabelle, which allows correctness proofs to be checked by machine and partially automated using tactics. In particular, tactics for type checking use the structure of programs to direct proofs. Type checking allows both the verification and derivation of programs, reducing specifications of correctness to sets of correctness conditions. These conditions can be proved in typed first-order logic, using well-known techniques of reasoning by induction and rewriting, and then lifted up to TPT. Examples of program termination are asserted and proved, using simple types. Behavioural specifications are expressed using dependent types, and the correctness of programs asserted and then proved. As a non-trivial example, a unification algorithm is specified and proved correct by machine.

The work in this dissertation clearly shows how a classical theory can be used to reason about program correctness, how general recursion can be reasoned about, and how programs can direct proofs of correctness.

UCAM-CL-TR-273

Innes A. Ferguson:

TouringMachines: an architecture for dynamic, rational, mobile agents

November 1992, 206 pages, PDF, PostScript
PhD thesis (Clare Hall, October 1992)

Abstract: It is becoming widely accepted that neither purely reactive nor purely deliberative control techniques are capable of producing the range of behaviours required of intelligent computational or robotic agents in dynamic, unpredictable, multi-agent worlds. We present a new architecture for controlling autonomous, mobile agents – building on previous work addressing reactive and deliberative control methods. The proposed multi-layered control architecture allows a resource-bounded, goal-directed agent to react promptly to unexpected changes in its environment; at the same time it enables the agent to reason predictively about potential conflicts by constructing and projecting causal models or theories which hypothesise other agents' goals and intentions.

The line of research adopted is very much a pragmatic one. A single, common architecture has been implemented which, being extensively parametrized, allows an experimenter to study functionally- and behaviourally-diverse agent configurations. A principal aim of this research is to understand the role different functional capabilities play in constraining an agent's behaviour under varying environmental conditions. To this end, we have constructed an experimental testbed comprising a simulated multi-agent world in which a variety of agent configurations and behaviours have been investigated. Experience with the new control architecture is described.

UCAM-CL-TR-274

Paul Curzon:

Of what use is a verified compiler specification?

November 1992, 23 pages, PDF

Abstract: Program verification is normally performed on source code. However, it is the object code which is executed and so which ultimately must be correct. The compiler used to produce the object code must not introduce bugs. The majority of the compiler correctness literature is concerned with the verification of compiler specifications rather than executable implementations. We discuss different ways that verified specifications can be used to obtain implementations with varying degrees of security. In particular we describe how a specification can be executed by proof. We discuss

how this method can be used in conjunction with an insecure production compiler so as to retain security without slowing the development cycle of application programs. A verified implementation of a compiler in a high-level language is not sufficient to obtain correct object code. The compiler must itself be compiled into a low level language before it can be executed. At first sight it appears we need an already verified compiler to obtain a secure low-level implementation of a compiler. We describe how a low-level implementation of a compiler can be securely obtained from a verified compiler implementation.

UCAM-CL-TR-275

Barney Pell:

Exploratory learning in the game of GO

18 pages, PostScript

Abstract: This paper considers the importance of exploration to game-playing programs which learn by playing against opponents. The central question is whether a learning program should play the move which offers the best chance of winning the present game, or if it should play the move which has the best chance of providing useful information for future games. An approach to addressing this question is developed using probability theory, and then implemented in two different learning methods. Initial experiments in the game of Go suggest that a program which takes exploration into account can learn better against a knowledgeable opponent than a program which does not.

UCAM-CL-TR-276

Barney Pell:

METAGAME: a new challenge for games and learning

15 pages, PostScript

Abstract: In most current approaches to Computer Game-Playing, including those employing some form of machine learning, the game analysis mainly is performed by humans. Thus, we are sidestepping largely the interesting (and difficult) questions. Human analysis also makes it difficult to evaluate the generality and applicability of different approaches.

To address these problems, we introduce a new challenge: Metagame. The idea is to write programs which take as input the rules of a set of new games within a pre-specified class, generated by a program which is publicly available. The programs compete against each other in many matches on each new game, and they can then be evaluated based on their overall performance and improvement through experience.

This paper discusses the goals, research areas, and general concerns for the idea of Metagame.

UCAM-CL-TR-277

Barney Pell:

METAGAME in symmetric chess-like games

30 pages, PostScript

Abstract: I have implemented a game generator that generates games from a wide but still restricted class. This class is general enough to include most aspects of many standard games, including Chess, Shogi, Chinese Chess, Checkers, Draughts, and many variants of Fairy Chess. The generator, implemented in Prolog is transparent and publicly available, and generates games using probability distributions for parameters such as piece complexity, types of movement, board size, and locality.

The generator is illustrated by means of a new game it produced, which is then subjected to a simple strategic analysis. This form of analysis suggests that programs to play Metagame well will either learn or apply very general game-playing principles. But because the class is still restricted, it may be possible to develop a naive but fast program which can outplay more sophisticated opponents. Performance in a tournament between programs is the deciding criterion.

UCAM-CL-TR-278

Monica Nesi:

A formalization of the process algebra CCS in high order logic

42 pages, PDF

Abstract: This paper describes a mechanization in higher order logic of the theory for a subset of Milner's CCS. The aim is to build a sound and effective tool to support verification and reasoning about process algebra specifications. To achieve this goal, the formal theory for pure CCS (no value passing) is defined in the interactive theorem prover HOL, and a set of proof tools, based on the algebraic presentation of CCS, is provided.

UCAM-CL-TR-279

Victor A. Carreño:

The transition assertions specification method

December 1992, 18 pages, PDF

Abstract: A modelling and specification method for real-time, reactive systems is described. Modelling is performed by constructing time dependent relations of the system parameters. A textual formal notation using higher order logic and a graphical notation are presented. The formal notation allows the use of rigorous mathematical methods on the specification, one of the primary sources of design errors. A cruise control case example is included in the paper and the HOL mechanised theorem prover is used to show that the specification comply with some top level requirements.

UCAM-CL-TR-280

Lawrence C. Paulson:

Introduction to Isabelle

January 1993, 61 pages, DVI

Abstract: Isabelle is a generic theorem prover, supporting formal proof in a variety of logics. Through a variety of examples, this paper explains the basic theory demonstrates the most important commands. It serves as the introduction to other Isabelle documentation.

UCAM-CL-TR-281

Sape J. Mullender, Ian M. Leslie, Derek McAuley:

Pegasus project description

September 1992, 23 pages, paper copy

UCAM-CL-TR-282

Ian M. Leslie, Derek McAuley, Sape J. Mullender:

Pegasus – Operating system support for distributed multimedia systems

December 1992, 14 pages, paper copy

UCAM-CL-TR-283

Lawrence C. Paulson:

The Isabelle reference manual

February 1993, 78 pages, DVI

Abstract: This manual is a comprehensive description of Isabelle, including all commands, functions and packages. It is intended for reference rather than for reading through, and is certainly not a tutorial. The manual assumes familiarity with the basic concepts explained in Introduction to Isabelle. Functions are organized by their purpose, by their operands (subgoals, tactics, theorems), and by their usefulness. In each section, basic functions appear first, then advanced functions, and finally esoteric functions.

UCAM-CL-TR-284

Claire Grover, John Carroll, Ted Briscoe:

The Alvey Natural Language Tools grammar (4th Release)

January 1993, 260 pages, paper copy

UCAM-CL-TR-285

Andrew Donald Gordon:

Functional programming and input/output

February 1993, 163 pages, paper copy
PhD thesis (King's College, August 1992)

UCAM-CL-TR-286

Lawrence C. Paulson:

Isabelle's object-logics

February 1993, 161 pages, DVI

Abstract: Several logics come with Isabelle. Many of them are sufficiently developed to serve as comfortable reasoning environments. They are also good starting points for defining new logics. Each logic is distributed with sample proofs, some of which are presented in the paper. The logics described include first-order logic, Zermelo-Fraenkel set theory, higher-order logic, constructive type theory, and the classical sequent calculus LK. A final chapter explains the fine points of defining logics in Isabelle.

UCAM-CL-TR-287

Andrew D. Gordon:

A mechanised definition of Silage in HOL

February 1993, 28 pages, DVI

Abstract: If formal methods of hardware verification are to have any impact on the practices of working engineers, connections must be made between the languages used in practice to design circuits, and those used for research into hardware verification. Silage is a simple dataflow language marketed for specifying digital signal processing circuits. Higher Order Logic (HOL) is extensively used for research into hardware verification. This paper presents a formal definition of a substantial subset of Silage, by mapping Silage declarations into HOL predicates. The definition has been mechanised in the HOL theorem prover to support the transformational design of Silage circuits as theorem proving in HOL.

UCAM-CL-TR-288

Rajeev Gore:

Cut-free sequent and tableau systems for propositional Diodorean modal logics

February 1993, 19 pages, PDF

Abstract: We present sound, (weakly) complete and cut-free tableau systems for the propositional normal modal logics S4.3, S4.3.1 and S4.14. When the modality $\Box$ is given a temporal interpretation, these logics respectively model time as a linear dense sequence of points; as a linear discrete sequence of points; and as a branching tree where each branch is a linear discrete sequence of points.

Although cut-free, the last two systems do not possess the subformula property. But for any given finite set of formulae X the "superformulae" involved are always bounded by a finite set of formulae X*L depending only on X and the logic L. Thus each system gives a nondeterministic decision procedure for the logic in question. The completeness proofs yield deterministic decision procedures for each logic because each proof is constructive.

Each tableau system has a cut-free sequent analogue proving that Gentzen's cut-elimination theorem holds for these logics. The techniques are due to Hintikka and Rautenberg.

UCAM-CL-TR-289

David Alan Howard Elworthy:

The semantics of noun phrase anaphora

February 1993, 191 pages, PDF
PhD thesis (Darwin College, February 1993)

Abstract: Anaphora is a linguistic phenomenon in which one expression, called an anaphor, gains some or all of its meaning from another, its antecedent. In this thesis, I study the semantics of one particular sort of anaphor, where both antecedent and anaphor are noun phrases. Most research in the past has dealt with singular anaphora; I also address plurals.

The two major theories of anaphora are Kamp's Discourse Representation Theory (DRT) and dynamic logics. While they have yielded many valuable insights into the phenomenon, I think it is time to subject them to some critical scrutiny. There are two main criticisms. Firstly, the interpretation assigned to the lingustic data is not always consistent with language users' intuitions about it. Secondly the current theories employ semantic formalisms which rely on either specific representational devices or on unconventional logics. I develop a new theory, TAI (Theory of Anaphoric Information), which attempts to rectify both problems.

This thesis starts with a critical re-examination of the linguistic data, and in particular of the so-called "donkey sentences", which exhibit complex interactions between quantification and anaphora. The following chapter examines DRT and dynamic logics in some detail, considering their successes and failings from both empirical and methodological perspectives.

TAI itself is presented in chapter 4. The theory starts from a conceptual model, which specifies the information needed to interpret anaphors correctly. A logic, L(GQA), is then developed, which derives both truth conditions and constraints on the anaphoric information from formulae derived from natural language sentences. The logic is static and does not rely on structured representations of the sort found in DRT. The translation procedure from linguistic input to L(GQA) formulae captures a significant part of the emprirical weight of the theory, and provides sufficient flexibility to make the required range of readings available.

The last chapter evaluates TAI from a variety of standpoints. The conceptual model is used as a baseline for comparing DRT, dynamic logics and TAI. The relation between semantic logics of TAI and pragmatic aspects of interpreting anaphors is considered. Computational aspects of TAI are also examined: how it relates to Webber's theory of anaphora, and how the logic could be implemented efficiently. Finally, some directions in which research based on TAI could proceed are identified.

UCAM-CL-TR-290

Karen Spärck Jones:

Discourse modelling for automatic summarising

February 1993, 30 pages, PDF

Abstract: Automatic abstracting is a challenging task for natural language processing. It depends not only on sentence interpretation and the local context representation this requires, but also on the recognition and use of large-scale discourse structure. This paper describes research investigating the nature of different approaches to discourse representation and their value for summarising. This work is focussed on comparative analysis, illustrated in the paper through the provision of different forms of representation, and different strategies for summary formation, for a short example text.

UCAM-CL-TR-291

J.R. Galliers, K. Spärck Jones:

Evaluating natural language processing systems

February 1993, 187 pages, PostScript

Abstract: This report presents a detailed analysis and review of NLP evaluation, in principle and in practice. Part 1 examines evaluation concepts and establishes a framework for NLP system evaluation. This makes use of experience in the related area of information retrieval and the analysis also refers to evaluation in speech processing. Part 2 surveys significant evaluation work done so far, for instance in machine translation, and discusses the particular problems of generic system evaluation. The conclusion is that evaluation strategies and techniques for NLP need much more development, in particular to take proper account of the influence of system tasks and settings. Part 3 develops a general approach to NLP evaluation, aimed at methodologically-sound strategies for test and evaluation motivated by comprehensive performance factor identification. The analysis throughout the report is supported by extensive illustrative examples.

UCAM-CL-TR-292

Cormac John Sreenan:

Synchronisation services for digital continuous media

March 1993, 123 pages, PostScript
PhD thesis (Christ's College, October 1992)

Abstract: The development of broadband ATM networking makes it attractive to use computer communication networks for the transport of digital audio and motion video. Coupled with advances in workstation technology, this creates the opportunity to integrate these continuous information media within a distributed computing system. Continuous media have an inherent temporal dimension, resulting in a set of synchronisation requirements which have real-time constraints. This dissertation identifies the role and position of synchronisation, in terms of the support which is necessary in an integrated distributed system. This work is supported by a set of experiments which were performed in an ATM inter-network using multi-media workstations, each equipped with an Olivetti Pandora Box.

UCAM-CL-TR-293

Jean Bacon, Ken Moody:

Objects and transactions for modelling distributed applications: concurrency control and commitment

April 1993, 39 pages, PDF

Abstract: The concepts of object and transaction form an ideal basis for reasoning about the behaviour of distributed applications. An object model allows the semantics of an application to be used to specify the required concurrency behaviour of each object. A transaction model covers multi-component computations where the components are distributed and therefore subject to concurrent execution and partial failure.

This tutorial establishes an object model for a distributed system in which transactions are used. It focusses on the alternative methods of concurrency control that might be employed and shows how each method might be appropriate for certain application characteristics and system behaviour. The background for this discussion is eatablished in [Bacon 1993].

UCAM-CL-TR-294

Ken Moody, Jean Bacon, Noha Adly, Mohamad Afshar, John Bates, Huang Feng, Richard Hayton, Sai Lai Lo, Scarlet Schwiderski, Robert Sultana, Zhixue Wu:

OPERA
Storage, programming and display of multimedia objects

April 1993, 9 pages, PDF

Abstract: This project aims to support the interactive display of synchronised multiple media types in workstation windows. This style of application needs high speed ATM networks and suitable protocols and operating systems; an infrastructure that exists at the University of Cambridge Computer Laboratory. Above this infrastructure we have designed and are building storage services (MSSA), a platform to support the creation and display of multimedia presentations (IMP) and a persistent programming language (PC++), for reliable and convenient programming of multimedia applications. This paper gives an overview of the work of the OPERA project in these three areas.

UCAM-CL-TR-295

Jean Bacon, John Bates, Sai Lai Lo, Ken Moody:

OPERA
Storage and presentation support for multimedia applications in a distributed, ATM network environment

April 1993, 12 pages, PDF

Abstract: We are building a display platform for multimedia applications above a multi-service storage architecture (MSSA). This style of application needs high speed ATM networks and suitable protocols and operating systems; an infrastructure that exists at the University of Cambridge Computer Laboratory.

An open storage architecture gives flexibility and extensibility. Conventional files, audio, video and structured objects are supported within a common architectural framework and composite objects, such as a display representation, may have components of any of these storage types. The two-level hierarchy of servers provides storage media and a byte-segment abstraction at the low level and a variety of abstractions at the high level. Quality of service guarantees, which are essential for continuous media file types, are supported by sessions and tickets. These are arranged via the high level servers and used directly with the low level servers.

A platform for the creation and interactive display of multimedia presentations (IMP) is being developed. A script language allows a multimedia presentation to be specified in terms of objects, the relationships between them and the (composite) events that drive it. Presentation data is stored on the structured data service of MSSA and component objects are stored on appropriate servers, and accepted and retrieved at guaranteed rates. The presentation requirements of an application are managed by applying a script to the data representing the presentation to create a display.

UCAM-CL-TR-296

Z. Wu, K. Moody, J. Bacon:

A persistent programming language for multimedia databases in the OPERA project

April 1993, 9 pages, PDF

Abstract: The advent of high bandwidth local area ATM networks has transformed the potential of distributed computing systems. At the Computer Laboratory we are moving towards a world in which multimedia displays are managed by editing, browsing and composing tools [Bates 93]. The recently completed Pandora project [Hopper 90] has given us some experience of multimedia applications, and an idea of their scope.

We have developed a persistent programming language PC++ [Wu 93], an extension of C++, to help programmers developing multimedia applications to make use of the MSSA. In this paper we present the design of PC++ and show how its special features meet the requiremets to effectively manage data in a distributed, real-time, context.

UCAM-CL-TR-297

Eike Ritter:

Categorical abstract machines for higher-order typed lambda calculi

April 1993, 149 pages, PDF
PhD thesis (Trinity College)

Abstract: We define in this thesis categorical abstract machines for the calculus of constructions, a special higher-order lambda-calculus. We start with the derivation of categorical combinators, i.e. an equational theory based on a categorical structure for the calculus. It turns out that only a generalization of Ehrhard's D-categories can be used for this purpose; all other categorical structures modelling the calculus yield only conditional equations or no equations at all. Next we orient the equations to obtain reduction rules. When we want to show that this reduction corresponds to reduction in the calculus, we run into difficulties in proving strong normalization. We can only show that any reduction that leads first to a combinator corresponding to a weak head-normal form is finite. These results are the key to formulate an eager and a lazy strategy for the reduction of a combinator to its normal form.

We then construct abstract machines for the eager and lazy strategy. Their correctness proof consists of an induction over the definition of the reduction strategies. These machines specialize to the CAM and Krivine's machine in the first order case respectively. The original construction of the CAM is based on cartesian closed categories (CCCs). They model both environments and terms by morphisms regardless of their conceptual difference, whereas the D-categories separate these two notions. Hence the correspondence between the D-categories and the abstract machines described in this thesis is closer than that between the CAM and the CCCs. We also obtain an abstract machine for type checking of these combinators, which uses the above reduction machines. Preliminary tests suggest that the abstract machines are quite efficient compared to other implementations.

UCAM-CL-TR-298

John Matthew Simon Doar:

Multicast in the asynchronous transfer mode environment

April 1993, 168 pages, PostScript
PhD thesis (St John's College, January 1993)

Abstract: In future multimedia communication networks, the ability to multicast information will be useful for many new and existing services. This dissertation considers the design of multicast switches for Asynchronous Transfer Mode (ATM) networks and

proposes one design based upon a slotted ring. Analysis and simulation studies of this design are presented and details of its implementation for an experimental ATM network (Project Fairisle) are described, together with the modifications to the existing multi-service protocol architecture necessary to provide multicast connections. Finally, a short study of the problem of multicast routing is presented, together with some simulations of the long-term effect upon the routing efficiency of modifying the number of destinations within a multicast group.

UCAM-CL-TR-299

Bjorn Gamback, Manny Rayner, Barney Pell:

Pragmatic reasoning in bridge

April 1993, 23 pages, PostScript

Abstract: In this paper we argue that bidding in the game of Contract Bridge can profitably be regarded as a micro-world suitable for experimenting with pragmatics. We sketch an analysis in which a "bidding system" is treated as the semantics of an artificial language, and show how this "language", despite its apparent simplicity, is capable of supporting a wide variety of common speech acts parallel to those in natural languages; we also argue that the reason for the relatively unsuccessful nature of previous attempts to write strong Bridge playing programs has been their failure to address the need to reason explicitly about knowledge, pragmatics, probabilities and plans. We give an overview of Pragma, a system currently under development, which embodies these ideas in concrete form, using a combination of rule-based inference, stochastic simulation, and "neural-net" learning. Examples are given illustrating the functionality of the system in its current form.

UCAM-CL-TR-300

Wai Wong:

Formal verification of VIPER's ALU

April 1993, 78 pages, PDF

Abstract: This research describes the formal verification of an arithmetic logic unit of the VIPER microprocessor. VIPER is one of the first processors designed using formal methods. A formal model in HOL has been created which models the ALU at two levels: on the higher level, the ALU is specified as a function taking two 32-bit operands and returning a result; on the lower level the ALU is implemented by a number of 4-bit slices which should take the same operands and return the same results. The ALU is capable of performing thirteen different operations. A formal proof of functional equivalence of these two levels has been completed successfully. The complete HOL text of the ALU formal model and details of the proof procedures are included in this report. It has demonstrated that the HOL system is powerful and efficient enough to perform formal verification of realistic hardware design.

UCAM-CL-TR-301

Zhixue Wu, Ken Moody, Jean Bacon:

The dual-level validation concurrency control method

June 1993, 24 pages, PDF

Abstract: Atomic data types permit maximum concurrency among transations by exploiting the semantics of object operations. Concurrency control is needed to ensure both object level atomicity and transaction level atomicity. It must be possible to regard each operation on an object as elementary. Recovery methods for transactions which are based on atomic objects must take into account that partial results of a transaction might be seen by other transactions.

This paper presents, formalises and verifies a protocol called the dual-level validation method which can be used to provide atomicity for atomic data types. It is optimistic and has a number of advantages over previous methods. It permits maximum concurrency at the low level by allowing non-conflicting operations to be scheduled concurrently. It allows applications to cope with very large objects by supporting multi-granularity shadowing. Transaction recovery is simple to implement. The method performs well, particularly when different transactions are unlikely to access the same (sub)objects concurrently. Finally, it is well suited to a distributed environment since validation and commit are not implemented atomically.

UCAM-CL-TR-302

Barney Pell:

Logic programming for general game-playing

June 1993, 15 pages, PostScript

Abstract: Meta-Game Playing is a new approach to games in Artificial Intelligence, where we construct programs to play new games in a well-defined class, which are output by an automatic game generator. As the specific games to be played are not known in advance, a degree of human bias is eliminated, and playing programs are required to perform any game-specific optimisations without human assistance.

The attempt to construct a general game-playing program is made difficult by the opposing goals of generality and efficiency. This paper shows how application of standard techniques in logic-programming (abstract interpretation and partial evaluation) makes it possible to achieve both of these goals. Using these

techniques, we can represent the semantics of a large class of games in a general and declarative way, but then have the program transform this representation into a more efficient version once it is presented with the rules of a new game. This process can be viewed as moving some of the responsibility for game analysis (that concerned with efficiency) from the researcher to the program itself.

UCAM-CL-TR-303

Andrew Kennedy:

Drawing trees — a case study in functional programming

June 1993, 9 pages, PDF

Abstract: This report describes the application of functional programming techniques to a problem previously studied by imperative programmers, that of drawing general trees automatically. We first consider the nature of the problem and the ideas behind its solution, independent of programming language implementation. The functional language implementation is described in a bottom up style starting with very general functions over trees and then narrowing in on the particular tree layout algorithm. Its correctness is considered informally. Finally we discuss the implementation's computational complexity and possible improvements.

UCAM-CL-TR-304

Lawrence C. Paulson:

Co-induction and co-recursion in higher-order logic

July 1993, 35 pages, PDF, PostScript, DVI

Abstract: A theory of recursive and corecursive definitions has been developed in higher-order logic (HOL) and mechanised using Isabelle. Least fixedpoints express inductive data types such as strict lists; greatest fixedpoints express co-inductive data types, such as lazy lists. Well-founded recursion expresses recursive functions over inductive data types; co-recursion expresses functions that yield elements of co-inductive data types. The theory rests on a traditional formalization of infinite trees. The theory is intended for use in specification and verification. It supports reasoning about a wide range of computable functions, but it does not formalize their operational semantics and can express non-computable functions also. The theory is demonstrated using lists and lazy lists as examples. The emphasis is on using co-recursion to define lazy list functions, and on using co-induction to reason about them.

UCAM-CL-TR-305

P.N. Benton:

Strong normalisation for the linear term calculus

July 1993, 13 pages, PDF

Abstract: We provide a strong normalisation result for the linear term calculus which was introduced in (Benton et al. 1992). Rather than prove the result from first principles, we give a translation of linear terms into terms in the second order polymorphic lambda calculus ($\lambda 2$) which allows the result to be proved by appealing to the well known strong normalisation property of $\lambda 2$. An interesting feature of the translation is that it makes use of the $\lambda 2$ coding of a coinductive datatype as the translation of the !-types (exponentials) of the linear calculus.

UCAM-CL-TR-306

Wai Wong:

Recording HOL proofs

July 1993, 57 pages, PDF

Abstract: This paper describes a text file format for recording HOL proofs. It is intended to become an interface between HOL and proof checkers. Modification to HOL-88 has been carried out to incorporate a proof recorder to generate a proof file in this format. The usage of this new feature is explained by a simple example. A more substantial proof has been recorded, and benchmark data is presented here.

UCAM-CL-TR-307

David D. Lewis, Karen Spärck Jones:

Natural language processing for information retrieval

July 1993, 22 pages, PostScript

Abstract: The paper summarizes the essential properties of document retrieval and reviews both conventional practice and research findings, the latter suggesting that simple statistical techniques can be effective. It then considers the new opportunities and challenges presented by the ability to search full text directly (rather than e.g. titles and abstracts), and suggests appropriate approaches to doing this, with a focus on the role of natural language processing. The paper also comments on possible connections with data and knowledge retrieval, and concludes by emphasizing the importance of rigorous performance testing.

UCAM-CL-TR-308

Jacob Frost:

A case study of co-induction in Isabelle HOL

August 1993, 27 pages, PDF, PostScript, DVI

Abstract: The consistency of the dynamic and static semantics for a small functional programming language was informally proved by R. Milner and M. Tofte. The notions of co-inductive definitions and the associated principle of co-induction played a pivotal role in the proof. With emphasis on co-induction, the work presented here deals with the formalisation of this result in the higher-order logic of the generic theorem prover Isabelle.

UCAM-CL-TR-309

Peter Nicholas Benton:

Strictness analysis of lazy functional programs

August 1993, 154 pages, PDF
PhD thesis (Pembroke College, December 1992)

Abstract: Strictness analysis is a compile-time analysis for lazy functional languages. The information gained by a strictness analyser can be used to improve code generation for both sequential and parallel implementations of such languages.

After reviewing the syntax and semantics of a simply typed lambda calculus with constants, we describe previous work on strictness analysis. We then give a new formulation of higher order strictness analysis, called strictness logic. This is inspired by previous work on static analysis by non-standard type inference, and by work on logic of domains. We investigate some proof theoretic and semantic properties of our logic, and relate it to the conventional approach using abstract interpretation. We also consider extending the logic with disjunction.

We then describe how to extend the simply typed lambda calculus with lazy algebraic datatyped. A new construction of lattices of strictness properties of such datatypes is described. This arises from the characterisation of the solutions to the recursive domain equations associated with these types as initial algebras.

Next we consider first order (ML-style) ploymorphism and show how Wadler's 'theorems for free' parametricity results may be obtained from a simple extension of the semantics of monomorphic language. We then prove a polymorphic invariance result relating the derivable strictness properties of different substitution instances of polymorphic terms.

UCAM-CL-TR-310

Noha Adly:

HARP: a hierarchical asynchronous replication protocol for massively replicated systems

August 1993, 34 pages, PostScript

Abstract: This paper presents a new asynchronous replication protocol that is especially suitable for wide area and mobile systems, and allows reads and writes to occur at any replica. Updates reach other replicas using a propagation scheme based on nodes organized into a logical hierarchy. The hierarchical structure enables the scheme to scale well for thousands of replicas, while ensuring reliable delivery. A new service interface is proposed that provides different levels of asynchrony, allowing strong consistency and weak consistency to be integrated into the same framework. Further, due to the hierarchical pattern of propagation, the scheme provides the ability to locate replicas that are more up-to-date than others, depending on the needs of various applications. Also, it allows a selection from a number of reconciliation techniques based on delivery order mechanisms. Restructuring operations are provided to build and reconfigure the hierarchy dynamically without disturbing normal operations. The scheme tolerates transmission failures and network partitions.

UCAM-CL-TR-311

Paul Curzon:

A verified Vista implementation

September 1993, 56 pages, PDF

Abstract: We describe the formal verification of a simple compiler using the HOL theorem proving system. The language and microprocessor considered are a subset of the structured assembly language Vista, and the Viper microprocessor, respectively. We describe how our work is directly applicable to a family of languages and compilers and discuss how the correctness theorem and verified compiler fit into a wider context of ensuring that object code is correct. We first show how the compiler correctness result can be formally combined with a proof system for application programs. We then show how our verified compiler, despite not being written in a traditional programming language, can be used to produce compiled code. We also discuss how a dependable implementation might be obtained.

UCAM-CL-TR-312

Lawrence C. Paulson:

Set theory for verification: II Induction and recursion

September 1993, 46 pages, PDF

Abstract: A theory of recursive definitions has been mechanized in Isabelle's Zermelo-Fraenkel (ZF) set theory. The objective is to support the formalization of particular recursive definitions for use in verification, semantics proofs and other computational reasoning.

Inductively defined sets are expressed as least fixedpoints, applying the Knaster-Tarski Theorem over a suitable set. Recursive functions are defined by well-founded recursion and its derivatives, such as transfinite recursion. Recursive data structures are expressed by applying the Knaster-Tarski Theorem to a set that is closed under Cartesian product and disjoint sum.

Worked examples include the transitive closure of a relation, lists, variable-branching trees and mutually recursive trees and forests. The Schröder-Bernstein Theorem and the soundness of propositional logic are proved in Isabelle sessions.

UCAM-CL-TR-313

Yves Bertot, Gilles Kahn, Laurent Théry:

Proof by pointing

October 1993, 27 pages, PDF

Abstract: A number of very powerful and elegant computer programs to assist in making formal proofs have been developed. While these systems incorporate ever more sophisticated tactics, proofs that can be carried out without any user directions are the exception. In this paper we present a principle called proof by pointing that allows the user to guide the proof process using the mouse in the user-interface. This idea is widely applicable and has been implemented by the authors in user-interfaces for several proof development systems.

UCAM-CL-TR-314

John Andrew Carroll:

Practical unification-based parsing of natural language

173 pages, PostScript
PhD thesis (September 1993)

Abstract: The thesis describes novel techniques and algorithms for the practical parsing of realistic Natural Language (NL) texts with a wide-coverage unification-based grammar of English. The thesis tackles two of the major problems in this area: firstly, the fact that parsing realistic inputs with such grammars can be computationally very expensive, and secondly, the observation that many analyses are often assigned to an input, only one of which usually forms the basis of the correct interpretation.

The thesis starts by presenting a new unification algorithm, justifies why it is well-suited to practical NL parsing, and describes a bottom-up active chart parser which employs this unification algorithm together with several other novel processing and optimisation techniques. Empirical results demonstrate that an implementation of this parser has significantly better practical performance than a comparable, state-of-the-art unification-based parser. Next, techniques for computing an LR table for a large unification grammar are described, a context free non-deterministic LR parsing algorithm is presented which has better time complexity than any previously reported using the same approach, and a unification-based version is derived. In experiments, the performance of an implementation of the latter is shown to exceed both the chart parser and also that of another efficient LR-like algorithm recently proposed.

Building on these methods, a system for parsing text taken from a given corpus is described which uses probabilistic techniques to identify the most plausible syntactic analyses for an input from the often large number licensed by the grammar. New techniques implemented include an incremental approach to semi-supervised training, a context-sensitive method of scoring sub-analyses, the accurate manipulation of probabilities during parsing, and the identification of the highest ranked analyses without exhaustive search. The system attains a similar success rate to approaches based on context-free grammar, but produces analyses which are more suitable for semantic processing.

The thesis includes detailed analyses of the worst-case space and time complexities of all the main algorithms described, and discusses the practical impact of the theoretical complexity results.

UCAM-CL-TR-315

Barney Darryl Pell:

Strategy generation and evaluation for meta-game playing

November 1993, 289 pages, PostScript
PhD thesis (Trinity College, August 1993)

Abstract: Meta-Game Playing (METAGAME) is a new paradigm for research in game-playing in which we design programs to take in the rules of unknown games and play those games without human assistance. Strong

performance in this new paradigm is evidence that the program, instead of its human designer, has performed the analysis of each specific game.

SCL-METAGAME is a concrete METAGAME research problem based around the class of symmetric chess-like games. The class includes the games of chess, checkers, noughts and crosses, Chinese-chess, and Shogi. An implemented game generator produces new games in this class, some of which are objects of interest in their own right.

METAGAMER is a program that plays SCL-METAGAME. The program takes as input the rules of a specific game and analyses those rules to construct for that game an efficient representation and an evaluation function, both for use with a generic search engine. The strategic analysis performed by the program relates a set of general knowledge sources to the details of the particular game. Among other properties, this analysis determines the relative value of the different pieces in a given game. Although METAGAMER does not learn from experience, the values resulting from its analysis are qualitatively similar to values used by experts on known games, and are sufficient to produce competitive performance the first time the program actually plays each game it is given. This appears to be the first program to have derived useful piece values directly from analysis of the rules of different games.

Experiments show that the knowledge implemented in METAGAMER is useful on games unknown to its programmer in advance of the competition and make it seem likely that future programs which incorporate learning and more sophisticated active-analysis techniques will have a demonstrable competitive advantage on this new problem. When playing the known games of chess and checkers against humans and specialised programs, METAGAMER has derived from more general principles some strategies which are familiar to players of those games and which are hard-wired in many game-specific programs.

UCAM-CL-TR-316

Ann Copestake:

The Compleat LKB

August 1993, 126 pages, PostScript

Abstract: This report is a full description of the lexical knowledge base system (LKB) and the representation language (LRL) developed on the Esprit ACQUILEX project. The LKB system is designed to allow the representation of multilingual lexical information in a way which integrates lexical semantics with syntax and formal semantics. The LRL is a typed feature structure language which makes it possible to represent the lexicon as a highly structured object and to capture relationships between individual word senses by (default) inheritance and by lexical rules. The extension to multilingual representation allows a concise and natural description of translation mismatches. Most of this report consists of a detailed formal description of the LRL — this is augmented with appendices containing the user manual, an implementation outline and a discussion of some of the algorithms used, and a bibliography of papers which describe the LKB and its use within ACQUILEX. (Some of this material has been published previously, but is included here to make this report a convenient reference source.)

UCAM-CL-TR-317

John Peter Van Tassel:

Femto-VHDL: the semantics of a subset of VHDL and its embedding in the HOL proof assistant

November 1993, 122 pages, PDF
PhD thesis (Gonville & Caius College, July 1993)

Abstract: The design of digital devices now resembles traditional computer programming. Components are specified in a specialised form of programming language known as a Hardware Description Language. Programs written in such languages are then executed to simulate the behaviour of the hardware they describe. These simulations cannot be exhaustive in most situations, so result in high, yet incomplete, confidence that the proper behaviour has been achieved.

The formal analysis of programming languages provides ways of mathematically proving properties of programs. These properties apply to behaviours resulting from all possible inputs rather then just a subset of them. The prerequisite for such an analysis is a formal understanding of the semantics of the language.

The Very High Speed Hardware Description Language (VHDL) is currently used to specify and simulate a wide range of digital devices. The language has no formal mathematical semantics as part of its definition, hence programs written in it have not been amenable to formal analysis.

The work presented here defines a structural operational semantics for a subset of VHDL. The semantics is then embedded in a mechanical proof assistant. This mechanisation allows one not only to reason about individual programs but also to express equivalences between programs. Examples which highlight the methodology used in this reasoning are provided as a series of case studies.

UCAM-CL-TR-318

Jim Grundy:

A method of program refinement

November 1993, 207 pages, PostScript
PhD thesis (Fitzwilliam College, November 1993)

Abstract: A method of specifying the desired behaviour of a computer program, and of refining such specifications into imperative programs is proposed. The refinement method has been designed with the intention of being amenable to tool support, and of being applicable to real-world refinement problems.

Part of the refinement method proposed involves the use of a style of transformational reasoning called 'window inference'. Window inference is particularly powerful because it allows the information inherent in the context of a subexpression to be used in its transformation. If the notion of transformational reasoning is generalised to include transformations that preserve relationships weaker than equality, then program refinement can be regarded as a special case of transformational reasoning. A generalisation of window inference is described that allows non-equivalence preserving transformations. Window inference was originally proposed independently from, and as an alternative to, traditional styles of reasoning. A correspondence between the generalised version of window inference and natural deduction is described. This correspondence forms the basis of a window inference tool that has been built on top of the HOL theorem proving system.

This dissertation adopts a uniform treatment of specifications and programs as predicates. A survey of the existing approaches to the treatment of programs as predicates is presented. A new approach is then developed based on using predicates of a three-valued logic. This new approach can distinguish more easily between specifications of terminating and nonterminating behaviour than can the existing approaches.

A method of program refinement is then described by combining the unified treatment of specifications and programs as three-valued predicates with the window inference style of transformational reasoning. The result is a simple method of refinement that is well suited to the provision of tool support.

The method of refinement includes a technique for developing recursive programs. The proof of such developments is usually complicated because little can be assumed about the form and termination properties of a partially developed program. These difficulties are side-stepped by using a simplified meaning for recursion that compels the development of terminating programs. Once the development of a program is complete, the simplified meaning for recursion is refined into the true meaning.

The dissertation concludes with a case study which presents the specification and development of a simple line-editor. The case study demonstrates the applicability of the refinement method to real-world problems. The line editor is a nontrivial example that contains features characteristic of large developments, including complex data structures and the use of data abstraction. Examination of the case study shows that window inference offers a convenient way of structuring large developments.

UCAM-CL-TR-319

Mark David Hayter:

A workstation architecture to support multimedia

November 1993, 99 pages, PostScript
PhD thesis (St John's College, September 1993)

Abstract: The advent of high speed networks in the wide and local area enables multimedia traffic to be easily carried between workstation class machines. The dissertation considers an architecture for a workstation to support such traffic effectively. In addition to presenting the information to a human user the architecture allows processing to be done on continuous media streams.

The proposed workstation architecture, known as the Desk Area Network (DAN), extends ideas from Asynchronous Transfer Mode (ATM) networks into the end-system. All processors and devices are connected to an ATM interconnect. The architecture is shown to be capable of supporting both multimedia data streams and more traditional CPU cache line traffic. The advocated extension of the CPU cache which allows caching of multimedia data streams is shown to provide a natural programming abstraction and a mechanism for synchronising the processor with the stream.

A prototype DAN workstation has been built. Experiments have been done to demonstrate the features of the architecture. In particular the use of the DAN as a processor-to-memory interconnect is closely studied to show the practicality of using ATM for cache line traffic in a real machine. Simple demonstrations of the stream cache ideas are used to show its utility in future applications.

UCAM-CL-TR-320

Lawrence C. Paulson:

A fixedpoint approach to implementing (co)inductive definitions (updated version)

July 1995, 29 pages, PDF, DVI

Abstract: Several theorem provers provide commands for formalizing recursive datatypes or inductively defined sets. This paper presents a new approach, based on fixedpoint definitions. It is unusually general: it admits all monotone inductive definitions. It is conceptually simple, which has allowed the easy implementation of mutual recursion and other conveniences. It also handles coinductive definitions: simply replace the least fixedpoint by a greatest fixedpoint. This represents the first automated support for coinductive definitions.

The method has been implemented in Isabelle's formalization of ZF set theory. It should be applicable to

any logic in which the Knaster-Tarski Theorem can be proved. The paper briefly describes a method of formalizing non-well-founded data structures in standard ZF set theory.

Examples include lists of n elements, the accessible part of a relation and the set of primitive recursive functions. One example of a coinductive definition is bisimulations for lazy lists. Recursive datatypes are examined in detail, as well as one example of a "codatatype": lazy lists. The appendices are simple user's manuals for this Isabelle/ZF package.

UCAM-CL-TR-321

Andrew M. Pitts:

Relational properties of domains

December 1993, 38 pages, PostScript

Abstract: New tools are presented for reasoning about properties of recursively defined domains. We work within a general, category-theoretic framework for various notions of 'relation' on domains and for actions of domain constructors on relations. Freyd's analysis of recursive types in terms of a property of mixed initiality/finality is transferred to a corresponding property of invariant relations. The existence of invariant relations is proved under completeness assumptions about the notion of relation. We show how this leads to simpler proofs of the computational adequacy of denotational semantics for functional programming languages with user-declared datatypes. We show how the initiality/finality property of invariant relations can be specialized to yield an induction principle for admissible subsets of recursively defined domains, generalizing the principle of structural induction for inductively defined sets. We also show how the initiality/finality property gives rise to the co-induction principle studied by the author (in UCAM-CL-TR-252), by which equalities between elements of recursively defined domains may be proved via an appropriate notion of 'bisimulation'.

UCAM-CL-TR-322

Guangxing Li:

Supporting distributed realtime computing

December 1993, 113 pages, PDF
PhD thesis (King's College, August 1993)

Abstract: Computers have been used for realtime systems for almost 50 years. However, it is only recently that computer research institutions are becoming interested in realtime computing, realizing the significance of realtime systems and their increasing practical importance. Realtime systems engineering still faces many challenges: current systems concepts and functions are unfavourable for the development of a general and consistent framework for realtime systems engineering. The realtime problem domain has also been further complicated by the rapid spread of distributed computing.

This dissertation is concerned with the design and construction of a distributed system environment for supporting realtime applications. The contributions range from high-level programming abstractions down to an operating system kernel interface through the detailed engineering tradeoffs required to create, implement, and integrate the mechanisms within the environment. The contributions consist of a realtime programming model, a timed RPC protocol, a temporal synchronisation facility and empirical validations.

The realtime programming model provides a framework to facilitate the enforcement of the stringent timing constraints found in distributed realtime applications. The model incorporates tasks and communication channels as its basic programming components. It synthesises aspects of resource requirements, resource allocation and resource scheduling into an object based programming paradigm.

The development of the timed RPC protocol allows a programmer to express and enforce reasonable timing requirements (representing different tradeoffs between consistency and strictness) with object invocations.

The definition and infrastructure support of the timed automata to provide a temporal synchronisation facility. This facility contributes to the understanding of temporal synchronisations in a distributed world.

A prototype implementation of the system environment has been constructed and used to evaluate the feasibility of the architectural concepts of the system.

UCAM-CL-TR-323

J. von Wright:

Representing higher-order logic proofs in HOL

January 1994, 28 pages, PDF

Abstract: When using a theorem prover based on classical logic, such as HOL [2], we are generally interested in the facts that are proved (the theorems) than in the way in which they were proved (the proofs). However we may be interested in checking the correctness of the proofs. Since machine-generated proofs are generaly very long we need a computer program, a proof checker, to do this. However, we would also want the correctness of the proof checker to be verified formally. One way of doing this is by specifying it in a mechanised logic (such as that of the HOL system) and then doing a correctness proof in that logic. While this may seem circular, it is acceptable provided we have a theory of proofs embedded in the logic.

This paper describes an attempt to formalise the notion of HOL proofs within HOL. The aim is to be able to verify (inside HOL) that what is claimed to be a proof really is a proof.

UCAM-CL-TR-324

J. von Wright:

Verifying modular programs in HOL

January 1994, 25 pages, PDF

Abstract: This paper describes a methodology for verifying imperative programs that are modular, i.e., built using separately defined functions and procedures.

The verification methodology is based on a simple programming notation with a weak precondition semantics. This notation has been semantically embedded in the HOL theorem prover [3] and a number of laws have been derived from the semantics.

These semantic laws are used to prove the correctness of functional procedures, by showing that a call to the procedure in question is equivalent to a call to the corresponding function as it is defined in the logic. This makes it possible to specify a program in an essentially functional style, but the functions are then implemented as imperative procedures (like user-defined functions in FORTRAN or Pascal).

We also show how to define non-functional procedures and calls to such procedures. Procedures may be recursive. Altogether, this gives us a basis for mechanical verification of modular imperative programs.

UCAM-CL-TR-325

Richard Crouch:

The temporal properties of English conditionals and modals

January 1994, 248 pages, PDF
PhD thesis (April 1993)

Abstract: This thesis deals with the patterns of temporal reference exhibited by conditional and modal sentences in English, and specifically with the way that past and present tenses can undergo deictic shift in these contexts. This shifting behaviour has consequences both for the semantics of tense and for the semantics of conditionals and modality.

Asymmetries in the behaviour of the past and present tenses under deictic shift are explained by positing a primary and secondary deictic centre for tenses. The two deictic centres, the assertion time and the verification time, are given independent motivation through an information based view of tense. This holds that the tense system not only serves to describe the way that the world changes over time, but also the way that information about the world changes. Information change takes place in two stages. First, it is asserted that some fact holds. And then, either at the same time or later, it is verified that is assertion is correct.

Typically, assertion and verification occur simultaneously, and most sentences convey verified information. Modals and conditionals allow delayed assertion and verification. "If A, then B" means roughly: suppose you were now to assert A; if and when A is verified, you will be in a position to assert B, and in due course this assertion will also be verified. Since A and B will both be tensed clauses, the shifting of the primary and secondary deictic centres leads to shifted interpretations of the two clauses.

The thesis presents a range of temporal properties of indicative and subjunctive conditionals that have not previously been discussed, and shows how they can be explained. A logic is presented for indicative conditionals, based around an extension of intuitionistic logic to allow for both verified and unverified assertions. This logic naturally gives rise to three forms of epistemic modality, corresponding to "must", "may" and "will".

UCAM-CL-TR-326

Sai-Lai Lo:

A modular and extensible network storage architecture

January 1994, 147 pages, PostScript
PhD thesis (Darwin College, November 1993)

Abstract: Most contemporary distributed file systems are not designed to be extensible. This work asserts that the lack of extensibility is a problem because:

– New data types, such as continuous-medium data and structured data, are significantly different from conventional unstructured data, such as text and binary, that contemporary distributed file systems are built to support.

– Value-adding clients can provide functional enhancements, such as convenient and reliable persistent programming and automatic and transparent file indexing, but cannot be integrated smoothly with contemporary distributed file systems.

– New media technologies, such as the optical jukebox and RAID disk, can extend the scale and performance of a storage service but contemporary distributed file systems do not have a clear framework to incorporate these new technologies and to provide the necessary user level transparency.

Motivated by these observations, the new network storage architecture (MSSA) presented in this dissertation, is designed to be extensible. Design modularity is taken as the key to achieve service extensibility. This dissertation examines a number of issues related to the design of the architecture. New ideas, such as a flexible access control mechanism based on temporary capabilities, a low level storage substrate that uses non-volatile memory to provide atomic update semantics at high performance, a concept of sessions to differentiate performance requirements of different data types, are introduced. Prototype implementations of the key components are evaluated.

UCAM-CL-TR-327

Siani L. Baker:

A new application for explanation-based generalisation within automated deduction

February 1994, 18 pages, PDF

Abstract: Generalisation is currently a major theorem-proving problem. This paper proposes a new method of generalisation, involving the use of explanation-based generalisation within a new domain, which may succeed when other methods fail. The method has been implemented for simple arithmetical examples.

UCAM-CL-TR-328

Paul Curzon:

The formal verification of the Fairisle ATM switching element: an overview

March 1994, 46 pages, PDF

Abstract: We give an overview of the formal verification of an implementation of a self routing ATM switching element. This verification was performed using the HOL90 theorem proving system so is fully machine checked. The switching element is in use in a real network, switching real data. Thus, this work constitutes a realistic formal verification case study. We give an informal overview of the switch and element and give a tutorial on the methods used. We outline how these techniques were applied to verify the switching element. We then discuss the time spent on the verification. This was comparable to the time spent designing and testing the element. Finally we describe the errors discovered.

UCAM-CL-TR-329

Paul Curzon:

The formal verification of the Fairisle ATM switching element

March 1994, 105 pages, PDF

Abstract: We describe the formal verification of an implementation of the switching element of the fairisle ATM switch. This verification was performed using the HOL90 theorem proving system so is fully machine-checked. We give here all the definitions used in the verification together with the main correctness theorems proved. Fairisle switches are in use in a working network, switching real data. Thus, this work constitutes a realistic formal verification case study.

UCAM-CL-TR-330

Pierre David Wellner:

Interacting with paper on the DigitalDesk

March 1994, 96 pages, PDF
PhD thesis (Clare Hall, October 1993)

Abstract: In the 1970's Xerox PARC developed the "desktop metaphor," which made computers easy to use by making them look and act like ordinary desks and paper. This led visionaries to predict the "paperless office" would dominate within a few years, but the trouble with this prediction is that people like paper too much. It is portable, tactile, universally accepted, and easier to read than a screen. Today, we continue to use paper, and computers produce more of it than they replace.

Instead of trying to use computers to replace paper, the DigitalDesk takes the opposite approach. It keeps the paper, but uses computers to make it more powerful. It provides a Computer Augmented Environment for paper.

The DigitalDesk is built around an ordinary physical desk and can be used as such, but it has extra capabilities. A video camera is mounted above the desk, pointing down at the work surface. This camera's output is fed through a system that can detect where the user is pointing, and it can read documents that are placed on the desk. A computer-driven electronic projector is also mounted above the desk, allowing the system to project electronic objects onto the work surface and onto real paper documents — something that can't be done with flat display panels or rear-projection. The system is called DigitalDesk because it allows pointing with the fingers.

Several applications have been prototyped on the DigitalDesk. The first was a calculator where a sheet of paper such as an annual report can be placed on the desk allowing the user to point at numbers with a finger or pen. The camera reads the numbers off the paper, recognizes them, and enters them into the display for further calculations. Another is a translation system which allows users to point at unfamiliar French words to get their English definitions projected down next to the paper. A third is a paper-based paint program (PaperPaint) that allows users to sketch on paper using traditional tools, but also be able to select and paste these sketches with the camera and projector to create merged paper and electronic documents. A fourth application is the DoubleDigitalDesk, which allows remote colleagues to "share" their desks, look at each other's paper documents and sketch on them remotely.

This dissertation introduces the concept of Computer Augmented Environments, describes the DigitalDesk and applications for it, and discusses some of the key implementation issues that need to be addressed to make this system work. It describes a toolkit

for building DigitalDesk applications, and it concludes with some more ideas for future work.

UCAM-CL-TR-331

Noha Adly, Akhil Kumar:

HPP: a hierarchical propagation protocol for large scale replication in wide area networks

March 1994, 24 pages, PDF

Abstract: This paper describes a fast, reliable, scalable and efficient propagation protocol for weak-consistency replica management. This protocol can be used to implement a bulletin board service such as the Usenet news on the Internet. It is based on organizing the nodes in a network into a logical hierarchy, and maintaining a limited amount of state information at each node. It ensures that messages are not lost due to failures or partitions once they are repaired and minimizes redundancy. Further the protocol allows messages to be diffused while nodes are down provided the parent and child nodes of a failed node are alive. Moreover the protocol allows nodes to be moved in the logical hierarchy, and the network to be restructured dynamically in order to improve performance while still ensuring that no messages are lost while the switch takes place and without disturbing normal operation.

UCAM-CL-TR-332

David Martin Evers:

Distributed computing with objects

March 1994, 154 pages, PDF
PhD thesis (Queens' College, September 1993)

Abstract: Distributed systems and object-based programming are now beginning to enter the mainstream of computing practice. These developments have the potential to simplify the distributed application programmer's task considerably, but current systems impose unnecessary burdens. Distributed operating systems provide palatable message passing between remote processes but leave the preparation and interpretation of the messages to application code. Remote procedure call systems use familiar language-level concepts to hide distribution, but the awkwardness of service creation and binding discourages the use of transient objects. Finally, object-based programming languages which support distribution often ignore the possibility of failures and do not efficiently accommodate heterogeneity.

This dissertation discusses the design, implementation and evaluation of a practical system for network objects which addresses these problems for a representative programming language (Modula-3) and distributed computing environment (the ANSA testbench). We propose that language level objects should explicitly represent bindings to potentially remote access points (interfaces), which are sufficiently lightweight that they can be used as transient handles for shared state. Our system uses local objects to stand for remote services and local method call to cause remote operation invocation. Within a process, concurrency control is provided by familiar language-level facilities. The local programming language's object type system is made to represent the global service type system in a natural way. We support dynamic creation of service interfaces and the transmission of network object references in invocations. We allow the dynamic types of network object references to propagate between separate programs. Finally we provide automatic, fault-tolerant and efficient distributed garbage collection of network objects. In each case, we discuss the requirements of a useful design and the tradeoffs necessary in a real implementation. Our implementation runs on stock systems connected by standard local and wide area networks and internetworking protocols. We believe our approach would support additional library-level tools for security, stable storage, distributed transactions and transparent service replication, though we have not pursued this.

The dissertation demonstrates that it is practical to retain many important amenities of modern programming languages when providing support for the construction of applications in a heterogeneous and evolving distributed system.

UCAM-CL-TR-333

G.M. Bierman:

What is a categorical model of intuitionistic linear logic?

April 1994, 15 pages, PDF

Abstract: This paper re-addresses the old problem of providing a categorical model for Intuitionistic Linear Logic (ILL). In particular we compare the new standard model proposed by Seely to the lesser known one proposed by Benton, Bierman, Hyland and de Paiva. Surprisingly we find that Seely's model is unsound in that it does not preserve equality of proofs — we shall give some examples of equal proofs which do not seem to be modelled as equal morphisms in the category. We shall propose how to adapt Seely's definition so as to correct these problems and consider how this compares with the model due to Benton et al.

UCAM-CL-TR-334

Lawrence C. Paulson:

A concrete final coalgebra theorem for ZF set theory

May 1994, 21 pages, PDF, PostScript, DVI

Abstract: A special final coalgebra theorem, in the style of Aczel (1988), is proved within standard Zermelo-Fraenkel set theory. Aczel's Anti-Foundation Axiom is replaced by a variant definition of function that admits non-well-founded constructions. Variant ordered pairs and tuples, of possibly infinite length, are special cases of variant functions. Analogues of Aczel's Solution and Substitution Lemmas are proved in the style of Rutten and Turi (1993).

The approach is less general than Aczel's; non-well-founded objects can be modelled only using the variant tuples and functions. But the treatment of non-well-founded objects is simple and concrete. The final coalgebra of a functor is its greatest fixedpoint. The theory is intended for machine implementation and a simple case of it is already implemented using the theorem prover Isabelle.

UCAM-CL-TR-335

G.J.F. Jones, J.T. Foote, K. Spärck Jones, S.J. Young:

Video mail retrieval using voice: report on keyword definition and data collection (deliverable report on VMR task No. 1)

April 1994, 38 pages, PDF

Abstract: This report describes the rationale, design, collection and basic statistics of the initial training and test database for the Cambridge Video Mail Retrieval (VMR) project. This database is intended to support both training for the wordspotting processes and testing for the document searching methods using these that are being developed for the project's message retrieval task.

UCAM-CL-TR-336

Barnaby P. Hilken:

Towards a proof theory of rewriting: the simply-typed 2-λ calculus

May 1994, 28 pages, PDF

Abstract: This paper describes the simply typed 2-λ-calculus, a language with three levels, types, terms and rewrites. The types and terms are those of the simply typed λ-calculus, and the rewrites are expressions denoting sequences of β-reductions and η-expansions. An equational theory is imposed on the rewrites, based on 2-categorical justifications, and the word problem for this theory is solved by finding a canonical expression in each equivalence class.

The canonical form of rewrites allows us to prove several properties of the calculus, including a strong form of confluence and a classification of the long-β-η-normal forms in terms of their rewrites. Finally we use these properties as the basic definitions of a theory of categorical rewriting, and find that the expected relationships between confluence, strong normalisation and normal forms hold.

UCAM-CL-TR-337

Richard John Boulton:

Efficiency in a fully-expansive theorem prover

May 1994, 126 pages, DVI
PhD thesis (Churchill College, December 1993)

Abstract: The HOL system is a fully-expansive theorem prover: Proofs generated in the system are composed of applications of the primitive inference rules of the underlying logic. This has two main advantages. First, the soundness of the system depends only on the implementations of the primitive rules. Second, users can be given the freedom to write their own proof procedures without the risk of making the system unsound. A full functional programming language is provided for this purpose. The disadvantage with the approach is that performance is compromised. This is partly due to the inherent cost of fully expanding a proof but, as demonstrated in this thesis, much of the observed inefficiency is due to the way the derived proof procedures are written.

This thesis seeks to identify sources of non-inherent inefficiency in the HOL system and proposes some general-purpose and some specialised techniques for eliminating it. One area that seems to be particularly amenable to optimisation is equational reasoning. This is significant because equational reasoning constitutes large portions of many proofs. A number of techniques are proposed that transparently optimise equational reasoning. Existing programs in the HOL system require little or no modification to work faster.

The other major contribution of this thesis is a framework in which part of the computation involved in HOL proofs can be postponed. This enables users to make better use of their time. The technique exploits a form of lazy evaluation. The critical feature is the separation of the code that generates the structure of a theorem from the code that justifies it logically. Delaying the justification allows some non-local optimisations to be performed in equational reasoning. None of the techniques sacrifice the security of the fully-expansive approach.

A decision procedure for a subset of the theory of linear arithmetic is used to illustrate many of the techniques. Decision procedures for this theory are commonplace in theorem provers due to the importance of arithmetic reasoning. The techniques described in the thesis have been implemented and execution times are given. The implementation of the arithmetic procedure is a major contribution in itself. For the first time, users of the HOL system are able to prove many arithmetic lemmas automatically in a practical amount of time (typically a second or two).

The applicability of the techniques to other fully-expansive theorem provers and possible extensions of the ideas are considered.

UCAM-CL-TR-338

Zhixue Wu:

A new approach to implementing atomic data types

May 1994, 170 pages, PDF
PhD thesis (Trinity College, October 1993)

Abstract: Many researchers have suggested the atomic data type approach to maintaining data consistency in a system. In this approach atomicity is ensured by the data objects that are shared by concurrent activities. By using the semantics of the operations of the shared objects, greater concurrency among activities can be permitted. In addition, by encapsulating synchronisation and recovery in the implementation of the shared objects, modularity can be enhanced. Existing systems support user-defined atomic data types in an explicit approach. They either permit limited semantics to be presented thus providing less concurrency, or permit a high level of semantics to be presented but in an encapsulated way, thus resulting in a complicated implementation. This research was done to make the implementation of user-defined atomic data types simple, efficient, while still permitting great concurrency.

The research aims to lessen the programmer's burden by supporting an implicit approach for implementing atomic data types. It permits a high level of semantics to be specified in a declarative way, which makes the implementation of user defined atomic data types as simple as in a sequential environment. A special concurrency control mechanism is implemented by the system. By using type inheritance, user-defined atomic data types can use the mechanism directly to provide local atomicity for their objects. A language has been developed for specifying the conflicts between object operations. Since the concurrency control mechanism can take operation semantics into account, the approach permits great concurrency.

To support the implicit approach, an appropriate concurrency control protocol must be proposed which can take advantage of operation semantics to increase concurrency and which can be implemented independently from user-defined atomic data types. Such a protocol, called the dual-level validation method, is presented and verified in this thesis. The method can make use of the parameters and results of object operations to achieve great concurrency. In addition, it also provides great internal concurrency by permitting operations to take place on an object concurrently.

The prototyping of the implicit approach in a persistent programming language called PC++ is described. The feasibility of the approach is shown by an application, namely a naming database for an active badge system. Some related issues are also addressed in the thesis such as remote object invocation, distributed transaction commitment and data persistence.

UCAM-CL-TR-339

Brian Logan, Steven Reece, Alison Cawsey, Julia Galliers, Karen Spärck Jones:

Belief revision and dialogue management in information retrieval

May 1994, 227 pages, PDF

Abstract: This report describes research to evaluate a theory of belief revision proposed by Galliers in the context of information-seeking interaction as modelled by Belkin, Brooks and Daniels and illustrated by user-librarian dialogues. The work covered the detailed assessment and development, and computational implementation and testing, of both the belief revision theory and the information retrieval model. Some features of the belief theory presented problems, and the original 'multiple expert' retrieval model had to be drastically modified to support rational dialogue management. But the experimental results showed that the characteristics of literature seeking interaction could be successfully captured by the belief theory, exploiting important elements of the retrieval model. Thus, though the system's knowledge and dialogue performance were very limited, it provides a useful base for further research. The report presents all aspects of the research in detail, with particular emphasis on the implementation of belief and intention revision, and the integration of revision with domain reasoning and dialogue interaction.

UCAM-CL-TR-340

Eoin Andrew Hyden:

Operating system support for quality of service

June 1994, 102 pages, PDF
PhD thesis (Wolfson College, February 1994)

Abstract: The deployment of high speed, multiservice networks within the local area has meant that it has become possible to deliver continuous media data to a general purpose workstation. This, in conjunction with the increasing speed of modern microprocessors, means that it is now possible to write application programs which manipulate continuous media in real-time. Unfortunately, current operating systems do not provide the resource management facilities which are required to ensure the timely execution of such applications.

This dissertation presents a flexible resource management paradigm, based on the notion of Quality of Service, with which it is possible to provide the scheduling support required by continuous media applications. The mechanisms which are required within an operating system to support this paradigm are described, and the design and implementation of a prototypical kernel which implements them is presented.

It is shown that, by augmenting the interface between an application and the operating system, the application can be informed of varying resource availabilities, and can make use of this information to vary the quality of its results. In particular an example decoder application is presented, which makes use of such information and exploits some of the fundamental properties of continuous media data to trade video image quality for the amount of processor time which it receives.

UCAM-CL-TR-341

John Bates:

Presentation support for distributed multimedia applications

June 1994, 140 pages, PostScript

Abstract: Distributed computing environments can now support digital continuous media (such as audio and video) in addition to still media (such as text and pictures). The work presented in this dissertation is motivated by the desire of application developers to create applications which utilise these multimedia environments. Many important application areas are emerging such as Computer-Aided Instruction (CAI) and Computer-Supported Cooperative Working (CSCW).

Building multimedia applications is currently a difficult and time consuming process. At run-time, an application must manage connections to a range of heterogeneous sevices to access data. Building applications directly on top of environment specific features roots them to those features. Continuous media introduces new problems into application management such as control of Quality of Service (QoS) and synchronisation of data items. An application may also be required to analyse, process or display data. Some multimedia applications are event-driven, i.e. they must perform actions in response to asynchronous run-time occurrences. They may also be required to control many workspaces and involve multiple users.

The thesis of this dissertation is based on two principles. Firstly, despite the heterogeneity between and within multimedia environments, that their functionality should be provided in a uniform way to application developers. By masking the control differences with generic abstractions, applications can easily be developed and ported. Secondly, that it is possible to develop such abstractions to support a wide range of multimedia applications. Extensible and configurable facilities can be provided to access, and present multimedia data and to support event-diven applications including cooperative ones.

The approach taken in this work is to provide a presentation support platform. To application developers this platform offers an authoring interface based on data modelling and specification using a script language. Using these facilities, the parts of an application involving interactive presentation of multimedia can be specified. Services have been built to support the runtime realisation of authored presentations on top of environments. Experiments show that a wide range of applications can be supported.

UCAM-CL-TR-342

Stephen Martin Guy Freeman:

An architecture for distributed user interfaces

July 1994, 127 pages, PDF
PhD thesis (Darwin College, 1994)

Abstract: Computing systems have changed rapidly since the first graphical user interfaces were developed. Hardware has become faster and software architectures have become more flexible and more open; a modern computing system consists of many communicating machines rather than a central host. Understanding of human-computer interaction has also become more sophisticated and places new demands on interactive software; these include, in particular, support for multi-user applications, continuous media, and 'ubiquitous' computing. The layer which binds user requirements and computing systems together, the user interface, has not changed as quickly; few user interface architectures can easily supportthe new requirements placed on them and few take advantage of the facilities offered by advanced computing systems.

Experiences of implementing systems with unusual user interfaces has shown that current window system models are only a special case of possible user interface architectures. These window systems are too strongly tied to assumptions about how users and computers interact to provide a suitable platform for further evolution. Users and application builders may reasonably expect to be able to use multiple input and output devices as their needs arise. Experimental applications show that flexible user interface architectures, which support multiple devices and users, can be built without excessive implementation and processing costs.

This dissertation describes Gemma, a model for a new generation of interactive systems that are not confined to virtual terminals but allows collections of independent devices to be bound together for the task at hand. It provides mediated shared access to basic devices and higher-level virtual devices so that people can share computational facilities in the real world, rather than in a virtual world. An example window system shows how these features may be exploited to provide a flexible, collaborative and mobile interactive environment.

UCAM-CL-TR-344

Martin John Turner:

The contour tree image encoding technique and file format

July 1994, 154 pages, PDF
PhD thesis (St John's College, April 1994)

Abstract: The process of contourization is presented which converts a raster image into a discrete plateaux of contours. These contours can be grouped into a hierarchical structure, defining total spacial inclusion called a contour tree. A contour coder has been developed which fully describes these contours in a compact and efficient manner and is the basis for an image compression method.

Simplification of the contour tree has been undertaken by merging contour tree nodes thus lowering the contour tree's entropy. This can be exploited by the contour coder to increase the image compression ratio. By applying general and simple rules derived from physiological experiments on the human vision system, lossy image compression can be achieved which minimises noticable artifacts in the simplified image.

The contour merging technique offers a complementary lossy compression system to the QDCT (Quantised Discrete Cosine Transform). The artifacts introduced by the two methods are very different; QDCT produces a general blurring and adds extra highlights in the form of overshoots, whereas contour merging sharpens edges, reduces highlights and introduces a degree of false contouring.

A format based on the contourization technique which caters for most image types is defined, called the contour tree image format. Image operations directly on this compressed format have been studied which for certain manipulations can offer significant operational speed increases over using a standard raster image format. A couple of examples of operations specific to the contour tree format are presented showing some of the features of the new format.

UCAM-CL-TR-345

Siani L. Baker:

A proof environment for arithmetic with the Omega rule

August 1994, 17 pages, PDF

Abstract: An important technique for investigating the derivability in formal systems of arithmetic has been to embed such systems into semi-formal systems with the ω-rule. This paper exploits this notion within the domain of automated theorem-proving and discusses the implementation of such a proof environment, namely the CORE system which implements a version of the primitive recursive ω-rule. This involves providing an appropriate representation for infinite proofs, and a means of verifying properties of such objects. By means of the CORE system, from a finite number of instances a conjecture of the proof of the universally quantified formula is automatically derived by an inductive inference algorithm, and checked for correctness. In addition, candidates for cut formulae may be generated by an explanation-based learning algorithm. This is an alternative approach to reasoning about inductively defined domains from traditionas structural induction, which may sometimes be more intuitive.

UCAM-CL-TR-346

G.M. Bierman:

On intuitionistic linear logic

August 1994, 191 pages, PDF
PhD thesis (Wolfson College, December 1993)

Abstract: In this thesis we carry out a detailed study of the (propositional) intuitionistic fragment of Girard's linear logic (ILL). Firstly we give sequent calculus, natural deduction and axiomatic formulations of ILL. In particular our natural deduction is different from others and has important properties, such as closure under substitution, which others lack. We also study the process of reduction in all three local formulations, including a detailed proof of cut elimination. Finally, we consider translations between Instuitionistic Logic (IL) and ILL.

We then consider the linear term calculus, which arises from applying the Curry-Howard correspondence to the natural deduction formulation. We show how the various proof theoretic formulations suggest reductions at the level of terms. The properties of strong normalization and confluence are proved for these reduction rules. We also consider mappings between the extended λ-calculus and the linear term calculus.

Next we consider a categorical model for ILL. We show how by considering the linear term calculus as an equational logic, we can derive a model: a linear category. We consider two alternative models: firstly,

one due to Seely and then one due to Lafont. Surprisingly, we find that Seely's model is not sound, in that equal terms are not modelled with equal morphisms. We show how after adapting Seely's model (by viewing it in a more abstract setting) it becomes a particular instance of a linear category. We show how Lafont's model can also be seen as another particular instance of a linear category. Finally we consider various categories of coalgebras, whose construction can be seen as a categorical equivalent of the translation of IL into ILL.

UCAM-CL-TR-347

Karen Spärck Jones:

Reflections on TREC

July 1994, 35 pages, PostScript, DVI

Abstract: This paper discusses the Text REtrieval Conferences (TREC) programme as a major enterprise in information retrieval research. It reviews its structure as an evaluation exercise, characterises the methods of indexing and retrieval being tested within it in terms of the approaches to system performance factors these represent; analyses the test results for solid, overall conclusions that can be drawn from them; and, in the light of the particular features of the test data, assesses TREC both for generally-applicable findings that emerge from it and for directions it offers for future research.

UCAM-CL-TR-348

Jane Louise Hunter:

Integrated sound synchronisation for computer animation

August 1994, 248 pages, paper copy
PhD thesis (Newnham College, August 1994)

UCAM-CL-TR-349

Brian Graham:

A HOL interpretation of Noden

September 1994, 78 pages, paper copy

UCAM-CL-TR-350

Jonathan P. Bowen, Michael G. Hinchey:

Ten commandments of formal methods

September 1994, 18 pages, PDF

Abstract: The formal methods community is in general very good at undertaking research into the mathematical aspects of formal methods, but not so good at promulgating the use of formal methods in an engineering environment and at an industrial scale. Technology transfer is an extremely important part of the overall effort necessary in the acceptance of formal techniques. This paper explores some of the more informal aspects of applying formal methods and presents some maxims with associated discussion that may help in the application of formal methods in an industrial setting. A significant bibliography is included providing pointers to more technical and detailed aspects.

UCAM-CL-TR-351

Subir Kumar Biswas:

Handling realtime traffic in mobile networks

September 1994, 198 pages, PostScript
PhD thesis (Darwin College, August 1994)

Abstract: The rapidly advancing technology of cellular communication and wireless LAN makes ubiquitous computing feasible where the mobile users can have access to the location independent information and the computing resources. Multimedia networking is another emerging technological trend of the 1990s and there is an increasing demand for supporting continuous media traffic in wireless personal communication environment. In order to guarantee the strict performance requirements of realtime traffic, the connection-oriented approaches are proving to be more efficient compared to the conventional datagram based networking. This dissertation deals with a network architecture and its design issues for implementing the connection-oriented services in a mobile radio environment.

The wired backbone of the proposed wireless LAN comprises of high speed ATM switching elements, connected in a modular fashion, where the new switches and the user devices can be dynamically added and reconnected for maintaining a desired topology. A dynamic reconfiguration protocol, which can cope with these changing network topologies, is proposed for the present network architecture. The details about a prototype implementation of the protocol and a simulation model for its performance evaluation are presented.

CSMA/AED, a single frequency and carrier sensing based protocol is proposed for the radio medium access operations. A simulation model is developed in order to investigate the feasibility of this statistical and reliable access scheme for the proposed radio network architecture. The effectiveness of a per-connection window based flow control mechanism, for the proposed radio LAN, is also investigated. A hybrid technique is used, where the medium access and the radio data-link layers are modelled using the mentioned simulator; an upper

layer end-to-end queueing model, involving flow dependent servers, is solved using an approximate Mean Value Analysis technique which is augmented for faster iterative convergence.

A distributed location server, for managing mobile users' location information and for aiding the mobile connection management tasks, is proposed. In order to hide the effects of mobility from the non-mobile network entities, the concept of a per-mobile software entity, known as a "representative", is introduced. A mobile connection management scheme is also proposed for handling the end-to-end network layer connections in the present mobile environment. The scheme uses the representatives and a novel connection caching technique for providing the necessary realtime traffic support functionalities.

A prototype system, comprising of the proposed location and the connection managers, has been built for demonstrating the feasibility of the presented architecture for transporting continuous media traffic. A set of experiments have been carried out in order to investigate the impacts of various design decisions and to identify the performance-critical parts of the design.

UCAM-CL-TR-352

P.N. Benton:

A mixed linear and non-linear logic: proofs, terms and models

October 1994, 65 pages, PDF

Abstract: Intuitionistic linear logic regains the expressive power of intuitionistic logic through the ! ('of course') modality. Benton, Bierman, Hyland and de Paiva have given a term assignment system for ILL and an associated notion of catagorical model in which the ! modality is modelled by a comonad satisfying certain extra conditions. Ordinary intuitionistic logic is then modelled in a cartesian closed category which arises as a full subcategory of the category of coalgebras for the comonad.

This paper attempts to explain the connection between ILL and IL more directly and symmetrically by giving a logic, term calculus and categorical model for a system in which the linear and non-linear worlds exist on an equal footing, with operations allowing one to pass in both directions. We start from the categorical model of ILL given by Benton, Bierman, Hyland and de Paiva and show that that this is equivalent to having a symmetric monoidal adjunction between a symmetric monoidal closed category and a cartesian closed category. We then derive both a sequent calculus and a natural deduction presentation of the logic corresponding to the new notion of model.

UCAM-CL-TR-353

Mike Gordon:

Merging HOL with set theory

November 1994, 40 pages, PDF

Abstract: Set theory is the standard foundation for mathematics, but the majority of general purpose mechanized proof assistants support versions of type theory (higher order logic). Examples include Alf, Automath, Coq, Ehdm, HOL, IMPS, Lambda, LEGO, Nuprl, PVS and Veritas. For many applications type theory works well and provides for specification the benefits of type-checking that are well known in programming. However, there are areas where types get in the way or seem unmotivated. Furthermore, most people with a scientific or engineering background already know set theory, whereas type theory may appear inaccessible and so be an obstacle to the uptake of proof assistants based on it. This paper describes some experiments (using HOL) in combining set theory and type theory; the aim is to get the best of both worlds in a single system. Three approaches have been tried, all based on an axiomatically specified type V of ZF-like sets: (i) HOL is used without any additions besides V; (ii) an embedding of the HOL logic into V is provided; (iii) HOL axiomatic theories are automatically translated into set-theoretic definitional theories. These approaches are illustrated with two examples: the construction of lists and a simple lemma in group theory.

UCAM-CL-TR-354

Sten Agerholm:

Formalising a model of the λ-calculus in HOL-ST

November 1994, 31 pages, PDF

Abstract: Many new theorem provers implement strong and complicated type theories which eliminate some of the limitations of simple type theories such as the HOL logic. A more accessible alternative might be to use a combination of set theory and simple type theory as in HOL-ST which is a version of the HOL system supporting a ZF-like set theory in addition to higher order logic. This paper presents a case study on the use of HOL-ST to build a model of the λ-calculus by formalising the inverse limit construction of domain theory. This construction is not possible in the HOL system itself, or in simple type theories in general.

UCAM-CL-TR-355

David Wheeler, Roger Needham:

Two cryptographic notes

November 1994, 6 pages, PDF

Abstract: A large block DES-like algorithm

DES was designed to be slow in software. We give here a DES type of code which applies directly to single blocks comprising two or more words of 32 bits. It is thought to be at least as secure as performing DES separately on two word blocks, and has the added advantage of not requiring chaining etc. It is about 8m/(12+2m) times as fast as DES for an m word block and has a greater gain for Feistel codes where the number of rounds is greater. We use the name GDES for the codes we discuss. The principle can be used on any Feistel code.

TEA, a Tiny Encryption Algorithm

We design a short program which will run on most machines and encypher safely. It uses a large number of iterations rather than a complicated program. It is hoped that it can easily be translated into most languages in a compatible way. The first program is given below. It uses little set up time and does a weak non linear iteration enough rounds to make it secure. There are no preset tables or long set up times. It assumes 32 bit words.

UCAM-CL-TR-356

S.E. Robertson, K. Spärck Jones:

Simple, proven approaches to text retrieval

December 1994, 8 pages, PDF

Abstract: This technical note describes straightforward techniques for document indexing and retrieval that have been solidly established through extensive testing and are easy to apply. They are useful for many different types of text material, are viable for very large files, and have the advantage that they do not require special skills or training for searching, but are easy for end users.

UCAM-CL-TR-357

Jonathan P. Bowen, Michael G. Hinchey:

Seven more myths of formal methods

December 1994, 12 pages, PDF

Abstract: For whatever reason, formal methods remain one of the more contentious techniques in industrial software engineering. Despite great increases in the number of organizations and projects applying formal methods, it is still the case that the vast majority of potential users of formal methods fail to become actual users. A paper by Hall in 1990 examined a number of 'myths' concerning formal methods, assumed by some to be valid. This paper considers a few more beliefs held by many and presents some counter examples.

UCAM-CL-TR-358

Simon William Moore:

Multithreaded processor design

February 1995, 125 pages, PDF
PhD thesis (Trinity Hall, October 1994)
This report was also published as a book of the same title (Kluwer/Springer-Verlag, 1996, ISBN 0-7923-9718-5).

Abstract: Multithreaded processors aim to improve upon both control-flow and data-flow processor models by forming some amalgam of the two. They combine sequential behaviour from the control-flow model with concurrent aspects from data-flow design.

Some multithreaded processor designs have added just a little concurrency to control-flow or limited sequential execution to data-flow. This thesis demonstrates that more significant benefits may be obtained by a more radical amalgamation of the two models. A data-driven microthread model is proposed where a microthread is a short control flow code sequence. To demonstrate the efficiency of this model, a suitable multithreaded processor called Anaconda is designed and evaluated.

Anaconda incorporates a scalable temporally predictable memory tree structure with distributed virtual address translation and memory protection. A temporally predictable cached direct-mapped matching store is provided to synchronise data to microthreads. Code is prefetched into an instruction cache before execution commences. Earliest-deadline-first or fixed-priority scheduling is supported via a novel hardware priority queue. Control-flow execution is performed by a modified Alpha 21064 styled pipeline which assists comparison with commerical processors.

UCAM-CL-TR-359

Jacob Frost:

A case study of co-induction in Isabelle

February 1995, 48 pages, PDF, PostScript, DVI

Abstract: The consistency of the dynamic and static semantics for a small functional programming language was informally proved by R. Milner and M. Tofte. The notions of co-inductive definitions and the associated principle of co-induction played a pivotal role in the proof. With emphasis on co-induction, the work presented here deals with the formalisation of this result in the generic theorem prover Isabelle.

UCAM-CL-TR-360

W.F. Clocksin:

On the calculation of explicit polymetres

March 1995, 12 pages, PDF

Abstract: Computer scientists take an interest in objects or events which can be counted, grouped, timed and synchronised. The computational problems involved with the interpretation and notation of musical rhythm are therefore of particular interest, as the most complex time-stamped structures yet devised by humankind are to be found in music notation. These problems are brought into focus when considering explicit polymetric notation, which is the concurrent use of different time signatures in music notation. While not in common use the notation can be used to specify complicated cross-rhythms, simple versus compound metres, and unequal note values without the need for tuplet notation. From a computational point of view, explicit polymetric notation is a means of specifying synchronisation relationships amongst multiple time-stamped streams. Human readers of explicit polymetic notation use the time signatures together with the layout of barlines and musical events as clues to determine the performance. However, if the aim is to lay out the notation (such as might be required by an automatic music notation processor), the location of barlines and musical events will be unknown, and it is necessary to calculate them given only the information conveyed by the time signatures. Similar problems arise when trying to perform the notation (i.e. animate the specification) in real-time. Some problems in the interpretation of explicit polymetric notation are identified and a solution is proposed. Two different interpretations are distinguished, and methods for their automatic calculation are given. The solution given may be applied to problems which involve the synchronisation or phase adjustment of multiple independent threads of time-stamped objects.

UCAM-CL-TR-361

Richard John Black:

Explicit network scheduling

April 1995, 121 pages, PostScript
PhD thesis (Churchill College, December 1994)

Abstract: This dissertation considers various problems associated with the scheduling and network I/O organisation found in conventional operating systems for effective support for multimedia applications which require Quality of Service.

A solution for these problems is proposed in a micro-kernel structure. The pivotal features of the proposed design are that the processing of device interrupts is performed by user-space processes which are scheduled by the system like any other, that events are used for both inter- and intra-process synchronisation, and the use of a specially developed high performance I/O buffer management system.

An evaluation of an experimental implementation is included. In addition to solving the scheduling and networking problems addressed, the prototype is shown to out-perform the Wanda system (a locally developed micro-kernel) on the same platform.

This dissertation concludes that it is possible to construct an operating system where the kernel provides only the fundamental job of fine grain sharing of the CPU between processes, and hence synchronisation between those processes. This enables processes to perform task specific optimisations; as a result system performance is enhanced, both with respect to throughput and the meeting of soft real-time guarantees.

UCAM-CL-TR-362

Mark Humphrys:

W-learning: competition among selfish Q-learners

April 1995, 30 pages, PostScript

Abstract: W-learning is a self-organising action-selection scheme for systems with multiple parallel goals, such as autonomous mobile robots. It uses ideas drawn from the subsumption architecture for mobile robots (Brooks), implementing them with the Q-learning algorithm from reinforcement learning (Watkins). Brooks explores the idea of multiple sensing-and-acting agents within a single robot, more than one of which is capable of controlling the robot on its own if allowed. I introduce a model where the agents are not only autonomous, but are in fact engaged in direct competition with each other for control of the robot. Interesting robots are ones where no agent achieves total victory, but rather the state-space is fragmented among different agents. Having the agents operate by Q-learning proves to be a way to implement this, leading to a local, incremental algorithm (W-learning) to resolve competition. I present a sketch proof that this algorithm converges when the world is a discrete, finite Markov decision process. For each state, competition is resolved with the most likely winner of the state being the agent that is most likely to suffer the most if it does not win. In this way, W-learning can be viewed as 'fair' resolution of competition. In the empirical section, I

show how W-learning may be used to define spaces of agent-collections whose action selection is learnt rather than hand-designed. This is the kind of solution-space that may be searched with a genetic algorithm.

UCAM-CL-TR-363

Ian Stark:

Names and higher-order functions

April 1995, 140 pages, PostScript, DVI
PhD thesis (Queens' College, December 1994)

Abstract: Many functional programming languages rely on the elimination of 'impure' features: assignment to variables, exceptions and even input/output. But some of these are genuinely useful, and it is of real interest to establish how they can be reintroducted in a controlled way. This dissertation looks in detail at one example of this: the addition to a functional language of dynamically generated "names". Names are created fresh, they can be compared with each other and passed around, but that is all. As a very basic example of "state", they capture the graduation between private and public, local and global, by their interaction with higher-order functions.

The vehicle for this study is the "nu-calculus", an extension of the simply-typed lambda-calculus. The nu-calculus is equivalent to a certain fragment of Standard ML, omitting side-effects, exceptions, datatypes and recursion. Even without all these features, the interaction of name creation with higher-order functions can be complex and subtle.

Various operational and denotational methods for reasoning about the nu-calculus are developed. These include a computational metalanguage in the style of Moggi, which distinguishes in the type system between values and computations. This leads to categorical models that use a strong monad, and examples are devised based on functor categories.

The idea of "logical relations" is used to derive powerful reasoning methods that capture some of the distinction between private and public names. These techniques are shown to be complete for establishing contextual equivalence between first-order expressions; they are also used to construct a correspondingly abstract categorical model.

All the work with the nu-calculus extends cleanly to Reduced ML, a larger language that introduces integer references: mutable storage cells that are dynamically allocated. It turns out that the step up is quite simple, and both the computational metalanguage and the sample categorical models can be reused.

UCAM-CL-TR-364

Ole Rasmussen:

The Church-Rosser theorem in Isabelle: a proof porting experiment

April 1995, 27 pages, PostScript

Abstract: This paper describes a proof of the Church-Rosser theorem for the pure lambda-calculus formalised in the Isabelle theorem prover. The initial version of the proof is ported from a similar proof done in the Coq proof assistant by Girard Huet, but a number of optimisations have been performed. The development involves the introduction of several inductive and recursive definitions and thus gives a good presentation of the inductive package of Isabelle.

UCAM-CL-TR-365

P.N. Benton, G.M. Bierman, V.C.V. de Paiva:

Computational types from a logical perspective I

May 1995, 19 pages, PDF

Abstract: Moggi's computational lambda calculus is a metalanguage for denotational semantics which arose from the observation that many different notions of computation have the categorical structure of a strong monad on a cartesian closed category. In this paper we show that the computational lambda calculus also arises naturally as the term calculus corresponding (by the Curry-Howard correspondence) to a novel intuitionistic modal propositional logic. We give natural deduction, sequent calculus and Hilbert-style presentations of this logic and prove a strong normalisation result.

UCAM-CL-TR-366

K. Spärck Jones, G.J.F. Jones, J.T. Foote, S.J. Young:

Retrieving spoken documents: VMR Project experiments

May 1995, 28 pages, PDF

Abstract: This paper describes initial work on an application for the retrieval of spoken documents in multimedia systems. Speech documents pose a particular problem for retrieval since the contents are unknown. The VMR project seeks to address this problem for a video mail application by combining state of the art speech recognition with established document retrieval technologies to provide an effective and efficient retrieval tool. Experiments with a small spoken message collection show that retrieval precision for the spoken file can reach 90% of that obtained when the same file is used, as a benchmark, in text transcription form.

UCAM-CL-TR-367

Andrew M. Pitts:

Categorical logic

May 1995, 94 pages, PostScript, DVI

Abstract: This document provides an introduction to the interaction between category theory and mathematical logic which is slanted towards computer scientists. It will be a chapter in the forthcoming Volume VI of: S. Abramsky, D. M. Gabbay, and T. S. E. Maibaum (eds), "Handbook of Logic in Computer Science", Oxford University Press.

UCAM-CL-TR-368

Burkhard Stiller:

CogPiT – configuration of protocols in TIP

June 1995, 73 pages, PostScript

Abstract: The variety of upcoming applications in terms of their performance and Quality-of-Service (QoS) requirements is increasing. Besides almost well-known applications, such as teleconferencing, audio- and video-transmissions, even more contemporary ones, such as medical imaging, Video-on-Demand, and interactive tutoring systems, are introduced and applied to existing networks. On the contrary, traditionally data-oriented applications, such as file transfer and remote login, are considerably different in terms of their QoS requirements. Therefore, the consequences of this evolution effect the architectures of end-systems, e.g., workstations that have to be capable of maintaining all different kinds of multi-media data, and intermediate-systems as well.

Therefore, a configuration approach of communication protocols has been developed to support the variety of applications. This approach offers the possibility to configure communication protocols automatically depending on the application requirements expressed in various QoS parameters. The result, an application-tailored communication protocol, matches the requested application requirements as far as possible. Additionally, network and system resources (NSR) are taken into account for a well-suited configuration.

The Configuration of Protocols in TIP is called CogPiT and is part of the Transport and Internetworking Package (TIP). As an example, in the TIP environment the transport protocol TEMPO is used for configuration purposes.

UCAM-CL-TR-369

Sten Agerholm:

A comparison of HOL-ST and Isabelle/ZF

July 1995, 23 pages, PDF

Abstract: The use of higher order logic (simple type theory) is often limited by its restrictive type system. Set theory allows many constructions on sets that are not possible on types in higher order logic. This paper presents a comparison of two theorem provers supporting set theory, namely HOL-ST and Isabelle/ZF, based on a formalization of the inverse limit construction of domain theory; this construction cannot be formalized in higher order logic directly. We argue that whilst the combination of higher order logic and set theory in HOL-ST has advantages over the first order set theory in Isabelle/ZF, the proof infrastructure of Isabelle/ZF has better support for set theory proofs than HOL-ST. Proofs in Isabelle/ZF are both considerably shorter and easier to write.

UCAM-CL-TR-370

Sten Agerholm:

A package for non-primitive recursive function definitions in HOL

July 1995, 36 pages, PDF

Abstract: This paper provides an approach to the problem of introducing non-primitive recursive function definitions in the HOL system. A recursive specification is translated into a domain theory version, where the recursive calls are treated as potentially non-terminating. Once we have proved termination, the original specification can be derived easity. Automated tools implemented in HOL88 are provided to support the definition of both partial recursive functions and total recursive functions which have well founded recursive specifications. There are constructions for building well-founded relations easily.

UCAM-CL-TR-371

Kim Ritter Wagner:

LIMINF convergence in Ω-categories

June 1995, 28 pages, PDF

Abstract: The aim of this article is twofold. From a mathematical perspective we present a notion of convergence which is suitably general such as to include the convergence of chains to their least upper bounds in preordered sets, and the convergence of Cauchy sequences to their metric limits in metric spaces. Rather than presenting this theory from a purely mathematical perspective however, we will use it to introduce a simple-minded domain theory based on a generic notion of approximation. Although this is the use which motivated the development of these concepts, it is hoped that this is not the only one.

UCAM-CL-TR-372

Stefan G. Hild:

A brief history of mobile telephony

January 1995, 19 pages, PDF

Abstract: Mobile telephony has gone through a decade of tremendous change and progress. Today, mobile phones are an indispensable tool to many professionals, and have great potential to become vital components in mobile data communication applications. In this survey we will attempt to present some of the milestones from the route which mobile telephony has taken over the past decades while developing from an experimental system with limited capabilities with to a mature technology (section 1), followd by a more detailed introduction into the modern pan-European GSM standard (section 2). Section 3 is devoted to the data communication services, covering two packet-oriented data only networks as well as data services planned for the GSM system. Section 4 covers some security issues and section 5 gives an insight into the realities today with details of some networks available in the UK. Finally, section 6 concludes this overview with a brief look into the future.

UCAM-CL-TR-373

Benjamín Macías, Stephen G. Pulman:

Natural-language processing and requirements specifications

July 1995, 73 pages, PDF

Abstract: This document reports on our participation in the MORSE ("A Method for Object Reuse in Safety-critical Environments") project. Our brief in the project was to investigate the role that natural-language processing (NLP) techniques can play in improving any of the aspects linking natural-language requirements specifications and formal specifications. The contents are as follows: We begin with a brief introduction to NLP in the context of requirements tasks, followed by an examination of some strategies to control the form of requirements specifications. We continue by describing an interface designed to correct some of the problems with known methods to control specifications, while employing current NLP to maximum advantage. We then show how to build a natural-langauge interface to a formal specification, and some aspects of the problem of paraphrasing formal expressions. We finish with the conclusions reached at the end of our participation in the project.

UCAM-CL-TR-374

Burkhard Stiller:

A framework for QoS updates in a networking environment

July 1995, PostScript

Abstract: The support of sufficient Quality-of-Service (QoS) for applications residing in a distributed environment and running on top of high performance networks is a demanding issue. Currently, the areas to provide this support adequately include communication protocols, operating systems support, and offered network services. A configurable approach of communication protocols offers the needed protocol flexibility to react accordingly on various different requirements.

Communication protocols and operating systems have to be parametrized using internal configuration parameters, such as window sizes, retry counters, or scheduling mechanisms, that rely closely on requested application-oriented or network-dependent QoS, such as bandwidth or delay. Moreover, these internal parameters have to be recalculated from time to time due to network changes (such as congestion or line breakdown) or due to application-specific alterations (such as enhanced bandwidth requirements or increased reliability) to adjust a temporary or semi-permanent "out-of-tune" service behavior.

Therefore, a rule-based evaluation and QoS updating framework for configuration parameters in a networking environment has been developed. The resulting "rulework" can be used within highly dynamic environments in a communication subsystem that offers the possibility to specify for every QoS parameter both a bounding interval of values and an average value. As an example, the framework has been integrated in the Function-based Communication Subsystem (F-CSS). Especially, an enhanced application service interface is offered, allowing for the specification of various QoS-parameters that are used to configure a sufficient application-tailored communication protocol.

UCAM-CL-TR-375

Feng Huang:

Restructuring virtual memory to support distributed computing environments

July 1995, 145 pages, PDF
PhD thesis (Clare Hall, July 1995)

Abstract: This dissertation considers the limitations of conventional memory and storage management approaches and proposes a coherent memory-mapped object system architecture for emerging distributed computing environments.

Conventionally, main memory and secondary storage management is based on the two-level store architecture, which provides one interface to access memory segments and another to access secondary storage objects. The quality and productivity of software development is impaired by two different views of volatile data and persistent data. Operating system performance is compromised because of mandatory data copying and unnecessary user/kernel boundary crossings. This is exacerbated in microkernel architectures, in which most of the user/kernel boundary crossings become context switches. Double paging may cause resources to be used inefficiently and the double paging anomaly may occur if a database system is implemented on top of this architecture. The work presented here seeks to tackle these problems by integrating main memory with secondary storage by using memory-mapping techniques. The different views of volatile and persistent data are unified; mandatory information copying and unnecessary user/kernel boundary crossings (or context switches in microkernels) are avoided; and double paging is eliminated.

Distributed Shared Memory (DSM) has been proposed as an attractive abstraction for constructing distributed applications bacause it is easier to program than the message-passing abstraction. However, the overhead for maintaining memory coherency in DSM systems is high. Also, existing DSM systems typically provide only one coherence protocol and there exists a potential mismatch between the supplied protocol and some applications' requirements. This work explores the architectural support for a flexible coherence mechanism, through which clients can choose the most suitable protocols for their applications to avoid coherency mismatch. Also low-level coherency control is integrated with high level concurrency control so that system-wide object coherency and synchronisation are realised without sacrificing performance.

In this dissertation, an architectural framework is proposed; various design issues are discussed and the design of a flexible coherence mechanism, which accommodates multiple coherence protocols, is detailed. A prototype implementation and performance measurements are then presented; and the use of the architecture is illustrated.

UCAM-CL-TR-376

Timothy Roscoe:

The structure of a multi-service operating system

August 1995, 113 pages, PostScript
PhD thesis (Queens' College, April 1995)

Abstract: Increases in processor speed and network bandwidth have led to workstations being used to process multimedia data in real time. These applications have requirements not met by existing operating systems, primarily in the area of resource control: there is a need to reserve resources, in particular the processor, at a fine granularity. Furthermore, guarantees need to be dynamically renegotiated to allow users to reassign resources when the machine is heavily loaded. There have been few attempts to provide the necessary facilities in traditional operating systems, and the internal structure of such systems makes the implementation of useful resource control difficult.

This dissertation presents a way of structuring an operating system to reduce crosstalk between applications sharing the machine, and enable useful resource guarantees to be made: instead of system services being located in the kernel or server processes, they are placed as much as possible in client protection domains and scheduled as part of the client, with communication between domains only occurring when necessary to enforce protection and concurrency control. This amounts to multiplexing the service at as low a level of abstraction as possible. A mechanism for sharing processor time between resources is also described. The prototype Nemesis operating system is used to demonstrate the ideas in use in a practical system, and to illustrate solutions to several implementation problems that arise.

Firstly, structuring tools in the form of typed interfaces within a single address space are used to reduce the complexity of the system from the programmer's viewpoint and enable rich sharing of text and data between applications.

Secondly, a scheduler is presented which delivers useful Quality of Service guarantees to applications in a highly efficient manner. Integrated with the scheduler is an inter-domain communication system which has minimal impact on resource guarantees, and a method of decoupling hardware interrupts from the execution of device drivers.

Finally, a framework for high-level inter-domain and inter-machine communication is described, which goes beyond object-based RPC systems to permit both Quality of Service negotiation when a communication binding is established, and services to be implemented straddling protection domain boundaries as well as locally and in remote processes.

UCAM-CL-TR-377

Larry Paulson, Krzysztof Grabczewski:

Mechanising set theory: cardinal arithmetic and the axiom of choice

July 1995, 33 pages, PDF, PostScript

Abstract: Fairly deep results of Zermelo-Fraenkel (ZF) set theory have been mechanised using the proof assistant Isabelle. The results concern cardinal arithmetic and the Axiom of Choice (AC). A key result about cardinal multiplication is K*K=K, where K is any infinite cardinal. Proving this result required developing theories of orders, order-isomorphisms, order types, ordinal arithmetic, cardinals, etc.; this covers most of Kunen, Set Theory, Chapter I. Furthermore, we have proved the equivalence of 7 formulations of the Well-ordering Theorem and 20 formulations of AC; this covers the first two chapters of Rubin and Rubin, Equivalents of the Axiom of Choice. The definitions used in the proofs are largely faithful in style to the original mathematics.

UCAM-CL-TR-378

Noha Adly:

Performance evaluation of HARP: a hierarchical asynchronous replication protocol for large scale system

August 1995, 94 pages, PostScript

Abstract: This report evaluates the performance of HARP, a hierarchical replication protocol based on nodes organised into a logical hierarchy. The scheme is based on communication with nearby replicas and scales well for thousands of replicas. It proposes a new service interface that provides different levels of asynchrony, allowing strong consistency and weak consistency to be integrated into the same framework. Further, it provides the ability to offer different levels of staleness, by querying from different levels of the hierarchy. We present results from a detailed simulation analysis evaluating the benefits and losses in performance resulting from using synchronous versus asynchronous operation within HARP under different system configurations and load mixes. Further, the performance is evaluated on different network topologies. An analytical solution based on the Open Queueing Network Model with Multiple Job Classes is carried out for the verification of the simulation model and the results are presented.

UCAM-CL-TR-379

Lawrence Paulson:

Proceedings of the First Isabelle Users Workshop

September 1995, 265 pages, paper copy

UCAM-CL-TR-380

Burkhard Stiller:

Quality-of-Service issues in networking environments

September 1995, 68 pages, PostScript

Abstract: Quality-of-Service (QoS) issues in networking environments cover various separate areas and topics. They include at least the specification of applications requirements, the definition of network services, QoS models, resource reservation methods, negotiation and transformation methods for QoS, and operating system support for guaranteed services. An embracing approach for handling, dealing with, and supporting QoS in different scenarios and technical set-ups is required to manage sufficiently forthcoming communication and networking tasks. Modern telecommunication systems require an integrated architecture for applications, communication subsystems, and network perspectives to overcome drawbacks of traditional communication architectures, such as redundant protocol functionality, weakly designed interfaces between the end-system and a network adapter, or impossibility of specifying and guaranteeing QoS parameter.

This work contains the discussion of a number of interconnected QoS issues, e.g., QoS mapping, QoS negotiation, QoS-based configuration of communication protocols, or QoS aspects in Asynchronous Transfer Mode (ATM) signaling protocols, which have been dealt with during a one-year research fellowship. This report is not intended to be a complete description of every technical detail, but tries to provide a brief overall picture of the emerging and explosively developing QoS issues in telecommunication systems. Additionally, investigations of some of these issues are undertaken in a more closer detail. It is mainly focussed on QoS mapping, negotiation, and updating in the communication protocol area.

UCAM-CL-TR-381

Uwe Michael Nimscheck:

Rendering for free form deformations

October 1995, 151 pages, PDF
PhD thesis (Wolfson College)

Abstract: Sederberg's Free Form Deformation (FFD) is an intuitive modelling technique that lets users sculpt and deform objects without having to worry about internal model representation issues. Unfortunately displaying these deformed objects is problematic and there exist no algorithms to display general FFD deformed

polygonal models. Based on deRose's Bezier composition algorithms we develop geometrically intuitive composition algorithms to find analytic expressions for deformed objects, which can then be rendered using standard rendering hardware. Alternatively, one can adaptively tessellate deformed objects into a mesh of triangles and display this deformed mesh. The finite element method provides us with a wealth of algorithms to mesh all types of objects. We show how to adapt these algorithms to computer graphics problems. The main problem is to define curvature measures to vary the mesh density according to the curvature of deformed objects. We find such measures and use them to develop a new meshing scheme, based on Lo's advancing front algorithm, to mesh and render FFD deformed objects. Our algorithm is superior to existing schemes both in the quality of the generated meshes and in the variety of solids it can be applied to.

The major contributions of this dissertation are: Firstly, the development of geometrically intuitive algorithms to determine closed form expressions of FFD deformed surfaces. Secondly, the transformation of tangent and normal vectors into deformed space. Thirdly, development of a new advancing front meshing algorithm that allows to mesh solids that have been deformed by non-uniform B-spline volumes. Finally, systematic experiments have been performed to assess the performance and limitations of the new meshing algorithm.

UCAM-CL-TR-382

Oliver M. Castle:

Synthetic image generation for a multiple-view autostereo display

October 1995, 184 pages, paper copy
PhD thesis (Wolfson College, April 1995)

UCAM-CL-TR-383

Noha Adly:

Management of replicated data in large scale systems

November 1995, 182 pages, PDF
PhD thesis (Corpus Christi College, August 1995)

Abstract: Data is replicated in distributed systems to improve system availability and performance. In recent years, the growth of internetworks and distributed applications has increased the need for large scale replicated systems. However, existing replication protocols do not address scale and autonomy issues adequately. Further, current applications require different degrees of of consistency, and therefore they should be given the ability to choose the level of consistency that is appropriate for their particular semantics. This dissertation presents a new scalable replication protocol (HARP) that is based on organising the replicas into a logical hierarchy. It is argues that adopting a hierarchical structure allows for exploiting localised communication, which is taken as the key to achieve scalability. Moreover it gives the ability to provide different degrees of consistency.

HARP provides an efficient and scalable propagation scheme where each node needs to communicate with a few nodes only while ensuring reliable delivery. A new service interface is proposed that gives the application the flexibility to choose between strong and weak consistency. Further the scheme provides the ability to offer different levels of staleness, depending on the needs of various applications. Dynamic restructuring operations are presented which allow the hierarchy to be built and reconfigured, including the restarting of failed nodes and re-merging partitioned networks. The operations produce low message traffic by exploiting localised communication, and do not disturb normal operations. This is achieved while ensuring no loss of messages.

Reconciliation methods based on delivery order mechanisms are provided to resolve temporary inconsistencies and an application can choose from them. A new algorithm that supports casual order delivery is proposed. The desirable characteristic of the algorithm is that, by relying on the hierarchical propagation of HARP, it cuts down the size of the timestamp required to verify causality significantly, and thus enhances scalability.

A detailed simulation study was carried out to evaluate the performance of HARP and to quantify the benefits and losses resulting from moving from strong consistency to weak consistency under different system configurations and load mixes. Further, a simulation study was conducted to compare the performance of HARP to another weak consistency replication protocol, the Time Stamped Anti Entropy.

An alternative hierarchical propagation protocol is proposed as an optimisation of HARP, called HPP. The main difference between HPP and HARP is that HPP avoids the exchange of global state information when reconfiguration or failiures occur. Therefore HPP is more scalable; however, it can tolerate only special patterns of failiure. The protocol is presented in detail and its strengths and limitations are analysed.

UCAM-CL-TR-384

Shaw-Cheng Chuang:

Securing ATM networks

January 1995, 30 pages, PostScript

Abstract: This is an interim report on the investigations into securing Asynchronous Transfer Mode (ATM) networks. We look at the challenge in providing such a secure ATM network and identify the important issues in achieving such goal. In this paper, we discuss the

issues and problems involved and outline some techniques to solving these problems. The network environment is first examined and we also consider the correct placement of security mechanism in such an environment. Following the analysis of the security requirement, we introduce and describe a key agile cryptographic device for ATM. The protection of the ATM data plane is extremely important to provide data confidentiality and data integrity. Techniques in providing synchronisation, dynamic key change, dynamic initialisation vector change and Message Authentication Code on ATM data, are also being considered. Next, we discuss the corresponding control functions. A few key exchange protocols are given as possible candidates for the establishment of the session key. The impact of such key exchange protocols on the design of an ATM signalling protocol has also been examined and security extension to an existing signalling protocol being discussed. We also talk about securing other control plane functions such as NNI routing, Inter-Domain Policy Routing, authorisation and auditing, firewall and intrusion detection, Byzantine robustness. Management plane functions are also being looked at, with discussions on bootstrapping, authenticated neighbour discovery, ILMI Security, PVC security, VPI security and ATM Forum management model.

UCAM-CL-TR-385

Sanjay Saraswat:

Performance evaluation of the Delphi machine

December 1995, 187 pages, paper copy
PhD thesis (St Edmund's College, October 1995)

UCAM-CL-TR-386

Andrew D. Gordon, Gareth D. Rees:

Bisimilarity for a first-order calculus of objects with subtyping

January 1996, 78 pages, paper copy

UCAM-CL-TR-387

Scarlet Schwiderski, Andrew Herbert, Ken Moody:

Monitoring composite events in distributed systems

February 1996, 20 pages, PDF

Abstract: One way of integrating heterogeneous, autonomous and distributed systems is to monitor their behaviour in terms of global compostite events. In specific applications, for example database, it is essential that global composite events can take account of general conditions such as the timing constraints on distributed system behaviour. In this paper the use of global composite events incorporating time events for expressing physical time is investigated. The detection of global composite events is complicated by the inherent features of distributed systems: lack of global time, message delays between sites and independent failures. Global event detectors are distributed to arbitrary sites. Relevant constituent events occur on remote sites and are signalled to corresponding global event detectors, where they are evaluated. Two different algorithms for the detection of global composite events are introduced which are based on the evaluation of trees: asynchronous and synchronous evaluation. Asynchronous evaluation provides fast but unreliable detection of global composite events, whereas synchronous evaluation is characterized by reliability and unpredictable delays.

UCAM-CL-TR-388

P.N. Benton:

A unified approach to strictness analysis and optimising transformations

February 1996, 21 pages, PDF

Abstract: We present an inference system for translating programs in a PCF-like source language into a variant of Moggi's computational lambda calculus. This translation combines a simple strictness analysis with its associated optimising transformations into a single system. The correctness of the translation is established using a logical relation between the denotational semantics of the source and target languages.

UCAM-CL-TR-389

Wai Wong:

A proof checked for HOL

March 1996, 165 pages, PDF

Abstract: Formal proofs generated mechanically by theorem provers are often vary large and shallow, and the theorem provers themselves very complex. Therefore, in certain application areas, such as safety-critical systems, it is necessary to have an independent means for ensuring the consistency of such formal proofs. This report describes an efficient proof checker for the HOL theorem prover. This proof checker has been tested

with practical proofs consisting of thousands of inference steps. It was implemented in Standard ML of New Jersey.

The first part of the report gives an overview of the program. It describes: the rationale of developing a proof checker; how to use the checker; and, how the checker works.

The second part of the report describes the program in detail. The complete source code is included in the description.

UCAM-CL-TR-390

Richard J. Boulton:

Syn: a single language for specifiying abstract syntax tress, lexical analysis, parsing and pretty-printing

March 1996, 25 pages, PostScript, DVI

Abstract: A language called Syn is described in which all aspects of context-free syntax can be specified without redundancy. The language is essentially an extended BNF grammar. Unusual features include high-level constructs for specifying lexical aspects of a language and specification of precedence by textual order. A system has been implemented for generating lexers, parsers, pretty-printers and abstract syntax tree representations from a Syn specification.

UCAM-CL-TR-391

Andrew John Kennedy:

Programming languages and dimensions

April 1996, 149 pages, PDF
PhD thesis (St Catherine's College, November 1995)

Abstract: Scientists and engineers must ensure that the equations and formulae which they use are dimensionally consistent, but existing programming languages treat all numeric values as dimensionless. This thesis investigates the extension of programming languages to support the notion of physical dimension.

A type system is presented similar to that of the programming language ML but extended with polymorphic dimension types. An algorithm which infers most general dimension types automatically is then described and proved correct.

The semantics of the language is given by a translation into an explicitlytyped language in which dimensions are passed as arguments to functions. The operational semantics of this language is specified in the usual way by an evaluation relation defined by a set of rules. This is used to show that if a program is well-typed then no dimension errors can occur during its evaluation.

More abstract properties of the language are investigated using a denotational semantics: these include a notion of invariance under changes in the units of measure used, analogous to parametricity in the polymorphic lambda calculus. Finally the dissertation is summarised and many possible directions for future research in dimension types and related type systems are described.

UCAM-CL-TR-392

Uwe Nestmann, Benjamin C. Pierce:

Decoding choice encodings

April 1996, 54 pages, PDF

Abstract: We study two encodings of the asynchronous π-calculus with input-guarded choice into its choice-free fragment. One encoding is divergence-free, but refines the atomic commitment of choice into gradual commitment. The other preserves atomicity, but introduces divergence. The divergent encoding is fully abstract with respect to weak bisumulation, but the more natural divergence-free encoding is not. Instead we shot that it is fully abstract with respect to coupled simulation, a slightly coarser -- but still coinductively defined -- equivalence that does not require bisimilarity of internal branching decisions. The correctness proofs for the two choice encodings exploit the properties of decodings from translations to source terms.

UCAM-CL-TR-393

Simon Andrew Crosby:

Performance management in ATM networks

April 1996, 215 pages, PostScript
PhD thesis (St John's College, May 1995)

Abstract: The Asynchronous Transfer Mode (ATM) has been identified as the technology of choice amongst high speed communication networks for its potential to integrate services with disparate resource needs and timing constraints. Before it can successfully deliver integrated services, however, significant problems remain to be solved. They centre around two major issues. First, there is a need for a simple, powerful network service interface capable of meeting the communications needs of new applications. Second, within the network there is a need to dynamically control a mix of diverse traffic types to ensure that they meet their performance criteria.

Addressing the first concern, this dissertation argues that a simple network control interface offers significant advantages over the traditional, heavyweight approach of the telecommunications industry. A network control architecture based on a distributed systems approach is presented which locates both the network

control functions and its services outside the network. The network service interface uses the Remote Procedure Call (RPC) paradigm and enables more complicated service offerings to be built from the basic primitives. A formal specification and verification of the user-network signalling protocol is presented. Implementations of the architecture, both on Unix and the Wanda micro-kernel, used on the Fairisle ATM switch, are described. The implementations demonstrate the feasibility of the architecture, and feature a high degree of experimental flexibility. This is exploited in the balance of the dissertation, which presents the results of a practical study of network performance under a range of dynamic control mechanisms.

Addressing the second concern, results are presented from a study of the cell delay variation suffered by ATM connections when multiplexed with real ATM traffic in an uncontrolled network, and from an investigation of the expansion of bursts of ATM traffic as a result of multiplexing. The results are compared with those of analytical models. Finally, results from a study of the performance delivered to delay sensitive traffic by priority and rate based cell scheduling algorithms, and the loss experienced by different types of traffic under several buffer allocation strategies are presented.

UCAM-CL-TR-394

Lawrence C. Paulson:

A simple formalization and proof for the mutilated chess board

April 1996, 11 pages, PDF, PostScript, DVI

Abstract: The impossibility of tiling the mutilated chess board has been formalized and verified using Isabelle. The formalization is concise because it is expressed using inductive definitions. The proofs are straightforward except for some lemmas concerning finite cardinalities. This exercise is an object lesson in choosing a good formalization. is applicable in a variety of domains.

UCAM-CL-TR-395

Torben Bräuner, Valeria de Paiva:

Cut-elimination for full intuitionistic linear logic

May 1996, 27 pages, PDF

Abstract: We describe in full detail a solution to the problem of proving the cut elimination theorem for FILL, a variant of (multiplicative and exponential-free) Linear Logic introduced by Hyland and de Paiva. Hyland and de Paiva's work used a term assignment system to describe FILL and barely sketched the proof of cut elimination. In this paper, as well as correcting a small mistake in their paper and extending the system to deal with exponentials, we introduce a different formal system describing the intuitionistic character of FILL and we provide a full proof of the cut elimination theorem. The formal system is based on a notion of dependency between formulae within a given proof and seems of independent interest. The procedure for cut elimination applies to (classical) multiplicative Linear Logic, and we can (with care) restrict our attention to the subsystem FILL. The proof, as usual with cut elimination proofs, is a little involved and we have not seen it published anywhere.

UCAM-CL-TR-396

Lawrence C. Paulson:

Generic automatic proof tools

May 1996, 28 pages, PDF, PostScript

Abstract: This paper explores a synthesis between two distinct traditions in automated reasoning: resolution and interaction. In particular it discusses Isabelle, an interactive theorem prover based upon a form of resolution. It aims to demonstrate the value of proof tools that, compared with traditional resolution systems, seem absurdly limited. Isabelle's classical reasoner searches for proofs using a tableau approach. The reasoner is generic: it accepts rules proved in applied theories, involving defined connectives. New constants are not reduced to first-order logic; the reasoner

UCAM-CL-TR-397

Borut Robič:

Optimal routing in 2-jump circulant networks

June 1996, 7 pages, PostScript

Abstract: An algorithm for routing a message along the shortest path between a pair of processors in 2-jump circulant (undirected double fixed step) network is given. The algorithm requires O(d) time for preprocessing, and l = O(d) routing steps, where l is the distance between the processors and d is the diameter of the network.

UCAM-CL-TR-398

N.A. Dodgson, J.R. Moore:

Design and implementation of an autostereoscopic camera system

June 1996, 20 pages, PDF

Abstract: An autostereoscopic display provides the viewer with a three-dimensional image without the need for special glasses, and allows the user to look around objects in the image by moving the head left-right. The time-multiplexed autostereo display developed at the University of Cambridge has been in operation since late 1991.

An autostereoscopic camera system has been designed and implemented. It is capable of taking video input from up to sixteen cameras, and multiplexing these into a video output stream with a pixel rate an order of magnitude faster than the individual input streams. Testing of the system with eight cameras and a Cambridge Autostereo Display has produced excellent live autostereoscopic video.

This report describes the design of this camera system which has been successfully implemented and demonstrated. Problems which arose during this process are discussed, and a comparison with similar systems made.

UCAM-CL-TR-399

Richard Hayton:

OASIS: An open architecture for secure interworking services

June 1996, 102 pages, PDF
PhD thesis (Fitzwilliam College, March 1996)

Abstract: An emerging requirement is for applications and distributed services to cooperate or inter-operate. Mechanisms have been devised to hide the heterogeneity of the host operating systems and abstract the issues of distribution and object location. However, in order for systems to inter-operate securely there must also be mechanisms to hide differences in security policy, or at least negotiate between them.

This would suggest that a uniform model of access control is required. Such a model must be extremely flexible with respect to the specification of policy, as different applications have radically different needs. In a widely distributed environment this situation is exacerbated by the differing requirements of different organisations, and in an open environment there is a need to interwork with organisations using alternative security mechanisms.

Other proposals for the interworking of security mechanisms have concentrated on the enforcement of access policy, and neglected the concerns of freedom of expression of this policy. For example it is common to associate each request with a user identity, and to use this as the only parameter when performing access control. This work describes an architectural approach to security. By reconsidering the role of the client and the server, we may reformulate access control issues in terms of client naming.

We think of a client as obtaining a name issued by a service; either based on credentials already held by the client, or by delegation from another client. A grammar has been devised that allows the conditions under which a client may assume a name to be specified, and the conditions under which use of the name will be revoked. This allows complex security policies to be specified that define how clients of a service may interact with each other (through election, delegation and revocation), how clients interact with a service (by invoking operations or receiving events) and how clients and services may inter-operate. (For example, a client of a Login service may become a client of a file service.)

This approach allows great flexibility when integrating a number of services, and reduces the mismatch of policies common in heterogeneous systems. A flexible security definition is meaningless if not backed by a robust and efficient implementation. In this thesis we present a systems architecture that can be implemented efficiently, but that allows individual services to 'fine tune' the trade-offs between security, efficiency and freedom of policy expression. The architecture is inherently distributed and scalable, and includes mechanisms for rapid and selective revocation of privileges which may cascade between services and organisations.

UCAM-CL-TR-400

Scarlet Schwiderski:

Monitoring the behaviour of distributed systems

July 1996, 161 pages, PDF
PhD thesis (Selwyn College, April 1996)

Abstract: Monitoring the behaviour of computing systems is an important task. In active database systems, a detected system behaviour leads to the triggering of an ECA (event-condition-action) rule. ECA rules are employed for supporting database management system functions as well as external applications. Although distributed database systems are becoming more commonplace, active database research has to date focussed on centralised systems. In distributed debugging systems, a detected system behaviour is compared with the expected system behaviour. Differences illustrate erroneous behaviour. In both application areas, system behaviours are specified in terms of events: primitive events represent elementary occurrences and composite events represent complex occurrence patterns. At system runtime, specified primitive and composite events are monitored and event occurrences are detected. However, in active database systems events are monitored in terms of physical time and in distributed debugging systems events are monitored in terms of logical time. The notion of physical time is difficult in distributed systems because of their special characteristics: no global time, network delays, etc.

This dissertation is concerned with monitoring the behaviour of distributed systems in terms of physical

time, i.e. the syntax, the semantics, the detection, and the implementation of events are considered.

The syntax of primitive and composite events is derived from the work of both active database systems and distributed debugging systems; differences and necessities are highlighted.

The semantics of primitive and composite events establishes when and where an event occurs; the semantics depends largely on the notion of physical time in distributed systems. Based on the model for an approximated global time base, the ordering of events in distributed systems is considered, and the structure and handling of timestamps are illustrated. In specific applications, a simplified version of the semantics can be applied which is easier and therefore more efficient to implement.

Algorithms for the detection of composite events at system runtime are developed; event detectors are distributed to arbitrary sites and composite events are evaluated concurrently. Two different evaluation policies are examined: asynchronous evaluation and synchronous evaluation. Asynchronous evaluation is characterised by the ad hoc consumption of signalled event occurrences. However, since the signalling of events involves variable delays, the events may not be evaluated in the system-wide order of their occurrence. On the other hand, synchronous evaluation enforces events to be evaluated in the system-wide order of their occurrence. But, due to site failures and network congestion, the evaluation may block on a fairly long-term basis.

The prototype implementation realises the algorithms for the detection of composite events with both asynchronous and synchronous evaluation. For the purpose of testing, primitive event occurrences are simulated by distributed event simulators. Several tests are performed illustrating the differences between asynchronous and synchronous evaluation: the first is 'fast and unreliable' whereas the latter is 'slow and reliable'.

UCAM-CL-TR-401

Gavin Bierman:

A classical linear λ-calculus

July 1996, 41 pages, PDF

Abstract: This paper proposes and studies a typed λ-calculus for classical linear logic. I shall give an explanation of a multiple-conclusion formulation for classical logic due to Parigot and compare it to more traditional treatments by Prawitz and others. I shall use Parigot's method to devise a natural deduction fomulation of classical linear logic. This formulation is compared in detail to the sequent calculus formulation. In an appendix I shall also demonstrate a somewhat hidden connection with the paradigm of control operators for functional languages which gives a new computational interpretation of Parigot's techniques.

UCAM-CL-TR-402

G.J.F. Jones, J.T. Foote, K. Spärck Jones, S.J. Young:

Video mail retrieval using voice: report on collection of naturalistic requests and relevance assessments

September 1996, 21 pages, PDF

Abstract: This report discusses the rationale, design, collection and initial statistics of a message request and retrieved document relevance assessment set for the Cambridge Video Mail Rewtrieval (VMR) Project. This data set is designed to complement the VMR Database 1 (VMR1) message set and was designed for the testing of document searching methods being investigated in the VMR project. The combined message and request set is referred to as VMR1b.

UCAM-CL-TR-403

Paul Ronald Barham:

Devices in a multi-service operating system

October 1996, 131 pages, PostScript, DVI
PhD thesis (Churchill College, June 1996)

Abstract: Increases in processor speed and network and device bandwidth have led to general purpose workstations being called upon to process continuous media data in real time. Conventional operating systems are unable to cope with the high loads and strict timing constraints introduced when such applications form part of a multi-tasking workload. There is a need for the operating system to provide fine-grained reservation of processor, memory and I/O resources and the ability to redistribute these resources dynamically. A small group of operating systems researchers have recently proposed a "vertically-structured" architecture where the operating system kernel provides minimal functionality and the majority of operating system code executes within the application itself. This structure greatly simplifies the task of accounting for processor usage by applications. The prototype Nemesis operating system embodies these principles and is used as the platform for this work.

This dissertation extends the provision of Quality of Service guarantees to the I/O system by presenting an architecture for device drivers which minimises crosstalk between applications. This is achieved by clearly separating the data-path operations, which require careful accounting and scheduling, and the infrequent control-path operations, which require protection and concurrency control. The approach taken is to abstract and multiplex the I/O data-path at the lowest level possible so as to simplify accounting, policing

and scheduling of I/O resources and enable application-specific use of I/O devices.

The architecture is applied to several representative classes of device including network interfaces, network connected peripherals, disk drives and framestores. Of these, disks and framestores are of particular interest since they must be shared at a very fine granularity but have traditionally been presented to the application via a window system or file-system with a high-level and coarse-grained interface.

A device driver for the framestore is presented which abstracts the device at a low level and is therefore able to provide each client with guaranteed bandwidth to the framebuffer. The design and implementation of a novel client-rendering window system is then presented which uses this driver to enable rendering code to be safely migrated into a shared library within the client.

A low-level abstraction of a standard disk drive is also described which efficiently supports a wide variety of file systems and other applications requiring persistent storage, whilst providing guaranteed rates of I/O to individual clients. An extent-based file system is presented which can provide guaranteed rate file access and enables clients to optimise for application-specific access patterns.

UCAM-CL-TR-404

Kam Hong Shum:

Adaptive parallelism for computing on heterogeneous clusters

November 1996, 147 pages, PDF
PhD thesis (Darwin College, August 1996)

Abstract: Until recent years most parallel machines have been made up of closely-coupled microprocessor-based computers. With the advent of high-performance workstations and high speed networking, the aggregate computational power and memory capacity of workstation clusters have become attractive and indispensable resources for parallel computing. Techniques to harness the power of workstation cluster computing, however, require the development of practical methods for controlling heterogeneous resources dynamically.

This dissertation proposes an integrated framework that comprises two related parts. The first part of the framework is a software structure that enables parallel applications to be adaptable to workload imbalances at runtime. To realize the adaptation, applications are partitioned into small components called tasks. The tasks are then grouped into grains; each grain is an object that facilitates execution of tasks on a workstation. An application can therefore optimize its performance by the reconfiguration of task-to-grain and grain-to-workstation mappings. Based on the software structure, the implementation and evaluation of workload distribution schemes for data-parallel and task-parallel applications are presented. The second part of the framework is a resource management system that allocates resources to parallel applications through competition. The applications respond to allocation decisions by dynamic reconfiguration. The objectives of the system are to maximise the speedup of the parallel applications and, at the same time, to allocate workstations fairly and efficiently to the applications. A prototype implementation which provides a testbed for studying the dynamics of competition is structured.

In addition a new structure for organizing replicated parallel applications is developed and an architecture for a multi-user, multi-parallel program environment based on the proposed framework is suggested. The effectiveness of the concept and the framework is demonstrated by the results of experiments conducted on the testbed. Ther parallel applications involved in the experiments consist of block-matrix multiplication, cycle-searching of a non-linear cryptographic function, and simulators of an ATM network.

UCAM-CL-TR-405

Richard J. Boulton:

A tool to support formal reasoning about computer languages

November 1996, 21 pages, PostScript, DVI

Abstract: A tool to support formal reasoning about computer languages and specific language texts is described. The intention is to provide a tool that can build a formal reasoning system in a mechanical theorem prover from two specifications, one for the syntax of the language and one for the semantics. A parser, pretty-printer and internal representations are generated from the former. Logical representations of syntax and semantics, and associated theorem proving tools, are generated from the combination of the two specifications. The main aim is to eliminate tedious work from the task of prototyping a reasoning tool for a computer language, but the abstract specifications of the language also assist the automation of proof.

UCAM-CL-TR-406

Lawrence C. Paulson:

Tool support for logics of programs

November 1996, 31 pages, PDF, PostScript

Abstract: Proof tools must be well designed if they are to be more effective than pen and paper. Isabelle supports a range of formalisms, two of which are described (higher-order logic and set theory). Isabelle's representation of logic is influenced by logic programming: its "logical variables" can be used to implement step-wise refinement. Its automatic proof procedures are based on search primitives that are directly available to users. While emphasizing basic concepts, the article also discusses applications such as an approach to the analysis of security protocols.

UCAM-CL-TR-407

Sebastian Schoenberg:

The L4 microkernel on Alpha Design and implementation

September 1996, 51 pages, PostScript

Abstract: The purpose of a microkernel is to cover the lowest level of the hardware and to provide a more general platform to operating systems and applications than the hardware itself. This has made microkernel development increasingly interesting. Different types of microkernels have been developed, ranging from kernels which merely deal with the hardware infterface (Windows NT HAL), kernels especially for embedded systems (RTEMS), to kernels for multimedia streams and real time support (Nemesis) and general purpose kernels (L4, Mach).

The common opinion that microkernels lead to deterioration in system performance has been disproved by recent research. L4 is an example of a fast and small, multi address space, message-based microkernel, developed originally for Intel systems only. Based on the L4 interface, which should be as similar as possible on different platforms, the L4 Alpha version has been developed.

This work describes design decisions, implementation and interfaces of the L4 version for 64-bit Alpha processors.

UCAM-CL-TR-408

John Robert Harrison:

Theorem proving with the real numbers

November 1996, 147 pages, PostScript, DVI
PhD thesis (Churchill College, June 1996)

Abstract: This thesis discusses the use of the real numbers in theorem proving. Typically, theorem provers only support a few 'discrete' datatypes such as the natural numbers. However the availability of the real numbers opens up many interesting and important application areas, such as the verification of floating point hardware and hybrid systems. It also allows the formalization of many more branches of classical mathematics, which is particularly relevant for attempts to inject more rigour into computer algebra systems.

Our work is conducted in a version of the HOL theorem prover. We describe the rigorous definitional construction of the real numbers, using a new version of Cantor's method, and the formalization of a significant portion of real analysis. We also describe an advanced derived decision procedure for the 'Tarski subset' of real algebra as well as some more modest but practically useful tools for automating explicit calculations and routine linear arithmetic reasoning.

Finally, we consider in more detail two interesting application areas. We discuss the desirability of combining the rigour of theorem provers with the power and convenience of computer algebra systems, and explain a method we have used in practice to achieve this. We then move on to the verification of floating point hardware. After a careful discussion of possible correctness specifications, we report on two case studies, one involving a transcendental function.

We aim to show that a theory of real numbers is useful in practice and interesting in theory, and that the 'LCF style' of theorem proving is well suited to the kind of work we describe. We hope also to convince the reader that the kind of mathematics needed for applications is well within the abilities of current theorem proving technology.

UCAM-CL-TR-409

Lawrence C. Paulson:

Proving properties of security protocols by induction

December 1996, 24 pages, PDF, PostScript, DVI

Abstract: Security protocols are formally specified in terms of traces, which may involve many interleaved protocol runs. Traces are defined inductively. Protocol descriptions model accidental key losses as well as attacks. The model spy can send spoof messages made up of components decrypted from previous traffic.

Correctness properties are verified using the proof tool Isabelle/HOL. Several symmetric-key protocols have been studied, including Needham-Schroeder, Yahalom and Otway-Rees. A new attack has been discovered in a variant of Otway-Rees (already broken by Mao and Boyd). Assertions concerning secrecy and authenticity have been proved.

The approach rests on a common theory of messages, with three operators. The operator "parts" denotes the components of a set of messages. The operator "analz" denotes those parts that can be decrypted with known keys. The operator "synth" denotes those messages that can be expressed in terms of given components. The three operators enjoy many algebraic laws that are invaluable in proofs.

UCAM-CL-TR-410

John Harrison:

Proof style

January 1997, 22 pages, PostScript, DVI

Abstract: We are concerned with how to communicate a mathematical proof to a computer theorem prover. This can be done in many ways, while allowing the machine to generate a completely formal proof object. The most obvious choice is the amount of guidance required

from the user, or from the machine perspective, the degree of automation provided. But another important consideration, which we consider particularly significant, is the bias towards a 'procedural' or 'declarative' proof style. We will explore this choice in depth, and discuss the strengths and weaknesses of declarative and procedural styles for proofs in pure mathematics and for verification applications. We conclude with a brief summary of our own experiments in trying to combine both approaches.

UCAM-CL-TR-411

Monica Nesi:

Formalising process calculi in Higher Order Logic

January 1997, 182 pages, PDF
PhD thesis (Girton College, April 1996)

Abstract: In the past few years, several methods and tools based on process calculi have been developed for verifying properties of concurrent and communicating systems. In this dissertation the interactive theorem prover HOL is used as a framework for supporting reasoning about process calculi based on all the various components of their formal theory. The aim is to build a sound and effective tool to allow both verification of process specifications and meta-theoretic reasoning. In particular the process calculus CCS is embedded in HOL logic. This is achieved by first addressing the pure subset of this calculus (no value passing) and then extending it to its value-passing version. The CCS theory is mechanised in HOL by following a purely definitional approach. This means that new objects are embedded in HOL using definition mechanisms which guarantee that no inconsistencies are introduced in the logic, and by deriving new facts from definitions and/or previously proved theorems by formal proof.

Pure CCS agent expressions are encoded as a type in the HOL logic, in which initially actions are represented as strings, agents with infinite behaviour are given through the rec-notation and agent summation is the usual binary operator. Recursive agents are then allowed to be defined through systems of recursive equations and to be parameterised. This makes the type of CCS expressions polymorphic and parameterised on the parameters' type. Operational and behavioural semantics and modal logic are defined and their properties and laws derived in HOL. Several proof tools, such as inference rules, conversions and tactics, are developed to enable users to carry out their proofs in an interactive way and to automate them whenever possible. Properties of infinite state systems, e.g. a counter which can expand indefinitely, can be formally verified in the resulting proof environment.

Then value-passing CCS is mechanised in HOL by translating value-passing expressions into pure ones. This entails a more general polymorphic type for pure agent expressions that includes an indexed summation operator. The translation is proved to be correct with respect to the semantics of value-passing CCS and then used at meta-level, together with the HOL formalisation for pure CCS, for developing behavioural theories for the value-passing calculus. A proof environment is thus derived, in which users will directly work on the value-passing specifications. A verification example illustrates how proofs about the data are neatly separated from proofs about the process behaviour and how ω-data-rules can be used in a practical way to reason about value-passing agents defined over an infinite value domain.

UCAM-CL-TR-412

G.M. Bierman:

Observations on a linear PCF (preliminary report)

January 1997, 30 pages, PDF

Abstract: This paper considers some theoretical and practical issues concerning the use of linear logic as a logical foundation of functional programming languages such as Haskell and SML. First I give an operational theory for a linear PCF: the (typed) linear λ-calculus extended with booleans, conditional and nontermination. An operational semantics is given which corresponds in a precise way to the process of β-reduction which originates from proof theory. Using this operational semantics I define notions of observational equivalence (sometimes called contextual equivalence). Surprisingly, the linearity of the language forces a reworking of the traditional notion of a context (the details are given in an appendix). A co-inductively defined notion, applicative bi-simularity, is developed and compared with observational equivalence using a variant of Howe's method. Interestingly the equivalence of these two notions is greatly complicated by the linearity of the language. These equivalences are used to study a call-by-name translation of PCF into linear PCF. It is shown that this translation is adequate but not fully abstract. Finally I show how Landin's SECD machine can be adpacted to execute linear PCF programs.

UCAM-CL-TR-413

Lawrence C. Paulson:

Mechanized proofs of security protocols: Needham-Schroeder with public keys

January 1997, 20 pages, PDF, PostScript, DVI

Abstract: The inductive approach to verifying security protocols, previously applied to shared-key encryption, is here applied to the public key version of the Needham-Schroeder protocol. As before, mechanized proofs are performed using Isabelle/HOL. Both the original, flawed version and Lowe's improved version are studied; the properties proved highlight the distinctions between the two versions. The results are compared with previous analyses of the same protocol. The analysis reported below required only 30 hours of the author's time. The proof scripts execute in under three minutes.

UCAM-CL-TR-414

Martín Abadi, Andrew D. Gordon:

A calculus for cryptographic protocols
The SPI calculus

January 1997, 105 pages, PostScript

Abstract: We introduce the spi calculus, an extension of the pi calculus designed for the description and analysis of cryptographic protocols. We show how to use the spi calculus, particularly for studying authentication protocols. The pi calculus (without extension) suffices for some abstract protocols; the spi calculus enables us to consider cryptographic issues in more detail. We represent protocols as processes in the spi calculus and state their security properties in terms of coarse-grained notions of protocol equivalence.

UCAM-CL-TR-415

Steven Leslie Pope:

Application support for mobile computing

February 1997, 145 pages, PDF
PhD thesis (Jesus College, October 1996)

Abstract: In recent years small, completely portable computers have become available on the marketplace. There is demand for such computers, termed walkstations, to access network services while retaining their mobility, and to operate effectively in a range of conditions. Future office environments are expected to support wireless networks with bandwidths which are several orders of magnitude greater than are available outdoors. In such environments there will be powerful compute servers available for a walkstation's use.

This dissertation describes a novel architecture called Notus and its support for applications operating in a mobile environment. The concept of the traded handoff is introduced where applications are able to participate in the handoff process, rebuilding connections to the most appropriate service. This is expected to benefit walkstations which roam over large distances, where connections to servers would otherwise be strained, and also between heterogeneous networks where cooperation between the networks in performing a handoff might be problematic. It is also proposed in this dissertation that applications could benefit from the ability to migrate onto compute servers as a walkstation moves into the office environment. This enables both the walkstation to conserve its own resources, and applications to improve the service provided to the end user. Finally by interleaving a traded handoff with the migration process it is possible for a migrating application to easily rebuild its connections as it moves to a new host.

The Notus architecture has been implemented, including a traded handoff service and a new application migration service. The new application migration service was designed since existing application migration services are unsuited to mobile environments and it enables applications to migrate between heterogeneous hosts with little disruption. Applications which use the service are written in a standard compiled langauge, and normal running applications suffer little overhead. A number of existing applications which are representative of a walkstation's interactive desk-top environment have been adapted to use the Notus architecture, and are evaluated.

In summary, this work describes how mobility awareness and the support from appropriate tools, can enable walkstation applications to better adapt to a changing mobile environment, particularly when the walkstation is carried between different network types or over great distances.

UCAM-CL-TR-416

Donald Syme:

DECLARE: a prototype declarative proof system for higher order logic

February 1997, 25 pages, PDF

Abstract: This report describes DECLARE, a prototype implementation of a declarative proof system for simple higher order logic. The purpose of DECLARE is to explore mechanisms of specification and proof that may be incorporated into other theorem provers. It has been developed to aid with reasoning about operational descriptions of systems and languages. Proofs in DECLARE are expressed as proof outlines, in a language that approximates written mathematics. The proof language includes specialised constructs for (co-)inductive types and relations. The system includes an abstract/article mechanism that provides a way of isolating the process of formalisation from what results, and simultaneously allow the efficient separate processing of work units. After describing the system we discuss our approach on two subsidiary issues: automation and the interactive environment provided to the user.

UCAM-CL-TR-417

Peter J.C. Brown:

Selective mesh refinement for interactive terrain rendering

February 1997, 18 pages, PDF

Abstract: Terrain surfaces are often approximated by geometric meshes to permit efficient rendering. This paper describes how the complexity of an approximating irregular mesh can be varied across its domain in order to minimise the number of displayed facets while ensuring that the rendered surface meets pre-determined resolution requirements. We first present a generalised scheme to represent a mesh over a continuous range of resolutions using the output from conventional single-resolution approximation methods. We then describe an algorithm which extracts a surface from this representation such that the resolution of the surface is enhanced only in specific areas of interest. We prove that the extracted surface is complete, minimal, satisfies the given resolution constraints and meets the Delaunay triangulation criterion if possible. In addition, we present a method of performing smooth visual transitions between selectively-refined meshes to permit efficient animation of a terrain scene.

A HTML version of that report is at https://www.cl.cam.ac.uk/research/rainbow/publications/pjcb/tr417/

UCAM-CL-TR-418

Lawrence C. Paulson:

Mechanized proofs for a recursive authentication protocol

March 1997, 30 pages, PDF, PostScript

Abstract: A novel protocol has been formally analyzed using the prover Isabelle/HOL, following the inductive approach described in earlier work. There is no limit on the length of a run, the nesting of messages or the number of agents involved. A single run of the protocol delivers session keys for all the agents, allowing neighbours to perform mutual authentication. The basic security theorem states that session keys are correctly delivered to adjacent pairs of honest agents, regardless of whether other agents in the chain are compromised. The protocol's complexity caused some difficulties in the specification and proofs, but its symmetry reduced the number of theorems to prove.

UCAM-CL-TR-419

James Quentin Stafford-Fraser:

Video-augmented environments

April 1997, 91 pages, PDF
PhD thesis (Gonville & Caius College, February 1996)

Abstract: In the future, the computer will be thought of more as an assistant than as a tool, and users will increasingly expect machines to make decisions on their behalf. As with a human assistant, a machine's ability to make informed choices will often depend on the extent of its knowledge of activities in the world around it. Equipping personal computers with a large number of sensors for monitoring their environment is, however, expensive and inconvenient, and a preferable solution would involve a small number of input devices with a broad scope of application. Video cameras are ideally suited to many realworld monitoring applications for this reason. In addition, recent reductions in the manufacturing costs of simple cameras will soon make their widespread deployment in the home and office economically viable. The use of video as an input device also allows the creation of new types of user-interface, more suitable in some circumstances than those afforded by the conventional keyboard and mouse.

This thesis examines some examples of these 'Video-Augmented Environments' and related work, and then describes two applications in detail. The first, a 'software cameraman', uses the analysis of one video stream to control the display of another. The second, 'BrightBoard', allows a user to control a computer by making marks on a conventional whiteboard, thus 'augmenting' the board with many of the facilities common to electronic documents, including the ability to fax, save, print and email the image of the board. The techniques which were found to be useful in the construction of these applications are common to many systems which monitor real-world video, and so they were combined in a toolkit called 'Vicar'. This provides an architecture for 'video plumbing', which allows standard videoprocessing components to be connected together under the control of a scripting language. It is a single application which can be programmed to create a variety of simple Video-Augmented Environments, such as those described above, without the need for any recompilation, and so should simplify the construction of such applications in the future. Finally, opportunities for further exploration on this theme are discussed.

UCAM-CL-TR-420

Jonathan Mark Sewell:

Managing complex models for computer graphics

April 1997, 206 pages, PDF
PhD thesis (Queens' College, March 1996)

Abstract: Three-dimensional computer graphics is becoming more common as increasing computational power becomes more readily available. Although the images that can be produced are becoming more complex, users' expectations continue to grow. This dissertation examines the changes in computer graphics software that will be needed to support continuing growth in complexity, and proposes techniques for tackling the problems that emerge.

Increasingly complex models will involve longer rendering times, higher memory requirements, longer data transfer periods and larger storage capacities. Furthermore, even greater demands will be placed on the constructors of such models. This dissertation aims to describe how to construct scalable systems which can be used to visualise models of any size without requiring dedicated hardware. This is achieved by controlling the quality of the results, and hence the costs incurred. In addition, the use of quality controls can become a tool to help users handle the large volume of information arising from complex models.

The underlying approach is to separate the model from the graphics application which uses it, so that the model exists independently. By doing this, an application is free to access only the data which is required at any given time. For the application to function in this manner, the data must be in an appropriate form. To achieve this, approximation hierarchies are defined as a suitable new model structure. These utilise multiple representations of both objects and groups of objects at all levels in the model.

In order to support such a structure, a novel method is proposed for rapidly constructing simplified representations of groups of complex objects. By calculating a few geometrical attributes, it is possible to generate replacement objects that preserve important aspects of the originals. Such objects, once placed into an approximation hierarchy, allow rapid loading and rendering of large portions of a model. Extensions to rendering algorithms are described that take advantage of this structure.

The use of multiple representations encompasses not only different quality levels, but also different storage formats and types of objects. It provides a framework within which such aspects are hidden from the user, facilitating the sharing and re-use of objects. A model manager is proposed as a means of encapsulating these mechanisms. This software gives, as far as possible, the illusion of direct access to the whole complex model, while at the same time making the best use of the limited resources available.

UCAM-CL-TR-421

Michael Norrish:

An abstract dynamic semantics for C

May 1997, 31 pages, PDF, PostScript, DVI

Abstract: This report is a presentation of a formal semantics for the C programming language. The semantics has been defined operationally in a structured semantics style and covers the bulk of the core of the language. The semantics has been developed in a theorem prover (HOL), where some expected consequences of the language definition

UCAM-CL-TR-422

Antony Rowstron:

Using the BONITA primitives: a case study

May 1997, 19 pages, PDF

Abstract: The co-ordination language Linda has been used for parallel processing for many years. Linda uses a shared tuple space and a number of primitives to provide access to the tuple space and thereby enabling communication between processes executing concurrently. Linda provides asynchronous communication between processes, but synchronous access between the processes and the tuple spaces. The Bonita primitives are a different set of primitives that provide asynchronous access to the tuple spaces. The Bonita priomitives can emulate the primary Linda primitives and therefore provides both asynchronous access and synchronous access to tuple spaces. It has been previously claimed that asynchronous tuple space access primitives are required to provide new co-ordination constructs and to improve performance for geographically distributed processes which are required to co-ordinate distributed processes (or agents).

In this paper a talk program is used as an example to demonstrate that the concept of tuple spaces are well suited for process co-ordination for distributed processes (or agents), and to provide a comparison between the use of Linda primitives and the Bonita primitives. It is shown that asynchronous tuple space access is essential for such process co-ordination.

UCAM-CL-TR-423

Karl F. MacDorman:

Symbol grounding: Learning categorical and sensorimotor predictions for coordination in autonomous robots

May 1997, 170 pages, PDF
PhD thesis (Wolfson College, March 1997)

Abstract: To act intelligently, agents must be able to adapt to changing behavioural possibilities. This dissertation proposes a model that enables them to do this.

An agent learns sensorimotor predictions from spatiotemporal correlations in sensory projections, motor signals, and physiological variables. Currently elicited predictions constitute its model of the world.

Agents learn predictions for mapping between different sensory modalities. In one example a robot records sensory projections as points in a multidimensional space. It coordinates hand-eye movements by using closest-point approximations to map between vision and proprioception. Thus, one modality elicits predictions more closely identifiable with another. In a different example, an agent generalizes about a car's sensorimotor relations by weighting sensorimotor variables according to their mutual influence: it learns to navigate without any a priori model of the car's dynamics.

With feedback from miscategorization, an agent can develop links between categorical representations and the relevant objects they distinguish. Wavelet analysis provides a neurologically plausible means of accentuating invariance that can subserve categorization. In some experiments, categorical representations, derived from inter-category invariance after wavelet analysis, proved to be efficient and accurate at distinguishing different species of mushrooms.

In a simulation of fish chemoreception, agents learn sensorimotor predictions that uncover salient invariance in their environment. Predictions are formed by quantizing a sensory subspace after each dimension has been weighted according to its impact on physiological variables. As these predictions also map from motor signals to likely changes in sensory projections, the agent can chain backwards from desired outcomes to form plans for their attainment.

UCAM-CL-TR-424

Fabio Massacci:

Simplification with renaming: a general proof technique for tableau and sequent-based provers

May 1997, 26 pages, DVI

Abstract: Tableau and sequent calculi are the basis for most popular interactive theorem provers for hardware and software verification.

Yet, when it comes to decision procedures or automatic proof search, tableaux are orders of magnitude slower than Davis-Putnam, SAT based procedures or other techniques based on resolution.

To meet this challenge, this paper proposes a theoretical innovation: the rule of simplification, which plays the same role for tableaux as subsumption does for resolution, and unit for Davis-Putman.

This technique gives a unifying view of a number of tableaux-like calculi such as DPLL, KE, HARP, hypertableaux etc. For instance the stand-alone nature of the first-order Davis-Putnam-Longeman-Loveland procedure can be explained away as a case of Smullyan tableau with propositional simplification.

Besides its computational effectiveness, the simplicity and generality of simplification make its extension possible in a uniform way. We define it for propositional and first order logic and a wide range of modal logics. For a full-fledged first order simplification we combine it with another technique, renaming, which subsumes the use of free universal variables in sequent and tableau calculi.

New experimental results are given for random SAT and the IFIP benchmarks for hardware verification.

UCAM-CL-TR-425

Leslie Lamport, Lawrence C. Paulson:

Should your specification language be typed?

May 1997, 30 pages, PDF

Abstract: Most specification languages have a type system. Type systems are hard to get right, and getting them wrong can lead to inconsistencies. Set theory can serve as the basis for a specification language without types. This possibility, which has been widely overlooked, offers many advantages. Untyped set theory is simple and is more flexible than any simple typed formalism. Polymorphism, overloading, and subtyping can make a type system more powerful, but at the cost of increased complexity, and such refinements can never attain the flexibility of having no types at all. Typed formalisms have advantages too, stemming from the power of mechanical type checking. While types serve little purpose in hand proofs, they do help with mechanized proofs. In the absence of verification, type checking can catch errors in specifications. It may be possible to have the best of both worlds by adding typing annotations to an untyped specification language.

We consider only specification languages, not programming languages.

UCAM-CL-TR-426

Mark Humphrys:

Action selection methods using reinforcement learning

June 1997, 195 pages, PostScript
PhD thesis (Trinity Hall)

Abstract: The Action Selection problem is the problem of run-time choice between conflicting and heterogenous goals, a central problem in the simulation of whole creatures (as opposed to the solution of isolated uninterrupted tasks). This thesis argues that Reinforcement Learning has been overlooked in the solution of the Action Selection problem. Considering a

decentralised model of mind, with internal tension and competition between selfish behaviors, this thesis introduces an algorithm called "W-learning", whereby different parts of the mind modify their behavior based on whether or not they are succeeding in getting the body to execute their actions. This thesis sets W-learning in context among the different ways of exploiting Reinforcement Learning numbers for the purposes of Action Selection. It is a 'Minimize the Worst Unhappiness' strategy. The different methods are tested and their strengths and weaknesses analysed in an artificial world.

UCAM-CL-TR-427

Don Syme:

Proving Java type soundness

June 1997, 35 pages, PDF

Abstract: This technical report describes a machine checked proof of the type soundness of a subset of the Java language called Java_s. A formal semantics for this subset has been developed by Drossopoulou and Eisenbach, and they have sketched an outline of the type soundness proof. The formulation developed here complements their written semantics and proof by correcting and clarifying significant details; and it demonstrates the utility of formal, machine checking when exploring a large and detailed proof based on operational semantics. The development also serves as a case study in the application of 'declarative' proof techniques to a major property of an operational system.

UCAM-CL-TR-428

John Harrison:

Floating point verification in HOL Light: the exponential function

June 1997, 112 pages, PostScript, DVI

Abstract: In that they often embody compact but mathematically sophisticated algorithms, operations for computing the common transcendental functions in floating point arithmetic seem good targets for formal verification using a mechanical theorem prover. We discuss some of the general issues that arise in verifications of this class, and then present a machine-checked verification of an algorithm for computing the exponential function in IEEE-754 standard binary floating point arithmetic. We confirm (indeed strengthen) the main result of a previously published error analysis, though we uncover a minor error in the hand proof and are forced to confront several subtle issues that might easily be overlooked informally.

Our main theorem connects the floating point exponential to its abstract mathematical counterpart. The specification we prove is that the function has the correct overflow behaviour and, in the absence of overflow, the error in the result is less than 0.54 units in the last place (0.77 if the answer is denormalized) compared against the exact mathematical exponential function. The algorithm is expressed in a simple formalized programming language, intended to be a subset of real programming and hardware description languages. It uses underlying floating point operations (addition, multiplication etc.) that are assumed to conform to the IEEE-754 standard for binary floating point arithmetic.

The development described here includes, apart from the proof itself, a formalization of IEEE arithmetic, a mathematical semantics for the programming language in which the algorithm is expressed, and the body of pure mathematics needed. All this is developed logically from first principles using the HOL Light prover, which guarantees strict adherence to simple rules of inference while allowing the user to perform proofs using higher-level derived rules. We first present the main ideas and conclusions, and then collect some technical details about the prover and the underlying mathematical theories in appendices.

UCAM-CL-TR-429

Andrew D. Gordon, Paul D. Hankin, Søren B. Lassen:

Compilation and equivalence of imperative objects

June 1997, 64 pages, PostScript, DVI

Abstract: We adopt the untyped imperative object calculus of Abadi and Cardelli as a minimal setting in which to study problems of compilation and program equivalence that arise when compiling object-oriented languages. We present both a big-step and a small-step substitution-based operational semantics for the calculus. Our first two results are theorems asserting the equivalence of our substitution-based semantics with a closure-based semantics like that given by Abadi and Cardelli. Our third result is a direct proof of the correctness of compilation to a stack-based abstract machine via a small-step decompilation algorithm. Our fourth result is that contextual equivalence of objects coincides with a form of Mason and Talcott's CIU equivalence; the latter provides a tractable means of establishing operational equivalences. Finally, we prove correct an algorithm, used in our prototype compiler, for statically resolving method offsets. This is the first study of correctness of an object-oriented abstract machine, and of operational equivalence for the imperative object calculus.

UCAM-CL-TR-430

G.J.F. Jones, J.T. Foote, K. Sparck Jones, S.J. Young:

Video mail retrieval using voice: Report on topic spotting (Deliverable report on VMR task no. 6)

July 1997, 73 pages, PDF

Abstract: This report describes research on topic spotting in audio document retrieval carried out in years 2 and 3 of the Cambridge Video Mail Retrieval (VMR) project. Topic spotting within VMR was concerned with ad-hoc querying of a message archive using classical information retrieval techniques developed from experience with text archives. The report describes experiments using three approaches to document indexing: fixed-vocabulary keyword spotting, open-vocabulary search term indexing using phone lattices, and message transcription using large vocabulary speech recognition. Additional experiments investigate the combination of these techniques for improved retrieval effectiveness.

UCAM-CL-TR-431

Martin Richards:

The MCPL programming manual and user guide

July 1997, 70 pages, PDF

Abstract: MCPL is a programming language that has been derived from BCPL by the inclusion of features found in ML, C and Prolog. Like BCPL, it is typeless, uses a contiguous runtime stack and has no builtin garbage collector, but it does make extensive use of ML-like pattern matching. The low level aspects of the language resemble those of BCPL and C. MCPL uses its own function calling sequence, however it is designed to allow MCPL and C functions to call each other.

Notable features of MCPL are its pattern matching facilities and the simple way in which data structures are handled.

This document gives the definition of the language, its library and how to obtain and install the system.

UCAM-CL-TR-432

Lawrence C. Paulson:

On two formal analyses of the Yahalom protocol

July 1997, 16 pages, PDF, PostScript, DVI

Abstract: The Yahalom protocol is one of those analyzed by Burrows et al. in the BAN paper. Based upon their analysis, they have proposed modifications to make the protocol easier to understand and analyze. Both versions of Yahalom have now been proved, using Isabelle/HOL, to satisfy strong security goals. The mathematical reasoning behind these machine proofs is presented informally.

The new proofs do not rely on a belief logic; they use an entirely different formal model, the inductive method. They confirm the BAN analysis and the advantages of the proposed modifications. The new proof methods detect more flaws than BAN and analyze protocols in finer detail, while remaining broadly consistent with the BAN principles. In particular, the proofs confirm the explicitness principle of Abadi and Needham.

UCAM-CL-TR-433

Martin Richards:

Backtracking algorithms in MCPL using bit patterns and recursion

July 1997, 80 pages, PDF

Abstract: This paper presents example programs, implemented in MCPL, that use bit pattern techniques and recursion for the efficient solution of various tree search problems.

UCAM-CL-TR-434

Martin Richards:

Demonstration programs for CTL and μ-calculus symbolic model checking

August 1997, 41 pages, PDF

Abstract: This paper presents very simple implementations of Symbolic Model Checkers for both Computational Tree Logic (CTL) and μ-calculus. They are intended to be educational rather than practical. The first program discovers, for a given non-deterministic finite state machine (NFSM), the states for which a given CTL formula holds. The second program does the same job for μ-calculus formulae.

For simplicity the number of states in the NFSM has been limited to 32 and a bit pattern representation is used to represent the boolean functions involved. It would be easy to extend both programs to use ordered binary decision diagrams more normally used in symbolic model checking.

The programs include lexical and syntax analysers for the formulae, the model checking algorithms and drivers to exercise them with respect to various simple machines. The programs are implemented in MCPL. A brief summary of MCPL is given at the end.

UCAM-CL-TR-435

Peter Sewell:

Global/local subtyping for a distributed π-calculus

August 1997, 57 pages, PostScript

Abstract: In the design of mobile agent programming languages there is a tension between the implementation cost and the expressiveness of the communication mechanisms provided. This paper gives a static type system for a distributed π-calculus in which the input and output of channels may be either global or local. This allows compile-time optimization where possible but retains the expressiveness of channel communication. Subtyping allows all communications to be invoked uniformly. Recursive types and products are included. The distributed π-calculus used integrates location and migration primitives from the Distributed Join Calculus with asynchronous π communication, taking a simple reduction semantics. Some alternative calculi are discussed.

UCAM-CL-TR-436

W.F. Clocksin:

A new method for estimating optical flow

November 1997, 20 pages, PDF

Abstract: Accurate and high density estimation of optical flow vectors in an image sequence is accomplished by a method that estimates the velocity distribution function for small overlapping regions of the image. Because the distribution is multimodal, the method can accurately estimate the change in velocity near motion contrast borders. Large spatiotemporal support without sacrificing spatial resolution is a feature of the method, so it is not necessary to smooth the resulting flow vectors in a subsequent operation, and there is a certain degree of resistance to aperture and aliasing effects. Spatial support also provides for the accurate estimation of long-range displacements, and subpixel accuracy is achieved by a simple weighted mean near the mode of the velocity distribution function.

The method is demonstrated using image sequences obtained from the analysis of ceramic and metal materials under stress. The performance of the system under degenerate conditions is also analysed to provide insight into the behaviour of optical flow methods in general.

UCAM-CL-TR-437

William S. Harbison:

Trusting in computer systems

December 1997, 95 pages, PDF
PhD thesis (Wolfson College, May 1997)

Abstract: We need to be able to reason about large systems, and not just about their components. For this we need new conceptual tools, and this dissertation therefore indicates the need for a new methodology which will allow us to better identify areas of possible conflict or lack of knowledge in a system.

In particular, it examines at the concept of trust, and how this can help us to understand the basic security aspects of a system. The main proposal of this present work is that systems are viewed in a manner which analyses the conditions under which they have been designed to perform, and the circumstances under which they have been implemented, and then compares the two. This problem is then examined from the point of what is being trusted in a system, or what it is being trusted for.

Starting from an approach developed in a military context, we demonstrate how this can lead to unanticipated risks when applied inappropriately. We further suggest that 'trust' be considered a relative concept, in contast to the more usual usage, and that it is not the result of knowledge but a substitute for it. The utility of these concepts is in their ability to quantify the risks associated with a specific participant, whether these are explicitly accepted by them, or not.

We finally propose a distinction between 'trust' and 'trustworthy' and demonstrate that most current uses of the term 'trust' are more appropriately viewed as statements of 'trustworthiness'. Ultimately, therefore, we suggest that the traditional "Orange Book" concept of trust resulting from knowledge can violate the security policy of a system.

UCAM-CL-TR-438

Feng Shi:

An architecture for scalable and deterministic video servers

November 1997, 148 pages, PDF
PhD thesis (Wolfson College, June 1997)

Abstract: A video server is a storage system that can provide a repository for continuous media (CM) data and sustain CM stream delivery (playback or recording) through networks. The voluminous nature of CM data demands a video server to be scalable in order to serve a large number of concurrent client requests. In addition, deterministic services can be provided by a video server for playback because the characteristics of

variable bit rate (VBR) video can be analysed in advance and used in run-time admission control (AC) and data retrieval.

Recent research has made gigabit switches a reality, and the cost/performance ratio of microprocessors and standard PCs is dropping steadily. It would be more cost effective and flexible to use off-the-shelf components inside a video server with a scalable switched network as the primary interconnect than to make a special purpose or massively parallel multiprocessor based video server. This work advocates and assumes such a scalable video server structure in which data is striped to multiple peripherals attached directly to a switched network.

However, most contemporary distributed file systems do not support data distribution across multiple networked nodes, let alone providing quality of service (QoS) to CM applications at the same time. It is the observation of this dissertation that the software system framework for network striped video servers is as important as the scalable hardware architecture itself. This leads to the development of a new system architecture, which is scalable, flexible and QoS aware, for scalable and deterministic video servers. The resulting srchitecture is called Cadmus from sCAlable and Deterministic MUlitmedia Servers.

Cadmus also provides integrated solutions to AC and actual QoS enforcement in storage nodes. This is achieved by considering resources such as CPU buffer, disk, and network, simultaneously but not independently and by including both real-time (RT) and non-real-time (NRT) activities, In addition, the potential to smooth the variability of VBR videos using read-ahead under client buffer constraints is identified. A new smoothing algorithm is presented, analysed, and incorporated into the Cadmus architecture.

A prototype implementation of Cadmus has been constructed based on distributed object computing and hardware modules directly connected to an Asynchronous Transfer Mode (ATM) network. Experiments were performed to evaluate the implementation and demonstrate the utility and feasibility of the architecture and its AC criteria.

UCAM-CL-TR-439

David A. Halls:

Applying mobile code to distributed systems

December 1997, 158 pages, PDF
PhD thesis (June 1997)

Abstract: Use of mobile code can make distributed systems and the abstractions they provide more flexible to build and use.

Richer functionality can be given to the interaction between processes by allowing code to be sent between them. More convenient, application level operations can be made over a network. By making higher order language features transmissible, distributed components can be tightly bound together when they communicate. At the same time familiar distributed systems can be built using mobile code.

Mobile code can make distributed systems adaptable to application needs. Rather than fixing the interface to a resource and the pattern of interaction with it, a minimal interface can be defined and code implementing higher level interfaces placed alongside it as and when required. These higher level interfaces can be application specific, allowing for interaction patterns that were unknownat the time the resource was made available. Sending code close to a resource can also reduce network usage because the point of interaction with it moves.

The combination of document markup supporting hypertext and a language supporting state-saving allows for stateful client-server sessions with stateless servers and lightweight clients. Putting dormant mobile code in documents provides an alternative to holding knowledge of application functionality on a server machine or running arbitrary code on a client machine.

Mobile code helps to support user mobility. Personalised environments that support state saving can follow a user between computers. Heterogeneous state-saving allows a user's programs to be relocated between computers. By using a mobile code system with language support for state-saving, applications can direct arbitrary component migration without priming program servers with specific support.

In summary, this dissertation supports the thesis that mobile code can be used to enhance distributed systems.

UCAM-CL-TR-440

Lawrence C. Paulson:

Inductive analysis of the internet protocol TLS

December 1997, 19 pages, PDF, PostScript

Abstract: Internet browsers use security protocols to protect confidential messages. An inductive analysis of TLS (a descendant of SSL 3.0) has been performed using the theorem prover Isabelle. Proofs are based on higher-order logic and make no assumptions concerning beliefs or finiteness. All the obvious security goals can be proved; session resumption appears to be secure even if old session keys have been compromised. The analysis suggests modest changes to simplify the protocol.

TLS, even at an abstract level, is much more complicated than most protocols that researchers have verified. Session keys are negotiated rather than distributed, and the protocol has many optional parts. Nevertheless, the resources needed to verify TLS are modest. The inductive approach scales up.

UCAM-CL-TR-441

Lawrence C. Paulson:

A generic tableau prover and its integration with Isabelle

January 1998, 16 pages, PDF, PostScript, DVI

Abstract: A generic tableau prover has been implemented and integrated with Isabelle. It is based on leantap but is much more complicated, with numerous modifications to allow it to reason with any supplied set of tableau rules. It has a higher-order syntax in order to support the binding operators of set theory; unification is first-order (extended for bound variables in obvious ways) instead of higher-order, for simplicity.

When a proof is found, it is returned to Isabelle as a list of tactics. Because Isabelle verifies the proof, the prover can cut corners for efficiency's sake without compromising soundness. For example, it knows almost nothing about types.

UCAM-CL-TR-442

Jacques Fleuriot, Lawrence C. Paulson:

A combination of nonstandard analysis and geometry theorem proving, with application to Newton's Principia

January 1998, 13 pages, PostScript, DVI

Abstract: The theorem prover Isabelle is used to formalise and reproduce some of the styles of reasoning used by Newton in his Principia. The Principia's reasoning is resolutely geometric in nature but contains "infinitesimal" elements and the presence of motion that take it beyond the traditional boundaries of Euclidean Geometry. These present difficulties that prevent Newton's proofs from being mechanised using only the existing geometry theorem proving (GTP) techniques.

Using concepts from Robinson's Nonstandard Analysis (NSA) and a powerful geometric theory, we introduce the concept of an infinitesimal geometry in which quantities can be infinitely small or infinitesimal. We reveal and prove new properties of this geometry that only hold because infinitesimal elements are allowed and use them to prove lemmas and theorems from the Principia.

UCAM-CL-TR-443

Lawrence C. Paulson:

The inductive approach to verifying cryptographic protocols

February 1998, 46 pages, PDF, PostScript

Abstract: Informal arguments that cryptographic protocols are secure can be made rigorous using inductive definitions. The approach is based on ordinary predicate calculus and copes with infinite-state systems. Proofs are generated using Isabelle/HOL. The human effort required to analyze a protocol can be as little as a week or two, yielding a proof script that takes a few minutes to run.

Protocols are inductively defined as sets of traces. A trace is a list of communication events, perhaps comprising many interleaved protocol runs. Protocol descriptions incorporate attacks and accidental losses. The model spy knows some private keys and can forge messages using components decrypted from previous traffic. Three protocols are analyzed below: Otway-Rees (which uses shared-key encryption), Needham-Schroeder (which uses public-key encryption), and a recursive protocol (which is of variable length).

One can prove that event ev always precedes event ev′ or that property P holds provided X remains secret. Properties can be proved from the viewpoint of the various principals: say, if A receives a final message from B then the session key it conveys is good.

UCAM-CL-TR-444

Peter Sewell:

From rewrite rules to bisimulation congruences

May 1998, 72 pages, PostScript

Abstract: The dynamics of many calculi can be most clearly defined by reduction semantics. To work with a calculus, however, an understanding of operational congruences is fundamental; these can often be given tractable definitions or characterisations using a labelled transition semantics. This paper considers calculi with arbitary reduction semantics of three simple classes, firstly ground term rewriting, then left-linear term rewriting, and then a class which is esentially the action calculi lacking substantive name binding. General definitions of labelled transitions are given in each case, uniformly in the set of rewrite rules, and without requiring the prescription of additional notions of observation. They give rise to bisimulation congruences. As a test of the theory it is shown that bisimulation for a fragment of CCS is recovered. The transitions generated for a fragment of the Ambient Calculus of Cardelli and Gordon, and for SKI combinators, are also discussed briefly.

UCAM-CL-TR-445

Michael Roe, Bruce Christianson, David Wheeler:

Secure sessions from weak secrets

July 1998, 12 pages, PDF

Abstract: Sometimes two parties who share a weak secret k (such as a password) wish to share a strong secret s (such as a session key) without revealing information about k to a (possibly active) attacker. We assume that both parties can generate strong random numbers and forget secrets, and present three protocols for secure strong secret sharing, based on RSA, Diffie-Hellman and El-Gamal. As well as being simpler and quicker than their predecessors, our protocols also have slightly stronger security properties: in particular, they make no cryptographic use of s and so impose no subtle restrictions upon the use which is made of s by other protocols.

UCAM-CL-TR-446

K. Spärck Jones, S. Walker, S.E. Robertson:

A probabilistic model of information and retrieval: development and status

August 1998, 74 pages, PostScript, DVI

Abstract: The paper combines a comprehensive account of the probabilistic model of retrieval with new systematic experiments on TREC Programme material. It presents the model from its foundations through its logical development to cover more aspects of retrieval data and a wider range of system functions. Each step in the argument is matched by comparative retrieval tests, to provide a single coherent account of a major line of research. The experiments demonstrate, for a large test collection, that the probabilistic model is effective and robust, and that it responds appropriately, with major improvements in performance, to key features of retrieval situations.

UCAM-CL-TR-447

Giampaolo Bella, Lawrence C. Paulson:

Are timestamps worth the effort? A formal treatment

September 1998, 12 pages, PDF

Abstract: Theorem proving provides formal and detailed support to the claim that timestamps can give better freshness guarantees than nonces do, and can simplify the design of crypto-protocols. However, since they rely on synchronised clocks, their benefits are still debatable. The debate should gain from our formal analysis, which is achieved through the comparison of a nonce-based crypto-protocol, Needham-Schroeder, with its natural modification by timestamps, Kerberos.

UCAM-CL-TR-448

G.M. Bierman:

A computational interpretation of the $\lambda\mu$ calculus

September 1998, 10 pages, PDF

Abstract: This paper proposes a simple computational interpretation of Parigot's $\lambda\mu$-calculus. The $\lambda\mu$-calculus is an extension of the typed λ-calculus which corresponds via the Curry-Howard correspondence to classical logic. Whereas other work has given computational interpretations by translating the $\lambda\mu$-calculus into other calculi, I wish to propose here that the $\lambda\mu$-calculus itself has a simple computational interpretation: it is a typed λ-calculus which is able to save and restore the runtime environment. This interpretation is best given as a single-step semantics which, in particular, leads to a relatively simple, but powerful, operational theory.

UCAM-CL-TR-449

Florian Kammüller, Markus Wenzel:

Locales A sectioning concept for Isabelle

October 1998, 16 pages, PDF

Abstract: Locales are a means to define local scopes for the interactive proving process of the theorem prover Isabelle. They delimit a range in which fixed assumptions are made, and theorems are proved which depend on these assumptions. A locale may also contain constants and associated with pretty printing syntax.

Locales can be seen as a simple form of modules. They are similar to sections as in Automath or Coq. Locales are used to enhance abstract reasoning. It also discusses some implementation issues.

UCAM-CL-TR-450

Jacobus Erasmus van der Merwe:

Open service support for ATM

November 1998, 164 pages, PDF
PhD thesis (St John's College, September 1997)

Abstract: Asynchronous Transfer Mode (ATM) technology provides superior transfer capabilities in an environment in which multiple services are provided and carried by a single network. Fully exploiting this potential is hampered by the assumption by standards bodies that a single control architecture, which was derived from a mono-service network, will filfil the needs of all applications in such a multi service environment.

While this weakness has been widely recognised, previous efforts to address it have met with limited success. This can be largely attributed to the fact that such attempts have often been proposed to replace one monolithic system with another. Avoiding this "one-size-fits-all" approach, this dissertation presents an Open Service Support Architecture (OSSA), in which multiple control architectures can be operational simultaneously in the same physical network. In this manner different control architectures, which provide diverse functionality and were designed to different models, can be accommodated.

A key concept of the OSSA is the partitioning of switch resources by a software entity called a Divider. The subset of switch resources is callled a switchlet, and the Divider allows each switchlet to be controlled by a separate control architecture. The divider polices the invocations of a control architecture to contain it in its allocated switchlet. Switchlets are combined into virtual networks, and a software entity called the Network Builder automates this process. The Network Builder allows virtual networks of arbitrary topology to by dynamically created and modified, and each virtual network is therefore controlled by a separate instance of a control architecture. The dissertation presents a proof of concept implementation of the OSSA, and reports on the efficiency of various implementations of crucial components.

The dynamic creation of virtual networks in the OSSA means that the usage of resources in an ATM network now needs to be considered on three time scales: short time scales for cell switching, longer time scales for connection creation, and even longer time scales for virtual network creation. The use of measurement based estimates of effective bandwidth to effect resource management at the two longer time scales of interest is investigated and the results presented.

Finally, the flexibility offered by the OSSA enables the use of service specific control architectures (SSCAs). An SSCA is a control architecture which utilises service specific knowledge in its manipulation of network resources, thereby providing a more efficient service than would be possible with a general purpose control architecture. The design and implementation of an SCCA for continuous media conferencing is presented.

UCAM-CL-TR-451

Sean Rooney:

The structure of open ATM control architectures

November 1998, 183 pages, PDF
PhD thesis (Wolfson College, February 1998)

Abstract: The design of networks capable of supporting a large number of different services is one of the principal areas of network research. ATM, by virtue of its ability to give resource guarantees to arbitrary services, is likely to become the transport protocol for high-speed service-independent networks. The ATM control plane — handling as it does the needs of many distinct services with diverse constraints — is necessarily complex. The approach adopted by industry has been to try and adopt the techniques used to control the telephony network to ATM.

This dissertation argues that current monolithic ATM signalling standards reduce the service support flexibility that was the principal motivating factor behind the introduction of ATM. It argues that a more open approach is required if ATM is to be able to meet the demands of a decentralised, deregulated service provision market.

A natural approach in handling complex systems is to divide them into simpler elements. This dissertation considers two types of separation. Firstly it shows how a clean separation can be made between the ATM control plane and the switch, allowing them to be implemented and evolve independently. Secondly, as a consequence of the clear separation of the controller from the switch, it demonstrates how several distinct control architectures can coexist simultaneously on the same physical network, removing the need for one single monolithic control architecture and allowing network operators to choose the control architecture most appropriate for their purposes.

The utility and practicality of this approach are demonstrated through the description of the structure of a switch-independent control architecture which efficiently implements a complete range of ATM control operations. Such a control architecture is more versatile than conventional signalling systems, while the environment in which it executes allows both standard and proprietary signalling systems to coexist.

Network robustness is of primary importance for large scale commercial networks. This dissertation shows how management of an open control network can be made less centralised and more adaptive. These qualities are particularly important in an environment in which there may be many network operators managing distinct networks simultaneously.

UCAM-CL-TR-452

Florian Kammüller, Lawrence C. Paulson:

A formal proof of Sylow's theorem An experiment in abstract algebra with Isabelle Hol

November 1998, 30 pages, PDF

Abstract: The theorem of Sylow is proved in Isabelle HOL. We follow the proof by Wielandt that is more general than the original and uses a non-trivial combinatorial identity. The mathematical proof is explained in some detail leading on to the mechanization of group theory and the necessary combinatorics in Isabelle. We present the mechanization of the proof in detail giving

reference to theorems contained in an appendix. Some weak points of the experiment with respect to a natural treatment of abstract algebraic reasoning give rise to a discussion of the use of module systems to represent abstract algebra in theorem provers. Drawing from that, we present tentative ideas for further research into a section concept for Isabelle.

UCAM-CL-TR-453

Michael Norrish:

C formalised in HOL

December 1998, 156 pages, PDF
PhD thesis (August 1998)

Abstract: We present a formal semantics of the C programming language, covering both the type system and the dynamic behaviour of programs. The semantics is wide-ranging, covering most of the language, with its most significant omission being the C library. Using a structural operational semantics we specify transition relations for C's expressions, statements and declarations in higher order logic.

The consistency of our definition is assured by its specification in the HOL theorem prover. With the theorem prover, we have used the semantics as the basis for a set of proofs of interesting theorems about C. We investigate properties of expressions and statements separately.

In our chapter of results about expressions, we begin with two results about the interaction between the type system and the dynamic semantics. We have both type preservation, that the values produced by expressions conform to the type predicted for them; and type safety, that typed expressions will not block, but will either evaluate to a value, or cause undefined behaviour. We then also show that two broad classes of expression are deterministic. This last result is of considerable practical value as it makes later verification proofs significantly easier.

In our chapter of results about statements, we prove a series of derived rules that provide C with Floyd-Hoare style "axiomatic" rules for verifying properties of programs. These rules are consequences of the original semantics, not independently stated axioms, so we can be sure of their soundness. This chapter also proves the correctness of an automatic tool for constructing post-conditions for loops with break and return statements.

Finally, we perform some simple verification case studies, going some way towards demonstrating practical utility for the semantics and accompanying tools.

This technical report is substantially the same as the PhD thesis I submitted in August 1998. The minor differences between that document and this are principally improvements suggested by my examiners Andy Gordon and Tom Melham, whom I thank for their help and careful reading.

UCAM-CL-TR-454

Andrew M. Pitts:

Parametric polymorphism and operational equivalence

December 1998, 39 pages, PDF

Abstract: Studies of the mathematical properties of impredicative polymorphic types have for the most part focused on the polymorphic lambda calculus of Girard-Reynolds, which is a calculus of total polymorphic functions. This paper considers polymorphic types from a functional programming perspective, where the partialness arising from the presence of fixpoint recursion complicates the nature of potentially infinite ('lazy') datatypes. An approach to Reynolds' notion of relational parametricity is developed that works directly on the syntax of a programming language, using a novel closure operator to relate operational behaviour to parametricity properties of types. Working with an extension of Plotkin's PCF with $\forall$-types, lazy lists and existential types, we show by example how the resulting logical relation can be used to prove properties of polymorphic types up to operational equivalence.

UCAM-CL-TR-455

G.M. Bierman:

Multiple modalities

December 1998, 26 pages, PDF

Abstract: Linear logic removes the structural rules of weakening and contraction and adds an S4-like modality (written !). Only formulae of the form $!\phi$ can be weakened or contracted. An interesting question is whether these two capabilities can be separated using two different modalities. This question was studied semantically in a comprehensive paper by Jacobs. This paper considers the question proof-theoretically, giving sequent calculus, natural deduction and axiomatic formulations.

UCAM-CL-TR-456

Joshua Robert Xavier Ross:

An evaluation based approach to process calculi

January 1999, 206 pages, PDF
PhD thesis (Clare College, July 1998)

Abstract: Process calculi have, starting with Milner's CCS, traditionally been expressed by specifying the operational semantics in terms of action-labelled transition relations between process expressions. Normally this has been done using transitions that are inductively defined by rules following the structure of the process expressions. This approach has been very successful but has suffered from certain problems. One of these is that the construction of weak, branching-time congruences has not been as simple as one might wish. In particular the natural weak bisimulations are not congruences, typically shown up by the introduction of summation. Secondly this method has not lent itself to the development of congruences for calculi that combine features of concurrency and higher-order functional languages. Another problem is more aesthetic. It is that in order to write these transition relations we need to use silent (τ) actions which are supposed to be unobservable. However, we need to represent them explicitly and make explicit reference to them in defining the congruence relations.

In this thesis, an approach to process calculi based on evaluation to committed forms is presented. In particular two process calculi are given. The first is a first-order CCS-like calculus, NCCS. This demonstrates the possibility of giving natural weak branching-time congruences, with such features as summation, without the use of explicit silent silent actions. Various bisumulations are defined on NCCS, and these are related to existing equivalences for CCS. The second is a higher order calculus, based on CML; a higher-order functional language extended with concurrent features. Again it is shown that a natural weak branching-time congruence exists. In both cases a transition relation is also given and the relationship between evaluation and transition is shown.

UCAM-CL-TR-457

Andrew D. Gordon, Paul D. Hankin:

A concurrent object calculus: reduction and typing

February 1999, 63 pages, PDF

Abstract: We obtain a new formalism for concurrent object-oriented languages by extending Abadi abd Cardelli's imperative object calculus with operators for concurrency from the μ-calculus and with operators for synchronisation based on mutexes. Our syntax of terms is extremely expressive; in a precise sense it unifies notions of expression, process, store, thread and configuration. We present a chemical-style reduction semantics, and prove it equivalent to a structural operational semantics. We identify a deterministic fragment that is closed under reduction and show that it includes the imperative object calculus. A collection of type systems for object oriented constructs is at the heart of Abadi and Cardelli's work. We recast one of Abadi and Cardelli's first-order type systems with object types and subtyping in the setting of our calculus and prove subject reduction. Since our syntax of terms includes both stores and running expressions, we avoid the need to separate store typing from typing of expressions. We translate communication channels and choice-free asynchronous μ-calculus into our calculus to illustrate its expressiveness; the types of read-only and write-only channels are supertypes of read-write channels.

UCAM-CL-TR-458

Lawrence C. Paulson:

Final coalgebras as greatest fixed points in ZF set theory

March 1999, 25 pages, PDF

Abstract: A special final coalgebra theorem, in the style of Aczel (1988), is proved within standard Zermelo-Fraenkel set theory. Aczel's Anti-Foundation Axiom is replaced by a variant definition of function that admits non-well-founded constructions. Variant ordered pairs and tuples, of possibly infinite length, are special cases of variant functions. Analogues of Aczel's solution and substitution lemmas are proved in the style of Rutten and Turi (1993). The approach is less general than Aczel's, but the treatment of non-well-founded objects is simple and concrete. The final coalgebra of a functor is its greatest fixedpoint. Compared with previous work (Paulson, 1995a), iterated substitutions and solutions are considered, as well as final coalgebras defined with respect to parameters. The disjoint sum construction is replaced by a smoother treatment of urelements that simplifies many of the derivations. The theory facilitates machine implementation of recursive definitions by letting both inductive and coinductive definitions be represented as fixedpoints. It has already been applied to the theorem prover Isabelle (Paulson, 1994).

UCAM-CL-TR-459

Mohamad Afshar:

An open parallel architecture for data-intensive applications

July 1999, 225 pages, PostScript, DVI
PhD thesis (King's College, December 1998)

Abstract: Data-intensive applications consist of both declarative data-processing parts and imperative computational parts. For applications such as climate modelling, scale hits both the computational aspects which are typically handled in a procedural programming language, and the data-processing aspects which are handled in a database query language. Although parallelism has been successfully exploited in the data-processing parts by parallel evaluation of

database queries associated with the application, current database query languages are poor at expressing the computational aspects, which are also subject to scale.

This thesis proposes an open architecture that delivers parallelism shared between the database, system and application, thus enabling the integration of the conventionally separated query and non-query components of a data-intensive application. The architecture is data-model independent and can be used in a variety of different application areas including decision-support applications, which are query based, and complex applications, which comprise procedural language statements with embedded queries. The architecture encompasses a unified model of parallelism and the realisation of this model in the form of a language within which it is possible to describe both the query and non-query components of data-intensive applications. The language enables the construction of parallel applications by the hierarchical composition of platform-independent parallel forms, each of which implements a form of task or data parallelism. These forms may be used to determine both query and non-query actions.

Queries are expressed in a declarative language based on "monoid comprehensions". The approach of using monoids to model data types and monoid homomorphisms to iterate over collection types enables mathematically provable compile-time optimisations whilst also facilitating multiple collection types and data type extensibility. Monoid comprehension programs are automatically transformed into parallel programs composed of applications of the parallel forms, one of which is the "monoid homomorphism". This process involves identifying the parts of a query where task and data parallelism are available and mapping that parallelism onto the most suitable form. Data parallelism in queries is mapped onto a form that implements combining tree parallelism for query evaluation and dividing tree parallelism to realise data partitioning. Task parallelism is mapped onto two separate forms that implement pipeline and independent parallelism. This translation process is applied to all comprehension queries including those in complex applications. The result is a skeleton program in which both the query and non-query parts are expressed within a single language. Expressions in this language are amenable to the application of optimising skeleton rewrite rules.

A complete prototype of the decision-support architecture has been constructed on a 128-cell MIMD parallel computer. A demonstration of the utility of the query framework is performed by modelling some of OQL and a substantial subset of SQL. The system is evaluated for query speedup with a number of hardware configurations using a large music catalogue database. The results obtained show that the implementation delivers the performance gains expected while offering a convenient definition of the parallel environment.

UCAM-CL-TR-460

Giampaolo Bella:

Message reception in the inductive approach

March 1999, 16 pages, PDF

Abstract: Cryptographic protocols can be formally analysed in great detail by means of Paulson's Inductive Approach, which is mechanised by the theorem prover Isabelle. The approach only relied on message sending (and noting) in order to keep the models simple. We introduce a new event, message reception, and show that the price paid in terms of runtime is negligible because old proofs can be reused. On the other hand, the new event enhances the global expressiveness, and makes it possible to define an accurate notion of agents' knowledge, which extends and replaces Paulson's notion of spy's knowledge. We have designed new guarantees to assure each agent that the peer does not know the crucial message items of the session. This work thus extends the scope of the Inductive approach. Finally, we provide general guidance on updating the protocols analysed so far, and give examples for some cases.

UCAM-CL-TR-461

Joe Hurd:

Integrating Gandalf and HOL

March 1999, 11 pages, PDF

Abstract: Gandalf is a first-order resolution theorem-prover, optimized for speed and specializing in manipulations of large clauses. In this paper I describe GANDALF TAC, a HOL tactic that proves goals by calling Gandalf and mirroring the resulting proofs in HOL. This call can occur over a network, and a Gandalf server may be set up servicing multiple HOL clients. In addition, the translation of the Gandalf proof into HOL fits in with the LCF model and guarantees logical consistency.

UCAM-CL-TR-462

Peter Sewell, Paweł T. Wojciechowski, Benjamin C. Pierce:

Location-independent communication for mobile agents: a two-level architecture

April 1999, 31 pages, PostScript

Abstract: We study communication primitives for interaction between mobile agents. They can be classified into two groups. At a low level there are location dependent primitives that require a programmer to know the current site of a mobile agent in order to communicate with it. At a high level there are location independent primitives that allow communication with a mobile agent irrespective of its current site and of any migrations. Implementation of these requires delicate distributed infrastructure. We propose a simple calculus of agents that allows implementation of such distributed infrastructure algorithms to be expressed as encodings, or compilations, of the whole calculus into the fragment with only location dependent communication. These encodings give executable descriptions of the algorithms, providing a clean implementation strategy for prototype languages. The calculus is equipped with a precise semantics, providing a solid basis for understanding the algorithms and reasoning about their correctness and robustness. Two sample infrastructure algorithms are presented as encodings.

UCAM-CL-TR-463

Peter Sewell, Jan Vitek:

Secure composition of insecure components

April 1999, 44 pages, PostScript

Abstract: Software systems are becoming heterogeneous: instead of a small number of large programs from well-established sources, a user's desktop may now consist of many smaller components that interact in intricate ways. Some components will be downloaded from the network from sources that are only partially trusted. A user would like to know that a number of security properties hold, e.g. that personal data is not leaked to the net, but it is typically infaesible to verify that such components are well-behaved. Instead they must be executed in a secure environment, or wrapper, that provides fine-grain control of the allowable interactions between them, and between components and other system resources.

In this paper we study such wrappers, focussing on how they can be expressed in a way that enables their security properties to be stated and proved rigorously. We introduce a model programming language, the box-π calculus, that supports composition of software components and the enforcement of security policies. Several example wrappers are expressed using the calculus; we explore the delicate security properties they guarantee.

UCAM-CL-TR-464

Boaz Lerner, William Clocksin, Seema Dhanjal, Maj Hultén, Christipher Bishop:

Feature representation for the automatic analysis of fluorescence in-situ hybridization images

May 1999, 36 pages, PDF

Abstract: Fast and accurate analysis of flourescence in-situ hybridization (FISH) images will depend mainly upon two components: a classifier to discriminate between artifacts and valid signal data, and well discriminating features to represent the signals. Our previous work has focused on the first component. To investigate the second component, we evaluate candidate feature sets by illustrating the probability density functions and scatter plots for the features. This analysis provides insight into dependencies between features, indicates the relative importancce of members of a feature set, and helps in identifying sources of potential classification errors. The analysis recommends several intensity and hue-based features for representing FISH signals. The recommendation is confirmed by the probability of misclassification using a two-layer neural network (NN), and also by a feature selection technique making use of a class separability criterion. Represented by these intensity and hue-based features, 90% of valid signals and artifacts are corrently classified using the NN.

UCAM-CL-TR-465

Boaz Lerner, Seema Dhanjal, Maj Hultén:

Gelfish – graphical environment for labelling FISH images

May 1999, 20 pages, PDF

Abstract: Dot counting in flourescence in-situ hybridization (FISH) images that relies on an automatic focusing method for obtaining clearly defined images is prone to errors. Our recently developed system has dispensed with automatic focusing, and insteaqd relies on a larger statistical sample of the specimen at a fixed focal plane. The system is based on well-discriminating features to represent the signals and a neural network classifier to discriminate between artifacts and valid signal data. Results showed that nearly 90% of valid signals and artifacts of two flourophores within 400 FISH images were correctly classified. To train the classifier, accurate labelling of the image is required. GELFISH is a Graphical Environment for Labelling FISH images that enables the labelling of FISH signals and the rejection of unanalysable nuclei simply and rapidly. Feedback provided by the environment allows the user to correct the results of labelling effortlessly by clicking GELFISH buttons using the mouse. Furthermore, GELFISH is flexible and can be modified easily for additional FISH applications. Implemented using popular software, the environment can be employed on any computer by any user.

UCAM-CL-TR-466

Boaz Lerner, William Clocksin, Seema Dhanjal, Maj Hultén, Christipher Bishop:

Automatic signal classification in fluorescence in-situ hybridization images

May 1999, 24 pages, PDF

Abstract: Previous systems for dot counting in fluorescence in-situ hybridization (FISH) images have relied on an automatic focusing method for obtaining a clearly defined image. Because signals are distributed in three dimensions within the nucleus and artifacts such as debris and background fluorescence can attract the focusing method , valid signals can be left unfocused or unseen. This leads to dot counting errors, which increase with the number of probes. The approach described here dispenses with automatic focusing, and instead relies on a larger statistical sample of the specimen at a fixed focal plane. Images across the specimen can be obtained in significantly less time if a fixed focal plane is used. A trainable classifier based on a neural network is used to discriminate between valid and artifact signals represented by a set of features. This improves on previous classification schemes that are based on non-adaptable decision boundaries and are trained using only examples of valid signals. Trained by examples of valid and artifact signals, three NN classifiers, two of them hierarchical, each achieve between 83% and 87% classification accuracy on unseen data. When data is pre-discriminated in this way, errors in dot counting can be significantly reduced.

UCAM-CL-TR-467

Lawrence C. Paulson:

Mechanizing UNITY in Isabelle

June 1999, 22 pages, PDF

Abstract: UNITY is an abstract formalism for proving properties of concurrent systems, which typically are expressed using guarded assignments [Chandy and Misra 1988]. UNITY has been mechanized in higher-order logic using Isabelle, a proof assistant. Safety and progress primitives, their weak forms (for the substitution axiom) and the program composition operator (union) have been formalized. To give a feel for the concrete syntax, the paper presents a few extracts from the Isabelle definitions and proofs. It discusses a small example, two-process mutual exclusion. A mechanical theory of unions of programs supports a degree of compositional reasoning. Original work on extending program states is presented and then illustrated through a simple example involving an array of processes.

UCAM-CL-TR-468

Stephen Paul Wilcox:

Synthesis of asynchronous circuits

July 1999, 250 pages, PDF
PhD thesis (Queens' College, December 1998)

Abstract: The majority of integrated circuits today are synchronous: every part of the chip times its operation with reference to a single global clock. As circuits become larger and faster, it becomes progressively more difficult to coordinate all actions of the chip to the clock. Asynchronous circuits do not suffer from this problem, because they do not require global synchronization; they also offer other benefits, such as modularity, lower power and automatic adaptation to physical conditions.

The main disadvantage of asynchronous circuits is that there are few tools to help with design. This thesis describes a new synthesis tool for asynchronous modules, which combines a number of novel ideas with existing methods for finite state machine synthesis. Connections between modules are assumed to have unbounded finite delays on all wires, but fundamental mode is used inside modules, rather than the pessimistic speed-independent or quasi-delay-insensitive models. Accurate technology-specific verification is performed to check that circuits work correctly.

Circuits are described using a language based upon the Signal Transition Graph, which is a well-known method for specifying asynchronous circuits. Concurrency reduction techniques are used to produce a large number of circuits that conform to a given specification. Circuits are verified using a simulation algorithm derived from the work of Brzozowski and Seger, and then performance estimations are obtained by a gate-level simulator utilising a new estimation of waveform slopes. Circuits can be ranked in terms of high speed, low power dissipation or small size, and then the best circuit for a particular task chosen.

Results are presented that show significant improvements over most circuits produced by other synthesis tools. Some circuits are twice as fast and dissipate half the power of equivalent speed-independent circuits. Specification examples are provided which show that the front-end specification is easier to use than current specification approaches. The price that must be paid for the improved performance is decreased reliability and technology dependence of the circuits produced; the proposed tool can also can a very long time to produce a result.

UCAM-CL-TR-469

Jacques Désiré Fleuriot:

A combination of geometry theorem proving and nonstandard analysis,

with application to Newton's Principia

August 1999, 135 pages, PDF
PhD thesis (Clare College, March 1999)

Abstract: Sir Isaac Newton's Philosophiae Naturalis Principia Mathematica (the Principia) was first published in 1687 and set much of the foundations that led to profound changes in modern science. Despite the influence of the work, the elegance of the geometrical techniques used by Newton is little known since the demonstrations of most of the theorems set out in it are usually done using calculus. Newton's reasoning also goes beyond the traditional boundaries of Euclidian geometry with the presence of both motion and infinitesimals.

This thesis describes the mechanization of lemmas and propositions from the Principia using formal tools developed in the generic theorem prover Isabelle. We discuss the formalisation of a geometry theory based on existing methods from automated geometry theorem proving. The theory contains extra geometric notions, including the definition of the ellipse and its tangent, that enables us to deal with the motion of bodies and other physical aspects.

We introduce the formalization of a theory of filters and ultrafilters, and the purely definitional construction of the hyperreal numbers of Nonstandard Analysis (NSA). The hyperreals form a proper field extension of the reals that contains new types of numbers including infinitesimals and infinite numbers.

By combining notions from NSA and geometry theorem proving, we propose an "infinitesimal" geometry in which quantities can be infinitely small. This approach then reveals the the new properties of the geometry that only hold because infinitesimal elements are allowed. We also mechanize some analytic geometry and use it to verify the geometry theories of Isabelle.

We then report on the main application of this framework. We discuss the formalization of several results from the Principia and give a detailed case study of one of its most important propositions: the Propositio Kepleriana. An anomaly is revealed in Newton's reasoning through our rigorous mechanization.

Finally we present the formalization of a portion of mathematical analysis using the nonstandard approach. We mechanize both standard and nonstandard definitions of familiar concepts, prove their equivalence, and use nonstandard arguments to provide intuitive yet rigorous proofs of many of their properties.

UCAM-CL-TR-470

Florian Kammüller:

Modular reasoning in Isabelle

August 1999, 128 pages, PDF
PhD thesis (Clare College, April 1999)

Abstract: This work is concerned with modules for higher order logic theorem provers, in particular Isabelle. Modules may be used to represent abstract mathematical structures. This is typical for applications in abstract algebra. In Chapter 1, we set out with the hypothesis that for an adequate representation of abstract structures we need modules that have a representation in the logic. We identify the aspects of locality and adequacy that are connected to the idea of modules in theorem provers.

In Chapter 2, we compare systems of interactive theorem provers and their applicability to abstract algebra. Furthermore we investigate a different family of proof systems based on type theory in Section 2.4.

We validate our hypothesis by performing a large case study in group theory: a mechanization of Sylow's theorem in Chapter 3.

Drawing from the experience gained by this large case study, we develop a concept of locales in Chapter 4 that captures local definitions, pretty printing syntax and local assumptions. This concept is implemented and released with Isabelle version 98-1.

However, this concept alone is not sufficient to describe abstract structures. For example, structures like groups and rings need a more explicit representation as objects in the logic. A mechanization of dependent Σ-types and Π-types as typed sets in higher order logic is produced in Chapter 5 to represent structures adequately.

In Chapter 6, we test our results by applying the two concepts we developed in combination. First, we reconsider the Sylow case study. Furthermore, we demonstrate more algebraic examples. Factorization of groups, direct product of groups, and ring automorphisms are constructions that form themselves groups, which is formally proved. We also discuss the proof of the full version of Tarski's fixed point theorem. Finally we consider how operations on modules can be realized by structures as dependent types. Locales are used in addition; we illustrate the reuse of theorems proved in a locale and the construction of a union of structures.

UCAM-CL-TR-471

Robert M. Brady, Ross J. Anderson, Robin C. Ball:

Murphy's law, the fitness of evolving species, and the limits of software reliability

September 1999, 14 pages, PDF

Abstract: We tackle two problems of interest to the software assurance community. Firstly, existing models of software development (such as the waterfall and spiral models) are oriented towards one-off software development projects, while the growth of mass market computing has led to a world in which most software consists of packages which follow an evolutionary development model. This leads us to ask whether anything

interesting and useful may be said about evolutionary development. We answer in the affirmative. Secondly, existing reliability growth models emphasise the Poisson distribution of individual software bugs, while the empirically observed reliability growth for large systems is asymptotically slower than this. We provide a rigorous explanation of this phenomenon. Our reliability growth model is inspired by statistical thermodynamics, but also applies to biological evolution. It is in close agreement with experimental measurements of the fitness of an evolving species and the reliability of commercial software products. However, it shows that there are significant differences between the evolution of software and the evolution of species. In particular, we establish maximisation properties corresponding to Murphy's law which work to the advantage of a biological species, but to the detriment of software reliability.

UCAM-CL-TR-472

Ben Y. Reis:

Simulating music learning with autonomous listening agents: entropy, ambiguity and context

September 1999, 200 pages, PDF
PhD thesis (Queens' College, July 1999)

Abstract: Music learning describes the gradual process of acculturation through which listeners in different cultures develop diverse sets of musical preferences and intuitions. This dissertation describes Maestro, a system designed over the course of this research to simulate certain aspects of music listening and learning.

In order to maintain the unbaised flexibility necessary for handling music from different styles, Maestro does not incorporate any a priori style-specific knowledge into its design. Instead, Maestro is based on a bottom up approach that maximises the use of perceptual information present in a performance.

Maestro's operation involves four stages: it first segments a musical performance on-line according to perceptual cues (segmentation) and constructs an appropriate model of the performance (modelling), based on the context modelling paradigm. This model is simultaneously used to generate expectations about upcoming events (prediction) and to interpret events once they have arrived (parsing).

Ambiguity is an essential part of music listening, especially in the context of learning, and can cause multiple hypotheses of interpretation to arise. A novel multi-agent methodology is developed and incorporated into Maestro for generating, maintaining, and reconciling these hypotheses. An information theoretic approach, based on measuring two types of entropy, is used to objectively evaluate the system's relative prediction performance. It is also found that entropy, along with a measure of agent activation, is useful for identifying and classifying different types of ambiguity.

Experiments performed with a collection of 100 Bach chorale melodies provides a basis for comparison with previous machine modelling reseasrch and with data from human subjects. A much larger collection of roughly 8,000 folk songs from different cultures enables significant large scale and panstylistic music learning experiments to be performed. Perceptually guided segmentation is argued to yield more cognitively realistic context models than other methods, and it is also empirically shown to yield more efficient models for prediction. Additionally, an adaptive modelling strategy allows appropriate multiple-step-ahead predictions to be generated. Finally a distributed, agent-based parsing methodology is developed and implemented.

The system provides insights into what implications certain theories from cognitive musicology have when put into practice. Maestro's flexible design together with the range of experiments performed and the diverse corpus of musical data enable a thorough and systematic machine-simulated study of key aspects of musical learning to be carried out.

UCAM-CL-TR-473

Clemens Ballarin:

Computer algebra and theorem proving

October 1999, 122 pages, PDF
PhD thesis (Darwin College)

Abstract: Is the use of computer algebra technology beneficial for mechanised reasoning in and about mathematical domains? Usually it is assumed that it is. Many works in this area, however, either have little reasoning content, or use symbolic computation only to simplify expressions. In work that has achieved more, the methods used do not scale up. They trust the computer algebra system either too much or too little.

Computer algebra systems are not as rigorous as many provers. They are not logically sound reasoning systems, but collections of algorithms. We classify soundness problems that occur in computer algebra systems. While many algorithms and their implementations are perfectly trustworthy the semantics of symbols is often unclear and leads to errors. On the other hand, more robust approaches to interface external reasoners to provers are not always practical because the mathematical depth of proof algorithms in computer algebra can be enormous.

Our own approach takes both trustworthiness of the overall system and efficiency into account. It relies on using only reliable parts of a computer algebra system which can be achieved by using a suitable library, and deriving specifications for these algorithms from their literature.

We design and implement an interface between the prover Isabelle and the computer algebra library Sumit and use it to prove non-trivial theorems from coding

theory. This is based on mechanisation of the algebraic theories of rings and polynomials. Coding theory is an area where proofs do have a substantial amount of computational content. Also it is realistic to assume that the verification of an encoding or decoding device cound be undertaken in, and indeed, be simplified by, such a system.

The reason why semantics of symbols is often unclear in current computer algebra systems is not mathematical difficulty, but the design of those systems. For Gaussian elimination we show how the soundness problem can be fixed by a small extention, and without using efficiency. This is a prerequisite for the efficient use of the algorithm in a prover.

UCAM-CL-TR-474

Boaz Lerner:

A Bayesian methodology and probability density estimation for fluorescence in-situ hybridization signal classification

October 1999, 31 pages, PDF

Abstract: Previous research has indicated the significance of accurate classification of flourescence in-situ hybridization (FISH) signals when images are captured in a fixed focal plane without relying on an autofocusing mechanism. Based on well-discriminating features and a trainable neural network (NN) classifier, a previous system enabled highly-accurate classification of valid signals and artifacts of two fluorophores. However, since training and optimisation of an NN require extensive resources and experimentation, we investigate in this work a simpler alternative for the NN classifier – the naive Bayesian classifier (NBC). The Bayesian methodology together with an independence assumption allow the NBC to predict the a posteriori probability of class membership using estimated class-conditional densities. Densities measured by three methods: single Gaussian estimation (SGE; parametric method), Gaussian mixture model (GMM; semi-parametic method) and kernel density estimation (KDE; non-parametric method) are evaluated for this purpose. The accuracy of the NBC employing data modelled by SGE is found to be similar to that based on GMM, slightly inferior to that based on KDE but widely inferior to that of the NN. Therefore, when supporting the two classifiers, the system enables a trade-off between the NN performance and the NBC simplicity. Finally, the evaluation of the NBC accuracy provides a mechanism for both model and feature selection.

UCAM-CL-TR-475

Boaz Lerner, Neil D. Lawrence:

A comparison of state-of-the-art classification techniques with application to cytogenetics

October 1999, 34 pages, PDF

Abstract: Several state of the art techniques: a neural network, Bayesian neural network, support vector machine and naive Bayesian classifier are experimentally evaluated in discriminating flourescence in-situ hybridization (FISH) signals. Highly-accurate classification of signals from real data and artifacts of two cytogenetic probes (colours) is required for detecting abnormalities in the data. More than 3100 FISH signals are classified by the techniques into colour and as real or artifact with accuracies of around 98% and 88%, respectively. The results of the comparison also show a trade-off between simplicity represented by the naive Bayesian classifier and high classification performance represented by the other techniques.

UCAM-CL-TR-476

Mark Staples:

Linking ACL2 and HOL

November 1999, 23 pages, PDF

Abstract: This report describes ACL2PII, a system which dynamically links the theorem provers ACL2 and Hol, using the PROSPER project's Plug-In interface to Hol. The focus of the system is on making ACL2 theorems available from within Hol. In a motivating example we show how to transfer results from ACL2's 'small machine' theory. This theory highlights two of ACL2's strengths: symbolic simulation and the fast execution of operationally defined functions. This allows ACL2 specifications to be readily validated against real world requirements. The ACL2PII system allows Hol users to capitalise on results about such ACL2 specifications. ACL2 and Hol are both general purpose theorem provers, but Hol is slightly more expressive, and has growing infrastructure for interoperability with other systems. This report assumes a passing knowledge of both ACL2 and Hol.

UCAM-CL-TR-477

Gian Luca Cattani, Glynn Winskel:

Presheaf models for CCS-like languages

November 1999, 46 pages, PDF

Abstract: The aim of this paper is to harness the mathematical machinery around presheaves for the purposes of process calculi. Joyal, Nielsen and Winskel proposed a general definition of bisimulation from open maps. Here we show that open-map bisimulations within a range of presheaf models are congruences for a general process language, in which CCS and related languages are easily encoded. The results are then transferred to traditional models for processes. By first establishing the congruence results for presheaf models, abstract, general proofs of congruence properties can be provided and the awkwardness caused through traditional models not always possessing the cartesian liftings, used in the break-down of process operations, are side-stepped. The abstract results are applied to show that hereditary history-preserving bisimulation is a congruence for CCS-like languages to which is added a refinement operator on event structures as proposed by van Glabbeek and Goltz.

UCAM-CL-TR-478

Peter Sewell, Jan Vitek:

Secure composition of untrusted code: wrappers and causality types

November 1999, 36 pages, PostScript

Abstract: We consider the problem of assembling concurrent software systems from untrusted or partially trusted off-the-shelf components, using wrapper programs to encapsulate components and enforce security policies. In previous work we introduced the box-π process calculus with constrained interaction to express wrappers and discussed the rigorous formulation of their security properties. This paper addresses the verification of wrapper information flow properties. We present a novel causal type system that statically captures the allowed flows between wrapped possibly-badly-typed components; we use it to prove that a unidirectional-flow wrapper enforces a causal flow property.

UCAM-CL-TR-479

Geraint Price:

The interaction between fault tolerance and security

December 1999, 144 pages, PDF
PhD thesis (Wolfson College, June 1999)

Abstract: This dissertation studies the effects on system design when including fault tolerance design principles within security services.

We start by looking at the changes made to the trust model within protocol design, and how moving away from trusted server design principles affects the structure of the protocol. Taking the primary results from this work, we move on to study how control in protocol execution can be used to increase assurances in the actions of legitimate participants. We study some examples, defining two new classes of attack, and note that by increasing client control in areas of protocol execution, it is possible to overcome certain vulnerabilities.

We then look at different models in fault tolerance, and how their adoption into a secure environment can change the design principles and assumptions made when applying the models.

We next look at the application of timing checks in protocols. There are some classes of timing attack that are difficult to thwart using existing techniques, because of the inherent unreliability of networked communication. We develop a method of converting the Quality of Service mechanisms built into ATM networks in order to achieve another layer of protection against timing attacks.

We then study the use of primary-backup mechanisms within server design, as previous work on server replication in security centres on the use of the state machine approach for replication, which provides a higher degree of assurance in system design, but adds complexity.

We then provide a design for a server to reliably and securely store objects across a loosely coupled, distributed environment. The main goal behind this design was to realise the ability for a client to exert control over the fault tolerance inherent in the service.

The main conclusions we draw from our research are that fault tolerance has a wider application within security than current practices, which are primarily based on replicating servers, and clients can exert control over the protocols and mechanisms to achieve resilience against differing classes of attack. We promote some new ideas on how, by challenging the prevailing model for client-server architectures in a secure environment, legitimate clients can have greater control over the services they use. We believe this to be a useful goal, given that the client stands to lose if the security of the server is undermined.

UCAM-CL-TR-480

Mike Gordon:

Programming combinations of deduction and BDD-based symbolic calculation

December 1999, 24 pages, PDF

Abstract: Theorem provers descended from LCF allow their users to write complex proof tools that provide high assurance that false theorems will not be proved. This paper describes some experiments in extending the 'LCF approach' to enable BDD-based symbolic algorithms to be programmed with a similar assurance. The

deduction is supplied by the HOL system and the BDD algorithms by the BuDDy package.

UCAM-CL-TR-481

Mike Gordon, Ken Friis Larsen:

Combining the Hol98 proof assistant with the BuDDy BDD package

December 1999, 71 pages, PDF

Abstract: Theorem provers descended from LCF allow their users to write complex proof tools with high assurance that false theorems will not be proved. This report describes an experimental system that extends the LCF approach to enable combinations of deduction and BDD-based symbolic calculation to be programmed with a similar assurance. The deduction is supplied by the Hol98 system and the BDD algorithms by Jørn Lind-Nielsen's BuDDy package.

The main idea is to provide LCF-style support to a set of inference rules for judgements ρ t $\mapsto$ b, where ρ is an order-inducing map from HOL variables to BDD variables, t is a HOL term and b is a BDD. A single oracle rule allows a HOL theorem $\vdash$ t to be deduced from ρ t $\mapsto$ TRUE.

This report is intended to serve as documentation for the Hol98 library HolBddLib. It is partly an exposition of standard results, partly tutorial and partly an account of research in combining deduction and symbolic state enumeration.

UCAM-CL-TR-482

John Daugman:

Biometric decision landscapes

January 2000, 15 pages, PDF

Abstract: This report investigates the "decision landscapes" that characterize several forms of biometric decision making. The issues discussed include: (i) Estimating the degrees-of-freedom associated with different biometrics, as a way of measuring the randomness and complexity (and therefore the uniqueness) of their templates. (ii) The consequences of combining more than one biometric test to arrive at a decision. (iii) The requirements for performing identification by large-scale exhaustive database search, as opposed to mere verification by comparison against a single template. (iv) Scenarios for Biometric Key Cryptography (the use of biometrics for encryption of messages). These issues are considered here in abstract form, but where appropriate, the particular example of iris recognition is used as an illustration. A unifying theme of all four sets of issues is the role of combinatorial complexity, and its measurement, in determining the potential decisiveness of biometric decision making.

UCAM-CL-TR-483

Hendrik Jaap Bos:

Elastic network control

January 2000, 184 pages, paper copy
PhD thesis (Wolfson College, August 1999)

UCAM-CL-TR-484

Richard Tucker:

Automatic summarising and the CLASP system

January 2000, 190 pages, PDF
PhD thesis (1999)

Abstract: This dissertation discusses summarisers and summarising in general, and presents CLASP, a new summarising system that uses a shallow semantic representation of the source text called a "predication cohesion graph".

Nodes in the graph are "simple predications" corresponding to events, states and entities mentioned in the text; edges indicate related or similar nodes. Summary content is chosen by selecting some of these predications according to criteria of "importance", "representativeness" and "cohesiveness". These criteria are expressed as functions on the nodes of a weighted graph. Summary text is produced either by extracting whole sentences from the source text, or by generating short, indicative "summary phrases" from the selected predications.

CLASP uses linguistic processing but no domain knowledge, and therefore does not restrict the subject matter of the source text. It is intended to deal robustly with complex texts that it cannot analyse completely accurately or in full. Experiments in summarising stories from the Wall Street Journal suggest there may be a benefit in identifying important material in a semantic representation rather than a surface one, but that, despite the robustness of the source representation, inaccuracies in CLASP's linguistic analysis can dramatically affect the readability of its summaries. I discuss ways in which this and other problems might be overcome.

UCAM-CL-TR-485

Daryl Stewart, Myra VanInwegen:

Three notes on the interpretation of Verilog

January 2000, 47 pages, PDF

Abstract: In order to simplify the many constructs available in the Verilog Hardware Description Language two methods were used to normalise code before analysis, scalarisation and hierarchy flattening.

A method for scalarising Verilog expressions is described and the replacement of expressions with scalarised versions is considered. This then forms the basis of an implementation of Verilog expression evaluation and normalization.

The organisation of hierarchical designs is described and an algorithm for flattening designs is derived from this.

UCAM-CL-TR-486

James Richard Thomas:

Stretching a point: aspect and temporal discourse

February 2000, 251 pages, paper copy
PhD thesis (Wolfson College, January 1999)

UCAM-CL-TR-487

Tanja Vos, Doaitse Swierstra:

Sequential program composition in UNITY

March 2000, 20 pages, PDF

Abstract: Large distributed applications are composed of basic blocks by using composition operators. In an ideal situation, one should be able to develop and verify each of these basic components by itself using compositionality theorems of the respective composition operators stating that properties of a composite program can be proved by proving properties of its components.

Generally two forms of distributed program composition can be distinguished: parallel composition and sequential composition. Parallel composition is standard in UNITY and is used when two distributed component-programs need to cooperate in one way or another. Sequential composition of UNITY programs is not part of core UNITY. It can however be very useful when we want a program to work with the results of another program. In this technical report we shall formally define and model sequential program composition within the HOL-UNITY embedding.

UCAM-CL-TR-488

Giampaolo Bella, Fabio Massacci, Lawrence Paulson, Piero Tramontano:

Formal verification of card-holder registration in SET

March 2000, 15 pages, PDF

Abstract: The first phase of the SET protocol, namely card-holder registration, has been modelled inductively. This phase is presented in outline and its formal model is described. A simple theorem has been proved using Isabelle/HOL, stating that a certification authority will certify a given key at most once. Many ambiguities, contradictions and omissions were noted when formalizing the protocol.

UCAM-CL-TR-489

Jong-Hyeon Lee:

Designing a reliable publishing framework

April 2000, 129 pages, PDF
PhD thesis (Wolfson College, January 2000)

Abstract: Due to the growth of the Internet and the widespread adoption of easy-to use web browsers, the web provides a new environment for conventional as well as new businesses. Publishing on the web is a fundamental and important means of supporting various activities on the Internet such as commercial transactions, personal home page publishing, medical information distribution, public key certification and academic scholarly publishing. Along with the dramatic growth of the web, the number of reported frauds is increasing sharply. Since the Internet was not originally designed for web publishing, it has some weaknesses that undermine its reliability.

How can we rely on web publishing? In order to resolve this question, we need to examine what makes people confident when reading conventional publications printed on paper, to investigate what attacks can erode confidence in web publishing, and to understand the nature of publishing in general.

In this dissertation, we examine security properties and policy models, and their applicability to publishing. We then investigate the nature of publishing so that we can extract its technical requirements. To help us understand the practical mechanisms which might satisfy these requirements, some applications of electronic publishing are discussed and some example mechanisms are presented.

We conclude that guaranteed integrity, verifiable authenticity and persistent availability of publications are required to make web publishing more reliable. Hence we design a framework that can support these properties. To analyse the framework, we define a security policy for web publishing that focuses on the guaranteed integrity and authenticity of web publications, and then describe some technical primitives that enable us to achieve our requirements. Finally, the Jikzi publishing system—an implementation of our framework—is presented with descriptions of its architecture and possible applications.

UCAM-CL-TR-490

Peter John Cameron Brown:

Selective mesh refinement for rendering

April 2000, 179 pages, paper copy
PhD thesis (Emmanuel College, February 1998)

Abstract: A key task in computer graphics is the rendering of complex models. As a result, there exist a large number of schemes for improving the speed of the rendering process, many of which involve displaying only a simplified version of a model. When such a simplification is generated selectively, i.e. detail is only removed in specific regions of a model, we term this selective mesh refinement.

Selective mesh refinement can potentially produce a model approximation which can be displayed at greatly reduced cost while remaining perceptually equivalent to a rendering of the original. For this reason, the field of selective mesh refinement has been the subject of dramatically increased interest recently. The resulting selective refinement methods, though, are restricted in both the types of model which they can handle and the form of output meshes which they can generate.

Our primary thesis is that a selectively refined mesh can be produced by combining fragments of approximations to a model without regard to the underlying approximation method. Thus we can utilise existing approximation techniques to produce selectively refined meshes in n-dimensions. This means that the capabilities and characteristics of standard approximation methods can be retained in our selectively refined models.

We also show that a selectively refined approximation produced in this manner can be smoothly geometrically morphed into another selective refinement in order to satisfy modified refinement criteria. This geometric morphing is necessary to ensure that detail can be added and removed from models which are selectively refined with respect to their impact on the current view frustum. For example, if a model is selectively refined in this manner and the viewer approaches the model then more detail may have to be introduced to the displayed mesh in order to ensure that it satisfies the new refinement criteria. By geometrically morphing this introduction of detail we can ensure that the viewer is not distracted by "popping" artifacts.

We have developed a novel framework within which these proposals have been verified. This framework consists of a generalised resolution-based model representation, a means of specifying refinement criteria and algorithms which can perform the selective refinement and geometric morphing tasks. The framework has allowed us to demonstrate that these twin tasks can be performed both on the output of existing approximation techniques and with respect to a variety of refinement criteria.

A HTML version of this thesis is at https://www.cl.cam.ac.uk/research/rainbow/publications/pjcb/thesis/

UCAM-CL-TR-491

Anna Korhonen, Genevive Gorrell, Diana McCarthy:

Is hypothesis testing useful for subcategorization acquisition?

May 2000, 9 pages, PDF

Abstract: Statistical filtering is often used to remove noise from automatically acquired subcategorization frames. In this paper, we compare three different approaches to filtering out spurious hypotheses. Two hypothesis tests perfrom poorly, compared to filtering frames on the basis of relative frequency. We discuss reasons for this and consider directions for future research.

UCAM-CL-TR-492

Paweł Tomasz Wojciechowski:

Nomadic Pict: language and infrastructure design for mobile computation

June 2000, 184 pages, PDF, PostScript
PhD thesis (Wolfson College, March 2000)

Abstract: Mobile agents – units of executing computation that can migrate between machines – are likely to become an important enabling technology for future distributed systems. We study the distributed infrastructures required for location-independent communication between migrating agents. These infrastructures are problematic: the choice or design of an infrastructure must be somewhat application-specific – any given algorithm will only have satisfactory performance for some range of migration and communication behaviour; the algorithms must be matched to the expected properties (and robustness demands) of applications and the failure characteristic of the communication medium. To study this problem we introduce an agent programming language – Nomadic Pict. It is designed to allow infrastructure algorithms to be expressed clearly, as translations from a high-level language to a lower level. The levels are based on rigorously-defined process calculi, which provide sharp levels of abstraction. In this dissertation we describe the language and use it to develop a distributed infrastructure for an example application. The language and examples have been implemented; we conclude with a description of the compiler and runtime system.

UCAM-CL-TR-493

Giampaolo Bella:

Inductive verification of cryptographic protocols

July 2000, 189 pages, PDF
PhD thesis (Clare College, March 2000)

Abstract: The dissertation aims at tailoring Paulson's Inductive Approach for the analysis of classical cryptographic protocols towards real-world protocols. The aim is pursued by extending the approach with new elements (e.g. timestamps and smart cards), new network events (e.g. message reception) and more expressive functions (e.g. agents' knowledge). Hence, the aim is achieved by analysing large protocols (Kerberos IV and Shoup-Rubin), and by studying how to specify and verify their goals.

More precisely, the modelling of timestamps and of a discrete time are first developed on BAN Kerberos, while comparing the outcomes with those of the BAN logic. The machinery is then applied to Kerberos IV, whose complicated use of session keys requires a dedicated treatment. Three new guarantees limiting the spy's abilities in case of compromise of a specific session key are established. Also, it is discovered that Kerberos IV is subject to an attack due to the weak guarantees of confidentiality for the protocol responder.

We develop general strategies to investigate the goals of authenticity, key distribution and non-injective agreement, which is a strong form of authentication. These strategies require formalising the agents' knowledge of messages. Two approaches are implemented. If an agent creates a message, then he knows all components of the message, including the cryptographic key that encrypts it. Alternatively, a broad definition of agents' knowledge can be developed if a new network event, message reception, is formalised.

The concept of smart card as a secure device that can store long-term secrets and perform easy computations is introduced. The model cards can be stolen and/or cloned by the spy. The kernel of their built-in algorithm works correctly, so they spy cannot acquire unlimited knowledge from their use. However, their functional interface is unreliable, so they send correct outputs in an unspecified order. The provably secure protocol based on smart cards designed by Shoup & Rubin is mechanised. Some design weaknesses (unknown to the authors' treatment by Bellare & Rogaway's approach) are unveiled, while feasible corrections are suggested and verified.

We realise that the evidence that a protocol achieves its goals must be available to the peers. In consequence, we develop a new a principle of prudent protocol design, goal availability, which holds of a protocol when suitable guarantees confirming its goals exist on assumptions that both peers can verify. Failure to observe our principle raises the risk of attacks, as is the case, for example, of the attack on Kerberos IV.

UCAM-CL-TR-494

Mark David Spiteri:

An architecture for the notification, storage and retrieval of events

July 2000, 165 pages, PDF
PhD thesis (Darwin College, January 2000)

Abstract: Event-driven and messaging infrastructures are emerging as the most flexible and feasible solution to enable rapid and dynamic integration and legacy and monolithic software applications into distributed systems. They also support deployment and enhancement of traditionally difficult-to-build active systems such as large scale collaborative environments and mobility aware architectures. However, complex systems issues like mobility, scalability, fedaration and persistence indicate a requirement for more advanced services within these infrastructures. The event notification paradigm is also applicable in emerging research areas such as modelling of business information flow within organisations, as well as workplace empowering through enchanced awareness of work practices relating to communication and interaction between individuals. In these areas further developments require complex interpretation and correlation of event information, highlighting the need for an event storage and retrieval service that provides the required groundwork.

It is the thesis of this this dissertation that the lack of a generic model for event representation and notification has restricted evolution within event driven applications. Furthermore, in order to empower existing applications and enable novel solutions, a crucial, and so-far-missing, service within event systems is capture, persistent storage, and meaningful retrieval of the messaging information driving these systems.

In order to address these issues, this dissertation defines a generic event model and presents a powerful event notification infrastructure that, amongst other structural contributions, embeds event storage functionality. An event repository architecture will then be presented that can capture and store events, as well as inject them back into distributed application components to simulate replay of sequences of activity. The general-purpose architecture presented is designed on the thesis that events are temporal indexing points for computing activities. Changes in the state of a distributed system can be captured in events, and replayed or reviewed at a later stage, supporting fault-tolerance, systems management, disconnected operation and mobility. The architecture delivers powerful querying of event histories, enabling extraction of simple and composite event patterns. This addresses the business requirements in several industries (such as finance, travel, news, retail and manufacturing) to locate temporal patterns of activity, as well as support applications like memory prosthesis tools and capture

of collaboration. The repository offers a selective store-and-forward functionality that enables messaging environments to scale and provide enhanced brokering and federation services.

In addition to enabling novel applications, the general-purpose infrastructure presented provides a more flexible approach to event notification, storage and retrieval, in areas where bespoke solutions had to be provided previously. The theoretical concepts illustrated in this dissertation are demonstrated through a working distributed implementation and deployment in several application scenarios.

UCAM-CL-TR-495

Mohammad S.M. Khorsheed:

Automatic recognition of words in Arabic manuscripts

July 2000, 242 pages, PDF
PhD thesis (Churchill College, June 2000)

Abstract: The need to transliterate large numbers of historic Arabic documents into machine-readable form has motivated new work on offline recognition of Arabic script. Arabic script presents two challenges: orthography is cursive and letter shape is context sensitive.

This dissertation presents two techniques to achieve high word recognition rates: the segmentation-free technique and the segmentation-based technique. The segmentation-free technique treats the word as a whole. The word image is first transformed into a normalised polar image. The two-dimensional Fourier transform is then applied to the polar image. This results in a Fourier spectrum that is invariant to dilation, translation, and rotation. The Fourier spectrum is used to form the word template, or train the word model in the template-based and the multiple hidden Markov model (HMM) recognition systems, respectively. The recognition of an input word image is based on the minimum distance measure from the word templates and the maximum likelihood probability for the word models.

The segmentation-based technique uses a single hidden Markov model, which is composed of multiple character-models. The technique implements the analytic approach in which words are segmented into smaller units, not necessarily characters. The word skeleton is decomposed into a number of links in orthographic order, it is then transferred into a sequence of discrete symbols using vector quantisation. the training of each character-model is performed using either: state assignment in the lexicon-driven configuration or the Baum-Welch method in the lexicon-free configuration. The observation sequence of the input word is given to the hidden Markov model and the Viterbi algorithm is applied to provide an ordered list of the candidate recognitions.

UCAM-CL-TR-496

Gian Luca Cattani, James J. Leifer, Robin Milner:

Contexts and embeddings for closed shallow action graphs

July 2000, 56 pages, PostScript

Abstract: Action calculi, which have a graphical presentation, were introduced to develop a theory shared among different calculi for interactive systems. The π-calculus, the λ-calculus, Petri nets, the Ambient calculus and others may all be represented as action calculi. This paper develops a part of the shared theory.

A recent paper by two of the authors was concerned with the notion of reactive system, essentially a category of process contexts whose behaviour is presented as a reduction relation. It was shown that one can, for any reactive system, uniformly derive a labelled transition system whose associated behavioural equivalence relations (e.g. trace equivalence or bisimilarity) will be congruential, under the condition that certain relative pushouts exist in the reactive system. In the present paper we treat closed, shallow action calculi (those with no free names and no nested actions) as a generic application of these results. We define a category of action graphs and embeddings, closely linked to a category of contexts which forms a reactive system. This connection is of independent interest; it also serves our present purpose, as it enables us to demonstrate that appropriate relative pushouts exist.

Complemented by work to be reported elsewhere, this demonstration yields labelled transition systems with behavioural congruences for a substantial class of action calculi. We regard this work as a step towards comparable results for the full class.

UCAM-CL-TR-497

G.M. Bierman, A. Trigoni:

Towards a formal type system for ODMG OQL

September 2000, 20 pages, PDF

Abstract: In this paper we consider in detail the type system of the object-oriented database query language, OQL, as defined by the ODMG. Our main technical contribution is a formal definition of the typing relation for OQL—surprisingly we could not find a complete definition in the literature. We have also uncovered a number of inaccuracies in the ODMG proposal, and other work.

UCAM-CL-TR-498

Peter Sewell:

Applied π – a brief tutorial

July 2000, 65 pages, PDF, PostScript

Abstract: This note provides a brief introduction to π-calculi and their application to concurrent and distributed programming. Chapter 1 introduces a simple π-calculus and discusses the choice of primitives, operational semantics (in terms of reductions and of indexed early labelled transitions), operational equivalences, Pict-style programming and typing. Chapter 2 goes on to discuss the application of these ideas to distributed systems, looking informally at the design of distributed π-calculi with grouping and interaction primitives. Chapter 3 returns to typing, giving precise definitions for a simple type system and soundness results for the labelled transition semantics. Finally, Chapters 4 and 5 provide a model development of the metatheory, giving first an outline and then detailed proofs of the results stated earlier. The note can be read in the partial order 1.(2+3+4.5).

UCAM-CL-TR-499

James Edward Gain:

Enhancing spatial deformation for virtual sculpting

August 2000, 161 pages, PDF
PhD thesis (St John's College, June 2000)

Abstract: The task of computer-based free-form shape design is fraught with practical and conceptual difficulties. Incorporating elements of traditional clay sculpting has long been recognised as a means of shielding a user from the complexities inherent in this form of modelling. The premise is to deform a mathematically-defined solid in a fashion that loosely simulates the physical moulding of an inelastic substance, such as modelling clay or silicone putty. Virtual sculpting combines this emulation of clay sculpting with interactive feedback.

Spatial deformations are a class of powerful modelling techniques well suited to virtual sculpting. They indirectly reshape an object by warping the surrounding space. This is analogous to embedding a flexible shape within a lump of jelly and then causing distortions by flexing the jelly. The user controls spatial deformations by manipulating points, curves or a volumetric hyperpatch. Directly Manipulated Free-Form Deformation (DMFFD), in particular, merges the hyperpatch- and point-based approaches and allows the user to pick and drag object points directly.

This thesis embodies four enhancements to the versatility and validity of spatial deformation:

1. We enable users to specify deformations by manipulating the normal vector and tangent plane at a point. A first derivative frame can be tilted, twisted and scaled to cause a corresponding distortion in both the ambient space and inset object. This enhanced control is accomplished by extending previous work on bivariate surfaces to trivariate hyperpatches.

2. We extend DMFFD to enable curve manipulation by exploiting functional composition and degree reduction. Although the resulting curve-composed DMFFD introduces some modest and bounded approximation, it is superior to previous curve-based schemes in other respects. Our technique combines all three forms of spatial deformation (hyperpatch, point and curve), can maintain any desired degree of derivative continuity, is amenable to the automatic detection and prevention of self-intersection, and achieves interactive update rates over the entire deformation cycle.

3. The approximation quality of a polygon-mesh object frequently degrades under spatial deformation to become either oversaturated or undersaturated with polygons. We have devised an efficient adaptive mesh refinement and decimation scheme. Our novel contributions include: incorporating fully symmetrical decimation, reducing the computation cost of the refinement/decimation trigger, catering for boundary and crease edges, and dealing with sampling problems.

4. The potential self-intersection of an object is a serious weakness in spatial deformation. We have developed a variant of DMFFD which guards against self-intersection by subdividing manipulations into injective (one-to-one) mappings. This depends on three novel contributions: analytic conditions for identifying self-intersection, and two injectivity tests (one exact but computationally costly and the other approximate but efficient).

UCAM-CL-TR-500

Jianxin Yan, Alan Blackwell, Ross Anderson, Alasdair Grant:

The memorability and security of passwords – some empirical results

September 2000, 13 pages, PDF

Abstract: There are many things that are 'well known' about passwords, such as that uers can't remember strong passwords and that the passwords they can remember are easy to guess. However, there seems to be a distinct lack of research on the subject that would pass muster by the standards of applied psychology.

Here we report a controlled trial in which, of four sample groups of about 100 first-year students, three were recruited to a formal experiment and of these two were given specific advice about password selection. The incidence of weak passwords was determined by cracking the password file, and the number of password resets was measured from system logs. We observed a number of phenomena which run counter to

the established wisdom. For example, passwords based on mnemonic phrases are just as hard to crack as random passwords yet just as easy to remember as naive user selections.

UCAM-CL-TR-501

David Ingram:

Integrated quality of service management

September 2000, 90 pages, paper copy
PhD thesis (Jesus College, August 2000)

UCAM-CL-TR-502

Thomas Marthedal Rasmussen:

Formalizing basic number theory

September 2000, 20 pages, PDF

Abstract: This document describes a formalization of basic number theory including two theorems of Fermat and Wilson.

Most of this has (in some context) been formalized before but we present a new generalized approach for handling some central parts, based on concepts which seem closer to the original mathematical intuition and likely to be useful in other (similar) developments.

Our formulation has been mechanized in the Isabelle/HOL system.

UCAM-CL-TR-503

Alan Mycroft, Richard Sharp:

Hardware/software co-design using functional languages

September 2000, 8 pages, PDF

Abstract: In previous work we have developed and prototyped a silicon compiler which translates a functional language (SAFL) into hardware. Here we present a SAFL-level program transformation which: (i) partitions a specification into hardware and software parts and (ii) generates a specialised architecture to execute the software part. The architecture consists of a number of interconnected heterogeneous processors. Our method allows a large design space to be explored by systematically transforming a single SAFL specification to investigate different points on the area-time spectrum.

UCAM-CL-TR-504

Oi Yee Kwong:

Word sense selection in texts: an integrated model

September 2000, 177 pages, PostScript
PhD thesis (Downing College, May 2000)

Abstract: Early systems for word sense disambiguation (WSD) often depended on individual tailor-made lexical resources, hand-coded with as much lexical information as needed, but of severely limited vocabulary size. Recent studies tend to extract lexical information from a variety of existing resources (e.g. machine-readable dictionaries, corpora) for broad coverage. However, this raises the issue of how to combine the information from different resources.

Thus while different types of resource could make different contribution to WSD, studies to date have not shown what contribution they make, how they should be combined, and whether they are equally relevant to all words to be disambiguated. This thesis proposes an Integrated Model as a framework to study the interrelatedness of three major parameters in WSD: Lexical Resource, Contextual Information, and Nature of Target Words. We argue that it is their interaction which shapes the effectiveness of any WSD system.

A generalised, structurally-based sense-mapping algorithm was designed to combine various types of lexical resource. This enables information from these resources to be used simultaneously and compatibly, while respecting their distinctive structures. In studying the effect of context on WSD, different semantic relations available from the combined resources were used, and a recursive filtering algorithm was designed to overcome combinatorial explosion. We then investigated, from two directions, how the target words themselves could affect the usefulness of different types of knowledge. In particular, we modelled WSD with the cloze test format, i.e. as texts with blanks and all senses for one specific word as alternative choices for filling the blank.

A full-scale combination of WordNet and Roget's Thesaurus was done, linking more than 30,000 senses. Using these two resources in combination, a range of disambiguation tests was done on more than 60,000 noun instances from corpus texts of different types, and 60 blanks from real cloze texts. Results show that combining resources is useful for enriching lexical information, and hence making WSD more effective though not completely. Also, different target words make different demand on contextual information, and this interaction is closely related to text types. Future work is suggested for expanding the analysis on target nature and making the combination of disambiguation evidence sensitive to the requirements of the word being disambiguated.

UCAM-CL-TR-505

Gian Luca Cattani, Peter Sewell:

Models for name-passing processes: interleaving and causal

September 2000, 42 pages, PDF, PostScript, DVI

Abstract: We study syntax-free models for name-passing processes. For interleaving semantics, we identify the indexing structure required of an early labelled transition system to support the usual π-calculus operations, defining Indexed Labelled Transition Systems. For noninterleaving causal semantics we define Indexed Labelled Asynchronous Transition Systems, smoothly generalizing both our interleaving model and the standard Asynchronous Transition Systems model for CCS-like calculi. In each case we relate a denotational semantics to an operational view, for bisimulation and causal bisimulation respectively. We establish completeness properties of, and adjunctions between, categories of the two models. Alternative indexing structures and possible applications are also discussed. These are first steps towards a uniform understanding of the semantics and operations of name-passing calculi.

UCAM-CL-TR-506

Peter Sewell:

Modules, abstract types, and distributed versioning

September 2000, 46 pages, PDF, PostScript, DVI

Abstract: In a wide-area distributed system it is often impractical to synchronise software updates, so one must deal with many coexisting versions. We study static typing support for modular wide-area programming, modelling separate compilation/linking and execution of programs that interact along typed channels. Interaction may involve communication of values of abstract types; we provide the developer with fine-grain versioning control of these types to support interoperation of old and new code. The system makes use of a second-class module system with singleton kinds; we give a novel operational semantics for separate compilation/linking and execution and prove soundness.

UCAM-CL-TR-507

Lawrence Paulson:

Mechanizing a theory of program composition for UNITY

November 2000, 28 pages, PDF

Abstract: Compositional reasoning must be better understood if non-trivial concurrent programs are to be verified. Chandy and Sanders [2000] have proposed a new approach to reasoning about composition, which Charpentier and Chandy [1999] have illustrated by developing a large example in the UNITY formalism. The present paper describes extensive experiments on mechanizing the compositionality theory and the example, using the proof tool Isabelle. Broader issues are discussed, in particular, the formalization of program states. The usual representation based upon maps from variables to values is contrasted with the alternatives, such as a signature of typed variables. Properties need to be transferred from one program component's signature to the common signature of the system. Safety properties can be so transferred, but progress properties cannot be. Using polymorphism, this problem can be circumvented by making signatures sufficiently flexible. Finally the proof of the example itself is outlined.

UCAM-CL-TR-508

James Leifer, Robin Milner:

Shallow linear action graphs and their embeddings

October 2000, 16 pages, PostScript

Abstract: In previous work, action calculus has been presented in terms of action graphs. Many calculi, or at least their salient features, can be expressed as specific action calculi; examples are Petri nets, λ-calculus, π-calculus, fusion calculus, ambient calculus and spi calculus.

We here offer linear action graphs as a primitive basis for action calculi. Linear action graphs have a simpler theory than the non-linear variety. This paper presents the category of embeddings of shallow linear action graphs (those without nesting), using a novel form of graphical reasoning which simplifies some otherwise complex manipulations in regular algebra. The work is done for undirected graphs, and adapted in a few lines to directed graphs.

The graphical reasoning used here will be applied in future work to develop behavioural congruences for action calculi.

UCAM-CL-TR-509

Wojciech Basalaj:

Proximity visualisation of abstract data

January 2001, 117 pages, PDF
PhD thesis (October 2000)

Abstract: Data visualisation is an established technique for exploration, analysis and presentation of data. A graphical presentation is generated from the data content, and viewed by an observer, engaging vision – the human sense with the greatest bandwidth, and the ability to recognise patterns subconciously. For instance, a correlation present between two variables can be elucidated with a scatter plot. An effective visualisation can be difficult to achieve for an abstract collection of objects, e.g. a database table with many attributes, or a set of multimedia documents, since there is no immediately obvious way of arranging the objects based on their content. Thankfully, similarity between pairs of elements of such a collection can be measured, and a good overview picture should respect this proximity information, by positioning similar elements close to one another, and far from dissimilar objects. The resulting proximity visualisation is a topology preserving map of the underlying data collection, and this work investigates various methods for generating such maps. A number of algorithms are devised, evaluated quantitatively by means of statistical inference, and qualitatively in a case study for each type of data collection. Other graphical representations for abstract data are surveyed and compared to proximity visualisation.

A standard method for modelling prximity relations is multidimensional scaling (MDS) analysis. The result is usually a two- or three-dimensional configuration of points – each representing a single element from a collection., with inter-point distances approximating the corresponding proximities. The quality of this approximation can be expressed as a loss function, and the optimal arrangement can be found by minimising it numerically – a procedure known as least-squares metric MDS. This work presents a number of algorithmic instances of this problem, using established function optimisation heuristics: Newton-Raphson, Tabu Search, Genetic Algorithm, Iterative Majorization, and Stimulated annealing. Their effectiveness at minimising the loss function is measured for a representative sample of data collections, and the relative ranking established. The popular classical scaling method serves as a benchmark for this study.

The computational cost of conventional MDS makes it unsuitable for visualising a large data collection. Incremental multidimensional scaling solves this problem by considering only a carefully chosen subset of all pairwise proximities. Elements that make up cluster diameters at a certain level of the single link cluster hierarchy are identified, and are subject to standard MDS, in order to establish the overall shape of the configuration. The remaining elements are positioned independently of one another with respect to this skeleton configuration. For very large collections the skeleton configuration can itself be built up incrementally. The incremental method is analysed for the compromise between solution quality and the proportion of proximities used, and compared to Principal Components Analysis on a number of large database tables.

In some applications it is convenient to represent individual objects by compact icons of fixed size, for example the use of thumbnails when visualising a set of images. Because the MDS analysis only takes the position of icons into account, and not their size, its direct use for visualisation may lead to partial or complete overlap of icons. Proximity grid – an analogue of MDS in a discrete domain – is proposed to overcome this deficiency. Each element of an abstract data collection is represented within a single cell of the grid, and thus considerable detail can be shown without overlap. The proximity relationships are preserved by clustering similar elements in the grid, and keeping dissimilar ones apart. Algorithms for generating such an arrangement are presented and compared in terms of output quality to one another as well as standard MDS.

UCAM-CL-TR-510

Richard Mortier, Rebecca Isaacs, Keir Fraser:

Switchlets and resource-assured MPLS networks

May 2000, 16 pages, PDF, PostScript

Abstract: MPLS (Multi-Protocol Label Switching) is a technology with the potential to support multiple control systems, each with guaranteed QoS (Quality of Service), on connectionless best-effort networks. However, it does not provide all the capabilities required of a multi-service network. In particular, although resource-assured VPNs (Virtual Private Networks) can be created, there is no provision for inter-VPN resource management. Control flexibility is limited because resources must be pinned down to be guaranteed, and best-effort flows in different VPNs compete for the same resources, leading to QoS crosstalk.

The contribution of this paper is an implementation on MPLS of a network control framework that supports inter-VPN resource management. Using resource partitions known as switchlets, it allows the creation of multiple VPNs with guaranteed resource allocations, and maintains isolation between these VPNs. Devolved control techniques permit each VPN a customised control system.

We motivate our work by discussing related efforts and example scenarios of effective deployment of our system. The implementation is described and evaluated, and we address interoperability with external IP control systems, in addition to interoperability of data across different layer 2 technologies.

UCAM-CL-TR-511

Calum Grant:

Software visualization in Prolog

December 1999, 193 pages, PDF
PhD thesis (Queens' College, 1999)

Abstract: Software visualization (SV) uses computer graphics to communicate the structure and behaviour of complex software and algorithms. One of the important issues in this field is how to specify SV, because existing systems are very cumbersome to specify and implement, which limits their effectiveness and hinders SV from being integrated into professional software development tools.

In this dissertation the visualization process is decomposed into a series of formal mappings, which provides a formal foundation, and allows separate aspects of visualization to be specified independently. The first mapping specifies the information content of each view. The second mapping specifies a graphical representation of the information, and a third mapping specifies the graphical components that make up the graphical representation. By combining different mappings, completely different views can be generated.

The approach has been implemented in Prolog to provide a very high level specification language for information visualization, and a knowledge engineering environment that allows data queries to tailor the information in a view. The output is generated by a graphical constraint solver that assembles the graphical components into a scene.

This system provides a framework for SV called Vmax. Source code and run-time data are analyzed by Prolog to provide access to information about the program structure and run-time data for a wide range of highly interconnected browsable views. Different views and means of visualization can be selected from menus. An automatic legend describes each view, and can be interactively modified to customize how data is presented. A text window for editing source code is synchronized with the graphical view. Vmax is a complete Java development environment and end user SV system.

Vmax compares favourably to existing SV systems in many taxonometric criteria, including automation, scope, information content, graphical output form, specification, tailorability, navigation, granularity and elision control. The performance and scalability of the new approach is very reasonable.

We conclude that Prolog provides a formal and high level specification language that is suitable for specifying all aspects of a SV system.

UCAM-CL-TR-512

Anthony Fox:

An algebraic framework for modelling and verifying microprocessors using HOL

March 2001, 24 pages, PDF

Abstract: This report describes an algebraic approach to the specification and verification of microprocessor designs. Key results are expressed and verified using the HOL proof tool. Particular attention is paid to the models of time and temporal abstraction, culminating in a number of one-step theorems. This work is then explained with a small but complete case study, which verifies the correctness of a datapath with microprogram control.

UCAM-CL-TR-513

Tetsuya Sakai, Karen Spärck Jones:

Generic summaries for indexing in information retrieval – Detailed test results

May 2001, 29 pages, PostScript

Abstract: This paper examines the use of generic summaries for indexing in information retrieval. Our main observations are that:

– With or without pseudo-relevance feedback, a summary index may be as effective as the corresponding fulltext index for precision-oriented search of highly relevant documents. But a reasonably sophisticated summarizer, using a compression ratio of 10–30%, is desirable for this purpose.

– In pseudo-relevance feedback, using a summary index at initial search and a fulltext index at final search is possibly effective for precision-oriented search, regardless of relevance levels. This strategy is significantly more effective than the one using the summary index only and probably more effective than using summaries as mere term selection filters. For this strategy, the summary quality is probably not a critical factor, and a compression ratio of 5–10% appears best.

UCAM-CL-TR-514

Asis Unyapoth:

Nomadic π-calculi: expressing and verifying communication infrastructure for mobile computation

June 2001, 316 pages, PDF, PostScript
PhD thesis (Pembroke College, March 2001)

Abstract: This thesis addresses the problem of verifying distributed infrastructure for mobile computation. In particular, we study language primitives for communication between mobile agents. They can be classified into two groups. At a low level there are "location dependent" primitives that require a programmer to know the current site of a mobile agent in order to communicate with it. At a high level there are "location independent" primitives that allow communication with a mobile agent irrespective of any migrations. Implementation of the high level requires delicate distributed infrastructure algorithms. In earlier work of Sewell, Wojciechowski and Pierce, the two levels were

made precise as process calculi, allowing such algorithms to be expressed as encodings of the high level into the low level; a distributed programming language "Nomadic Pict" has been built for experimenting with such encodings.

This thesis turns to semantics, giving a definition of the core language (with a type system) and proving correctness of an example infrastructure. This involves extending the standard semantics and proof techniques of process calculi to deal with the new notions of sites and agents. The techniques adopted include labelled transition semantics, operational equivalences and preorders (e.g., expansion and coupled simulation), "up to" equivalences, and uniform receptiveness. We also develop two novel proof techniques for capturing the design intuitions regarding mobile agents: we consider "translocating" versions of operational equivalences that take migration into account, allowing compositional reasoning; and "temporary immobility", which captures the intuition that while an agent is waiting for a lock somewhere in the system, it will not migrate.

The correctness proof of an example infrastructure is non-trivial. It involves analysing the possible reachable states of the encoding applied to an arbitrary high-level source program. We introduce an intermediate language for factoring out as many 'house-keeping' reduction steps as possible, and focusing on the partially-committed steps.

UCAM-CL-TR-515

Andrei Serjantov, Peter Sewell, Keith Wansbrough:

The UDP calculus: rigorous semantics for real networking

July 2001, 70 pages, PostScript

Abstract: Network programming is notoriously hard to understand: one has to deal with a variety of protocols (IP, ICMP, UDP, TCP, etc.), concurrency, packet loss, host failure, timeouts, the complex sockets interface to the protocols, and subtle protability issues. Moreover, the behavioural properties of operating systems and the network are not well documented.

A few of these issues have been addressed in the process calculus and distributed algorithm communities, but there remains a wide gulf between what has been captured in semantic models and what is required for a precise understanding of the behaviour of practical distributed programs that use these protocols.

In this paper we demonstrate (in a preliminary way) that the gulf can be bridged. We give an operational model for socket programming with a substantial fraction of UDP and ICMP, including loss and failure. The model has been validated by experiment against actual systems. It is not tied to a particular programming language, but can be used with any language equipped with an operational semantics for system calls – here we give such a language binding for an OCaml fragment. We illustrate the model with a few small network programs.

UCAM-CL-TR-516

Rebecca Isaacs:

Dynamic provisioning of resource-assured and programmable virtual private networks

September 2001, 145 pages, PostScript
PhD thesis (Darwin College, December 2000)

Abstract: Virtual Private Networks (VPNs) provide dedicated connectivity to a closed group of users on a shared network. VPNs have traditionally been deployed for reasons of economy of scale, but have either been statically defined, requiring manual configuration, or else unable to offer any quality of service (QoS) guarantees.

This dissertation describes VServ, a service offering dynamic and resource-assured VPNs that can be acquired and modified on demand. In VServ, a VPN is both a subset of physical resources, such as bandwidth and label space, together with the means to perform fine-grained management of those resources. This network programmability, combined with QoS guarantees, enables the multiservice network – a single universal network that can support all types of service and thus be efficient, cost-effective and flexible.

VServ is deployed over a network control framework known as Tempest. The Tempest explicitly distinguishes between inter- and intra-VPN resource management mechanisms. This makes the dynamic resource reallocation capabilities of VServ viable, whilst handling highly dynamic VPNs or a large number of VPNs. Extensions to the original implementation of the Tempest to support dynamically reconfigurable QoS are detailed.

A key part of a dynamic and responsive VPN service is fully automated VPN provisioning. A notation for VPN specification is described, together with mechanisms for incorporating policies of the service provider and the current resource availability in the network into the design process. The search for a suitable VPN topology can be expressed as a optimisation problem that is not computationally tractable except for very small networks. This dissertation describes how the search is made practical by tailoring it according to the characteristics of the desired VPN.

Availability of VServ is addressed with a proposal for distributed VPN creation. A resource revocation protocol exploits the dynamic resource management capabilities of VServ to allow adaptation in the control plane on a per-VPN basis. Managed resource revocation supports highly flexible resource allocation and

reallocation policies, allowing VServ to efficiently provision for short-lived or highly dynamic VPNs.

UCAM-CL-TR-517

Karen Spärck Jones, P. Jourlin, S.E. Johnson, P.C. Woodland:

The Cambridge Multimedia Document Retrieval Project: summary of experiments

July 2001, 30 pages, PostScript, DVI

Abstract: This report summarises the experimental work done under the Multimedia Document Retrieval (MDR) project at Cambridge from 1997-2000, with selected illustrations. The focus is primarily on retrieval studies, and on speech tests directly related to retrieval, not on speech recognition itself. The report draws on the many and varied tests done during the project, but also presents a new series of results designed to compare strategies across as many different data sets as possible by using consistent system parameter settings.

The project tests demonstrate that retrieval from files of audio news material transcribed using a state of the art speech recognition system can match the reference level defined by human transcriptions; and that expansion techniques, especially when applied to queries, can be very effective means for improving basic search performance.

UCAM-CL-TR-518

Jeff Jianxin Yan, Yongdong Wu:

An attack on a traitor tracing scheme

July 2001, 14 pages, PDF

Abstract: In Crypto'99, Boneh and Franklin proposed a public key traitor tracing scheme, which was believed to be able to catch all traitors while not accusing any innocent users (i.e., full-tracing and error-free). Assuming that Decision Diffie-Hellman problem is unsolvable in Gq, Boneh and Franklin proved that a decoder cannot distinguish valid ciphertexts from invalid ones that are used for tracing. However, our novel pirate decoder P3 manages to make some invalid ciphertexts distinguishable without violating their assumption, and it can also frame innocent user coalitions to fool the tracer. Neither the single-key nor arbitrary pirate tracing algorithm presented in [1] can identify all keys used by P3 as claimed. Instead, it is possible for both algorithms to catch none of the traitors. We believe that the construction of our novel pirate also demonstrates a simple way to defeat some other black-box traitor tracing schemes in general.

UCAM-CL-TR-519

Martin Choquette:

Local evidence in document retrieval

August 2001, 177 pages, paper copy
PhD thesis (Fitzwilliam College, November 2002)

UCAM-CL-TR-520

Mohamed Hassan, Neil A. Dodgson:

Ternary and three-point univariate subdivision schemes

September 2001, 8 pages, PDF

Abstract: The generating function formalism is used to analyze the continuity properties of univariate ternary subdivision schemes. These are compared with their binary counterparts.

UCAM-CL-TR-521

James Leifer:

Operational congruences for reactive systems

September 2001, 144 pages, PostScript
PhD thesis (Trinity College, March 2001)

Abstract: The dynamics of process calculi, eg. CCS, have often been defined using a labelled transaction system (LTS). More recently it has become common when defining dynamics to use reaction rules –ie. unlabelled transition rules– together with a structural congruence. This form, which I call a reactive system, is highly expressive but is limited in an important way: LTSs lead more naturally to operational equivalences and preorders.

So one would like to derive from reaction rules a suitable LTS. This dissertation shows how to derive an LTS for a wide range of reactive systems. A label for an agent (process), a, is defined to be any context, F, which intuitively is just large enough so that the agent Fa ("a in context F") is able to perform a reaction. The key contribution of my work is the precise definition of "just large enough", in terms of the categorical notation of relative pushout (RPO), which ensures that several operational equivalences and preorders (strong bisimulation, weak bisimulation, the traces preorder, and the failures preorder) are congruences when sufficient RPOs exist.

I present a substantial example of a family of reactive systems based on closed, shallow action calculi (those with no free names and no nesting). I prove that RPOs exist for a category of such contexts. The proof is carried out indirectly in terms of a category of action

graphs and embeddings and gives precise (necessary and sufficient) conditions for the existance of RPOs. I conclude by arguing that these conditions are satisfied for a wide class of reaction rules. The thrust of this dissertation is, therefore, towards easing the burden of exploring new models of computation by providing a general method for achieving useful operational congruences.

UCAM-CL-TR-522

Mark F.P. Gillies:

Practical behavioural animation based on vision and attention

September 2001, 187 pages, PDF, PostScript

Abstract: The animation of human like characters is a vital aspect of computer animation. Most animations rely heavily on characters of some sort or other. This means that one important aspect of computer animation research is to improve the animation of these characters both by making it easier to produce animations and by improving the quality of animation produced. One approach to animating characters is to produce a simulation of the behaviour of the characters which will automatically animate the character.

The dissertation investigates the simulation of behaviour in practical applications. In particular it focuses on models of visual perception for use in simulating human behaviour. A simulation of perception is vital for any character that interacts with its surroundings. Two main aspects of the simulation of perception are investigated:

– The use of psychology for designing visual algorithms.

– The simulation of attention in order to produce both behaviour and gaze patterns.

Psychological theories are a useful starting point for designing algorithms for simulating visual perception. The dissertation investigates their use and presents some algorithms based on psychological theories.

Attention is the focusing of a person's perception on a particular object. The dissertation presents a simulation of what a character is attending to (looking at). This is used to simulate behaviour and for animating eye movements.

The algorithms for the simulation of vision and attention are applied to two tasks in the simulation of behaviour. The first is a method for designing generic behaviour patterns from simple pieces of motion. The second is a behaviour pattern for navigating a cluttered environment. The simulation of vision and attention gives advantages over existing work on both problems. The approaches to the simulation of perception will be evaluated in the context of these examples.

UCAM-CL-TR-523

Robin Milner:

Bigraphical reactive systems: basic theory

September 2001, 87 pages, PDF

Abstract: A notion of bigraph is proposed as the basis for a model of mobile interaction. A bigraph consists of two independent structures: a topograph representing locality and a monograph representing connectivity. Bigraphs are equipped with reaction rules to form bigraphical reactive systems (BRSs), which include versions of the π-calculus and the ambient calculus. Bigraphs are shown to be a special case of a more abstract notion, wide reactive systems (WRSs), not assuming any particular graphical or other structure but equipped with a notion of width, which expresses that agents, contexts and reactions may all be widely distributed entities.

A behavioural theory is established for WRSs using the categorical notion of relative pushout; it allows labelled transition systems to be derived uniformly, in such a way that familiar behavioural preorders and equivalences, in particular bisimilarity, are congruential under certain conditions. Then the theory of bigraphs is developed, and they are shown to meet these conditions. It is shown that, using certain functors, other WRSs which meet the conditions may also be derived; these may, for example, be forms of BRS with additional structure.

Simple examples of bigraphical systems are discussed; the theory is developed in a number of ways in preparation for deeper application studies.

UCAM-CL-TR-524

Giampaolo Bella, Fabio Massacci, Lawrence C. Paulson:

Verifying the SET purchase protocols

November 2001, 14 pages, PDF

Abstract: The Secure Electronic Transaction (SET) protocol has been proposed by a consortium of credit card companies and software corporations to guarantee the authenticity of e-commerce transactions and the confidentiality of data. When the customer makes a purchase, the SET dual signature keeps his account details secret from the merchant and his choice of goods secret from the bank. This paper reports verification results for the purchase step of SET, using the inductive method. The credit card details do remain confidential. The customer, merchant and bank can confirm most details of a transaction even when some of those details are kept from them. The usage of dual signatures requires repetition in protocol messages, making proofs

more difficult but still feasible. The formal analysis has revealed a significant defect. The dual signature lacks explicitness, giving rise to potential vulnerabilities.

UCAM-CL-TR-525

Timothy L. Harris:

Extensible virtual machines

December 2001, 209 pages, PDF
PhD thesis

Abstract: Virtual machines (VMs) have enjoyed a resurgence as a way of allowing the same application program to be used across a range of computer systems. This flexibility comes from the abstraction that the provides over the native interface of a particular computer. However, this also means that the application is prevented from taking the features of particular physical machines into account in its implementation.

This dissertation addresses the question of why, where and how it is useful, possible and practicable to provide an application with access to lower-level interfaces. It argues that many aspects of implementation can be devolved safely to untrusted applications and demonstrates this through a prototype which allows control over run-time compilation, object placement within the heap and thread scheduling. The proposed architecture separates these application-specific policy implementations from the application itself. This allows one application to be used with different policies on different systems and also allows naïve or premature optimizations to be removed.

UCAM-CL-TR-526

Andrew J. Penrose:

Extending lossless image compression

December 2001, 137 pages, PDF
PhD thesis (Gonville & Caius College, January 2001)

Abstract: "It is my thesis that worthwhile improvements can be made to lossless image compression schemes, by considering the correlations between the spectral, temporal and interview aspects of image data, in extension to the spatial correlations that are traditionally exploited."

Images are an important part of today's digital world. However, due to the large quantity of data needed to represent modern imagery the storage of such data can be expensive. Thus, work on efficient image storage (image compression) has the potential to reduce storage costs and enable new applications.

Many image compression schemes are lossy; that is they sacrifice image informationto achieve very compact storage. Although this is acceptable for many applications, some environments require that compression not alter the image data. This lossless image compression has uses in medical, scientific and professional video processing applications.

Most of the work on lossless image compression has focused on monochrome images and has made use of the spatial smoothness of image data. Only recently have researchers begun to look specifically at the lossless compression of colour images and video. By extending compression schemes for colour images and video, the storage requirements for these important classes of image data can be further reduced.

Much of the previous research into lossless colour image and video compression has been exploratory. This dissertation studies the problem in a structured way. Spatial, spectral and temporal correlations are all considered to facilitate improved compression. This has lead to a greater data reduction than many existing schemes for lossless colour image and colour video compression.

Furthermore, this work has considered the application of extended lossless image coding to more recent image types, such as multiview imagery. Thus, systems that use multiple views of the same scene to provide 3D viewing, have beenprovided with a completely novel solution for the compression of multiview colour video.

UCAM-CL-TR-527

Umar Saif:

Architectures for ubiquitous systems

January 2002, 271 pages, PDF
PhD thesis

Abstract: Advances in digital electronics over the last decade have made computers faster, cheaper and smaller. This coupled with the revolution in communication technology has led to the development of sophisticated networked appliances and handheld devices. "Computers" are no longer boxes sitting on a desk, they are all around us, embedded in every nook and corner of our environment. This increasing complexity in our environment leads to the desire to design a system that could allow this pervasive functionality to disappear in the infrastructure, automatically carrying out everyday tasks of the users.

Such a system would enable devices embedded in the environment to cooperate with one another to make a wide range of new and useful applications possible, not originally conceived by the manufacturer, to achieve greater functionality, flexibility and utility.

The compelling question then becomes "what software needs to be embedded in these devices to enable them to participate in such a ubiquitous system"? This is the question addressed by the dissertation.

Based on the experience with home automation systems, as part of the AutoHAN project, the dissertation presents two compatible but different architectures; one to enable dumb devices to be controlled by the system and the other to enable intelligent devices to control, extend and program the system.

Control commands for dumb devices are managed using an HTTP-based publish/subscribe/notify architecture; devices publish their control commands to the system as XML-typed discrete messages, applications discover and subscribe interest in these events to send and receive control commands from these devices, as typed messages, to control their behavior. The architecture handles mobility and failure of devices by using soft-state, redundent subscriptions and "care-of" nodes. The system is programmed with event scripts that encode automation rules as condition-action bindings. Finally, the use of XML and HTTP allows devices to be controlled by a simple Internet browser.

While the publish/subscribe/notify defines a simple architecture to enable interoperability of limited capability devices, intelligent devices can afford more complexity that can be utilized to support user applications and services to control, manage and program the system. However, the operating system embedded in these devices needs to address the heterogeneity, longevity, mobility and dynamism of the system.

The dissertation presents the architecture of an embedded distributed operating system that lends itself to safe context-driven adaptation. The operating system is instrumented with four artifacts to address the challenges posed by a ubiquitous system. 1) An XML-based directory service captures and notifies the applications and services about changes in the device context, as resources move, fail, leave or join the system, to allow context-driven adaptation. 2) A Java-based mobile agent system allows new software to be injected in the system and moved and replicated with the changing characteristics of the system to define a self-organizing system. 3) A subscribe/notify interface allows context-specific extensions to be dynamically added to the operating system to enable it to efficiently interoperate in its current context according to application requirements. 4) Finally, a Dispatcher module serves as the context-aware system call interface for the operating system; when requested to invoke a service, the Dispatcher invokes the resource that best satisfies the requirements given the characteristics of the system.

Definition alone is not sufficient to prove the validity of an architecture. The dissertation therefore describes a prototype implementation of the operating system and presents both a quantitative comparison of its performance with related systems and its qualitative merit by describing new applications made possible by its novel architecture.

UCAM-CL-TR-528

Andrew William Moore:

Measurement-based management of network resources

April 2002, 273 pages, PDF
PhD thesis

Abstract: Measurement-Based Estimators are able to characterise data flows, enabling improvements to existing management techniques and access to previously impossible management techniques. It is the thesis of this dissertation that in addition to making practical adaptive management schemes, measurement-based estimators can be practical within current limitations of resource.

Examples of network management include the characterisation of current utilisation for explicit admission control and the configuration of a scheduler to divide link-capacity among competing traffic classes. Without measurements, these management techniques have relied upon the accurate characterisation of traffic – without accurate traffic characterisation, network resources may be under or over utilised.

Embracing Measurement-Based Estimation in admission control, Measurement-Based Admission Control (MBAC) algorithms have allowed characterisation of new traffic flows while adapting to changing flow requirements. However, there have been many MBAC algorithms proposed, often with no clear differentiation between them. This has motivated the need for a realistic, implementation-based comparison in order to identify an ideal MBAC algorithm.

This dissertation reports on an implementation-based comparison of MBAC algorithms conducted using a purpose built test environment. The use of an implementation-based comparison has allowed the MBAC algorithms to be tested under realistic conditions of traffic load and realistic limitations on memory, computational resources and measurements. Alongside this comparison is a decomposition of a group of MBAC algorithms, illustrating the relationship among MBAC algorithm components, as well as highlighting common elements among different MBAC algorithms.

The MBAC algorithm comparison reveals that, while no single algorithm is ideal, the specific resource demands, such as computation overheads, can dramatically impact on the MBAC algorithm's performance. Further, due to the multiple timescales present in both traffic and management, the estimator of a robust MBAC algorithm must base its estimate on measurements made over a wide range of timescales. Finally, a reliable estimator must account for the error resulting from random properties of measurements.

Further identifying that the estimator components used in MBAC algorithms need not be tied to the admission control problem, one of the estimators (originally constructed as part of an MBAC algorithm) is used to continuously characterise resource requirements for a number of classes of traffic. Continuous characterisation of traffic, whether requiring similar or orthogonal resources, leads to the construction and demonstration of a network switch that is able to provide differentiated service while being adaptive to the demands of each traffic class. The dynamic allocation of resources is an approach unique to a measurement-based technique that would not be possible if resources were based upon static declarations of requirement.

UCAM-CL-TR-529

Neil Johnson:

The triVM intermediate language reference manual

February 2002, 83 pages, PDF

This research was sponsored by a grant from ARM Limited.

Abstract: The triVM intermediate language has been developed as part of a research programme concentrating on code space optimization. The primary aim in developing triVM is to provide a language that removes the complexity of high-level languages, such as C or ML, while maintaining sufficient detail, at as simple a level as possible, to support reseach and experimentation into code size optimization. The basic structure of triVM is a notional Static Single Assignment-based three-address machine. A secondary aim is to develop an intermediate language that supports graph-based translation, using graph rewrite rules, in a textual, human-readable format. Experience has shown that text-format intermediate files are much easier to use for experimentation, while the penalty in translating this human-readable form to the internal data structures used by the software is negligible. Another aim is to provide a flexible language in which features and innovations can be evaluated; for example, this is one of the first intermediate languages directly based on the Static Single Assignment technique, and which explicitly exposes the condition codes as a result of arithmetic operations. While this paper is concerned solely with the description of triVM, we present a brief summary of other research-orientated intermediate languages.

UCAM-CL-TR-530

Anna Korhonen:

Subcategorization acquisition

February 2002, 189 pages, PDF

PhD thesis (Trinity Hall, September 2001)

Abstract: Manual development of large subcategorised lexicons has proved difficult because predicates change behaviour between sublanguages, domains and over time. Yet access to a comprehensive subcategorization lexicon is vital for successful parsing capable of recovering predicate-argument relations, and probabilistic parsers would greatly benefit from accurate information concerning the relative likelihood of different subcategorisation frames SCFs of a given predicate. Acquisition of subcategorization lexicons from textual corpora has recently become increasingly popular. Although this work has met with some success, resulting lexicons indicate a need for greater accuracy. One significant source of error lies in the statistical filtering used for hypothesis selection, i.e. for removing noise from automatically acquired SCFs.

This thesis builds on earlier work in verbal subcategorization acquisition, taking as a starting point the problem with statistical filtering. Our investigation shows that statistical filters tend to work poorly because not only is the underlying distribution zipfian, but there is also very little correlation between conditional distribution of SCFs specific to a verb and unconditional distribution regardless of the verb. More accurate back-off estimates are needed for SCF acquisition than those provided by unconditional distribution.

We explore whether more accurate estimates could be obtained by basing them on linguistic verb classes. Experiments are reported which show that in terms of SCF distributions, individual verbs correlate more closely with syntactically similar verbs and even more closely with semantically similar verbs, than with all verbs in general. On the basis of this result, we suggest classifying verbs according to their semantic classes and obtaining back-off estimates specific to these classes.

We propose a method for obtaining such semantically based back-off estimates, and a novel approach to hypothesis selection which makes use of these estimates. This approach involves automatically identifying the semantic class of a predicate, using subcategorization acquisition machinery to hypothesise conditional SCF distribution for the predicate, smoothing the conditional distribution with the back-off estimates of the respective semantic verb class, and employing a simple method for filtering, which uses a threshold on the estimates from smoothing. Adopting Briscoe and Carroll's (1997) system as a framework, we demonstrate that this semantically-driven approach to hypothesis selection can significantly improve the accuracy of large-scale subcategorization acquisition.

UCAM-CL-TR-531

Giampaolo Bella, Fabio Massacci, Lawrence C. Paulson:

Verifying the SET registration protocols

March 2002, 24 pages, PDF

Abstract: SET (Secure Electronic Transaction) is an immense e-commerce protocol designed to improve the security of credit card purchases. In this paper we focus on the initial bootstrapping phases of SET, whose objective is the registration of customers and merchants with a SET certification authority. The aim of registration is twofold: getting the approval of the cardholder's or merchant's bank, and replacing traditional credit card numbers with electronic credentials that customers can present to the merchant, so that their privacy is protected. These registration sub-protocols present a number of challenges to current formal verification methods. First, they do not assume that each agent knows

the public keys of the other agents. Key distribution is one of the protocols' tasks. Second, SET uses complex encryption primitives (digital envelopes) which introduce dependency chains: the loss of one secret key can lead to potentially unlimited losses. Building upon our previous work, we have been able to model and formally verify SET's registration with the inductive method in Isabelle/HOL solving its challenges with very general techniques.

UCAM-CL-TR-532

Richard Mortier:

Internet traffic engineering

April 2002, 129 pages, PDF
PhD thesis (Churchill College, October 2001)

Abstract: Due to the dramatically increasing popularity of the services provided over the public Internet, problems with current mechanisms for control and management of the Internet are becoming apparent. In particular, it is increasingly clear that the Internet and other networks built on the Internet protocol suite do not provide sufficient support for the efficient control and management of traffic, i.e. for Traffic Engineering.

This dissertation addresses the problem of traffic engineering in the Internet. It argues that traffic management techniques should be applied at multiple timescales, and not just at data timescales as is currently the case. It presents and evaluates mechanisms for traffic engineering in the Internet at two further timescales: flow admission control and control of per-flow packet marking, enabling control timescale traffic engineering; and support for load based inter-domain routeing in the Internet, enabling management timescale traffic engineering.

This dissertation also discusses suitable policies for the application of the proposed mechanisms. It argues that the proposed mechanisms are able to support a wide range of policies useful to both users and operators. Finally, in a network of the size of the Internet consideration must also be given to the deployment of proposed solutions. Consequently, arguments for and against the deployment of these mechanisms are presented and the conclusion drawn that there are a number of feasible paths toward deployment.

The work presented argues the following: firstly, it is possible to implement mechanisms within the Internet framework that enable traffic engineering to be carried out by operators; secondly, that applying these mechanisms with suitable policies can ease the management problems faced by operators and at the same time improve the efficiency with which the network can be run; thirdly, that these improvements can correspond to increased network performance as viewed by the user; and finally, that not only the resulting deployment but also the deployment process itself are feasible.

UCAM-CL-TR-533

Aline Villavicencio:

The acquisition of a unification-based generalised categorial grammar

April 2002, 223 pages, PDF
PhD thesis (Hughes Hall, September 2001)

Abstract: The purpose of this work is to investigate the process of grammatical acquisition from data. In order to do that, a computational learning system is used, composed of a Universal Grammar with associated parameters, and a learning algorithm, following the Principles and Parameters Theory. The Universal Grammar is implemented as a Unification-Based Generalised Categorial Grammar, embedded in a default inheritance network of lexical types. The learning algorithm receives input from a corpus of spontaneous child-directed transcribed speech annotated with logical forms and sets the parameters based on this input. This framework is used as a basis to investigate several aspects of language acquisition. In this thesis I concentrate on the acquisition of subcategorisation frames and word order information, from data. The data to which the learner is exposed can be noisy and ambiguous, and I investigate how these factors affect the learning process. The results obtained show a robust learner converging towards the target grammar given the input data available. They also show how the amount of noise present in the input data affects the speed of convergence of the learner towards the target grammar. Future work is suggested for investigating the developmental stages of language acquisition as predicted by the learning model, with a thorough comparison with the developmental stages of a child. This is primarily a cognitive computational model of language learning that can be used to investigate and gain a better understanding of human language acquisition, and can potentially be relevant to the development of more adaptive NLP technology.

UCAM-CL-TR-534

Austin Donnelly:

Resource control in network elements

April 2002, 183 pages, PDF
PhD thesis (Pembroke College, January 2002)

Abstract: Increasingly, substantial data path processing is happening on devices within the network. At or near the edges of the network, data rates are low enough that commodity workstations may be used to process packet flows. However, the operating systems such machines use are not suited to the needs of data-driven processing. This dissertation shows why this is a problem, how current work fails to address it, and proposes a new approach.

The principal problem is that crosstalk occurs in the processing of different data flows when they contend for a shared resource and their accesses to this resource are not scheduled appropriately; typically the shared resource is located in a server process. Previous work on vertically structured operating systems reduces the need for such shared servers by making applications responsible for performing as much of their own processing as possible, protecting and multiplexing devices at the lowest level consistent with allowing untrusted user access.

However, shared servers remain on the data path in two circumstances: firstly, dumb network adaptors need non-trivial processing to allow safe access by untrusted user applications. Secondly, shared servers are needed wherever trusted code must be executed for security reasons.

This dissertation presents the design and implementation of Expert, an operating system which avoids crosstalk by removing the need for such servers.

This dissertation describes how Expert handles dumb network adaptors to enable applications to access them via a low-level interface which is cheap to implement in the kernel, and retains application responsibility for the work involved in running a network stack.

Expert further reduces the need for application-level shared servers by introducing paths which can trap into protected modules of code to perform actions which would otherwise have to be implemented within a server.

Expert allows traditional compute-bound tasks to be freely mixed with these I/O-driven paths in a single system, and schedules them in a unified manner. This allows the processing performed in a network element to be resource controlled, both for background processing tasks such as statistics gathering, and for data path processing such as encryption.

UCAM-CL-TR-535

Claudia Faggian, Martin Hyland:

Designs, disputes and strategies

May 2002, 21 pages, PDF

Abstract: Important progresses in logic are leading to interactive and dynamical models. Geometry of Interaction and Games Semantics are two major examples. Ludics, initiated by Girard, is a further step in this direction.

The objects of Ludics which correspond to proofs are designs. A design can be described as the skeleton of a sequent calculus derivation, where we do not manipulate formulas, but their location (the address where the formula is stored). To study the traces of the interactions between designs as primitive leads to an alternative presentation, which is to describe a design as the set of its possible interactions, called disputes. This presentation has the advantage to make precise the correspondence between the basic notions of Ludics (designs, disputes and chronicles) and the basic notions of Games semantics (strategies, plays and views).

UCAM-CL-TR-536

Sergei Skorobogatov:

Low temperature data remanence in static RAM

June 2002, 9 pages, PDF

Abstract: Security processors typically store secret key material in static RAM, from which power is removed if the device is tampered with. It is commonly believed that, at temperatures below −20 °C, the contents of SRAM can be 'frozen'; therefore, many devices treat temperatures below this threshold as tampering events. We have done some experiments to establish the temperature dependency of data retention time in modern SRAM devices. Our experiments show that the conventional wisdom no longer holds.

UCAM-CL-TR-537

Mantsika Matooane:

Parallel systems in symbolic and algebraic computation

June 2002, 139 pages, PDF
PhD thesis (Trinity College, August 2001)

Abstract: This report describes techniques to exploit distributed memory massively parallel supercomputers to satisfy the peak memory demands of some very large computer algebra problems (over 10 GB). The memory balancing is based on a randomized hashing algorithm for dynamic data distribution. Fine grained partitioning is used to provide flexibility in the memory allocation, at the cost of higher communication cost. The main problem areas are multivariate polynomial algebra, and linear algebra with polynomial matrices. The system was implemented and tested on a Hitachi SR2201 supercomputer.

UCAM-CL-TR-538

Mark Ashdown, Peter Robinson:

The Escritoire: A personal projected display for interacting with documents

June 2002, 12 pages, PDF

Abstract: The Escritoire is a horizontal desk interface that uses two projectors to create a foveal display. Items such as images, documents, and the interactive displays of other conventional computers, can be manipulated on the desk using pens in both hands. The periphery covers the desk, providing ample space for laying out the objects relevant to a task, allowing them to be identified at a glance and exploiting human spatial memory for rapid retrieval. The fovea is a high resolution focal area that can be used to view any item in detail. The projected images are continuously warped with commodity graphics hardware before display, to reverse the effects of misaligned projectors and ensure registration between fovea and periphery. The software is divided into a hardware-specific client driving the display, and a platform-independent server imposing control.

UCAM-CL-TR-539

N.A. Dodgson, M.A. Sabin, L. Barthe, M.F. Hassan:

Towards a ternary interpolating subdivision scheme for the triangular mesh

July 2002, 12 pages, PDF

Abstract: We derive a ternary interpolating subdivision scheme which works on the regular triangular mesh. It has quadratic precision and fulfils the standard necessary conditions for C2 continuity. Further analysis is required to determine its actual continuity class and to define its behaviour around extraordinary points.

UCAM-CL-TR-540

N.A. Dodgson, J.R. Moore:

The use of computer graphics rendering software in the analysis of a novel autostereoscopic display design

August 2002, 6 pages, PDF

Abstract: Computer graphics 'ray tracing' software has been used in the design and evaluation of a new autostereoscopic 3D display. This software complements the conventional optical design software and provides a cost-effective method of simulating what is actually seen by a viewer of the display. It may prove a useful tool in similar design problems.

UCAM-CL-TR-541

L. Barthe, N.A. Dodgson, M.A. Sabin, B. Wyvill, V. Gaildrat:

Different applications of two-dimensional potential fields for volume modeling

August 2002, 26 pages, PDF

Abstract: Current methods for building models using implicit volume techniques present problems defining accurate and controllable blend shapes between implicit primitives. We present new methods to extend the freedom and controllability of implicit volume modeling. The main idea is to use a free-form curve to define the profile of the blend region between implicit primitives.

The use of a free-form implicit curve, controlled point-by-point in the Euclidean user space, allows us to group boolean composition operators with sharp transitions or smooth free-form transitions in a single modeling metaphor. This idea is generalized for the creation, sculpting and manipulation of volume objects, while providing the user with simplicity, controllability and freedom in volume modeling.

Bounded volume objects, known as "Soft objects" or "Metaballs", have specific properties. We also present binary Boolean composition operators that gives more control on the form of the transition when these objects are blended.

To finish, we show how our free-form implicit curves can be used to build implicit sweep objects.

UCAM-CL-TR-542

I.P. Ivrissimtzis, N.A. Dodgson, M.A. Sabin:

A generative classification of mesh refinement rules with lattice transformations

September 2002, 13 pages, PDF
An updated, improved version of this report has been published in Computer Aided Geometric Design 22(1):99–109, January 2004 [doi:10.1016/j.cagd.2003.08.001]. The classification scheme is slightly different in the CAGD version. Please refer to the CAGD version in any work which you produce.

Abstract: We give a classification of the subdivision refinement rules using sequences of similar lattices. Our work expands and unifies recent results in the classification of primal triangular subdivision [Alexa, 2001], and results on the refinement of quadrilateral lattices [Sloan, 1994, 1989]. In the examples we concentrate on the cases with low ratio of similarity and find new

univariate and bivariate refinement rules with the lowest possible such ratio, showing that this very low ratio usually comes at the expense of symmetry.

UCAM-CL-TR-543

Kerry Rodden:

Evaluating similarity-based visualisations as interfaces for image browsing

September 2002, 248 pages, PDF
PhD thesis (Newnham College, 11 October 2001)

Abstract: Large collections of digital images are becoming more and more common, and the users of these collections need computer-based systems to help them find the images they require. Digital images are easy to shrink to thumbnail size, allowing a large number of them to be presented to the user simultaneously. Generally, current image browsing interfaces display thumbnails in a two-dimensional grid, in some default order, and there has been little exploration of possible alternatives to this model.

With textual document collections, information visualisation techniques have been used to produce representations where the documents appear to be clustered according to their mutual similarity, which is based on the words they have in common. The same techniques can be applied to images, to arrange a set of thumbnails according to a defined measure of similarity. In many collections, the images are manually annotated with descriptive text, allowing their similarity to be measured in an analogous way to textual documents. Alternatively, research in content-based image retrieval has made it possible to measure similarity based on low-level visual features, such as colour.

The primary goal of this research was to investigate the usefulness of such similarity-based visualisations as interfaces for image browsing. We concentrated on visual similarity, because it is applicable to any image collection, regardless of the availability of annotations. Initially, we used conventional information retrieval evaluation methods to compare the relative performance of a number of different visual similarity measures, both for retrieval and for creating visualisations.

Thereafter, our approach to evaluation was influenced more by human-computer interaction: we carried out a series of user experiments where arrangements based on visual similarity were compared to random arrangements, for different image browsing tasks. These included finding a given target image, finding a group of images matching a generic requirement, and choosing subjectively suitable images for a particular purpose (from a shortlisted set). As expected, we found that similarity-based arrangements are generally more helpful than random arrangements, especially when the user already has some idea of the type of image she is looking for.

Images are used in many different application domains; the ones we chose to study were stock photography and personal photography. We investigated the organisation and browsing of personal photographs in some depth, because of the inevitable future growth in usage of digital cameras, and a lack of previous research in this area.

UCAM-CL-TR-544

I.P. Ivrissimtzis, M.A. Sabin, N.A. Dodgson:

On the support of recursive subdivision

September 2002, 20 pages, PDF
An updated, improved version of this report has been published in ACM Trans. Graphics 23(4):1043–1060, October 2004 [doi:10.1145/1027411.1027417]

Abstract: We study the support of subdivision schemes, that is, the area of the subdivision surface that will be affected by the displacement of a single control point. Our main results cover the regular case, where the mesh induces a regular Euclidean tessellation of the parameter space. If n is the ratio of similarity between the tessellation at step k and step $k-1$ of the subdivision, we show that this number determines if the support is polygonal or fractal. In particular if n=2, as it is in the most schemes, the support is a polygon whose vertices can be easily determined. If n is not equal to two as, for example, in the square root of three scheme, the support is usually fractal and on its boundary we can identify sets like the classic ternary Cantor set.

UCAM-CL-TR-545

Anthony C.J. Fox:

A HOL specification of the ARM instruction set architecture

June 2001, 45 pages, PDF

Abstract: This report gives details of a HOL specification of the ARM instruction set architecture. It is shown that the HOL proof tool provides a suitable environment in which to model the architecture. The specification is used to execute fragments of ARM code generated by an assembler. The specification is based primarily around the third version of the ARM architecture, and the intent is to provide a target semantics for future microprocessor verifications.

UCAM-CL-TR-546

Jonathan David Pfautz:

Depth perception in computer graphics

September 2002, 182 pages, PDF
PhD thesis (Trinity College, May 2000)

Abstract: With advances in computing and visual display technology, the interface between man and machine has become increasingly complex. The usability of a modern interactive system depends on the design of the visual display. This dissertation aims to improve the design process by examining the relationship between human perception of depth and three-dimensional computer-generated imagery (3D CGI).

Depth is perceived when the human visual system combines various different sources of information about a scene. In Computer Graphics, linear perspective is a common depth cue, and systems utilising binocular disparity cues are of increasing interest. When these cues are inaccurately and inconsistently presented, the effectiveness of a display will be limited. Images generated with computers are sampled, meaning they are discrete in both time and space. This thesis describes the sampling artefacts that occur in 3D CGI and their effects on the perception of depth. Traditionally, sampling artefacts are treated as a Signal Processing problem. The approach here is to evaluate artefacts using Human Factors and Ergonomics methodology; sampling artefacts are assessed via performance on relevant visual tasks.

A series of formal and informal experiments were performed on human subjects to evaluate the effects of spatial and temporal sampling on the presentation of depth in CGI. In static images with perspective information, the relative size of an object can be inconsistently presented across depth. This inconsistency prevented subjects from making accurate relative depth judgements. In moving images, these distortions were most visible when the object was moving slowly, pixel size was large, the object was located close to the line of sight and/or the object was located a large virtual distance from the viewer. When stereo images are presented with perspective cues, the sampling artefacts found in each cue interact. Inconsistencies in both size and disparity can occur as the result of spatial and temporal sampling. As a result, disparity can vary inconsistently across an object. Subjects judged relative depth less accurately when these inconsistencies were present. An experiment demonstrated that stereo cues dominated in conflict situations for static images. In moving imagery, the number of samples in stereo cues is limited. Perspective information dominated the perception of depth for unambiguous (i.e., constant in direction and velocity) movement.

Based on the experimental results, a novel method was developed that ensures the size, shape and disparity of an object are consistent as it moves in depth. This algorithm manipulates the edges of an object (at the expense of positional accuracy) to enforce consistent size, shape and disparity. In a time-to-contact task using only stereo and perspective depth cues, velocity was judged more accurately using this method. A second method manipulated the location and orientation of the viewpoint to maximise the number of samples of perspective and stereo depth in a scene. This algorithm was tested in a simulated air traffic control task. The experiment demonstrated that knowledge about where the viewpoint is located dominates any benefit gained in reducing sampling artefacts.

This dissertation provides valuable information for the visual display designer in the form of task-specific experimental results and computationally inexpensive methods for reducing the effects of sampling.

UCAM-CL-TR-547

Agathoniki Trigoni:

Semantic optimization of OQL queries

October 2002, 171 pages, PDF
PhD thesis (Pembroke College, October 2001)

Abstract: This work explores all the phases of developing a query processor for OQL, the Object Query Language proposed by the Object Data Management Group (ODMG 3.0). There has been a lot of research on the execution of relational queries and their optimization using syntactic or semantic transformations. However, there is no context that has integrated and tested all the phases of processing an object query language, including the use of semantic optimization heuristics. This research is motivated by the need for query execution tools that combine two valuable properties: i) the expressive power to encompass all the features of the object-oriented paradigm and ii) the flexibility to benefit from the experience gained with relational systems, such as the use of semantic knowledge to speed up query execution.

The contribution of this work is twofold. First, it establishes a rigorous basis for OQL by defining a type inference model for OQL queries and proposing a complete framework for their translation into calculus and algebraic representations. Second, in order to enhance query execution it provides algorithms for applying two semantic optimization heuristics: constraint introduction and constraint elimination techniques. By taking into consideration a set of association rules with exceptions, it is possible to add or remove predicates from an OQL query, thus transforming it to a more efficient form.

We have implemented this framework, which enables us to measure the benefits and the cost of exploiting semantic knowledge during query execution. The experiments showed significant benefits, especially in the application of the constraint introduction technique. In contexts where queries are optimized once

and are then executed repeatedly, we can ignore the cost of optimization, and it is always worth carrying out the proposed transformation. In the context of adhoc queries the cost of the optimization becomes an important consideration. We have developed heuristics to estimate the cost as well as the benefits of optimization. The optimizer will carry out a semantic transformation only when the overhead is less than the expected benefit. Thus transformations are performed safely even with adhoc queries. The framework can often speed up the execution of an OQL query to a considerable extent.

UCAM-CL-TR-548

Anthony Fox:

Formal verification of the ARM6 micro-architecture

November 2002, 59 pages, PDF

Abstract: This report describes the formal verification of the ARM6 micro-architecture using the HOL theorem prover. The correctness of the microprocessor design compares the micro-architecture with an abstract, target instruction set semantics. Data and temporal abstraction maps are used to formally relate the state spaces and to capture the timing behaviour of the processor. The verification is carried out in HOL and one-step theorems are used to provide the framework for the proof of correctness. This report also describes the formal specification of the ARM6's three stage pipelined micro-architecture.

UCAM-CL-TR-549

Ross Anderson:

Two remarks on public key cryptology

December 2002, 7 pages, PDF

Abstract: In some talks I gave in 1997-98, I put forward two observations on public-key cryptology, concerning forward-secure signatures and compatible weak keys. I did not publish a paper on either of them as they appeared to be rather minor footnotes to public key cryptology. But the work has occasionally been cited, and I've been asked to write a permanent record.

UCAM-CL-TR-550

Karen Spärck Jones:

Computer security – a layperson's guide, from the bottom up

June 2002, 23 pages, PDF

Abstract: Computer security as a technical matter is complex, and opaque for those who are not themselves computer professionals but who encounter, or are ultimately responsible for, computer systems. This paper presents the essentials of computer security in non-technical terms, with the aim of helping people affected by computer systems to understand what security is about and to withstand the blinding with science mantras that too often obscure the real issues.

UCAM-CL-TR-551

Lawrence C. Paulson:

The relative consistency of the axiom of choice — mechanized using Isabelle/ZF

December 2002, 63 pages, PDF

Abstract: The proof of the relative consistency of the axiom of choice has been mechanized using Isabelle/ZF. The proof builds upon a previous mechanization of the reflection theorem. The heavy reliance on metatheory in the original proof makes the formalization unusually long, and not entirely satisfactory: two parts of the proof do not fit together. It seems impossible to solve these problems without formalizing the metatheory. However, the present development follows a standard textbook, Kunen's "Set Theory", and could support the formalization of further material from that book. It also serves as an example of what to expect when deep mathematics is formalized.

UCAM-CL-TR-552

Keir A. Fraser, Steven M. Hand, Timothy L. Harris, Ian M. Leslie, Ian A. Pratt:

The Xenoserver computing infrastructure

January 2003, 11 pages, PDF

Abstract: The XenoServer project will build a public infrastructure for wide-area distributed computing. We envisage a world in which XenoServer execution platforms will be scattered across the globe and available for any member of the public to submit code for execution. Crucially, the code's sponsor will be billed for all the resources used or reserved during its execution. This will encourage load balancing, limit congestion, and make the platform self-financing.

Such a global infrastructure is essential to address the fundamental problem of communication latency. By enabling principals to run programs at points throughout the network they can ensure that their code executes close to the entities with which it interacts. As

well as reducing latency this can be used to avoid network bottlenecks, to reduce long-haul network charges and to provide a network presence for transiently-connected mobile devices.

This project will build and deploy a global XenoServer test-bed and make it available to authenticated external users; initially members of the scientific community and ultimately of the general public. In this environment accurate resource accounting and pricing is critical – whether in an actual currency or one that is fictitious. As with our existing work on OS resource management, pricing provides the feedback necessary for applications that can adapt, and prevents over-use by applications that cannot.

UCAM-CL-TR-553

Paul R. Barham, Boris Dragovic,
Keir A. Fraser, Steven M. Hand,
Timothy L. Harris, Alex C. Ho,
Evangelos Kotsovinos,
Anil V.S. Madhavapeddy, Rolf Neugebauer,
Ian A. Pratt, Andrew K. Warfield:

Xen 2002

January 2003, 15 pages, PDF

Abstract: This report describes the design of Xen, the hypervisor developed as part of the XenoServer wide-area computing project. Xen enables the hardware resources of a machine to be virtualized and dynamically partitioned such as to allow multiple different 'guest' operating system images to be run simultaneously.

Virtualizing the machine in this manner provides flexibility, allowing different users to choose their preferred operating system (Windows, Linux, NetBSD), and also enables use of the platform as a testbed for operating systems research. Furthermore, Xen provides secure partitioning between these 'domains', and enables better resource accounting and QoS isolation than can be achieved within a conventional operating system. We show these benefits can be achieved at negligible performance cost.

We outline the design of Xen's main sub-systems, and the interface exported to guest operating systems. Initial performance results are presented for our most mature guest operating system port, Linux 2.4. This report covers the initial design of Xen, leading up to our first public release which we plan to make available for download in April 2003. Further reports will update the design as our work progresses and present the implementation in more detail.

UCAM-CL-TR-554

Jon Crowcroft:

Towards a field theory for networks

January 2003, 9 pages, PDF

Abstract: It is often claimed that Internet Traffic patterns are interesting because the Internet puts few constraints on sources. This leads to innovation. It also makes the study of Internet traffic, what we might cal the search for the Internet Erlang, very difficult. At the same time, traffic control (congestion control) and engineering are both hot topics.

What if "flash crowds" (a.k.a. slashdot), cascades, epidemics and so on are the norm? What if the trend continues for network link capacity to become flatter, with more equal capacity in the access and core, or even more capacity in the access than the core (as in the early 1980s with 10Mbps LANs versus Kbps links in the ARPANET)? How could we cope?

This is a paper about the use of field equations (e.g. gravitational, electrical, magnetic, strong and weak atomic and so forth) as a future model for managing network traffic. We believe that in the future, one could move from this model to a very general prescriptive technique for designing network control on different timescales, including traffic engineering and the set of admission and congestion control laws. We also speculate about the use of the same idea in wireless networks.

UCAM-CL-TR-555

Jon Crowcroft, Richard Gibbens,
Stephen Hailes:

BOURSE – Broadband Organisation of Unregulated Radio Systems through Economics

January 2003, 10 pages, PDF

Abstract: This is a technical report about an idea for research in the intersection of active nets, cognitive radio and power laws of network topologies.

UCAM-CL-TR-556

Jon Crowcroft:

Turing Switches – Turing machines for all-optical Internet routing

January 2003, 7 pages, PDF

Abstract: This is technical report outlining an idea for basic long term research into the architectures for programmable all-optical Internet routers.

We are revisiting some of the fundamental tenets of computer science to carry out this work, and so it is necessarily highly speculative.

Currently, the processing elements in all-electronic routers are typically fairly conventional von-Neumann architecture computers with processors that have large, complex instruction sets (even RISC is relatively complex compared with the actual requirements for packet processing) and Random Access Memory.

As the need for speed increases, first this architecture, and then the classical computing hardware components, and finally, electronics cease to be able to keep up.

At this time, optical device technology is making great strides, and we see the availability of gates, as well as a plethora of invention in providing buffering mechanisms.

However, a critical problem we foresee is the ability to re-program devices for different packet processing functions such as classification and scheduling. This proposal is aimed at researching one direction for adding optical domain programmability.

UCAM-CL-TR-557

G.M. Bierman, P. Sewell:

Iota: A concurrent XML scripting language with applications to Home Area Networking

January 2003, 32 pages, PDF

Abstract: Iota is a small and simple concurrent language that provides native support for functional XML computation and for typed channel-based communication. It has been designed as a domain-specific language to express device behaviour within the context of Home Area Networking.

In this paper we describe Iota, explaining its novel treatment of XML and describing its type system and operational semantics. We give a number of examples including Iota code to program Universal Plug 'n' Play (UPnP) devices.

UCAM-CL-TR-558

Yolanta Beresnevichiene:

A role and context based security model

January 2003, 89 pages, PDF
PhD thesis (Wolfson College, June 2000)

Abstract: Security requirements approached at the enterprise level initiate the need for models that capture the organisational and distributed aspects of information usage. Such models have to express organisation-specific security policies and internal controls aiming to protect information against unauthorised access and modification, and against usage of information for unintended purposes. This technical report describes a systematic approach to modelling the security requirements from the perspective of job functions and tasks performed in an organisation. It deals with the design, analysis, and management of security abstractions and mechanisms in a unified framework.

The basis of access control policy in this framework is formulated around a semantic construct of a role. Roles are granted permissions according to the job functions that exist in an organisation, and then users are assigned to roles on basis of their specific job responsibilities. In order to ensure that permissions included in the roles are used by users only for purposes corresponding to the organisation's present business needs, a novel approach of "active" context-based access control is proposed. The usage of role permissions in this approach is controlled according to the emerging context associated with progress of various tasks in the organisation.

The work explores formally the security properties of the established model, in particular, support for separation of duty and least privilege principles that are important requirements in many commercial systems. Results have implications for understanding different variations of separation of duty policy that are currently used in the role-based access control.

Finally, a design architecture of the defined security model is presented detailing the components and processing phases required for successful application of the model to distributed computer environments. The model provides opportunities for the implementers, based on application requirements, to choose between several alternative design approaches.

UCAM-CL-TR-559

Eiko Yoneki, Jean Bacon:

Pronto: MobileGateway with publish-subscribe paradigm over wireless network

February 2003, 22 pages, PDF

Abstract: This paper presents the design, implementation, and evaluation of Pronto, a middleware system for mobile applications with messaging as a basis. It provides a solution for mobile application specific problems such as resource constraints, network characteristics, and data optimization. Pronto consists of three main functions: 1) MobileJMS Client, a lightweight client of Message Oriented Middleware (MOM) based on Java Message Service (JMS), 2) Gateway for reliable and efficient transmission between mobile devices and a server with pluggable components, and 3) Serverless JMS based on IP multicast. The publish-subscribe paradigm is ideal for mobile applications, as mobile devices are commonly used for data collection under conditions of frequent disconnection and changing numbers of recipients. This paradigm provides greater flexibility due to the decoupling of publisher and subscriber. Adding a gateway as a message hub to transmit information in real-time or with store-and-forward messaging provides powerful optimization and data transformation. Caching is an essential function of the gateway, and SmartCaching is designed for generic caching

in an N-tier architecture. Serverless JMS aims at a decentralized messaging model, which supports an ad-hoc network, as well as creating a high-speed messaging BUS. Pronto is an intelligent MobileGateway, providing a useful MOM intermediary between a server and mobile devices over a wireless network.

UCAM-CL-TR-560

Mike Bond, Piotr Zieliński:

Decimalisation table attacks for PIN cracking

February 2003, 14 pages, PDF

Abstract: We present an attack on hardware security modules used by retail banks for the secure storage and verification of customer PINs in ATM (cash machine) infrastructures. By using adaptive decimalisation tables and guesses, the maximum amount of information is learnt about the true PIN upon each guess. It takes an average of 15 guesses to determine a four digit PIN using this technique, instead of the 5000 guesses intended. In a single 30 minute lunch-break, an attacker can thus discover approximately 7000 PINs rather than 24 with the brute force method. With a £300 withdrawal limit per card, the potential bounty is raised from £7200 to £2.1 million and a single motivated attacker could withdraw £30–50 thousand of this each day. This attack thus presents a serious threat to bank security.

UCAM-CL-TR-561

Paul B. Menage:

Resource control of untrusted code in an open network environment

March 2003, 185 pages, PDF
PhD thesis (Magdalene College, June 2000)

Abstract: Current research into Active Networks, Open Signalling and other forms of mobile code have made use of the ability to execute user-supplied code at locations within the network infrastructure, in order to avoid the inherent latency associated with wide area networks or to avoid sending excessive amounts of data across bottleneck links or nodes. Existing research has addressed the design and evaluation of programming environments, and testbeds have been implemented on traditional operating systems. Such work has deferred issues regarding resource control; this has been reasonable, since this research has been conducted in a closed environment.

In an open environment, which is required for widespread deployment of such technologies, the code supplied to the network nodes may not be from a trusted source. Thus, it cannot be assumed that such code will behave non-maliciously, nor that it will avoid consuming more than its fair share of the available system resources.

The computing resources consumed by end-users on programmable nodes within a network are not free, and must ultimately be paid for in some way. Programmable networks allow users substantially greater complexity in the way that they may consume network resources. This dissertation argues that, due to this complexity, it is essential to be able control and account for the resources used by untrusted user-supplied code if such technology is to be deployed effectively in a wide-area open environment.

The Resource Controlled Active Node Environment (RCANE) is presented to facilitate the control of untrusted code. RCANE supports the allocation, scheduling and accounting of the resources available on a node, including CPU and network I/O scheduling, memory allocation, and garbage collection overhead.

UCAM-CL-TR-562

Carsten Moenning, Neil A. Dodgson:

Fast Marching farthest point sampling

April 2003, 16 pages, PDF

Abstract: Using Fast Marching for the incremental computation of distance maps across the sampling domain, we obtain an efficient farthest point sampling technique (FastFPS). The method is based on that of Eldar et al. (1992, 1997) but extends more naturally to the case of non-uniform sampling and is more widely applicable. Furthermore, it can be applied to both planar domains and curved manifolds and allows for weighted domains in which different cost is associated with different points on the surface. We conclude with considering the extension of FastFPS to the sampling of point clouds without the need for prior surface reconstruction.

UCAM-CL-TR-563

G.M. Bierman, M.J. Parkinson, A.M. Pitts:

MJ: An imperative core calculus for Java and Java with effects

April 2003, 53 pages, PDF

Abstract: In order to study rigorously object-oriented languages such as Java or C#, a common practice is to define lightweight fragments, or calculi, which are sufficiently small to facilitate formal proofs of key properties. However many of the current proposals for calculi lack important language features. In this paper we propose Middleweight Java, MJ, as a contender for a minimal imperative core calculus for Java. Whilst compact, MJ models features such as object identity, field

assignment, constructor methods and block structure. We define the syntax, type system and operational semantics of MJ, and give a proof of type safety. In order to demonstrate the usefulness of MJ to reason about operational features, we consider a recent proposal of Greenhouse and Boyland to extend Java with an effects system. This effects system is intended to delimit the scope of computational effects within a Java program. We define an extension of MJ with a similar effects system and instrument the operational semantics. We then prove the correctness of the effects system; a question left open by Greenhouse and Boyland. We also consider the question of effect inference for our extended calculus, detail an algorithm for inferring effects information and give a proof of correctness.

UCAM-CL-TR-564

Ulrich Lang:

Access policies for middleware

May 2003, 138 pages, PDF
PhD thesis (Wolfson College, March 2003)

Abstract: This dissertation examines how the architectural layering of middleware constrains the design of a middleware security architecture, and analyses the complications that arise from that. First, we define a precise notion of middleware that includes its architecture and features. Our definition is based on the Common Object Request Broker Architecture (CORBA), which is used throughout this dissertation both as a reference technology and as a basis for a proof of concept implementation. In several steps, we construct a security model that fits to the described middleware architecture. The model facilitates conceptual reasoning about security. The results of our analysis indicate that the cryptographic identities available on the lower layers of the security model are only of limited use for expressing fine-grained security policies, because they are separated from the application layer entities by the middleware layer. To express individual application layer entities in access policies, additional more fine-grained descriptors are required. To solve this problem for the target side (i.e., the receiving side of an invocation), we propose an improved middleware security model that supports individual access policies on a per-target basis. The model is based on so-called "resource descriptors", which are used in addition to cryptographic identities to describe application layer entities in access policies. To be useful, descriptors need to fulfil a number of properties, such as local uniqueness and persistency. Next, we examine the information available at the middleware layer for its usefulness as resource descriptors, in particular the interface name and the instance information inside the object reference. Unfortunately neither fulfils all required properties. However, it is possible to obtain resource descriptors on the target side through a mapping process that links target instance information to an externally provided descriptor. We describe both the mapping configuration when the target is instantiated and the mapping process at invocation time. A proof of concept implementation, which contains a number of technical improvements over earlier attempts to solve this problem, shows that this approach is useable in practice, even for complex architectures, such as CORBA and CORBASec (the security services specified for CORBA). Finally, we examine the security approaches of several related middleware technologies that have emerged since the specification of CORBA and CORBASec, and show the applicability of the resource descriptor mapping.

UCAM-CL-TR-565

Carsten Moenning, Neil A. Dodgson:

Fast Marching farthest point sampling for point clouds and implicit surfaces

May 2003, 15 pages, PDF

Abstract: In a recent paper (Moenning and Dodgson, 2003), the Fast Marching farthest point sampling strategy (FastFPS) for planar domains and curved manifolds was introduced. The version of FastFPS for curved manifolds discussed in the paper deals with surface domains in triangulated form only. Due to a restriction of the underlying Fast Marching method, the algorithm further requires the splitting of any obtuse into acute triangles to ensure the consistency of the Fast Marching approximation. In this paper, we overcome these restrictions by using Memoli and Sapiro's (Memoli and Sapiro, 2001 and 2002) extension of the Fast Marching method to the handling of implicit surfaces and point clouds. We find that the extended FastFPS algorithm can be applied to surfaces in implicit or point cloud form without the loss of the original algorithm's computational optimality and without the need for any preprocessing.

UCAM-CL-TR-566

Joe Hurd:

Formal verification of probabilistic algorithms

May 2003, 154 pages, PDF
PhD thesis (Trinity College, December 2001)

Abstract: This thesis shows how probabilistic algorithms can be formally verified using a mechanical theorem prover.

We begin with an extensive foundational development of probability, creating a higher-order logic formalization of mathematical measure theory. This allows the definition of the probability space we use to model a random bit generator, which informally is a

stream of coin-flips, or technically an infinite sequence of IID Bernoulli(1/2) random variables.

Probabilistic programs are modelled using the state-transformer monad familiar from functional programming, where the random bit generator is passed around in the computation. Functions remove random bits from the generator to perform their calculation, and then pass back the changed random bit generator with the result.

Our probability space modelling the random bit generator allows us to give precise probabilistic specifications of such programs, and then verify them in the theorem prover.

We also develop technical support designed to expedite verification: probabilistic quantifiers; a compositional property subsuming measurability and independence; a probabilistic while loop together with a formal concept of termination with probability 1. We also introduce a technique for reducing properties of a probabilistic while loop to properties of programs that are guaranteed to terminate: these can then be established using induction and standard methods of program correctness.

We demonstrate the formal framework with some example probabilistic programs: sampling algorithms for four probability distributions; some optimal procedures for generating dice rolls from coin flips; the symmetric simple random walk. In addition, we verify the Miller-Rabin primality test, a well-known and commercially used probabilistic algorithm. Our fundamental perspective allows us to define a version with strong properties, which we can execute in the logic to prove compositeness of numbers.

UCAM-CL-TR-567

Joe Hurd:

Using inequalities as term ordering constraints

June 2003, 17 pages, PDF

Abstract: In this paper we show how linear inequalities can be used to approximate Knuth-Bendix term ordering constraints, and how term operations such as substitution can be carried out on systems of inequalities. Using this representation allows an off-the-shelf linear arithmetic decision procedure to check the satisfiability of a set of ordering constraints. We present a formal description of a resolution calculus where systems of inequalities are used to constrain clauses, and implement this using the Omega test as a satisfiability checker. We give the results of an experiment over problems in the TPTP archive, comparing the practical performance of the resolution calculus with and without inherited inequality constraints.

UCAM-CL-TR-568

Gavin Bierman, Michael Hicks, Peter Sewell, Gareth Stoyle, Keith Wansbrough:

Dynamic rebinding for marshalling and update, with destruct-time λ

February 2004, 85 pages, PDF

Abstract: Most programming languages adopt static binding, but for distributed programming an exclusive reliance on static binding is too restrictive: dynamic binding is required in various guises, for example when a marshalled value is received from the network, containing identifiers that must be rebound to local resources. Typically it is provided only by ad-hoc mechanisms that lack clean semantics.

In this paper we adopt a foundational approach, developing core dynamic rebinding mechanisms as extensions to the simply-typed call-by-value λ-calculus. To do so we must first explore refinements of the call-by-value reduction strategy that delay instantiation, to ensure computations make use of the most recent versions of rebound definitions. We introduce redex-time and destruct-time strategies. The latter forms the basis for a λ-marsh calculus that supports dynamic rebinding of marshalled values, while remaining as far as possible statically-typed. We sketch an extension of λ-marsh with concurrency and communication, giving examples showing how wrappers for encapsulating untrusted code can be expressed. Finally, we show that a high-level semantics for dynamic updating can also be based on the destruct-time strategy, defining a λ-update calculus with simple primitives to provide type-safe updating of running code. We thereby establish primitives and a common semantic foundation for a variety of real-world dynamic rebinding requirements.

UCAM-CL-TR-569

James J. Leifer, Gilles Peskine, Peter Sewell, Keith Wansbrough:

Global abstraction-safe marshalling with hash types

June 2003, 86 pages, PDF

Abstract: Type abstraction is a key feature of ML-like languages for writing large programs. Marshalling is necessary for writing distributed programs, exchanging values via network byte-streams or persistent stores. In this paper we combine the two, developing compile-time and run-time semantics for marshalling, that guarantee abstraction-safety between separately-built programs.

We obtain a namespace for abstract types that is global, ie meaningful between programs, by hashing module declarations. We examine the scenarios in

which values of abstract types are communicated from one program to another, and ensure, by constructing hashes appropriately, that the dynamic and static notions of type equality mirror each other. We use singleton kinds to express abstraction in the static semantics; abstraction is tracked in the dynamic semantics by coloured brackets. These allow us to prove preservation, erasure, and coincidence results. We argue that our proposal is a good basis for extensions to existing ML-like languages, pragmatically straightforward for language users and for implementors.

UCAM-CL-TR-570

Ole Høgh Jensen, Robin Milner:

Bigraphs and mobile processes

July 2003, 121 pages, PDF

Abstract: A bigraphical reactive system (BRS) involves bigraphs, in which the nesting of nodes represents locality, independently of the edges connecting them; it also allows bigraphs to reconfigure themselves. BRSs aim to provide a uniform way to model spatially distributed systems that both compute and communicate. In this memorandum we develop their static and dynamic theory.

In Part I we illustrate bigraphs in action, and show how they correspond to to process calculi. We then develop the abstract (non-graphical) notion of wide reactive system (WRS), of which BRSs are an instance. Starting from reaction rules —often called rewriting rules— we use the RPO theory of Leifer and Milner to derive (labelled) transition systems for WRSs, in a way that leads automatically to behavioural congruences.

In Part II we develop bigraphs and BRSs formally. The theory is based directly on graphs, not on syntax. Key results in the static theory are that sufficient RPOs exist (enabling the results of Part I to be applied), that parallel combinators familiar from process calculi may be defined, and that a complete algebraic theory exists at least for pure bigraphs (those without binding). Key aspects in the dynamic theory —the BRSs— are the definition of parametric reaction rules that may replicate or discard parameters, and the full application of the behavioural theory of Part I.

In Part III we introduce a special class: the simple BRSs. These admit encodings of many process calculi, including the π-calculus and the ambient calculus. A still narrower class, the basic BRSs, admits an easy characterisation of our derived transition systems. We exploit this in a case study for an asynchronous π-calculus. We show that structural congruence of process terms corresponds to equality of the representing bigraphs, and that classical strong bisimilarity corresponds to bisimilarity of bigraphs. At the end, we explore several directions for further work.

UCAM-CL-TR-571

James Hall:

Multi-layer network monitoring and analysis

July 2003, 230 pages, PDF
PhD thesis (King's College, April 2003)

Abstract: Passive network monitoring offers the possibility of gathering a wealth of data about the traffic traversing the network and the communicating processes generating that traffic. Significant advantages include the non-intrusive nature of data capture and the range and diversity of the traffic and driving applications which may be observed. Conversely there are also associated practical difficulties which have restricted the usefulness of the technique: increasing network bandwidths can challenge the capacity of monitors to keep pace with passing traffic without data loss, and the bulk of data recorded may become unmanageable.

Much research based upon passive monitoring has in consequence been limited to that using a sub-set of the data potentially available, typically TCP/IP packet headers gathered using Tcpdump or similar monitoring tools. The bulk of data collected is thereby minimised, and with the possible exception of packet filtering, the monitor's available processing power is available for the task of collection and storage. As the data available for analysis is drawn from only a small section of the network protocol stack, detailed study is largely confined to the associated functionality and dynamics in isolation from activity at other levels. Such lack of context severely restricts examination of the interaction between protocols which may in turn lead to inaccurate or erroneous conclusions.

The work described in this report attempts to address some of these limitations. A new passive monitoring architecture — Nprobe — is presented, based upon 'off the shelf' components and which, by using clusters of probes, is scalable to keep pace with current high bandwidth networks without data loss. Monitored packets are fully captured, but are subject to the minimum processing in real time needed to identify and associate data of interest across the target set of protocols. Only this data is extracted and stored. The data reduction ratio thus achieved allows examination of a wider range of encapsulated protocols without straining the probe's storage capacity.

Full analysis of the data harvested from the network is performed off-line. The activity of interest within each protocol is examined and is integrated across the range of protocols, allowing their interaction to be studied. The activity at higher levels informs study of the lower levels, and that at lower levels infers detail of the higher. A technique for dynamically modelling TCP connections is presented, which, by using data from both the transport and higher levels of the protocol stack, differentiates between the effects of network and end-process activity.

The balance of the report presents a study of Web traffic using Nprobe. Data collected from the IP, TCP, HTTP and HTML levels of the stack is integrated to identify the patterns of network activity involved in downloading whole Web pages: by using the links contained in HTML documents observed by the monitor, together with data extracted from the HTML headers of downloaded contained objects, the set of TCP connections used, and the way in which browsers use them, are studied as a whole. An analysis of the degree and distribution of delay is presented and contributes to the understanding of performance as perceived by the user. The effects of packet loss on whole page download times are examined, particularly those losses occurring early in the lifetime of connections before reliable estimations of round trip times are established. The implications of such early packet losses for pages downloads using persistent connections are also examined by simulations using the detailed data available.

UCAM-CL-TR-572

Tim Harris:

Design choices for language-based transactions

August 2003, 7 pages, PDF

Abstract: This report discusses two design choices which arose in our recent work on introducing a new 'atomic' keyword as an extension to the Java programming language. We discuss the extent to which programs using atomic blocks should be provided with an explicit 'abort' operation to roll-back the effects of the current block. We also discuss mechanisms for supporting blocks that perform I/O operations or external database transactions.

UCAM-CL-TR-573

Lawrence C. Paulson:

Mechanizing compositional reasoning for concurrent systems: some lessons

August 2003, 20 pages, PDF

Abstract: The paper reports on experiences of mechanizing various proposals for compositional reasoning in concurrent systems. The work uses the UNITY formalism and the Isabelle proof tool. The proposals investigated include existential/universal properties, guarantees properties and progress sets. The paper mentions some alternative proposals that are also worth of investigation. The conclusions are that many of these methods work and are suitable candidates for further development.

UCAM-CL-TR-574

Ivan Edward Sutherland:

Sketchpad: A man-machine graphical communication system

September 2003, 149 pages, PDF
PhD thesis (Massachusetts Institute of Technology, January 1963)
New preface by Alan Blackwell and Kerry Rodden.

Abstract: The Sketchpad system uses drawing as a novel communication medium for a computer. The system contains input, output, and computation programs which enable it to interpret information drawn directly on a computer display. It has been used to draw electrical, mechanical, scientific, mathematical, and animated drawings; it is a general purpose system. Sketchpad has shown the most usefulness as an aid to the understanding of processes, such as the notion of linkages, which can be described with pictures. Sketchpad also makes it easy to draw highly repetitive or highly accurate drawings and to change drawings previously drawn with it. The many drawings in this thesis were all made with Sketchpad.

A Sketchpad user sketches directly on a computer display with a "light pen." The light pen is used both to position parts of the drawing on the display and to point to them to change them. A set of push buttons controls the changes to be made such as "erase," or "move." Except for legends, no written language is used.

Information sketched can include straight line segments and circle arcs. Arbitrary symbols may be defined from any collection of line segments, circle arcs, and previously defined symbols. A user may define and use as many symbols as he wishes. Any change in the definition of a symbol is at once seen wherever that symbol appears.

Sketchpad stores explicit information about the topology of a drawing. If the user moves one vertex of a polygon, both adjacent sides will be moved. If the user moves a symbol, all lines attached to that symbol will automatically move to stay attached to it. The topological connections of the drawing are automatically indicated by the user as he sketches. Since Sketchpad is able to accept topological information from a human being in a picture language perfectly natural to the human, it can be used as an input program for computation programs which require topological data, e.g., circuit simulators.

Sketchpad itself is able to move parts of the drawing around to meet new conditions which the user may apply to them. The user indicates conditions with the light pen and push buttons. For example, to make two lines parallel, he successively points to the lines with the light pen and presses a button. The conditions themselves are displayed on the drawing so that they may be erased or changed with the light pen language. Any combination

of conditions can be defined as a composite condition and applied in one step.

It is easy to add entirely new types of conditions to Sketchpad's vocabulary. Since the conditions can involve anything computable, Sketchpad can be used for a very wide range of problems. For example, Sketchpad has been used to find the distribution of forces in the members of truss bridges drawn with it.

Sketchpad drawings are stored in the computer in a specially designed "ring" structure. The ring structure features rapid processing of topological information with no searching at all. The basic operations used in Sketchpad for manipulating the ring structure are described.

UCAM-CL-TR-575

Tim Granger:

Reconfigurable wavelength-switched optical networks for the Internet core

November 2003, 184 pages, PDF
PhD thesis (King's College, August 2003)

Abstract: With the quantity of data traffic carried on the Internet doubling each year, there is no let up in the demand for ever increasing network capacity. Optical fibres have a theoretical capacity of many tens of terabits per second. Currently six terabits per second has been achieved using Dense Wavelength Division Multiplexing: multiple signals at different wavelengths carried on the same fibre.

This large available bandwidth moves the performance bottlenecks to the processing required at each network node to receive, buffer, route, and transmit each individual packet. For the last 10 years the speed of the electronic routers has been, in relative terms, increasing slower than optical capacity. The space required and power consumed by these routers is also becoming a significant limitation.

One solution examined in this dissertation is to create a virtual topology in the optical layer by using all-optical switches to create lightpaths across the network. In this way nodes that are not directly connected can appear to be a single virtual hop away, and no per-packet processing is required at the intermediate nodes. With advances in optical switches it is now possible for the network to reconfigure lightpaths dynamically. This allows the network to share the resources available between the different traffic streams flowing across the network, and track changes in traffic volumes by allocating bandwidth on demand.

This solution is inherently a circuit-switched approach, but taken into account are characteristics of optical switching, in particular waveband switching (where we switch a contiguous range of wavelengths as a single unit) and latency required to achieve non disruptive switching.

This dissertation quantifies the potential gain from such a system and how that gain is related to the frequency of reconfiguration. It outlines possible network architectures which allow reconfiguration and, through simulation, measures the performance of these architectures. It then discusses the possible interactions between a reconfiguring optical layer and higher-level network layers.

This dissertation argues that the optical layer should be distinct from higher network layers, maintaining stable full-mesh connectivity, and dynamically reconfiguring the sizes and physical routes of the virtual paths to take advantage of changing traffic levels.

UCAM-CL-TR-576

David R. Spence:

An implementation of a coordinate based location system

November 2003, 12 pages, PDF

Abstract: This paper explains the co-ordinate based location system built for XenoSearch, a resource discovery system in the XenoServer Open Platform. The system is builds on the work of GNP, Lighthouse and many more recent schemes. We also present results from various combinations of algorithms to perform the actual co-ordinate calculation based on GNP, Lighthouse and spring based systems and show our implementations of the various algorithms give similar prediction errors.

UCAM-CL-TR-577

Markus G. Kuhn:

Compromising emanations: eavesdropping risks of computer displays

December 2003, 167 pages, PDF
PhD thesis (Wolfson College, June 2002)

Abstract: Electronic equipment can emit unintentional signals from which eavesdroppers may reconstruct processed data at some distance. This has been a concern for military hardware for over half a century. The civilian computer-security community became aware of the risk through the work of van Eck in 1985. Military "Tempest" shielding test standards remain secret and no civilian equivalents are available at present. The topic is still largely neglected in security textbooks due to a lack of published experimental data.

This report documents eavesdropping experiments on contemporary computer displays. It discusses the nature and properties of compromising emanations for both cathode-ray tube and liquid-crystal monitors. The detection equipment used matches the capabilities to

be expected from well-funded professional eavesdroppers. All experiments were carried out in a normal unshielded office environment. They therefore focus on emanations from display refresh signals, where periodic averaging can be used to obtain reproducible results in spite of varying environmental noise.

Additional experiments described in this report demonstrate how to make information emitted via the video signal more easily receivable, how to recover plaintext from emanations via radio-character recognition, how to estimate remotely precise video-timing parameters, and how to protect displayed text from radio-frequency eavesdroppers by using specialized screen drivers with a carefully selected video card. Furthermore, a proposal for a civilian radio-frequency emission-security standard is outlined, based on path-loss estimates and published data about radio noise levels.

Finally, a new optical eavesdropping technique is demonstrated that reads CRT displays at a distance. It observes high-frequency variations of the light emitted, even after diffuse reflection. Experiments with a typical monitor show that enough video signal remains in the light to permit the reconstruction of readable text from signals detected with a fast photosensor. Shot-noise calculations provide an upper bound for this risk.

UCAM-CL-TR-578

Robert Ennals, Richard Sharp, Alan Mycroft:

Linear types for packet processing (extended version)

January 2004, 31 pages, PDF

Abstract: We present PacLang: an imperative, concurrent, linearly-typed language designed for expressing packet processing applications. PacLang's linear type system ensures that no packet is referenced by more than one thread, but allows multiple references to a packet within a thread. We argue (i) that this property greatly simplifies compilation of high-level programs to the distributed memory architectures of modern Network Processors; and (ii) that PacLang's type system captures that style in which imperative packet processing programs are already written. Claim (ii) is justified by means of a case-study: we describe a PacLang implementation of the IPv4 unicast packet forwarding algorithm.

PacLang is formalised by means of an operational semantics and a Unique Ownership theorem formalises its correctness with respect to the type system.

UCAM-CL-TR-579

Keir Fraser:

Practical lock-freedom

February 2004, 116 pages, PDF
PhD thesis (King's College, September 2003)

Abstract: Mutual-exclusion locks are currently the most popular mechanism for interprocess synchronisation, largely due to their apparent simplicity and ease of implementation. In the parallel-computing environments that are increasingly commonplace in high-performance applications, this simplicity is deceptive: mutual exclusion does not scale well with large numbers of locks and many concurrent threads of execution. Highly-concurrent access to shared data demands a sophisticated 'fine-grained' locking strategy to avoid serialising non-conflicting operations. Such strategies are hard to design correctly and with good performance because they can harbour problems such as deadlock, priority inversion and convoying. Lock manipulations may also degrade the performance of cache-coherent multiprocessor systems by causing coherency conflicts and increased interconnect traffic, even when the lock protects read-only data.

In looking for solutions to these problems, interest has developed in lock-free data structures. By eschewing mutual exclusion it is hoped that more efficient and robust systems can be built. Unfortunately the current reality is that most lock-free algorithms are complex, slow and impractical. In this dissertation I address these concerns by introducing and evaluating practical abstractions and data structures that facilitate the development of large-scale lock-free systems.

Firstly, I present an implementation of two useful abstractions that make it easier to develop arbitrary lock-free data structures. Although these abstractions have been described in previous work, my designs are the first that can be practically implemented on current multiprocessor systems.

Secondly, I present a suite of novel lock-free search structures. This is interesting not only because of the fundamental importance of searching in computer science and its wide use in real systems, but also because it demonstrates the implementation issues that arise when using the practical abstractions I have developed.

Finally, I evaluate each of my designs and compare them with existing lock-based and lock-free alternatives. To ensure the strongest possible competition, several of the lock-based alternatives are significant improvements on the best-known solutions in the literature. These results demonstrate that it is possible to build useful data structures with all the perceived benefits of lock-freedom and with performance better than sophisticated lock-based designs. Furthermore, and contrary to popular belief, this work shows that existing hardware primitives are sufficient to build practical lock-free implementations of complex data structures.

UCAM-CL-TR-580

Ole Høgh Jensen, Robin Milner:

Bigraphs and mobile processes (revised)

February 2004, 131 pages, PDF

Abstract: A bigraphical reactive system (BRS) involves bigraphs, in which the nesting of nodes represents locality, independently of the edges connecting them; it also allows bigraphs to reconfigure themselves. BRSs aim to provide a uniform way to model spatially distributed systems that both compute and communicate. In this memorandum we develop their static and dynamic theory.

In Part I we illustrate bigraphs in action, and show how they correspond to to process calculi. We then develop the abstract (non-graphical) notion of wide reactive system (WRS), of which BRSs are an instance. Starting from reaction rules —often called rewriting rules— we use the RPO theory of Leifer and Milner to derive (labelled) transition systems for WRSs, in a way that leads automatically to behavioural congruences.

In Part II we develop bigraphs and BRSs formally. The theory is based directly on graphs, not on syntax. Key results in the static theory are that sufficient RPOs exist (enabling the results of Part I to be applied), that parallel combinators familiar from process calculi may be defined, and that a complete algebraic theory exists at least for pure bigraphs (those without binding). Key aspects in the dynamic theory —the BRSs— are the definition of parametric reaction rules that may replicate or discard parameters, and the full application of the behavioural theory of Part I.

In Part III we introduce a special class: the simple BRSs. These admit encodings of many process calculi, including the π-calculus and the ambient calculus. A still narrower class, the basic BRSs, admits an easy characterisation of our derived transition systems. We exploit this in a case study for an asynchronous π-calculus. We show that structural congruence of process terms corresponds to equality of the representing bigraphs, and that classical strong bisimilarity corresponds to bisimilarity of bigraphs. At the end, we explore several directions for further work.

UCAM-CL-TR-581

Robin Milner:

Axioms for bigraphical structure

February 2004, 26 pages, PDF

Abstract: This paper axiomatises the structure of bigraphs, and proves that the resulting theory is complete. Bigraphs are graphs with double structure, representing locality and connectivity. They have been shown to represent dynamic theories for the π-calculus, mobile ambients and Petri nets, in a way that is faithful to each of those models of discrete behaviour. While the main purpose of bigraphs is to understand mobile systems, a prerequisite for this understanding is a well-behaved theory of the structure of states in such systems. The algebra of bigraph structure is surprisingly simple, as the paper demonstrates; this is because bigraphs treat locality and connectivity orthogonally.

UCAM-CL-TR-582

Piotr Zieliński:

Latency-optimal Uniform Atomic Broadcast algorithm

February 2004, 28 pages, PDF

Abstract: We present a new asynchronous Uniform Atomic Broadcast algorithm with a delivery latency of two communication steps in optimistic settings, which is faster than any other known algorithm and has been shown to be the lower bound. It also has the weakest possible liveness requirements (the Ω failure detector and a majority of correct processes) and achieves three new lower bounds presented in this paper. Finally, we introduce a new notation and several new abstractions, which are used to construct and present the algorithm in a clear and modular way.

UCAM-CL-TR-583

Cédric Gérot, Loïc Barthe, Neil A. Dodgson, Malcolm A. Sabin:

Subdivision as a sequence of sampled Cp surfaces and conditions for tuning schemes

March 2004, 68 pages, PDF

Abstract: We deal with practical conditions for tuning a subdivision scheme in order to control its artifacts in the vicinity of a mark point. To do so, we look for good behaviour of the limit vertices rather than good mathematical properties of the limit surface. The good behaviour of the limit vertices is characterised with the definition of C2-convergence of a scheme. We propose necessary explicit conditions for C2-convergence of a scheme in the vicinity of any mark point being a vertex of valency n or the centre of an n-sided face with n greater or equal to three. These necessary conditions concern the eigenvalues and eigenvectors of subdivision matrices in the frequency domain. The components of these matrices may be complex. If we could guarantee that they were real, this would simplify numerical analysis of the eigenstructure of the matrices, especially in the context of scheme tuning where we manipulate symbolic terms. In this paper we show that an appropriate choice of the parameter space combined with a substitution of vertices lets us transform these matrices into pure real ones. The substitution consists in replacing some vertices by linear combinations of themselves. Finally, we explain how to derive conditions on the eigenelements of the real matrices which are necessary for the C2-convergence of the scheme.

UCAM-CL-TR-584

Stephen Brooks:

Concise texture editing

March 2004, 164 pages, PDF
PhD thesis (Jesus College, October 2003)

Abstract: Many computer graphics applications remain in the domain of the specialist. They are typically characterized by complex user-directed tasks, often requiring proficiency in design, colour spaces, computer interaction and file management. Furthermore, the demands of this skill set are often exacerbated by an equally complex collection of image or object manipulation commands embedded in a variety of interface components. The complexity of these graphic editing tools often requires that the user possess a correspondingly high level of expertise.

Concise Texture Editing is aimed at addressing the over-complexity of modern graphics tools and is based on the intuitive notion that the human user is skilled at high level decision making while the computer is proficient at rapid computation. This thesis has focused on the development of interactive editing tools for 2D texture images and has led to the development of a novel texture manipulation system that allows:

– the concise painting of a texture;
– the concise cloning of textures;
– the concise alteration of texture element size.

The system allows complex operations to be performed on images with minimal user interaction. When applied to the domain of image editing, this implies that the user can instruct the system to perform complex changes to digital images without having to specify copious amounts of detail. In order to reduce the user's workload, the inherent self-similarity of textures is exploited to interactively replicate editing operations globally over an image. This unique image system thereby reduces the user's workload through semi-automation, resulting in an acutely concise user interface.

UCAM-CL-TR-585

Mark S. D. Ashdown:

Personal projected displays

March 2004, 150 pages, PDF
PhD thesis (Churchill College, September 2003)

Abstract: Since the inception of the personal computer, the interface presented to users has been defined by the monitor screen, keyboard, and mouse, and by the framework of the desktop metaphor. It is very different from a physical desktop which has a large horizontal surface, allows paper documents to be arranged, browsed, and annotated, and is controlled via continuous movements with both hands. The desktop metaphor will not scale to such a large display; the continuing profusion of paper, which is used as much as ever, attests to its unsurpassed affordances as a medium for manipulating documents; and despite its proven manual and cognitive benefits, two-handed input is still not used in computer interfaces.

I present a system called the Escritoire that uses a novel configuration of overlapping projectors to create a large desk display that fills the area of a conventional desk and also has a high resolution region in front of the user for precise work. The projectors need not be positioned exactly—the projected imagery is warped using standard 3D video hardware to compensate for rough projector positioning and oblique projection. Calibration involves computing planar homographies between the 2D co-ordinate spaces of the warped textures, projector framebuffers, desk, and input devices.

The video hardware can easily perform the necessary warping and achieves 30 frames per second for the dual-projector display. Oblique projection has proved to be a solution to the problem of occlusion common to front-projection systems. The combination of an electromagnetic digitizer and an ultrasonic pen allows simultaneous input with two hands. The pen for the non-dominant hand is simpler and coarser than that for the dominant hand, reflecting the differing roles of the hands in bimanual manipulation. I give a new algorithm for calibrating a pen, that uses piecewise linear interpolation between control points. I also give an algorithm to calibrate a wall display at distance using a device whose position and orientation are tracked in three dimensions.

The Escritoire software is divided into a client that exploits the video hardware and handles the input devices, and a server that processes events and stores all of the system state. Multiple clients can connect to a single server to support collaboration. Sheets of virtual paper on the Escritoire can be put in piles which can be browsed and reordered. As with physical paper this allows items to be arranged quickly and informally, avoiding the premature work required to add an item to a hierarchical file system. Another interface feature is pen traces, which allow remote users to gesture to each other. I report the results of tests with individuals and with pairs collaborating remotely. Collaborating participants found an audio channel and the shared desk surface much more useful than a video channel showing their faces.

The Escritoire is constructed from commodity components, and unlike multi-projector display walls its cost is feasible for an individual user and it fits into a normal office setting. It demonstrates a hardware configuration, calibration algorithm, graphics warping process, set of interface features, and distributed architecture that can make personal projected displays a reality.

UCAM-CL-TR-586

András Belokosztolszki:

Role-based access control policy administration

March 2004, 170 pages, PDF
PhD thesis (King's College, November 2003)

Abstract: The wide proliferation of the Internet has set new requirements for access control policy specification. Due to the demand for ad-hoc cooperation between organisations, applications are no longer isolated from each other; consequently, access control policies face a large, heterogeneous, and dynamic environment. Policies, while maintaining their main functionality, go through many minor adaptations, evolving as the environment changes.

In this thesis we investigate the long-term administration of role-based access control (RBAC) – in particular OASIS RBAC – policies.

With the aim of encapsulating persistent goals of policies we introduce extensions in the form of meta-policies. These meta-policies, whose expected lifetime is longer than the lifetime of individual policies, contain extra information and restrictions about policies. It is expected that successive policy versions are checked at policy specification time to ensure that they comply with the requirements and guidelines set by meta-policies.

In the first of the three classes of meta-policies we group together policy components by annotating them with context labels. Based on this grouping and an information flow relation on context labels, we limit the way in which policy components may be connected to other component groups. We use this to partition conceptually disparate portions of policies, and reference these coherent portions to specify policy restrictions and policy enforcement behaviour.

In our second class of meta-policies – compliance policies – we specify requirements on an abstract policy model. We then use this for static policy checking. As compliance tests are performed at policy specification time, compliance policies may include restrictions that either cannot be included in policies, or whose inclusion would result in degraded policy enforcement performance. We also indicate how to use compliance policies to provide information about organisational policies without disclosing sensitive information.

The final class of our meta-policies, called interface policies, is used to help set up and maintain cooperation among organisations by enabling them to use components from each other's policies. Being based on compliance policies, they use an abstract policy component model, and can also specify requirements for both component exporters and importers. Using such interface policies we can reconcile compatibility issues between cooperating parties automatically.

Finally, building on our meta-policies, we consider policy evolution and self-administration, according to which we treat RBAC policies as distributed resources to which access is specified with the help of RBAC itself. This enables environments where policies are maintained by many administrators who have varying levels of competence, trust, and jurisdiction.

We have tested all of these concepts in Desert, our proof of concept implementation.

UCAM-CL-TR-587

Paul Alexander Cunningham:

Verification of asynchronous circuits

April 2004, 174 pages, PDF
PhD thesis (Gonville and Caius College, January 2002)

Abstract: The purpose of this thesis is to introduce proposition-oriented behaviours and apply them to the verification of asynchronous circuits. The major contribution of proposition-oriented behaviours is their ability to extend existing formal notations to permit the explicit use of both signal levels and transitions.

This thesis begins with the formalisation of proposition-oriented behaviours in the context of gate networks, and with the set-theoretic extension of both regular-expressions and trace-expressions to reason over proposition-oriented behaviours. A new trace-expression construct, referred to as biased composition, is also introduced. Algorithmic realisation of these set-theoretic extensions is documented using a special form of finite automata called proposition automata. A verification procedure for conformance of gate networks to a set of proposition automata is described in which each proposition automaton may be viewed either as a constraint or a specification. The implementation of this procedure as an automated verification program called Veraci is summarised, and a number of example Veraci programs are used to demonstrate contributions of proposition-oriented behaviour to asynchronous circuit design. These contributions include level-event unification, event abstraction, and relative timing assumptions using biased composition. The performance of Veraci is also compared to an existing event-oriented verification program called Versify, the result of this comparison being a consistent performance gain using Veraci over Versify.

This thesis concludes with the design and implementation of a 2048 bit dual-rail asynchronous Montgomery exponentiator, MOD_EXP, in a 0.18μm standard-cell process. The application of Veraci to the design of MOD_EXP is summarised, and the practical benefits of proposition-oriented verification are discussed.

UCAM-CL-TR-588

Panit Watcharawitch:

MulTEP: A MultiThreaded Embedded Processor

May 2004, 190 pages, PDF
PhD thesis (Newnham College, November 2003)

Abstract: Conventional embedded microprocessors have traditionally followed the footsteps of high-end processor design to achieve high performance. Their underlying architectures prioritise tasks by time-critical interrupts and rely on software to perform scheduling tasks. Single threaded execution relies on instruction-based probabilistic techniques, such as speculative execution and branch prediction, which are unsuitable for embedded systems when real-time performance guarantees need to be met. Multithreading appears to be a feasible solution for embedded processors. Thread-level parallelism has a potential to overcome the limitations of insufficient instruction-level parallelism to hide the increasing memory latencies. MulTEP is designed to provide high performance thread-level parallelism, real-time characteristics, a flexible number of threads and low incremental cost per thread for the embedded system. In its architecture, a matching-store synchronisation mechanism allows a thread to wait for multiple data items. A tagged up/down dynamic-priority hardware scheduler is provided for real-time scheduling. Pre-loading, pre-fetching and colour-tagging techniques are implemented to allow context switches without any overhead. The architecture provides four additional multithreading instructions for programmers and advance compilers to create programs with low-overhead multithreaded operations. Experimental results demonstrate that multithreading can be effectively used to improve performance and system utilisation. Latency operations that would otherwise stall the pipeline are hidden by the execution of the other threads. The hardware scheduler provides priority scheduling, which is suitable for real-time embedded applications.

UCAM-CL-TR-589

Glynn Winskel, Francesco Zappa Nardelli:

new-HOPLA — a higher-order process language with name generation

May 2004, 16 pages, PDF

Abstract: This paper introduces new-HOPLA, a concise but powerful language for higher-order nondeterministic processes with name generation. Its origins as a metalanguage for domain theory are sketched but for the most part the paper concentrates on its operational semantics. The language is typed, the type of a process describing the shape of the computation paths it can perform. Its transition semantics, bisimulation, congruence properties and expressive power are explored. Encodings of π-calculus and HOπ are presented.

UCAM-CL-TR-590

Peter R. Pietzuch:

Hermes: A scalable event-based middleware

June 2004, 180 pages, PDF
PhD thesis (Queens' College, February 2004)

Abstract: Large-scale distributed systems require new middleware paradigms that do not suffer from the limitations of traditional request/reply middleware. These limitations include tight coupling between components, a lack of information filtering capabilities, and support for one-to-one communication semantics only. We argue that event-based middleware is a scalable and powerful new type of middleware for building large-scale distributed systems. However, it is important that an event-based middleware platform includes all the standard functionality that an application programmer expects from middleware.

In this thesis we describe the design and implementation of Hermes, a distributed, event-based middleware platform. The power and flexibility of Hermes is illustrated throughout for two application domains: Internet-wide news distribution and a sensor-rich, active building. Hermes follows a type- and attribute-based publish/subscribe model that places particular emphasis on programming language integration by supporting type-checking of event data and event type inheritance. To handle dynamic, large-scale environments, Hermes uses peer-to-peer techniques for autonomic management of its overlay network of event brokers and for scalable event dissemination. Its routing algorithms, implemented on top of a distributed hash table, use rendezvous nodes to reduce routing state in the system, and include fault-tolerance features for repairing event dissemination trees. All this is achieved without compromising scalability and efficiency, as is shown by a simulational evaluation of Hermes routing.

The core functionality of an event-based middleware is extended with three higher-level middleware services that address different requirements in a distributed computing environment. We introduce a novel congestion control service that avoids congestion in the overlay broker network during normal operation and recovery after failure, and therefore enables a resource-efficient deployment of the middleware. The expressiveness of subscriptions in the event-based middleware is enhanced with a composite event service that performs the distributed detection of complex event patterns, thus taking the burden away from clients. Finally, a security service adds access control to Hermes according to a secure publish/subscribe model. This model supports fine-grained access control decisions so that separate trust domains can share the same overlay broker network.

UCAM-CL-TR-591

Silas S. Brown:

Conversion of notations

June 2004, 159 pages, PDF
PhD thesis (St John's College, November 2003)

Abstract: Music, engineering, mathematics, and many other disciplines have established notations for writing their documents. The effectiveness of each of these notations can be hampered by the circumstances in which it is being used, or by a user's disability or cultural background. Adjusting the notation can help, but the requirements of different cases often conflict, meaning that each new document will have to be transformed between many versions. Tools that support the programming of such transformations can also assist by allowing the creation of new notations on demand, which is an under-explored option in the relief of educational difficulties.

This thesis reviews some programming tools that can be used to manipulate the tree-like structure of a notation in order to transform it into another. It then describes a system "4DML" that allows the programmer to create a "model" of the desired result, from which the transformation is derived. This is achieved by representing the structure in a geometric space with many dimensions, where the model acts as an alternative frame of reference.

Example applications of 4DML include the transcription of songs and musical scores into various notations, the production of specially-customised notations to assist a sight-impaired person in learning Chinese, an unusual way of re-organising personal notes, a "website scraping" system for extracting data from on-line services that provide only one presentation, and an aid to making mathematics and diagrams accessible to people with severe print disabilities. The benefits and drawbacks of the 4DML approach are evaluated, and possible directions for future work are explored.

UCAM-CL-TR-592

Mike Bond, Daniel Cvrček,
Steven J. Murdoch:

Unwrapping the Chrysalis

June 2004, 15 pages, PDF

Abstract: We describe our experiences reverse engineering the Chrysalis-ITS Luna CA^3 – a PKCS#11 compliant cryptographic token. Emissions analysis and security API attacks are viewed by many to be simpler and more efficient than a direct attack on an HSM. But how difficult is it to actually "go in the front door"? We describe how we unpicked the CA^3 internal architecture and abused its low-level API to impersonate a CA^3 token in its cloning protocol – and extract PKCS#11 private keys in the clear. We quantify the effort involved in developing and applying the skills necessary for such a reverse-engineering attack. In the process, we discover that the Luna CA^3 has far more undocumented code and functionality than is revealed to the end-user.

UCAM-CL-TR-593

Piotr Zieliński:

Paxos at war

June 2004, 30 pages, PDF

Abstract: The optimistic latency of Byzantine Paxos can be reduced from three communication steps to two, without using public-key cryptography. This is done by making a decision when more than (n+3f)/2 acceptors report to have received the same proposal from the leader, with n being the total number of acceptors and f the number of the faulty ones. No further improvement in latency is possible, because every Consensus algorithm must take at least two steps even in benign settings. Moreover, if the leader is correct, our protocol achieves the latency of at most three steps, even if some other processes fail. These two properties make this the fastest Byzantine agreement protocol proposed so far.

By running many instances of this algorithm in parallel, we can implement Vector Consensus and Byzantine Atomic Broadcast in two and three steps, respectively, which is two steps faster than any other known algorithm.

UCAM-CL-TR-594

George Danezis:

Designing and attacking anonymous communication systems

July 2004, 150 pages, PDF
PhD thesis (Queens' College, January 2004)

Abstract: This report contributes to the field of anonymous communications over widely deployed communication networks. It describes novel schemes to protect anonymity; it also presents powerful new attacks and new ways of analysing and understanding anonymity properties.

We present Mixminion, a new generation anonymous remailer, and examine its security against all known passive and active cryptographic attacks. We use the secure anonymous replies it provides, to describe a pseudonym server, as an example of the anonymous protocols that mixminion can support. The security of mix systems is then assessed against a compulsion threat model, in which an adversary can request the decryption of material from honest nodes. A new construction, the fs-mix, is presented that makes tracing messages by such an adversary extremely expensive.

Moving beyond the static security of anonymous communication protocols, we define a metric based

on information theory that can be used to measure anonymity. The analysis of the pool mix serves as an example of its use. We then create a framework within which we compare the traffic analysis resistance provided by different mix network topologies. A new topology, based on expander graphs, proves to be efficient and secure. The rgb-mix is also presented; this implements a strategy to detect flooding attacks against honest mix nodes and neutralise them by the use of cover traffic.

Finally a set of generic attacks are studied. Statistical disclosure attacks model the whole anonymous system as a black box, and are able to uncover the relationships between long-term correspondents. Stream attacks trace streams of data travelling through anonymizing networks, and uncover the communicating parties very quickly. They both use statistical methods to drastically reduce the anonymity of users. Other minor attacks are described against peer discovery and route reconstruction in anonymous networks, as well as the naïve use of anonymous replies.

UCAM-CL-TR-595

Pablo J. Arrighi:

Representations of quantum operations, with applications to quantum cryptography

July 2004, 157 pages, PDF
PhD thesis (Emmanuel College, 23 September 2003)

Abstract: Representations of quantum operations – We start by introducing a geometrical representation (real vector space) of quantum states and quantum operations. To do so we exploit an isomorphism from positive matrices to a subcone of the Minkowski future light-cone. Pure states map onto certain light-like vectors, whilst the axis of revolution encodes the overall probability of occurrence for the state. This extension of the Generalized Bloch Sphere enables us to cater for non-trace-preserving quantum operations, and in particular to view the per-outcome effects of generalized measurements. We show that these consist of the product of an orthogonal transform about the axis of the cone of revolution and a positive real symmetric linear transform. In the case of a qubit the representation becomes all the more interesting since it elegantly associates, to each measurement element of a generalized measurement, a Lorentz transformation in Minkowski space. We formalize explicitly this correspondence between 'observation of a quantum system' and 'special relativistic change of inertial frame'. To end this part we review the state-operator correspondence, which was successfully exploited by Choi to derive the operator-sum representation of quantum operations. We go further and show that all of the important theorems concerning quantum operations can in fact be derived as simple corollaries of those concerning quantum states. Using this methodology we derive novel composition laws upon quantum states and quantum operations, Schmidt-type decompositions for bipartite pure states and some powerful formulae relating to the correspondence.

Quantum cryptography – The key principle of quantum cryptography could be summarized as follows. Honest parties communicate using quantum states. To the eavesdropper these states are random and non-orthogonal. In order to gather information she must measure them, but this may cause irreversible damage. Honest parties seek to detect her mischief by checking whether certain quantum states are left intact. Thus tradeoff between the eavesdropper's information gain, and the disturbance she necessarily induces, can be viewed as the power engine behind quantum cryptographic protocols. We begin by quantifying this tradeoff in the case of a measure distinguishing two non-orthogonal equiprobable pure states. A formula for this tradeoff was first obtained by Fuchs and Peres, but we provide a shorter, geometrical derivation (within the framework of the above mentioned conal representation). Next we proceed to analyze the Information gain versus disturbance tradeoff in a scenario where Alice and Bob interleave, at random, pairwise superpositions of two message words within their otherwise classical communications. This work constitutes one of the few results currently available regarding d-level systems quantum cryptography, and seems to provide a good general primitive for building such protocols. The proof crucially relies on the state-operator correspondence formulae derived in the first part, together with some methods by Banaszek. Finally we make use of this analysis to prove the security of a 'blind quantum computation' protocol, whereby Alice gets Bob to perform some quantum algorithm for her, but prevents him from learning her input to this quantum algorithm.

UCAM-CL-TR-596

Keir Fraser, Steven Hand, Rolf Neugebauer, Ian Pratt, Andrew Warfield, Mark Williamson:

Reconstructing I/O

August 2004, 16 pages, PDF

Abstract: We present a next-generation architecture that addresses problems of dependability, maintainability, and manageability of I/O devices and their software drivers on the PC platform. Our architecture resolves both hardware and software issues, exploiting emerging hardware features to improve device safety. Our high-performance implementation, based on the Xen virtual machine monitor, provides an immediate transition opportunity for today's systems.

UCAM-CL-TR-597

Advaith Siddharthan:

Syntactic simplification and text cohesion

August 2004, 195 pages, PDF
PhD thesis (Gonville and Caius College, November 2003)

Abstract: Syntactic simplification is the process of reducing the grammatical complexity of a text, while retaining its information content and meaning. The aim of syntactic simplification is to make text easier to comprehend for human readers, or process by programs. In this thesis, I describe how syntactic simplification can be achieved using shallow robust analysis, a small set of hand-crafted simplification rules and a detailed analysis of the discourse-level aspects of syntactically rewriting text. I offer a treatment of relative clauses, apposition, coordination and subordination.

I present novel techniques for relative clause and appositive attachment. I argue that these attachment decisions are not purely syntactic. My approaches rely on a shallow discourse model and on animacy information obtained from a lexical knowledge base. I also show how clause and appositive boundaries can be determined reliably using a decision procedure based on local context, represented by part-of-speech tags and noun chunks.

I then formalise the interactions that take place between syntax and discourse during the simplification process. This is important because the usefulness of syntactic simplification in making a text accessible to a wider audience can be undermined if the rewritten text lacks cohesion. I describe how various generation issues like sentence ordering, cue-word selection, referring-expression generation, determiner choice and pronominal use can be resolved so as to preserve conjunctive and anaphoric cohesive-relations during syntactic simplification.

In order to perform syntactic simplification, I have had to address various natural language processing problems, including clause and appositive identification and attachment, pronoun resolution and referring-expression generation. I evaluate my approaches to solving each problem individually, and also present a holistic evaluation of my syntactic simplification system.

UCAM-CL-TR-598

James J. Leifer, Robin Milner:

Transition systems, link graphs and Petri nets

August 2004, 64 pages, PDF

Abstract: A framework is defined within which reactive systems can be studied formally. The framework is based upon s-categories, a new variety of categories, within which reactive systems can be set up in such a way that labelled transition systems can be uniformly extracted. These lead in turn to behavioural preorders and equivalences, such as the failures preorder (treated elsewhere) and bisimilarity, which are guaranteed to be congruential. The theory rests upon the notion of relative pushout previously introduced by the authors. The framework is applied to a particular graphical model known as link graphs, which encompasses a variety of calculi for mobile distributed processes. The specific theory of link graphs is developed. It is then applied to an established calculus, namely condition-event Petri nets. In particular, a labelled transition system is derived for condition-event nets, corresponding to a natural notion of observable actions in Petri net theory. The transition system yields a congruential bisimilarity coinciding with one derived directly from the observable actions. This yields a calibration of the general theory of reactive systems and link graphs against known specific theories.

UCAM-CL-TR-599

Mohamed F. Hassan:

Further analysis of ternary and 3-point univariate subdivision schemes

August 2004, 9 pages, PDF

Abstract: The precision set, approximation order and Hölder exponent are derived for each of the univariate subdivision schemes described in Technical Report UCAM-CL-TR-520.

UCAM-CL-TR-600

Brian Ninham Shand:

Trust for resource control: Self-enforcing automatic rational contracts between computers

August 2004, 154 pages, PDF
PhD thesis (Jesus College, February 2004)

Abstract: Computer systems need to control access to their resources, in order to give precedence to urgent or important tasks. This is increasingly important in networked applications, which need to interact with other machines but may be subject to abuse unless protected from attack. To do this effectively, they need an explicit resource model, and a way to assess others' actions in terms of it. This dissertation shows how the actions can be represented using resource-based computational

contracts, together with a rich trust model which monitors and enforces contract compliance.

Related research in the area has focused on individual aspects of this problem, such as resource pricing and auctions, trust modelling and reputation systems, or resource-constrained computing and resource-aware middleware. These need to be integrated into a single model, in order to provide a general framework for computing by contract.

This work explores automatic computerized contracts for negotiating and controlling resource usage in a distributed system. Contracts express the terms under which client and server promise to exchange resources, such as processor time in exchange for money, using a constrained language which can be automatically interpreted. A novel, distributed trust model is used to enforce these promises, and this also supports trust delegation through cryptographic certificates. The model is formally proved to have appropriate properties of safety and liveness, which ensure that cheats cannot systematically gain resources by deceit, and that mutually profitable contracts continue to be supported.

The contract framework has many applications, in automating distributed services and in limiting the disruptiveness of users' programs. Applications such as resource-constrained sandboxes, operating system multimedia support and automatic distribution of personal address book entries can all treat the user's time as a scarce resource, to trade off computational costs against user distraction. Similarly, commercial Grid services can prioritise computations with contracts, while a cooperative service such as distributed composite event detection can use contracts for detector placement and load balancing. Thus the contract framework provides a general purpose tool for managing distributed computation, allowing participants to take calculated risks and rationally choose which contracts to perform.

UCAM-CL-TR-601

Hasan Amjad:

Combining model checking and theorem proving

September 2004, 131 pages, PDF
PhD thesis (Trinity College, March 2004)

Abstract: We implement a model checker for the modal mu-calculus as a derived rule in a fully expansive mechanical theorem prover, without causing an unacceptable performance penalty.

We use a restricted form of a higher order logic representation calculus for binary decision diagrams (BDDs) to interface the model checker to a high-performance BDD engine. This is used with a formalised theory of the modal mu-calculus (which we also develop) for model checking in which all steps of the algorithm are justified by fully expansive proof. This provides a fine-grained integration of model checking and theorem proving using a mathematically rigourous interface. The generality of our theories allows us to perform much of the proof offline, in contrast with earlier work. This substantially reduces the inevitable performance penalty of doing model checking by proof.

To demonstrate the feasibility of our approach, optimisations to the model checking algorithm are added. We add naive caching and also perform advanced caching for nested non-alternating fixed-point computations.

Finally, the usefulness of the work is demonstrated. We leverage our theory by proving translations to simpler logics that are in more widespread use. We then implement an executable theory for counterexample-guided abstraction refinement that also uses a SAT solver. We verify properties of a bus architecture in use in industry as well as a pedagogical arithmetic and logic unit. The benchmarks show an acceptable performance penalty, and the results are correct by construction.

UCAM-CL-TR-602

Hasan Amjad:

Model checking the AMBA protocol in HOL

September 2004, 27 pages, PDF

Abstract: The Advanced Microcontroller Bus Architecture (AMBA) is an open System-on-Chip bus protocol for high-performance buses on low-power devices. In this report we implement a simple model of AMBA and use model checking and theorem proving to verify latency, arbitration, coherence and deadlock freedom properties of the implementation.

UCAM-CL-TR-603

Robin Milner:

Bigraphs whose names have multiple locality

September 2004, 15 pages, PDF

Abstract: The previous definition of binding bigraphs is generalised so that local names may be located in more than one region, allowing more succinct and flexible presentation of bigraphical reactive systems. This report defines the generalisation, verifies that it retains relative pushouts, and introduces a new notion of bigraph extension; this admits a wider class of parametric reaction rules. Extension is shown to be well-behaved algebraically; one consequence is that, as in the original definition of bigraphs, discrete parameters are sufficient to generate all reactions.

UCAM-CL-TR-604

Andrei Serjantov:

On the anonymity of anonymity systems

October 2004, 162 pages, PDF
PhD thesis (Queens' College, March 2003)

Abstract: Anonymity on the Internet is a property commonly identified with privacy of electronic communications. A number of different systems exist which claim to provide anonymous email and web browsing, but their effectiveness has hardly been evaluated in practice. In this thesis we focus on the anonymity properties of such systems. First, we show how the anonymity of anonymity systems can be quantified, pointing out flaws with existing metrics and proposing our own. In the process we distinguish the anonymity of a message and that of an anonymity system.

Secondly, we focus on the properties of building blocks of mix-based (email) anonymity systems, evaluating their resistance to powerful blending attacks, their delay, their anonymity under normal conditions and other properties. This leads us to methods of computing anonymity for a particular class of mixes – timed mixes – and a new binomial mix.

Next, we look at the anonymity of a message going through an entire anonymity system based on a mix network architecture. We construct a semantics of a network with threshold mixes, define the information observable by an attacker, and give a principled definition of the anonymity of a message going through such a network.

We then consider low latency connection-based anonymity systems, giving concrete attacks and describing methods of protection against them. In particular, we show that Peer-to-Peer anonymity systems provide less anonymity against the global passive adversary than ones based on a "classic" architecture.

Finally, we give an account of how anonymity can be used in censorship resistant systems. These are designed to provide availability of documents, while facing threats from a powerful adversary. We show how anonymity can be used to hide the identity of the servers where each of the documents are stored, thus making them harder to remove from the system.

UCAM-CL-TR-605

Peter Sewell, James J. Leifer, Keith Wansbrough, Mair Allen-Williams, Francesco Zappa Nardelli, Pierre Habouzit, Viktor Vafeiadis:

Acute: High-level programming language design for distributed computation

October 2004, 193 pages, PDF

Abstract: This paper studies key issues for distributed programming in high-level languages. We discuss the design space and describe an experimental language, Acute, which we have defined and implemented.

Acute extends an OCaml core to support distributed development, deployment, and execution, allowing type-safe interaction between separately-built programs. It is expressive enough to enable a wide variety of distributed infrastructure layers to be written as simple library code above the byte-string network and persistent store APIs, disentangling the language runtime from communication.

This requires a synthesis of novel and existing features:

(1) type-safe marshalling of values between programs;

(2) dynamic loading and controlled rebinding to local resources;

(3) modules and abstract types with abstraction boundaries that are respected by interaction;

(4) global names, generated either freshly or based on module hashes: at the type level, as runtime names for abstract types; and at the term level, as channel names and other interaction handles;

(5) versions and version constraints, integrated with type identity;

(6) local concurrency and thread thunkification; and

(7) second-order polymorphism with a namecase construct.

We deal with the interplay among these features and the core, and develop a semantic definition that tracks abstraction boundaries, global names, and hashes throughout compilation and execution, but which still admits an efficient implementation strategy.

UCAM-CL-TR-606

Nicholas Nethercote:

Dynamic binary analysis and instrumentation

November 2004, 177 pages, PDF
PhD thesis (Trinity College, November 2004)

Abstract: Dynamic binary analysis (DBA) tools such as profilers and checkers help programmers create better software. Dynamic binary instrumentation (DBI) frameworks make it easy to build new DBA tools. This dissertation advances the theory and practice of dynamic binary analysis and instrumentation, with an emphasis on the importance of the use and support of metadata.

The dissertation has three main parts.

The first part describes a DBI framework called Valgrind which provides novel features to support heavyweight DBA tools that maintain rich metadata, especially location metadata—the shadowing of every register and memory location with a metavalue. Location metadata is used in shadow computation, a kind of DBA where every normal operation is shadowed by an abstract operation.

The second part describes three powerful DBA tools. The first tool performs detailed cache profiling. The second tool does an old kind of dynamic analysis—bounds-checking—in a new way. The third tool produces dynamic data flow graphs, a novel visualisation that cuts to the essence of a program's execution. All three tools were built with Valgrind, and rely on Valgrind's support for heavyweight DBA and rich metadata, and the latter two perform shadow computation.

The third part describes a novel system of semi-formal descriptions of DBA tools. It gives many example descriptions, and also considers in detail exactly what dynamic analysis is.

The dissertation makes six main contributions.

First, the descriptions show that metadata is the key component of dynamic analysis; in particular, whereas static analysis predicts approximations of a program's future, dynamic analysis remembers approximations of a program's past, and these approximations are exactly what metadata is.

Second, the example tools show that rich metadata and shadow computation make for powerful and novel DBA tools that do more than the traditional tracing and profiling.

Third, Valgrind and the example tools show that a DBI framework can make it easy to build heavyweight DBA tools, by providing good support for rich metadata and shadow computation.

Fourth, the descriptions are a precise and concise way of characterising tools, provide a directed way of thinking about tools that can lead to better implementations, and indicate the theoretical upper limit of the power of DBA tools in general.

Fifth, the three example tools are interesting in their own right, and the latter two are novel.

Finally, the entire dissertation provides many details, and represents a great deal of condensed experience, about implementing DBI frameworks and DBA tools.

UCAM-CL-TR-607

Neil E. Johnson:

Code size optimization for embedded processors

November 2004, 159 pages, PDF
PhD thesis (Robinson College, May 2004)

Abstract: This thesis studies the problem of reducing code size produced by an optimizing compiler. We develop the Value State Dependence Graph (VSDG) as a powerful intermediate form. Nodes represent computation, and edges represent value (data) and state (control) dependencies between nodes. The edges specify a partial ordering of the nodes—sufficient ordering to maintain the I/O semantics of the source program, while allowing optimizers greater freedom to move nodes within the program to achieve better (smaller) code. Optimizations, both classical and new, transform the graph through graph rewriting rules prior to code generation. Additional (semantically inessential) state edges are added to transform the VSDG into a Control Flow Graph, from which target code is generated.

We show how procedural abstraction can be advantageously applied to the VSDG. Graph patterns are extracted from a program's VSDG. We then select repeated patterns giving the greatest size reduction, generate new functions from these patterns, and replace all occurrences of the patterns in the original VSDG with calls to these abstracted functions. Several embedded processors have load- and store-multiple instructions, representing several loads (or stores) as one instruction. We present a method, benefiting from the VSDG form, for using these instructions to reduce code size by provisionally combining loads and stores before code generation. The final contribution of this thesis is a combined register allocation and code motion (RACM) algorithm. We show that our RACM algorithm formulates these two previously antagonistic phases as one combined pass over the VSDG, transforming the graph (moving or cloning nodes, or spilling edges) to fit within the physical resources of the target processor.

We have implemented our ideas within a prototype C compiler and suite of VSDG optimizers, generating code for the Thumb 32-bit processor. Our results show improvements for each optimization and that we can achieve code sizes comparable to, and in some cases better than, that produced by commercial compilers with significant investments in optimization technology.

UCAM-CL-TR-608

Walt Yao:

Trust management for widely distributed systems

November 2004, 191 pages, PDF
PhD thesis (Jesus College, February 2003)

Abstract: In recent years, we have witnessed the evolutionary development of a new breed of distributed systems. Systems of this type share a number of characteristics – highly decentralized, of Internet-grade scalability, and autonomous within their administrative domains. Most importantly, they are expected to operate collaboratively across both known and unknown

domains. Prime examples include peer-to-peer applications and open web services. Typically, authorization in distributed systems is identity-based, e.g. access control lists. However, approaches based on predefined identities are unsuitable for the new breed of distributed systems because of the need to deal with unknown users, i.e. strangers, and the need to manage a potentially large number of users and/or resources. Furthermore, effective administration and management of authorization in such systems requires: (1) natural mapping of organizational policies into security policies; (2) managing collaboration of independently administered domains/organizations; (3) decentralization of security policies and policy enforcement.

This thesis describes Fidelis, a trust management framework designed to address the authorization needs for the next-generation distributed systems. A trust management system is a term coined to refer to a unified framework for the specification of security policies, the representation of credentials, and the evaluation and enforcement of policy compliances. Based on the concept of trust conveyance and a generic abstraction for trusted information as trust statements, Fidelis provides a generic platform for building secure, trust-aware distributed applications. At the heart of the Fidelis framework is a language for the specification of security policies, the Fidelis Policy Language (FPL), and the inference model for evaluating policies expressed in FPL. With the policy language and its inference model, Fidelis is able to model recommendation-style policies and policies with arbitrarily complex chains of trust propagation.

Web services have rapidly been gaining significance both in industry and research as a ubiquitous, next-generation middleware platform. The second half of the thesis describes the design and implementation of the Fidelis framework for the standard web service platform. The goal of this work is twofold: first, to demonstrate the practical feasibility of Fidelis, and second, to investigate the use of a policy-driven trust management framework for Internet-scale open systems. An important requirement in such systems is trust negotiation that allows unfamiliar principals to establish mutual trust and interact with confidence. Addressing this requirement, a trust negotiation framework built on top of Fidelis is developed.

This thesis examines the application of Fidelis in three distinctive domains: implementing generic role-based access control, trust management in the World Wide Web, and an electronic marketplace comprising unfamiliar and untrusted but collaborative organizations.

UCAM-CL-TR-609

Eleanor Toye, Anil Madhavapeddy, Richard Sharp, David Scott, Alan Blackwell, Eben Upton:

Using camera-phones to interact with context-aware mobile services

December 2004, 23 pages, PDF

Abstract: We describe an interaction technique for controlling site-specific mobile services using commercially available camera-phones, public information displays and visual tags. We report results from an experimental study validating this technique in terms of pointing speed and accuracy. Our results show that even novices can use camera-phones to "point-and-click" on visual tags quickly and accurately. We have built a publicly available client/server software framework for rapid development of applications that exploit our interaction technique. We describe two prototype applications that were implemented using this framework and present findings from user-experience studies based on these applications.

UCAM-CL-TR-610

Tommy Ingulfsen:

Influence of syntax on prosodic boundary prediction

December 2004, 49 pages, PDF
MPhil thesis (Churchill College, July 2004)

Abstract: In this thesis we compare the effectiveness of different syntactic features and syntactic representations for prosodic boundary prediction, setting out to clarify which representations are most suitable for this task. The results of a series of experiments show that it is not possible to conclude that a single representation is superior to all others. Three representations give rise to similar experimental results. One of these representations is composed only of shallow features, which were originally thought to have less predictive power than deep features. Conversely, one of the deep representations that seemed to be best suited for our purposes (syntactic chunks) turns out not to be among the three best.

UCAM-CL-TR-611

Neil A. Dodgson:

An heuristic analysis of the classification of bivariate subdivision schemes

December 2004, 18 pages, PDF

Abstract: Alexa [*] and Ivrissimtzis et al. [+] have proposed a classification mechanism for bivariate subdivision schemes. Alexa considers triangular primal schemes, Ivrissimtzis et al. generalise this both to quadrilateral and hexagonal meshes and to dual and mixed schemes. I summarise this classification and then proceed to analyse it in order to determine which classes of subdivision scheme are likely to contain useful members. My aim is to ascertain whether there are any potentially useful classes which have not yet been investigated or whether we can say, with reasonable confidence, that all of the useful classes have already been considered. I apply heuristics related to the mappings of element types (vertices, face centres, and midedges) to one another, to the preservation of symmetries, to the alignment of meshes at different subdivision levels, and to the size of the overall subdivision mask. My conclusion is that there are only a small number of useful classes and that most of these have already been investigated in terms of linear, stationary subdivision schemes. There is some space for further work, particularly in the investigation of whether there are useful ternary linear, stationary subdivision schemes, but it appears that future advances are more likely to lie elsewhere.

[*] M. Alexa. Refinement operators for triangle meshes. Computer Aided Geometric Design, 19(3):169-172, 2002.

[+] I. P. Ivrissimtzis, N. A. Dodgson, and M. A. Sabin. A generative classification of mesh refinement rules with lattice transformations. Computer Aided Geometric Design, 22(1):99-109, 2004.

UCAM-CL-TR-612

Alastair R. Beresford:

Location privacy in ubiquitous computing

January 2005, 139 pages, PDF
PhD thesis (Robinson College, April 2004)

Abstract: The field of ubiquitous computing envisages an era when the average consumer owns hundreds or thousands of mobile and embedded computing devices. These devices will perform actions based on the context of their users, and therefore ubiquitous systems will gather, collate and distribute much more personal information about individuals than computers do today. Much of this personal information will be considered private, and therefore mechanisms which allow users to control the dissemination of these data are vital. Location information is a particularly useful form of context in ubiquitous computing, yet its unconditional distribution can be very invasive.

This dissertation develops novel methods for providing location privacy in ubiquitous computing. Much of the previous work in this area uses access control to enable location privacy. This dissertation takes a different approach and argues that many location-aware applications can function with anonymised location data and that, where this is possible, its use is preferable to that of access control.

Suitable anonymisation of location data is not a trivial task: under a realistic threat model simply removing explicit identifiers does not anonymise location information. This dissertation describes why this is the case and develops two quantitative security models for anonymising location data: the mix zone model and the variable quality model.

A trusted third-party can use one, or both, models to ensure that all location events given to untrusted applications are suitably anonymised. The mix zone model supports untrusted applications which require accurate location information about users in a set of disjoint physical locations. In contrast, the variable quality model reduces the temporal or spatial accuracy of location information to maintain user anonymity at every location.

Both models provide a quantitative measure of the level of anonymity achieved; therefore any given situation can be analysed to determine the amount of information an attacker can gain through analysis of the anonymised data. The suitability of both these models is demonstrated and the level of location privacy available to users of real location-aware applications is measured.

UCAM-CL-TR-613

David J. Scott:

Abstracting application-level security policy for ubiquitous computing

January 2005, 186 pages, PDF
PhD thesis (Robinson College, September 2004)

Abstract: In the future world of Ubiquitous Computing, tiny embedded networked computers will be found in everything from mobile phones to microwave ovens. Thanks to improvements in technology and software engineering, these computers will be capable of running sophisticated new applications constructed from mobile agents. Inevitably, many of these systems will contain application-level vulnerabilities; errors caused by either unanticipated mobility or interface behaviour. Unfortunately existing methods for applying security policy – network firewalls – are inadequate to control and protect the hordes of vulnerable mobile devices. As more and more critical functions are handled by these systems, the potential for disaster is increasing rapidly.

To counter these new threats, this report champions the approach of using new application-level security policy languages in combination to protect vulnerable applications. Policies are abstracted from main application code, facilitating both analysis and future

maintenance. As well as protecting existing applications, such policy systems can help as part of a security-aware design process when building new applications from scratch.

Three new application-level policy languages are contributed each addressing a different kind of vulnerability. Firstly, the policy language MRPL allows the creation of Mobility Restriction Policies, based on a unified spatial model which represents both physical location of objects as well as virtual location of mobile code. Secondly, the policy language SPDL-2 protects applications against a large number of common errors by allowing the specification of per-request/response validation and transformation rules. Thirdly, the policy language SWIL allows interfaces to be described as automata which may be analysed statically by a model-checker before being checked dynamically in an application-level firewall. When combined together, these three languages provide an effective means for preventing otherwise critical application-level vulnerabilities.

Systems implementing these policy languages have been built; an implementation framework is described and encouraging performance results and analysis are presented.

UCAM-CL-TR-614

Robin Milner:

Pure bigraphs

January 2005, 66 pages, PDF

Abstract: Bigraphs are graphs whose nodes may be nested, representing locality, independently of the edges connecting them. They may be equipped with reaction rules, forming a bigraphical reactive system (Brs) in which bigraphs can reconfigure themselves. Brss aim to unify process calculi, and to model applications — such as pervasive computing— where locality and mobility are prominent. The paper is devoted to the theory of pure bigraphs, which underlie various more refined forms. It begins by developing a more abstract structure, a wide reactive system Wrs, of which a Brs is an instance; in this context, labelled transitions are defined in such a way that the induced bisimilarity is a congruence.

This work is then specialised to Brss, whose graphical structure allows many refinements of the dynamic theory. Elsewhere it is shown that behavioural analysis for Petri nets, π-calculus and mobile ambients can all be recovered in the uniform framework of bigraphs. The latter part of the paper emphasizes the parts of bigraphical theory that are common to these applications, especially the treatment of dynamics via labelled transitions. As a running example, the theory is applied to finite pure CCS, whose resulting transition system and bisimilarity are analysed in detail.

The paper also discusses briefly the use of bigraphs to model both pervasive computing and biological systems.

UCAM-CL-TR-615

Evangelos Kotsovinos:

Global public computing

January 2005, 229 pages, PDF
PhD thesis (Trinity Hall, November 2004)

Abstract: High-bandwidth networking and cheap computing hardware are leading to a world in which the resources of one machine are available to groups of users beyond their immediate owner. This trend is visible in many different settings. Distributed computing, where applications are divided into parts that run on different machines for load distribution, geographical dispersion, or robustness, has recently found new fertile ground. Grid computing promises to provide a common framework for scheduling scientific computation and managing the associated large data sets. Proposals for utility computing envision a world in which businesses rent computing bandwidth in server farms on-demand instead of purchasing and maintaining servers themselves.

All such architectures target particular user and application groups or deployment scenarios, where simplifying assumptions can be made. They expect centralised ownership of resources, cooperative users, and applications that are well-behaved and compliant to a specific API or middleware. Members of the public who are not involved in Grid communities or wish to deploy out-of-the-box distributed services, such as game servers, have no means to acquire resources on large numbers of machines around the world to launch their tasks.

This dissertation proposes a new distributed computing paradigm, termed global public computing, which allows any user to run any code anywhere. Such platforms price computing resources, and ultimately charge users for resources consumed. This dissertation presents the design and implementation of the XenoServer Open Platform, putting this vision into practice. The efficiency and scalability of the developed mechanisms are demonstrated by experimental evaluation; the prototype platform allows the global-scale deployment of complex services in less than 45 seconds, and could scale to millions of concurrent sessions without presenting performance bottlenecks.

To facilitate global public computing, this work addresses several research challenges. It introduces reusable mechanisms for representing, advertising, and supporting the discovery of resources. To allow flexible and federated control of resource allocation by all stakeholders involved, it proposes a novel role-based resource management framework for expressing and combining distributed management policies. Furthermore, it implements effective service deployment models for launching distributed services on large numbers of machines around the world easily, quickly, and efficiently. To keep track of resource consumption and pass

charges on to consumers, it devises an accounting and charging infrastructure.

UCAM-CL-TR-616

Donnla Nic Gearailt:

Dictionary characteristics in cross-language information retrieval

February 2005, 158 pages, PDF
PhD thesis (Gonville and Caius College, February 2003)

Abstract: In the absence of resources such a as suitable MT system, translation in Cross-Language Information Retrieval (CLIR) consists primarily of mapping query terms to a semantically equivalent representation in the target language. This can be accomplished by looking up each term in a simple bilingual dictionary. The main problem here is deciding which of the translations provided by the dictionary for each query term should be included in the query translation. We tackled this problem by examining different characteristics of the system dictionary. We found that dictionary properties such as scale (the average number of translations per term), translation repetition (providing the same translation for a term more than once in a dictionary entry, for example, for different senses of a term), and dictionary coverage rate (the percentage of query terms for which the dictionary provides a translation) can have a profound effect on retrieval performance. Dictionary properties were explored in a series of carefully controlled tests, designed to evaluate specific hypotheses. These experiments showed that (a) contrary to expectation, smaller scale dictionaries resulted in better performance than large-scale ones, and (b) when appropriately managed e.g. through strategies to ensure adequate translational coverage, dictionary-based CLIR could perform as well as other CLIR methods discussed in the literature. Our experiments showed that it is possible to implement an effective CLIR system with no resources other than the system dictionary itself, provided this dictionary is chosen with careful examination of its characteristics, removing any dependency on outside resources.

UCAM-CL-TR-617

Augustin Chaintreau, Pan Hui, Jon Crowcroft, Christophe Diot, Richard Gass, James Scott:

Pocket Switched Networks: Real-world mobility and its consequences for opportunistic forwarding

February 2005, 26 pages, PDF

Abstract: Opportunistic networks make use of human mobility and local forwarding in order to distribute data. Information can be stored and passed, taking advantage of the device mobility, or forwarded over a wireless link when an appropriate contact is met. Such networks fall into the fields of mobile ad-hoc networking and delay-tolerant networking. In order to evaluate forwarding algorithms for these networks, accurate data is needed on the intermittency of connections.

In this paper, the inter-contact time between two transmission opportunities is observed empirically using four distinct sets of data, two having been specifically collected for this work, and two provided by other research groups.

We discover that the distribution of inter-contact time follows an approximate power law over a large time range in all data sets. This observation is at odds with the exponential decay expected by many currently used mobility models. We demonstrate that opportunistic transmission schemes designed around these current models have poor performance under approximate power-law conditions, but could be significantly improved by using limited redundant transmissions.

UCAM-CL-TR-618

Mark R. Shinwell:

The Fresh Approach: functional programming with names and binders

February 2005, 111 pages, PDF
PhD thesis (Queens' College, December 2004)

Abstract: This report concerns the development of a language called Fresh Objective Caml, which is an extension of the Objective Caml language providing facilities for the manipulation of data structures representing syntax involving α-convertible names and binding operations.

After an introductory chapter which includes a survey of related work, we describe the Fresh Objective Caml language in detail. Next, we proceed to formalise a small core language which captures the essence of Fresh Objective Caml; we call this Mini-FreshML. We provide two varieties of operational semantics for this language and prove them equivalent. Then in order to prove correctness properties of representations of syntax in the language we introduce a new variety of domain theory called FM-domain theory, based on the permutation model of name binding from Pitts and Gabbay. We show how classical domain-theoretic constructions—including those for the solution of recursive domain equations—fall naturally into this setting, where they are augmented by new constructions to handle name-binding.

After developing the necessary domain theory, we demonstrate how it may be exploited to give a monadic

denotational semantics to Mini-FreshML. This semantics in itself is quite novel and demonstrates how a simple monad of continuations is sufficient to model dynamic allocation of names. We prove that our denotational semantics is computationally adequate with respect to the operational semantics—in other words, equality of denotation implies observational equivalence. After this, we show how the denotational semantics may be used to prove our desired correctness properties.

In the penultimate chapter, we examine the implementation of Fresh Objective Caml, describing detailed issues in the compiler and runtime systems. Then in the final chapter we close the report with a discussion of future avenues of research and an assessment of the work completed so far.

UCAM-CL-TR-619

James R. Bulpin:

Operating system support for simultaneous multithreaded processors

February 2005, 130 pages, PDF
PhD thesis (King's College, September 2004)

Abstract: Simultaneous multithreaded (SMT) processors are able to execute multiple application threads in parallel in order to improve the utilisation of the processor's execution resources. The improved utilisation provides a higher processor-wide throughput at the expense of the performance of each individual thread.

Simultaneous multithreading has recently been incorporated into the Intel Pentium 4 processor family as "Hyper-Threading". While there is already basic support for it in popular operating systems, that support does not take advantage of any knowledge about the characteristics of SMT, and therefore does not fully exploit the processor.

SMT presents a number of challenges to operating system designers. The threads' dynamic sharing of processor resources means that there are complex performance interactions between threads. These interactions are often unknown, poorly understood, or hard to avoid. As a result such interactions tend to be ignored leading to a lower processor throughput.

In this dissertation I start by describing simultaneous multithreading and the hardware implementations of it. I discuss areas of operating system support that are either necessary or desirable.

I present a detailed study of a real SMT processor, the Intel Hyper-Threaded Pentium 4, and describe the performance interactions between threads. I analyse the results using information from the processor's performance monitoring hardware.

Building on the understanding of the processor's operation gained from the analysis, I present a design for an operating system process scheduler that takes into account the characteristics of the processor and the workloads in order to improve the system-wide throughput. I evaluate designs exploiting various levels of processor-specific knowledge.

I finish by discussing alternative ways to exploit SMT processors. These include the partitioning onto separate simultaneous threads of applications and hardware interrupt handling. I present preliminary experiments to evaluate the effectiveness of this technique.

UCAM-CL-TR-620

Eleftheria Katsiri:

Middleware support for context-awareness in distributed sensor-driven systems

February 2005, 176 pages, PDF
PhD thesis (Clare College, January 2005)

Abstract: Context-awareness concerns the ability of computing devices to detect, interpret and respond to aspects of the user's local environment. Sentient Computing is a sensor-driven programming paradigm which maintains an event-based, dynamic model of the environment which can be used by applications in order to drive changes in their behaviour, thus achieving context-awareness. However, primitive events, especially those arising from sensors, e.g., that a user is at position {x,y,z} are too low-level to be meaningful to applications. Existing models for creating higher-level, more meaningful events, from low-level events, are insufficient to capture the user's intuition about abstract system state. Furthermore, there is a strong need for user-centred application development, without undue programming overhead. Applications need to be created dynamically and remain functional independently of the distributed nature and heterogeneity of sensor-driven systems, even while the user is mobile. Both issues combined necessitate an alternative model for developing applications in a real-time, distributed sensor-driven environment such as Sentient Computing.

This dissertation describes the design and implementation of the SCAFOS framework. SCAFOS has two novel aspects. Firstly, it provides powerful tools for inferring abstract knowledge from low-level, concrete knowledge, verifying its correctness and estimating its likelihood. Such tools include Hidden Markov Models, a Bayesian Classifier, Temporal First-Order Logic, the theorem prover SPASS and the production system CLIPS. Secondly, SCAFOS provides support for simple application development through the XML-based SCALA language. By introducing the new concept of a generalised event, an abstract event, defined as a notification of changes in abstract system state, expressiveness compatible with human intuition is achieved when using SCALA. The applications that are created

through SCALA are automatically integrated and operate seamlessly in the various heterogeneous components of the context-aware environment even while the user is mobile or when new entities or other applications are added or removed in SCAFOS.

UCAM-CL-TR-621

Mark R. Shinwell, Andrew M. Pitts:

Fresh Objective Caml user manual

February 2005, 21 pages, PDF

Abstract: This technical report is the user manual for the Fresh Objective Caml system, which implements a functional programming language incorporating facilities for manipulating syntax involving names and binding operations.

UCAM-CL-TR-622

Jörg H. Lepler:

Cooperation and deviation in market-based resource allocation

March 2005, 173 pages, PDF
PhD thesis (St John's College, November 2004)

Abstract: This thesis investigates how business transactions are enhanced through competing strategies for economically motivated cooperation. To this end, it focuses on the setting of a distributed, bilateral allocation protocol for electronic services and resources. Cooperative efforts like these are often threatened by transaction parties who aim to exploit their competitors by deviating from so-called cooperative goals. We analyse this conflict between cooperation and deviation by presenting the case of two novel market systems which use economic incentives to solve the complications that arise from cooperation.

The first of the two systems is a pricing model which is designed to address the problematic resource market situation, where supply exceeds demand and perfect competition can make prices collapse to level zero. This pricing model uses supply functions to determine the optimal Nash-Equilibrium price. Moreover, in this model the providers' market estimations are updated with information about each of their own transactions. Here, we implement the protocol in a discrete event simulation, to show that the equilibrium prices are above competitive levels, and to demonstrate that deviations from the pricing model are not profitable.

The second of the two systems is a reputation aggregation model, which seeks the subgroup of raters that (1) contains the largest degree of overall agreement and (2) derives the resulting reputation scores from their comments. In order to seek agreement, this model assumes that not all raters in the system are equally able to foster an agreement. Based on the variances of the raters' comments, the system derives a notion of the reputation for each rater, which is in turn fed back into the model's recursive scoring algorithm. We demonstrate the convergence of this algorithm, and show the effectiveness of the model's ability to discriminate between poor and strong raters. Then with a series of threat models, we show how resilient this model is in terms of finding agreement, despite large collectives of malicious raters. Finally, in a practical example, we apply the model to the academic peer review process in order to show its versatility at establishing a ranking of rated objects.

UCAM-CL-TR-623

Keith Wansbrough:

Simple polymorphic usage analysis

March 2005, 364 pages, PDF
PhD thesis (Clare Hall, March 2002)

Abstract: Implementations of lazy functional languages ensure that computations are performed only when they are needed, and save the results so that they are not repeated. This frees the programmer to describe solutions at a high level, leaving details of control flow to the compiler.

This freedom however places a heavy burden on the compiler; measurements show that over 70% of these saved results are never used again. A usage analysis that could statically detect values used at most once would enable these wasted updates to be avoided, and would be of great benefit. However, existing usage analyses either give poor results or have been applied only to prototype compilers or toy languages.

This thesis presents a sound, practical, type-based usage analysis that copes with all the language features of a modern functional language, including type polymorphism and user-defined algebraic data types, and addresses a range of problems that have caused difficulty for previous analyses, including poisoning, mutual recursion, separate compilation, and partial application and usage dependencies. In addition to well-typing rules, an inference algorithm is developed, with proofs of soundness and a complexity analysis.

In the process, the thesis develops simple polymorphism, a novel approach to polymorphism in the presence of subtyping that attempts to strike a balance between pragmatic concerns and expressive power. This thesis may be considered an extended experiment into this approach, worked out in some detail but not yet conclusive.

The analysis described was designed in parallel with a full implementation in the Glasgow Haskell Compiler, leading to informed design choices, thorough coverage of language features, and accurate measurements of its potential and effectiveness when used on real code. The latter demonstrate that the analysis yields moderate benefit in practice.

UCAM-CL-TR-624

Steve Bishop, Matthew Fairbairn, Michael Norrish, Peter Sewell, Michael Smith, Keith Wansbrough:

TCP, UDP, and Sockets: rigorous and experimentally-validated behavioural specification Volume 1: Overview

March 2005, 88 pages, PDF

Abstract: We have developed a mathematically rigorous and experimentally-validated post-hoc specification of the behaviour of TCP, UDP, and the Sockets API. It characterises the API and network-interface interactions of a host, using operational semantics in the higher-order logic of the HOL automated proof assistant. The specification is detailed, covering almost all the information of the real-world communications: it is in terms of individual TCP segments and UDP datagrams, though it abstracts from the internals of IP. It has broad coverage, dealing with arbitrary API call sequences and incoming messages, not just some well-behaved usage. It is also accurate, closely based on the de facto standard of (three of) the widely-deployed implementations. To ensure this we have adopted a novel experimental semantics approach, developing test generation tools and symbolic higher-order-logic model checking techniques that let us validate the specification directly against several thousand traces captured from the implementations.

The resulting specification, which is annotated for the non-HOL-specialist reader, may be useful as an informal reference for TCP/IP stack implementors and Sockets API users, supplementing the existing informal standards and texts. It can also provide a basis for high-fidelity automated testing of future implementations, and a basis for design and formal proof of higher-level communication layers. More generally, the work demonstrates that it is feasible to carry out similar rigorous specification work at design-time for new protocols. We discuss how such a design-for-test approach should influence protocol development, leading to protocol specifications that are both unambiguous and clear, and to high-quality implementations that can be tested directly against those specifications.

This document (Volume 1) gives an overview of the project, discussing the goals and techniques and giving an introduction to the specification. The specification itself is given in the companion Volume 2 (UCAM-CL-TR-625), which is automatically typeset from the (extensively annotated) HOL source. As far as possible we have tried to make the work accessible to four groups of intended readers: workers in networking (implementors of TCP/IP stacks, and designers of new protocols); in distributed systems (implementors of software above the Sockets API); in distributed algorithms (for whom this may make it possible to prove properties about executable implementations of those algorithms); and in semantics and automated reasoning.

UCAM-CL-TR-625

Steve Bishop, Matthew Fairbairn, Michael Norrish, Peter Sewell, Michael Smith, Keith Wansbrough:

TCP, UDP, and Sockets: rigorous and experimentally-validated behavioural specification Volume 2: The Specification

March 2005, 386 pages, PDF

Abstract: See Volume 1 (UCAM-CL-TR-624).

UCAM-CL-TR-626

Meng How Lim, Adam Greenhalgh, Julian Chesterfield, Jon Crowcroft:

Landmark Guided Forwarding: A hybrid approach for Ad Hoc routing

March 2005, 28 pages, PDF

Abstract: Wireless Ad Hoc network routing presents some extremely challenging research problems, trying to optimize parameters such as energy conservation vs connectivity and global optimization vs routing overhead scalability. In this paper we focus on the problems of maintaining network connectivity in the presence of node mobility whilst providing globally efficient and robust routing. The common approach among existing wireless Ad Hoc routing solutions is to establish a global optimal path between a source and a destination. We argue that establishing a globally optimal path is both unreliable and unsustainable as the network diameter, traffic volume, number of nodes all increase in the presence of moderate node mobility. To address this we propose Landmark Guided Forwarding (LGF), a protocol that provides a hybrid solution of topological and geographical routing algorithms. We demonstrate that LGF is adaptive to unstable connectivity and scalable to large networks. Our results indicate therefore that Landmark Guided Forwarding converges much faster, scales better and adapts well within a dynamic wireless Ad Hoc environment in comparison to existing solutions.

UCAM-CL-TR-627

Keith Vertanen:

Efficient computer interfaces using continuous gestures, language models, and speech

March 2005, 46 pages, PDF
MPhil thesis (Darwin College, July 2004)

Abstract: Despite advances in speech recognition technology, users of dictation systems still face a significant amount of work to correct errors made by the recognizer. The goal of this work is to investigate the use of a continuous gesture-based data entry interface to provide an efficient and fun way for users to correct recognition errors. Towards this goal, techniques are investigated which expand a recognizer's results to help cover recognition errors. Additionally, models are developed which utilize a speech recognizer's n-best list to build letter-based language models.

UCAM-CL-TR-628

Moritz Y. Becker:

A formal security policy for an NHS electronic health record service

March 2005, 81 pages, PDF

Abstract: The ongoing NHS project for the development of a UK-wide electronic health records service, also known as the 'Spine', raises many controversial issues and technical challenges concerning the security and confidentiality of patient-identifiable clinical data. As the system will need to be constantly adapted to comply with evolving legal requirements and guidelines, the Spine's authorisation policy should not be hard-coded into the system but rather be specified in a high-level, general-purpose, machine-enforceable policy language.

We describe a complete authorisation policy for the Spine and related services, written for the trust management system Cassandra, and comprising 375 formal rules. The policy is based on the NHS's Output-based Specification (OBS) document and deals with all requirements concerning access control of patient-identifiable data, including legitimate relationships, patients restricting access, authenticated express consent, third-party consent, and workgroup management.

UCAM-CL-TR-629

Meng How Lim, Adam Greenhalgh, Julian Chesterfield, Jon Crowcroft:

Hybrid routing: A pragmatic approach to mitigating position uncertainty in geo-routing

April 2005, 26 pages, PDF

Abstract: In recent years, research in wireless Ad Hoc routing seems to be moving towards the approach of position based forwarding. Amongst proposed algorithms, Greedy Perimeter Stateless Routing has gained recognition for guaranteed delivery with modest network overheads. Although this addresses the scaling limitations with topological routing, it has limited tolerance for position inaccuracy or stale state reported by a location service. Several researchers have demonstrated that the inaccuracy of the positional system could have a catastrophic effect on position based routing protocols. In this paper, we evaluate how the negative effects of position inaccuracy can be countered by extending position based forwarding with a combination of restrictive topological state, adaptive route advertisement and hybrid forwarding. Our results show that a hybrid of the position and topology approaches used in Landmark Guided Forwarding yields a high goodput and timely packet delivery, even with 200 meters of position error.

UCAM-CL-TR-630

Sergei P. Skorobogatov:

Semi-invasive attacks – A new approach to hardware security analysis

April 2005, 144 pages, PDF
PhD thesis (Darwin College, September 2004)

Abstract: Semiconductor chips are used today not only to control systems, but also to protect them against security threats. A continuous battle is waged between manufacturers who invent new security solutions, learning their lessons from previous mistakes, and the hacker community, constantly trying to break implemented protections. Some chip manufacturers do not pay enough attention to the proper design and testing of protection mechanisms. Even where they claim their products are highly secure, they do not guarantee this and do not take any responsibility if a device is compromised. In this situation, it is crucial for the design engineer to have a convenient and reliable method of testing secure chips.

This thesis presents a wide range of attacks on hardware security in microcontrollers and smartcards. This

includes already known non-invasive attacks, such as power analysis and glitching, and invasive attacks, such as reverse engineering and microprobing. A new class of attacks – semi-invasive attacks – is introduced. Like invasive attacks, they require depackaging the chip to get access to its surface. But the passivation layer remains intact, as these methods do not require electrical contact to internal lines. Semi-invasive attacks stand between non-invasive and invasive attacks. They represent a greater threat to hardware security, as they are almost as effective as invasive attacks but can be low-cost like non-invasive attacks.

This thesis' contribution includes practical fault-injection attacks to modify SRAM and EEPROM content, or change the state of any individual CMOS transistor on a chip. This leads to almost unlimited capabilities to control chip operation and circumvent protection mechanisms. A second contribution consist of experiments on data remanence, which show that it is feasible to extract information from powered-off SRAM and erased EPROM, EEPROM and Flash memory devices.

A brief introduction to copy protection in microcontrollers is given. Hardware security evaluation techniques using semi-invasive methods are introduced. They should help developers to make a proper selection of components according to the required level of security. Various defence technologies are discussed, from low-cost obscurity methods to new approaches in silicon design.

UCAM-CL-TR-631

Wenjun Hu, Jon Crowcroft:

MIRRORS: An integrated framework for capturing real world behaviour for models of ad hoc networks

April 2005, 16 pages, PDF

Abstract: The simulation models used in mobile ad hoc network research have been criticised for lack of realism. While credited with ease of understanding and implementation, they are often based on theoretical models, rather than real world observations. Criticisms have centred on radio propagation or mobility models.

In this work, we take an integrated approach to modelling the real world that underlies a mobile ad hoc network. While pointing out the correlations between the space, radio propagation and mobility models, we use mobility as a focal point to propose a new framework, MIRRORS, that captures real world behaviour. We give the formulation of a specific model within the framework and present simulation results that reflect topology properties of the networks synthesised. Compared with the existing models studied, our model better represent real world topology properties and presents a wider spectrum of variation in the metrics examined, due to the model encapsulating more detailed dynamics. While the common approach is to focus on performance evaluation of existing protocols using these models, we discuss protocol design opportunities across layers in view of the simulation results.

UCAM-CL-TR-632

R.I. Tucker, K. Spärck Jones:

Between shallow and deep: an experiment in automatic summarising

April 2005, 34 pages, PDF

Abstract: This paper describes an experiment in automatic summarising using a general-purpose strategy based on a compromise between shallow and deep processing. The method combines source text analysis into simple logical forms with the use of a semantic graph for representation and operations on the graph to identify summary content.

The graph is based on predications extracted from the logical forms, and the summary operations apply three criteria, namely importance, representativeness, and cohesiveness, in choosing node sets to form the content representation for the summary. This is used in different ways for output summaries. The paper presents the motivation for the strategy, details of the CLASP system, and the results of initial testing and evaluation on news material.

UCAM-CL-TR-633

Alex Ho, Steven Smith, Steven Hand:

On deadlock, livelock, and forward progress

May 2005, 8 pages, PDF

Abstract: Deadlock and livelock can happen at many different levels in a distributed system. We unify both around the concept of forward progress and standstill. We describe a framework capable of detecting the lack of forward progress in distributed systems. Our prototype can easily solve traditional deadlock problems where synchronization is via a customer network protocol; however, many interesting research challenges remain.

UCAM-CL-TR-634

Kasim Rehman:

Visualisation, interpretation and use of location-aware interfaces

May 2005, 159 pages, PDF
PhD thesis (St Catharine's College, November 2004)

Abstract: Ubiquitous Computing (Ubicomp), a term coined by Mark Weiser in the early 1990's, is about transparently equipping the physical environment and everyday objects in it with computational, sensing and networking abilities. In contrast with traditional desktop computing the "computer" moves into the background, unobtrusively supporting users in their everyday life.

One of the instantiations of Ubicomp is location-aware computing. Using location sensors, the "computer" reacts to changes in location of users and everyday objects. Location changes are used to infer user intent in order to give the user the most appropriate support for the task she is performing. Such support can consist of automatically providing information or configuring devices and applications deemed adequate for the inferred user task.

Experience with these applications has uncovered a number of usability problems that stem from the fact that the "computer" in this paradigm has become unidentifiable for the user. More specifically, these arise from lack of feedback from, loss of user control over, and the inability to provide a conceptual model of the "computer".

Starting from the proven premise that feedback is indispensable for smooth human-machine interaction, a system that uses Augmented Reality in order to visually provide information about the state of a location-aware environment and devices in it, is designed and implemented.

Augmented Reality (AR) as it is understood for the purpose of this research uses a see-through head-mounted display, trackers and 3-dimensional (3D) graphics in order to give users the illusion that 3-dimensional graphical objects specified and generated on a computer are actually located in the real world.

The system described in this thesis can be called a Graphical User Interface (GUI) for a physical environment. Properties of GUIs for desktop environments are used as a valuable resource in designing a software architecture that supports interactivity in a location-aware environment, understanding how users might conceptualise the "computer" and extracting design principles for visualisation in a Ubicomp environment.

Most importantly this research offers a solution to fundamental interaction problems in Ubicomp environments. In doing so this research presents the next step from reactive environments to interactive environments.

UCAM-CL-TR-635

John Daugman:

Results from 200 billion iris cross-comparisons

June 2005, 8 pages, PDF

Abstract: Statistical results are presented for biometric recognition of persons by their iris patterns, based on 200 billion cross-comparisons between different eyes. The database consisted of 632,500 iris images acquired in the Middle East, in a national border-crossing protection programme that uses the Daugman algorithms for iris recognition. A total of 152 different nationalities were represented in this database. The set of exhaustive cross-comparisons between all possible pairings of irises in the database shows that with reasonable acceptance thresholds, the False Match rate is less than 1 in 200 billion. Recommendations are given for the numerical decision threshold policy that would enable reliable identification performance on a national scale in the UK.

UCAM-CL-TR-636

Rana Ayman el Kaliouby:

Mind-reading machines: automated inference of complex mental states

July 2005, 185 pages, PDF
PhD thesis (Newnham College, March 2005)

Abstract: People express their mental states all the time, even when interacting with machines. These mental states shape the decisions that we make, govern how we communicate with others, and affect our performance. The ability to attribute mental states to others from their behaviour, and to use that knowledge to guide one's own actions and predict those of others is known as theory of mind or mind-reading.

The principal contribution of this dissertation is the real time inference of a wide range of mental states from head and facial displays in a video stream. In particular, the focus is on the inference of complex mental states: the affective and cognitive states of mind that are not part of the set of basic emotions. The automated mental state inference system is inspired by and draws on the fundamental role of mind-reading in communication and decision-making.

The dissertation describes the design, implementation and validation of a computational model of mind-reading. The design is based on the results of a number of experiments that I have undertaken to analyse the facial signals and dynamics of complex mental states. The resulting model is a multi-level probabilistic graphical model that represents the facial events in a raw video stream at different levels of spatial and temporal abstraction. Dynamic Bayesian Networks model observable head and facial displays, and corresponding hidden mental states over time.

The automated mind-reading system implements the model by combining top-down predictions of mental state models with bottom-up vision-based processing of the face. To support intelligent human-computer interaction, the system meets three important criteria.

These are: full automation so that no manual preprocessing or segmentation is required, real time execution, and the categorization of mental states early enough after their onset to ensure that the resulting knowledge is current and useful.

The system is evaluated in terms of recognition accuracy, generalization and real time performance for six broad classes of complex mental states—agreeing, concentrating, disagreeing, interested, thinking and unsure, on two different corpora. The system successfully classifies and generalizes to new examples of these classes with an accuracy and speed that are comparable to that of human recognition.

The research I present here significantly advances the nascent ability of machines to infer cognitive-affective mental states in real time from nonverbal expressions of people. By developing a real time system for the inference of a wide range of mental states beyond the basic emotions, I have widened the scope of human-computer interaction scenarios in which this technology can be integrated. This is an important step towards building socially and emotionally intelligent machines.

UCAM-CL-TR-637

Shishir Nagaraja, Ross Anderson:

The topology of covert conflict

July 2005, 15 pages, PDF

Abstract: Often an attacker tries to disconnect a network by destroying nodes or edges, while the defender counters using various resilience mechanisms. Examples include a music industry body attempting to close down a peer-to-peer file-sharing network; medics attempting to halt the spread of an infectious disease by selective vaccination; and a police agency trying to decapitate a terrorist organisation. Albert, Jeong and Barabási famously analysed the static case, and showed that vertex-order attacks are effective against scale-free networks. We extend this work to the dynamic case by developing a framework based on evolutionary game theory to explore the interaction of attack and defence strategies. We show, first, that naive defences don't work against vertex-order attack; second, that defences based on simple redundancy don't work much better, but that defences based on cliques work well; third, that attacks based on centrality work better against clique defences than vertex-order attacks do; and fourth, that defences based on complex strategies such as delegation plus clique resist centrality attacks better than simple clique defences. Our models thus build a bridge between network analysis and evolutionary game theory, and provide a framework for analysing defence and attack in networks where topology matters. They suggest definitions of efficiency of attack and defence, and may even explain the evolution of insurgent organisations from networks of cells to a more virtual leadership that facilitates operations rather than directing them. Finally, we draw some conclusions and present possible directions for future research.

UCAM-CL-TR-638

Piotr Zieliński:

Optimistic Generic Broadcast

July 2005, 22 pages, PDF

Abstract: We consider an asynchronous system with the Ω failure detector, and investigate the number of communication steps required by various broadcast protocols in runs in which the leader does not change. Atomic Broadcast, used for example in state machine replication, requires three communication steps. Optimistic Atomic Broadcast requires only two steps if all correct processes receive messages in the same order. Generic Broadcast requires two steps if no messages conflict. We present an algorithm that subsumes both of these approaches and guarantees two-step delivery if all conflicting messages are received in the same order, and three-step delivery otherwise. Internally, our protocol uses two new algorithms. First, a Consensus algorithm which decides in one communication step if all proposals are the same, and needs two steps otherwise. Second, a method that allows us to run infinitely many instances of a distributed algorithm, provided that only finitely many of them are different. We assume that fewer than a third of all processes are faulty ($n > 3f$).

UCAM-CL-TR-639

Chris Purcell, Tim Harris:

Non-blocking hashtables with open addressing

September 2005, 23 pages, PDF

Abstract: We present the first non-blocking hashtable based on open addressing that provides the following benefits: it combines good cache locality, accessing a single cacheline if there are no collisions, with short straight-line code; it needs no storage overhead for pointers and memory allocator schemes, having instead an overhead of two words per bucket; it does not need to periodically reorganise or replicate the table; and it does not need garbage collection, even with arbitrary-sized keys. Open problems include resizing the table and replacing, rather than erasing, entries. The result is a highly-concurrent set algorithm that approaches or outperforms the best externally-chained implementations we tested, with fixed memory costs and no need to select or fine-tune a garbage collector or locking strategy.

UCAM-CL-TR-640

Feng Hao, Ross Anderson, John Daugman:

Combining cryptography with biometrics effectively

July 2005, 17 pages, PDF

Abstract: We propose the first practical and secure way to integrate the iris biometric into cryptographic applications. A repeatable binary string, which we call a biometric key, is generated reliably from genuine iris codes. A well-known difficulty has been how to cope with the 10 to 20% of error bits within an iris code and derive an error-free key. To solve this problem, we carefully studied the error patterns within iris codes, and devised a two-layer error correction technique that combines Hadamard and Reed-Solomon codes. The key is generated from a subject's iris image with the aid of auxiliary error-correction data, which do not reveal the key, and can be saved in a tamper-resistant token such as a smart card. The reproduction of the key depends on two factors: the iris biometric and the token. The attacker has to procure both of them to compromise the key. We evaluated our technique using iris samples from 70 different eyes, with 10 samples from each eye. We found that an error-free key can be reproduced reliably from genuine iris codes with a 99.5% success rate. We can generate up to 140 bits of biometric key, more than enough for 128-bit AES. The extraction of a repeatable binary string from biometrics opens new possible applications, where a strong binding is required between a person and cryptographic operations. For example, it is possible to identify individuals without maintaining a central database of biometric templates, to which privacy objections might be raised.

UCAM-CL-TR-641

Ross Anderson, Mike Bond, Jolyon Clulow, Sergei Skorobogatov:

Cryptographic processors – a survey

August 2005, 19 pages, PDF

Abstract: Tamper-resistant cryptographic processors are becoming the standard way to enforce data-usage policies. Their history began with military cipher machines, and hardware security modules used to encrypt the PINs that bank customers use to authenticate themselves to ATMs. In both cases, the designers wanted to prevent abuse of data and key material should a device fall into the wrong hands. From these specialist beginnings, cryptoprocessors spread into devices such as prepayment electricity meters, and the vending machines that sell credit for them. In the 90s, tamper-resistant smartcards became integral to GSM mobile phone identification and to key management in pay-TV set-top boxes, while secure microcontrollers were used in remote key entry devices for cars. In the last five years, dedicated crypto chips have been embedded in devices from games console accessories to printer ink cartridges, to control product and accessory aftermarkets. The 'Trusted Computing' initiative will soon embed cryptoprocessors in PCs so that they can identify each other remotely.

This paper surveys the range of applications of tamper-resistant hardware, and the array of attack and defence mechanisms which have evolved in the tamper-resistance arms race.

UCAM-CL-TR-642

Gavin Bierman, Alisdair Wren:

First-class relationships in an object-oriented language

August 2005, 53 pages, PDF

Abstract: In this paper we investigate the addition of first-class relationships to a prototypical object-oriented programming language (a "middleweight" fragment of Java). We provide language-level constructs to declare relationships between classes and to manipulate relationship instances. We allow relationships to have attributes and provide a novel notion of relationship inheritance. We formalize our language giving both the type system and operational semantics and prove certain key safety properties.

UCAM-CL-TR-643

Nathan E. Dimmock:

Using trust and risk for access control in Global Computing

August 2005, 145 pages, PDF
PhD thesis (Jesus College, April 2005)

Abstract: Global Computing is a vision of a massively networked infrastructure supporting a large population of diverse but cooperating entities. Similar to ubiquitous computing, entities of global computing will operate in environments that are dynamic and unpredictable, requiring them to be capable of dealing with unexpected interactions and previously unknown principals using an unreliable infrastructure.

These properties will pose new security challenges that are not adequately addressed by existing security models and mechanisms. Traditionally privileges are statically encoded as security policy, and while rôle-based access control introduces a layer of abstraction between privilege and identity, rôles, privileges and context must still be known in advance of any interaction taking place.

Human society has developed the mechanism of trust to overcome initial suspicion and gradually evolve privileges. Trust successfully enables collaboration amongst human agents — a computational model of trust ought to be able to enable the same in computational agents. Existing research in this area has concentrated on developing trust management systems that permit the encoding of, and reasoning about, trust beliefs, but the relationship between these and privilege is still hard coded. These systems also omit any explicit reasoning about risk, and its relationship to privilege, nor do they permit the automated evolution of trust over time.

This thesis examines the relationship between trust, risk and privilege in an access control system. An outcome-based approach is taken to risk modelling, using explicit costs and benefits to model the relationship between risk and privilege. This is used to develop a novel model of access control — trust-based access control (TBAC) — firstly for the limited domain of collaboration between Personal Digital Assistants (PDAs), and later for more general global computing applications using the SECURE computational trust framework.

This general access control model is also used to extend an existing rôle-based access control system to explicitly reason about trust and risk. A further refinement is the incorporation of the economic theory of decision-making under uncertainty by expressing costs and benefits as utility, or preference-scaling, functions. It is then shown how Bayesian trust models can be used in the SECURE framework, and how these models enable a better abstraction to be obtained in the access control policy. It is also shown how the access control model can be used to take such decisions as whether the cost of seeking more information about a principal is justified by the risk associated with granting the privilege, and to determine whether a principal should respond to such requests upon receipt. The use of game theory to help in the construction of policies is also briefly considered.

Global computing has many applications, all of which require access control to prevent abuse by malicious principals. This thesis develops three in detail: an information sharing service for PDAs, an identity-based spam detector and a peer-to-peer collaborative spam detection network. Given the emerging nature of computational trust systems, in order to evaluate the effectiveness of the TBAC model, it was first necessary to develop an evaluation methodology. This takes the approach of a threat-based analysis, considering possible attacks at the component and system level, to ensure that components are correctly integrated, and system-level assumptions made by individual components are valid. Applying the methodology to the implementation of the TBAC model demonstrates its effectiveness in the scenarios chosen, with good promise for further, untested, scenarios.

UCAM-CL-TR-644

Paul Youn, Ben Adida, Mike Bond,
Jolyon Clulow, Jonathan Herzog,
Amerson Lin, Ronald L. Rivest,
Ross Anderson:

Robbing the bank with a theorem prover

August 2005, 26 pages, PDF

Abstract: We present the first methodology for analysis and automated detection of attacks on security application programming interfaces (security APIs) – the interfaces to hardware cryptographic services used by developers of critical security systems, such as banking applications. Taking a cue from previous work on the formal analysis of security protocols, we model APIs purely according to specifications, under the assumption of ideal encryption primitives. We use a theorem prover tool and adapt it to the security API context. We develop specific formalization and automation techniques that allow us to fully harness the power of a theorem prover. We show how, using these techniques, we were able to automatically re-discover all of the pure API attacks originally documented by Bond and Anderson against banking payment networks, since their discovery of this type of attack in 2000. We conclude with a note of encouragement: the complexity and unintuiveness of the modelled attacks make a very strong case for continued focus on automated formal analysis of cryptographic APIs.

UCAM-CL-TR-645

Frank Stajano:

RFID is X-ray vision

August 2005, 10 pages, PDF

Abstract: Making RFID tags as ubiquitous as barcodes will enable machines to see and recognize any tagged object in their vicinity, better than they ever could with the smartest image processing algorithms. This opens many opportunities for "sentient computing" applications.

However, in so far as this new capability has some of the properties of X-ray vision, it opens the door to abuses. To promote discussion, I won't elaborate on low level technological solutions; I shall instead discuss a simple security policy model that addresses most of the privacy issues. Playing devil's advocate, I shall also indicate why it is currently unlikely that consumers will enjoy the RFID privacy that some of them vociferously demand.

UCAM-CL-TR-647

Sam Staton:

An agent architecture for simulation of end-users in programming-like tasks

October 2005, 12 pages, PDF

Abstract: We present some motivation and technical details for a software simulation of an end-user performing programming-like tasks. The simulation uses an agent/agenda model by breaking tasks down into subgoals, based on work of A. Blackwell. This document was distributed at the CHI 2002 workshop on Cognitive Models of Programming-Like Processes.

UCAM-CL-TR-648

Moritz Y. Becker:

Cassandra: flexible trust management and its application to electronic health records

October 2005, 214 pages, PDF
PhD thesis (Trinity College, September 2005)

Abstract: The emergence of distributed applications operating on large-scale, heterogeneous and decentralised networks poses new and challenging problems of concern to society as a whole, in particular for data security, privacy and confidentiality. Trust management and authorisation policy languages have been proposed to address access control and authorisation in this context. Still, many key problems have remained unsolved. Existing systems are often not expressive enough, or are so expressive that access control becomes undecidable; their semantics is not formally specified; and they have not been shown to meet the requirements set by actual real-world applications.

This dissertation addresses these problems. We present Cassandra, a role-based language and system for expressing authorisation policy, and the results of a substantial case study, a policy for a national electronic health record (EHR) system, based on the requirements of the UK National Health Service's National Programme for Information Technology (NPfIT).

Cassandra policies are expressed in a language derived from Datalog with constraints. Cassandra supports credential-based authorisation (eg between administrative domains), and rules can refer to remote policies (for credential retrieval and trust negotiation). The expressiveness of the language (and its computational complexity) can be tuned by choosing an appropriate constraint domain. The language is small and has a formal semantics for both query evaluation and the access control engine.

There has been a lack of real-world examples of complex security policies: our NPfIT case study fills this gap. The resulting Cassandra policy (with 375 rules) demonstrates that the policy language is expressive enough for a real-world application. We thus demonstrate that a general-purpose trust management system can be designed to be highly flexible, expressive, formally founded and meet the complex requirements of real-world applications.

UCAM-CL-TR-649

Mark Grundland, Neil A. Dodgson:

The decolorize algorithm for contrast enhancing, color to grayscale conversion

October 2005, 15 pages, PDF

Abstract: We present a new contrast enhancing color to grayscale conversion algorithm which works in real-time. It incorporates novel techniques for image sampling and dimensionality reduction, sampling color differences by Gaussian pairing and analyzing color differences by predominant component analysis. In addition to its speed and simplicity, the algorithm has the advantages of continuous mapping, global consistency, and grayscale preservation, as well as predictable luminance, saturation, and hue ordering properties. We give an extensive range of examples and compare our method with other recently published algorithms.

UCAM-CL-TR-650

Rashid Mehmood, Jon Crowcroft:

Parallel iterative solution method for large sparse linear equation systems

October 2005, 22 pages, PDF

Abstract: Solving sparse systems of linear equations is at the heart of scientific computing. Large sparse systems often arise in science and engineering problems. One such problem we consider in this paper is the steady-state analysis of Continuous Time Markov Chains (CTMCs). CTMCs are a widely used formalism for the performance analysis of computer and communication systems. A large variety of useful performance measures can be derived from a CTMC via the computation of its steady-state probabilities. A CTMC may be represented by a set of states and a transition rate matrix containing state transition rates as coefficients, and can be analysed using probabilistic model checking. However, CTMC models for realistic systems are very large. We address this largeness problem in this paper, by considering parallelisation of symbolic methods. In particular, we consider Multi-Terminal Binary Decision Diagrams (MTBDDs) to store CTMCs, and,

using Jacobi iterative method, present a parallel method for the CTMC steady-state solution. Employing a 24-node processor bank, we report results of the sparse systems with over a billion equations and eighteen billion nonzeros.

UCAM-CL-TR-651

Rob Hague:

End-user programming in multiple languages

October 2005, 122 pages, PDF
PhD thesis (Fitzwilliam College, July 2004)

Abstract: Advances in user interface technology have removed the need for the majority of users to program, but they do not allow the automation of repetitive or indirect tasks. End-user programming facilities solve this problem without requiring users to learn and use a conventional programming language, but must be tailored to specific types of end user. In situations where the user population is particularly diverse, this presents a problem.

In addition, studies have shown that the performance of tasks based on the manipulation and interpretation of data depends on the way in which the data is represented. Different representations may facilitate different tasks, and there is not necessarily a single, optimal representation that is best for all tasks. In many cases, the choice of representation is also constrained by other factors, such as display size. It would be advantageous for an end-user programming system to provide multiple, interchangeable representations of programs.

This dissertation describes an architecture for providing end-user programming facilities in the networked home, a context with a diverse user population, and a wide variety of input and output devices. The Media Cubes language, a novel end-user programming language, is introduced as the context that lead to the development of the architecture. A framework for translation between languages via a common intermediate form is then described, with particular attention paid to the requirements of mappings between languages and the intermediate form. The implementation of Lingua Franca, a system realizing this framework in the given context, is described.

Finally, the system is evaluated by considering several end-user programming languages implemented within this system. It is concluded that translation between programming languages, via a common intermediate form, is viable for systems within a limited domain, and the wider applicability of the technique is discussed.

UCAM-CL-TR-652

Roongroj Nopsuwanchai:

Discriminative training methods and their applications to handwriting recognition

November 2005, 186 pages, PDF
PhD thesis (Downing College, August 2004)

Abstract: This thesis aims to improve the performance of handwriting recognition systems by introducing the use of discriminative training methods. Discriminative training methods use data from all competing classes when training the recogniser for each class. We develop discriminative training methods for two popular classifiers: Hidden Markov Models (HMMs) and a prototype-based classifier. At the expense of additional computations in the training process, discriminative training has demonstrated significant improvements in recognition accuracies from the classifiers that are not discriminatively optimised. Our studies focus on isolated character recognition problems with an emphasis on, but not limited to, off-line handwritten Thai characters.

The thesis is organised as followed. First, we develop an HMM-based classifier that employs a Maximum Mutual Information (MMI) discriminative training criterion. HMMs have an increasing number of applications to character recognition in which they are usually trained by Maximum Likelihood (ML) using the Baum-Welch algorithm. However, ML training does not take into account the data of other competing categories, and thus is considered non-discriminative. By contrast, MMI provides an alternative training method with the aim of maximising the mutual information between the data and their correct categories. One of our studies highlights the efficiency of MMI training that improves the recognition results from ML training, despite being applied to a highly constrained system (tied-mixture density HMMs). Various aspects of MMI training are investigated, including its optimisation algorithms and a set of optimised parameters that yields maximum discriminabilities.

Second, a system for Thai handwriting recognition based on HMMs and MMI training is introduced. In addition, novel feature extraction methods using block-based PCA and composite images are proposed and evaluated. A technique to improve generalisation of the MMI-trained systems and the use of N-best lists to efficiently compute the probabilities are described. By applying these techniques, the results from extensive experiments are compelling, showing up to 65% relative error reduction, compared to conventional ML training without the proposed features. The best results are comparable to those achieved by other high performance systems.

Finally, we focus on the Prototype-Based Minimum Error Classifier (PBMEC), which uses a discriminative Minimum Classification Error (MCE) training

method to generate the prototypes. MCE tries to minimise recognition errors during the training process using data from all classes. Several key findings are revealed, including the setting of smoothing parameters and a proposed clustering method that are more suitable for PBMEC than using the conventional methods. These studies reinforce the effectiveness of discriminative training and are essential as a foundation for its application to the more difficult problem of cursive handwriting recognition.

UCAM-CL-TR-653

Richard Clayton:

Anonymity and traceability in cyberspace

November 2005, 189 pages, PDF
PhD thesis (Darwin College, August 2005)

Abstract: Traceability is the ability to map events in cyberspace, particularly on the Internet, back to real-world instigators, often with a view to holding them accountable for their actions. Anonymity is present when traceability fails.

I examine how traceability on the Internet actually works, looking first at a classical approach from the late 1990s that emphasises the rôle of activity logging and reporting on the failures that are known to occur. Failures of traceability, with consequent unintentional anonymity, have continued as the technology has changed. I present an analysis that ascribes these failures to the mechanisms at the edge of the network being inherently inadequate for the burden that traceability places upon them. The underlying reason for this continuing failure is a lack of economic incentives for improvement. The lack of traceability at the edges is further illustrated by a new method of stealing another person's identity on an Ethernet Local Area Network that existing tools and procedures would entirely fail to detect.

Preserving activity logs is seen, especially by Governments, as essential for the traceability of illegal cyberspace activity. I present a new and efficient method of processing email server logs to detect machines sending bulk unsolicited email "spam" or email infected with "viruses". This creates a clear business purpose for creating logs, but the new detector is so effective that the logs can be discarded within days, which may hamper general traceability.

Preventing spam would be far better than tracing its origin or detecting its transmission. Many analyse spam in economic terms, and wish to levy a small charge for sending each email. I consider an oft-proposed approach using computational "proof-of-work" that is elegant and anonymity preserving. I show that, in a world of high profit margins and insecure end-user machines, it is impossible to find a payment level that stops the spam without affecting legitimate usage of email.

Finally, I consider a content-blocking system with a hybrid design that has been deployed by a UK Internet Service Provider to inhibit access to child pornography. I demonstrate that the two-level design can be circumvented at either level, that content providers can use the first level to attack the second, and that the selectivity of the first level can be used as an "oracle" to extract a list of the sites being blocked. Although many of these attacks can be countered, there is an underlying failure that cannot be fixed. The system's database holds details of the traceability of content, as viewed from a single location at a single time. However, a blocking system may be deployed at many sites and must track content as it moves in space and time; functions which traceability, as currently realized, cannot deliver.

UCAM-CL-TR-654

Matthew J. Parkinson:

Local reasoning for Java

November 2005, 120 pages, PDF
PhD thesis (Churchill College, August 2005)

Abstract: This thesis develops the local reasoning approach of separation logic for common forms of modularity such as abstract datatypes and objects. In particular, this thesis focuses on the modularity found in the Java programming language.

We begin by developing a formal semantics for a core imperative subset of Java, Middleweight Java (MJ), and then adapt separation logic to reason about this subset. However, a naive adaption of separation logic is unable to reason about encapsulation or inheritance: it provides no support for modularity.

First, we address the issue of encapsulation with the novel concept of an abstract predicate, which is the logical analogue of an abstract datatype. We demonstrate how this method can encapsulate state, and provide a mechanism for ownership transfer: the ability to transfer state safely between a module and its client. We also show how abstract predicates can be used to express the calling protocol of a class.

However, the encapsulation provided by abstract predicates is too restrictive for some applications. In particular, it cannot reason about multiple datatypes that have shared read-access to state, for example list iterators. To compensate, we alter the underlying model to allow the logic to express properties about read-only references to state. Additionally, we provide a model that allows both sharing and disjointness to be expressed directly in the logic.

Finally, we address the second modularity issue: inheritance. We do this by extending the concept of abstract predicates to abstract predicate families. This extension allows a predicate to have multiple definitions that are indexed by class, which allows subclasses to have a different internal representation while remaining behavioural subtypes. We demonstrate the usefulness of this concept by verifying a use of the visitor design pattern.

UCAM-CL-TR-655

Karen Spärck Jones:

Wearing proper combinations

November 2005, 27 pages, PDF

Abstract: This paper discusses the proper treatment of multiple indexing fields, representations, or streams, in document retrieval. Previous experiments by Robertson and his colleagues have shown that, with a widely used type of term weighting and fields that share keys, document scores should be computed using term frequencies over fields rather than by combining field scores. Here I examine a wide range of document and query indexing situations, and consider their implications for this approach to document scoring.

UCAM-CL-TR-656

Pablo Vidales:

Seamless mobility in 4G systems

November 2005, 141 pages, PDF
PhD thesis (Girton College, May 2005)

Abstract: The proliferation of radio access technologies, wireless networking devices, and mobile services has encouraged intensive nomadic computing activity. When travelling, mobile users experience connectivity disturbances, particularly when they handoff between two access points that belong to the same wireless network and when they change from one access technology to another. Nowadays, an average mobile user might connect to many different wireless networks in the course of a day to obtain diverse services, whilst demanding transparent operation. Current protocols offer portability and transparent mobility. However, they fail to cope with huge delays caused by different link-layer characteristics when roaming between independent disparate networks. In this dissertation, I address this deficiency by introducing and evaluating practical methods and solutions that minimise connection disruptions and support transparent mobility in future communication systems.

UCAM-CL-TR-657

Hyun-Jin Choi:

Security protocol design by composition

January 2006, 155 pages, PDF
PhD thesis (Churchill College, December 2004)

Abstract: The aim of this research is to present a new methodology for the systematic design of compound protocols from their parts. Some security properties can be made accumulative, i.e. can be put together without interfering with one another, by carefully selecting the mechanisms which implement them. Among them are authentication, secrecy and non-repudiation. Based on this observation, a set of accumulative protocol mechanisms called protocol primitives are proposed and their correctness is verified. These protocol primitives are obtained from common mechanisms found in many security protocols such as challenge and response. They have been carefully designed not to interfere with each other. This feature makes them flexible building blocks in the proposed methodology. Equipped with these protocol primitives, a scheme for the systematic construction of a complicated protocol from simple protocol primitives is presented, namely, design by composition. This design scheme allows the combination of several simple protocol parts into a complicated protocol without destroying the security properties established by each independent part. In other words, the composition framework permits the specification of a complex protocol to be decomposed into the specifications of simpler components, and thus makes the design and verification of the protocol easier to handle. Benefits of this approach are similar to those gained when using a modular approach to software development.

The applicability and practicality of the proposed methodology are validated through many design examples of protocols found in many different environments and with various initial assumptions. The method is not aimed to cover all existent design issues, but a reasonable range of protocols is addressed.

UCAM-CL-TR-658

Carsten Moenning:

Intrinsic point-based surface processing

January 2006, 166 pages, PDF
PhD thesis (Queens' College, January 2005)

Abstract: The need for the processing of surface geometry represents an ubiquitous problem in computer graphics and related disciplines. It arises in numerous important applications such as computer-aided design, reverse engineering, rapid prototyping, medical imaging, cultural heritage acquisition and preservation, video gaming and the movie industry. Existing surface processing techniques predominantly follow an extrinsic approach using combinatorial mesh data structures in the embedding Euclidean space to represent, manipulate and visualise the surfaces. This thesis advocates, firstly, the intrinsic processing of surfaces, i.e. processing directly across the surface rather than in its embedding space. Secondly, it continues the trend towards the use of point primitives for the processing and representation of surfaces.

The discussion starts with the design of an intrinsic point sampling algorithm template for surfaces. This is followed by the presentation of a module library of template instantiations for surfaces in triangular mesh or point cloud form. The latter is at the heart of the intrinsic meshless surface simplification algorithm also put forward. This is followed by the introduction of intrinsic meshless surface subdivision, the first intrinsic meshless surface subdivision scheme and a new method for the computation of geodesic centroids on manifolds. The meshless subdivision scheme uses an intrinsic neighbourhood concept for point-sampled geometry also presented in this thesis. Its main contributions can therefore be summarised as follows:

– An intrinsic neighbourhood concept for point-sampled geometry.

– An intrinsic surface sampling algorithm template with sampling density guarantee.

– A modular library of template instantiations for the sampling of planar domains and surfaces in triangular mesh or point cloud form.

– A new method for the computation of geodesic centroids on manifolds.

– An intrinsic meshless surface simplification algorithm.

– The introduction of the notion of intrinsic meshless surface subdivision.

– The first intrinsic meshless surface subdivision scheme.

The overall result is a set of algorithms for the processing of point-sampled geometry centering around a generic sampling template for surfaces in the most widely-used forms of representation. The intrinsic nature of these point-based algorithms helps to overcome limitations associated with the more traditional extrinsic, mesh-based processing of surfaces when dealing with highly complex point-sampled geometry as is typically encountered today.

UCAM-CL-TR-659

Viktor Vafeiadis, Maurice Herlihy,
Tony Hoare, Marc Shapiro:

A safety proof of a lazy concurrent list-based set implementation

January 2006, 19 pages, PDF

Abstract: We prove the safety of a practical concurrent list-based implementation due to Heller et al. It exposes an interface of an integer set with methods contains, add, and remove. The implementation uses a combination of fine-grain locking, optimistic and lazy synchronisation. Our proofs are hand-crafted. They use rely-guarantee reasoning and thereby illustrate its power and applicability, as well as some of its limitations. For each method, we identify the linearisation point, and establish its validity. Hence we show that the methods are safe, linearisable and implement a high-level specification. This report is a companion document to our PPoPP 2006 paper entitled "Proving correctness of highly-concurrent linearisable objects".

UCAM-CL-TR-660

Jeremy Singer:

Static program analysis based on virtual register renaming

February 2006, 183 pages, PDF
PhD thesis (Christ's College, March 2005)

Abstract: Static single assignment form (SSA) is a popular program intermediate representation (IR) for static analysis. SSA programs differ from equivalent control flow graph (CFG) programs only in the names of virtual registers, which are systematically transformed to comply with the naming convention of SSA. Static single information form (SSI) is a recently proposed extension of SSA that enforces a greater degree of systematic virtual register renaming than SSA. This dissertation develops the principles, properties, and practice of SSI construction and data flow analysis. Further, it shows that SSA and SSI are two members of a larger family of related IRs, which are termed virtual register renaming schemes (VRRSs). SSA and SSI analyses can be generalized to operate on any VRRS family member. Analysis properties such as accuracy and efficiency depend on the underlying VRRS.

This dissertation makes four significant contributions to the field of static analysis research.

First, it develops the SSI representation. Although SSI was introduced five years ago, it has not yet received widespread recognition as an interesting IR in its own right. This dissertation presents a new SSI definition and an optimistic construction algorithm. It also sets SSI in context among the broad range of IRs for static analysis.

Second, it demonstrates how to reformulate existing data flow analyses using new sparse SSI-based techniques. Examples include liveness analysis, sparse type inference and program slicing. It presents algorithms, together with empirical results of these algorithms when implemented within a research compiler framework.

Third, it provides the only major comparative evaluation of the merits of SSI for data flow analysis. Several qualitative and quantitative studies in this dissertation compare SSI with other similar IRs.

Last, it identifies the family of VRRSs, which are all CFGs with different virtual register naming conventions. Many extant IRs are classified as VRRSs. Several new IRs are presented, based on a consideration of previously unspecified members of the VRRS family. General analyses can operate on any family member. The required level of accuracy or efficiency can be selected by working in terms of the appropriate family member.

UCAM-CL-TR-661

Anna Ritchie:

Compatible RMRS representations from RASP and the ERG

March 2006, 41 pages, PDF

Abstract: Various applications could potentially benefit from the integration of deep and shallow processing techniques. A universal representation, compatible between deep and shallow parsers, would enable such integration, allowing the advantages of both to be combined. This paper describes efforts to make RMRS such a representation. This work was done as part of DeepThought, funded under the 5th Framework Program of the European Commission (contract reference IST-2001-37836).

UCAM-CL-TR-662

Ted Briscoe:

An introduction to tag sequence grammars and the RASP system parser

March 2006, 30 pages, PDF

Abstract: This report describes the tag sequence grammars released as part of the Robust Accurate Statistical Parsing (RASP) system. It is intended to help users of RASP understand the linguistic and engineering rationale behind the grammars and prepare them to customise the system for their application. It also contains a fairly exhaustive list of references to extant work utilising the RASP parser.

UCAM-CL-TR-663

Richard Bergmair:

Syntax-driven analysis of context-free languages with respect to fuzzy relational semantics

March 2006, 49 pages, PDF

Abstract: A grammatical framework is presented that augments context-free production rules with semantic production rules that rely on fuzzy relations as representations of fuzzy natural language concepts. It is shown how the well-known technique of syntax-driven semantic analysis can be used to infer from an expression in a language defined in such a semantically augmented grammar a weak ordering on the possible worlds it describes. Considering the application of natural language query processing, we show how to order elements in the domain of a relational database scheme according to the degree to which they fulfill the intuition behind a given natural language statement like "Carol lives in a small city near San Francisco".

UCAM-CL-TR-664

Alan F. Blackwell:

Designing knowledge: An interdisciplinary experiment in research infrastructure for shared description

April 2006, 18 pages, PDF

Abstract: The report presents the experimental development, evaluation and refinement of a method for doing adventurous design work, in contexts where academics must work in collaboration with corporate and public policy strategists and researchers. The intention has been to do applied social science, in which a reflective research process has resulted in a "new social form", as expressed in the title of the research grant that funded the project. The objective in doing so is not simply to produce new theories, or to enjoy interdisciplinary encounters (although both of those have been side effects of this work). My purpose in doing the work and writing this report is purely instrumental – working as a technologist among social scientists, the outcome described in this report is intended for adoption as a kind of social technology. I have given this product a name: the "Blackwell-Leach Process" for interdisciplinary design. The Blackwell-Leach process has since been applied and proven useful in several novel situations, and I believe is now sufficiently mature to justify publication of the reports that describe both the process and its development.

UCAM-CL-TR-665

Huiyun Li:

Security evaluation at design time for cryptographic hardware

April 2006, 81 pages, PDF
PhD thesis (Trinity Hall, December 2005)

Abstract: Consumer security devices are becoming ubiquitous, from pay-TV through mobile phones, PDA, prepayment gas meters to smart cards. There are many ongoing research efforts to keep these devices secure from opponents who try to retrieve key information by observation or manipulation of the chip's components. In common industrial practise, it is after the chip has been manufactured that security evaluation is performed. Due to design time oversights, however, weaknesses are often revealed in fabricated chips. Furthermore, post manufacture security evaluation is time consuming, error prone and very expensive. This evokes the need of "design time security evaluation" techniques in order to identify avoidable mistakes in design.

This thesis proposes a set of "design time security evaluation" methodologies covering the well-known non-invasive side-channel analysis attacks, such as power analysis and electromagnetic analysis attacks. The thesis also covers the recently published semi-invasive optical fault injection attacks. These security evaluation technologies examine the system under test by reproducing attacks through simulation and observing its subsequent response.

The proposed "design time security evaluation" methodologies can be easily implemented into the standard integrated circuit design flow, requiring only commonly used EDA tools. So it adds little non-recurrent engineering (NRE) cost to the chip design but helps identify the security weaknesses at an early stage, avoids costly silicon re-spins, and helps succeed in industrial evaluation for faster time-to-market.

UCAM-CL-TR-666

Mike Bond, George Danezis:

A pact with the Devil

June 2006, 14 pages, PDF

Abstract: We study malware propagation strategies which exploit not the incompetence or naivety of users, but instead their own greed, malice and short-sightedness. We demonstrate that interactive propagation strategies, for example bribery and blackmail of computer users, are effective mechanisms for malware to survive and entrench, and present an example employing these techniques. We argue that in terms of propagation, there exists a continuum between legitimate applications and pure malware, rather than a quantised scale.

UCAM-CL-TR-667

Piotr Zieliński:

Minimizing latency of agreement protocols

June 2006, 239 pages, PDF
PhD thesis (Trinity Hall, September 2005)

Abstract: Maintaining consistency of fault-tolerant distributed systems is notoriously difficult to achieve. It often requires non-trivial agreement abstractions, such as Consensus, Atomic Broadcast, or Atomic Commitment. This thesis investigates implementations of such abstractions in the asynchronous model, extended with unreliable failure detectors or eventual synchrony. The main objective is to develop protocols that minimize the number of communication steps required in failure-free scenarios but remain correct if failures occur. For several agreement problems and their numerous variants, this thesis presents such low-latency algorithms and lower-bound theorems proving their optimality.

The observation that many agreement protocols share the same round-based structure helps to cope with a large number of agreement problems in a uniform way. One of the main contributions of this thesis is "Optimistically Terminating Consensus" (OTC) – a new lightweight agreement abstraction that formalizes the notion of a round. It is used to provide simple modular solutions to a large variety of agreement problems, including Consensus, Atomic Commitment, and Interactive Consistency. The OTC abstraction tolerates malicious participants and has no latency overhead; agreement protocols constructed in the OTC framework require no more communication steps than their ad-hoc counterparts.

The attractiveness of this approach lies in the fact that the correctness of OTC algorithms can be tested automatically. A theory developed in this thesis allows us to quickly evaluate OTC algorithm candidates without the time-consuming examination of their entire state space. This technique is then used to scan the space of possible solutions in order to automatically discover new low-latency OTC algorithms. From these, one can now easily obtain new implementations of Consensus and similar agreement problems such as Atomic Commitment or Interactive Consistency.

Because of its continuous nature, Atomic Broadcast is considered separately from other agreement abstractions. I first show that no algorithm can guarantee a latency of less than three communication steps in all failure-free scenarios. Then, I present new Atomic Broadcast algorithms that achieve the two-step latency in some special cases, while still guaranteeing three steps for other failure-free scenarios. The special cases considered here are: Optimistic Atomic Broadcast, (Optimistic) Generic Broadcast, and closed-group Atomic Broadcast. For each of these, I present an appropriate algorithm and prove its latency to be optimal.

UCAM-CL-TR-668

Piotr Zieliński:

Optimistically Terminating Consensus

June 2006, 35 pages, PDF

Abstract: Optimistically Terminating Consensus (OTC) is a variant of Consensus that decides if all correct processes propose the same value. It is surprisingly easy to implement: processes broadcast their proposals and decide if sufficiently many processes report the same proposal. This paper shows an OTC-based framework which can reconstruct all major asynchronous Consensus algorithms, even in Byzantine settings, with no overhead in latency or the required number of processes. This result does not only deepen our understanding of Consensus, but also reduces the problem of designing new, modular distributed agreement protocols to choosing the parameters of OTC.

UCAM-CL-TR-669

David M. Eyers:

Active privilege management for distributed access control systems

June 2006, 222 pages, PDF
PhD thesis (King's College, June 2005)

Abstract: The last decade has seen the explosive uptake of technologies to support true Internet-scale distributed systems, many of which will require security.

The policy dictating authorisation and privilege restriction should be decoupled from the services being protected: (1) policy can be given its own independent language syntax and semantics, hopefully in an application independent way; (2) policy becomes portable – it can be stored away from the services it protects; and (3) the evolution of policy can be effected dynamically.

Management of dynamic privileges in wide-area distributed systems is a challenging problem. Supporting fast credential revocation is a simple example of dynamic privilege management. More complex examples include policies that are sensitive to the current state of a principal, such as dynamic separation of duties.

The Open Architecture for Secure Interworking Services (OASIS), an expressive distributed role-based access control system, is traced to the development of the Clinical and Biomedical Computing Limited (CBCL) OASIS implementation. Two OASIS deployments are discussed – an Electronic Health Record framework, and an inter-organisational distributed courseware system.

The Event-based Distributed Scalable Authorisation Control architecture for the 21st century (EDSAC21, or just EDSAC) is then presented along with its four design layers. It builds on OASIS, adding support for the collaborative enforcement of distributed dynamic constraints, and incorporating publish/subscribe messaging to allow scalable and flexible deployment. The OASIS policy language is extended to support delegation, dynamic separation of duties, and obligation policies.

An EDSAC prototype is examined. We show that our architecture is ideal for experiments performed into location-aware access control. We then demonstrate how event-based features specific to EDSAC facilitate integration of an ad hoc workflow monitor into an access control system.

The EDSAC architecture is powerful, flexible and extensible. It is intended to have widespread applicability as the basis for designing next-generation security middleware and implementing distributed, dynamic privilege management.

UCAM-CL-TR-670

Sarah Thompson:

On the application of program analysis and transformation to high reliability hardware

July 2006, 215 pages, PDF
PhD thesis (St Edmund's College, April 2006)

Abstract: Safety- and mission-critical systems must be both correct and reliable. Electronic systems must behave as intended and, where possible, do so at the first attempt – the fabrication costs of modern VLSI devices are such that the iterative design/code/test methodology endemic to the software world is not financially feasible. In aerospace applications it is also essential to establish that systems will, with known probability, remain operational for extended periods, despite being exposed to very low or very high temperatures, high radiation, large G-forces, hard vacuum and severe vibration.

Hardware designers have long understood the advantages of formal mathematical techniques. Notably, model checking and automated theorem proving both gained acceptance within the electronic design community at an early stage, though more recently the research focus in validation and verification has drifted toward software. As a consequence, the newest and most powerful techniques have not been significantly applied to hardware; this work seeks to make a modest contribution toward redressing the imbalance.

An abstract interpretation-based formalism is introduced, transitional logic, that supports formal reasoning about dynamic behaviour of combinational asynchronous circuits. The behaviour of majority voting circuits with respect to single-event transients is analysed, demonstrating that such circuits are not SET-immune. This result is generalised to show that SET immunity is impossible for all delay-insensitive circuits.

An experimental hardware partial evaluator, HarPE, is used to demonstrate the 1st Futamura projection in hardware – a small CPU is specialised with respect to a ROM image, yielding results that are equivalent to compiling the program into hardware. HarPE is then used alongside an experimental non-clausal SAT solver to implement an automated transformation system that is capable of repairing FPGAs that have suffered cosmic ray damage. This approach is extended to support automated configuration, dynamic testing and dynamic error recovery of reconfigurable spacecraft wiring harnesses.

UCAM-CL-TR-671

Piotr Zieliński:

Low-latency Atomic Broadcast in the presence of contention

July 2006, 23 pages, PDF

Abstract: The Atomic Broadcast algorithm described in this paper can deliver messages in two communication steps, even if multiple processes broadcast at the same time. It tags all broadcast messages with the local real time, and delivers all messages in order of these timestamps. The Ω-elected leader simulates processes it suspects to have crashed (◇S). For fault-tolerance, it uses a new cheap Generic Broadcast algorithm that requires only a majority of correct processes ($n > 2f$) and, in failure-free runs, delivers all non-conflicting messages in two steps. The main algorithm satisfies several new lower bounds, which are proved in this paper.

UCAM-CL-TR-672

Calicrates Policroniades-Borraz:

Decomposing file data into discernible items

August 2006, 230 pages, PDF
PhD thesis (Hughes Hall, December 2005)

Abstract: The development of the different persistent data models shows a constant pattern: the higher the level of abstraction a storage system exposes the greater the payoff for programmers. The file API offers a simple storage model that is agnostic of any structure or data types in file contents. As a result, developers employ substantial programming effort in writing persistent code. At the other extreme, orthogonally persistent programming languages reduce the impedance mismatch between the volatile and the persistent data spaces by exposing persistent data as conventional programming objects. Consequently, developers spend considerably less effort in developing persistent code.

This dissertation addresses the lack of ability in the file API to exploit the advantages of gaining access to the logical composition of file content. It argues that the trade-off between efficiency and ease of programmability of persistent code in the context of the file API is unbalanced. Accordingly, in this dissertation I present and evaluate two practical strategies to disclose structure and type in file data.

First, I investigate to what extent it is possible to identify specific portions of file content in diverse data sets through the implementation and evaluation of techniques for data redundancy detection. This study is interesting not only because it characterises redundancy levels in storage systems content, but also because redundant portions of data at a sub-file level can be an indication of internal file data structure. Although these techniques have been used by previous work, my analysis of data redundancy is the first that makes an in-depth comparison of them and highlights the trade-offs in their employment.

Second, I introduce a novel storage system API, called Datom, that departs from the view of file content as a monolithic object. Through a minimal set of commonly-used abstract data types, it discloses a judicious degree of structure and type in the logical composition of files and makes the data access semantics of applications explicit. The design of the Datom API weighs the addition of advanced functionality and the overheads introduced by their employment, taking into account the requirements of the target application domain. The implementation of the Datom API is evaluated according to different criteria such as usability, impact at the source-code level, and performance. The experimental results demonstrate that the Datom API reduces work-effort and improves software quality by providing a storage interface based on high-level abstractions.

UCAM-CL-TR-673

Judita Preiss:

Probabilistic word sense disambiguation
Analysis and techniques for combining knowledge sources

August 2006, 108 pages, PDF
PhD thesis (Trinity College, July 2005)

Abstract: This thesis shows that probabilistic word sense disambiguation systems based on established statistical methods are strong competitors to current state-of-the-art word sense disambiguation (WSD) systems.

We begin with a survey of approaches to WSD, and examine their performance in the systems submitted to the SENSEVAL-2 WSD evaluation exercise. We discuss existing resources for WSD, and investigate the amount of training data needed for effective supervised WSD.

We then present the design of a new probabilistic WSD system. The main feature of the design is that it combines multiple probabilistic modules using both Dempster-Shafer theory and Bayes Rule. Additionally, the use of Lidstone's smoothing provides a uniform mechanism for weighting modules based on their accuracy, removing the need for an additional weighting scheme.

Lastly, we evaluate our probabilistic WSD system using traditional evaluation methods, and introduce a novel task-based approach. When evaluated on the gold standard used in the SENSEVAL-2 competition, the performance of our system lies between the first and second ranked WSD system submitted to the English all words task.

Task-based evaluations are becoming more popular in natural language processing, being an absolute measure of a system's performance on a given task. We present a new evaluation method based on subcategorization frame acquisition. Experiments with our probabilistic WSD system give an extremely high correlation between subcategorization frame acquisition performance and WSD performance, thus demonstrating

the suitability of SCF acquisition as a WSD evaluation task.

UCAM-CL-TR-674

Meng How Lim:

Landmark Guided Forwarding

October 2006, 109 pages, PDF
PhD thesis (St Catharine's College, September 2006)

Abstract: Wireless mobile ad hoc network routing presents some extremely challenging research problems. While primarily trying to provide connectivity, algorithms may also be designed to minimise resource consumption such as power, or to trade off global optimisation against the routing protocol overheads. In this thesis, we focus on the problems of maintaining network connectivity in the presence of node mobility whilst providing a balance between global efficiency and robustness. The common design goal among existing wireless ad hoc routing solutions is to search for an optimal topological path between a source and a destination for some shortest path metric. We argue that the goal of establishing an end to end globally optimal path is unsustainable as the network diameter, traffic volume and number of nodes all increase in the presence of moderate node mobility.

Some researchers have proposed using geographic position-based forwarding, rather than a topological-based approach. In position-based forwarding, besides knowing about its own geographic location, every node also acquires the geographic position of its surrounding neighbours. Packet delivery in general is achieved by first learning the destination position from a location service. This is followed by addressing the packet with the destination position before forwarding the packet on to a neighbour that, amongst all other neighbours, is geographically nearest to the destination. It is clear that in the ad hoc scenario, forwarding only by geodesic position could result in situations that prevent the packet from advancing further. To resolve this, some researchers propose improving delivery guarantees by routing the packet along a planar graph constructed from a Gabriel (GG) or a Relative Neighbour Graph (RNG). This approach however has been shown to fail frequently when position information is inherently inaccurate, or neighbourhood state is stale, such as is the case in many plausible deployment scenarios, e.g. due to relative mobility rates being higher than location service update frequency.

We propose Landmark Guided Forwarding (LGF), an algorithm that harnesses the strengths of both topological and geographical routing algorithms. LGF is a hybrid scheme that leverages the scaling property of the geographic approach while using local topology knowledge to mitigate location uncertainty. We demonstrate through extensive simulations that LGF is suited both to situations where there are high mobility rates, and deployment when there is inherently less accurate position data. Our results show that Landmark Guided Forwarding converges faster, scales better and is more flexible in a range of plausible mobility scenarios than representative protocols from the leading classes of existing solutions, namely GPSR, AODV and DSDV.

UCAM-CL-TR-675

Paula J. Buttery:

Computational models for first language acquisition

November 2006, 176 pages, PDF
PhD thesis (Churchill College, March 2006)

Abstract: This work investigates a computational model of first language acquisition; the Categorial Grammar Learner or CGL. The model builds on the work of Villavicenio, who created a parametric Categorial Grammar learner that organises its parameters into an inheritance hierarchy, and also on the work of Buszkowski and Kanazawa, who demonstrated the learnability of a k-valued Classic Categorial Grammar (which uses only the rules of function application) from strings. The CGL is able to learn a k-valued General Categorial Grammar (which uses the rules of function application, function composition and Generalised Weak Permutation). The novel concept of Sentence Objects (simple strings, augmented strings, unlabelled structures and functor-argument structures) are presented as potential points from which learning may commence. Augmented strings (which are strings augmented with some basic syntactic information) are suggested as a sensible input to the CGL as they are cognitively plausible objects and have greater information content than strings alone. Building on the work of Siskind, a method for constructing augmented strings from unordered logic forms is detailed and it is suggested that augmented strings are simply a representation of the constraints placed on the space of possible parses due to a strings associated semantic content. The CGL makes crucial use of a statistical Memory Module (constructed from a Type Memory and Word Order Memory) that is used to both constrain hypotheses and handle data which is noisy or parametrically ambiguous. A consequence of the Memory Module is that the CGL learns in an incremental fashion. This echoes real child learning as documented in Browns Stages of Language Development and also as alluded to by an included corpus study of child speech. Furthermore, the CGL learns faster when initially presented with simpler linguistic data; a further corpus study of child-directed speech suggests that this echos the input provided to children. The CGL is demonstrated to learn from real data. It is evaluated against previous parametric learners (the Triggering Learning Algorithm of Gibson and Wexler and the Structural Triggers Learner of Fodor and Sakas) and is found to be more efficient.

UCAM-CL-TR-676

R.J. Gibbens, Y. Saacti:

Road traffic analysis using MIDAS data: journey time prediction

December 2006, 35 pages, PDF
Department for Transport Horizons Research Programme "Investigating the handling of large transport related datasets" (project number H05-217)

Abstract: The project described in this report was undertaken within the Department for Transport's second call for proposals in the Horizons research programme under the theme of "Investigating the handling of large transport related datasets". The project looked at the variability of journey times across days in three day categories: Mondays, midweek days and Fridays. Two estimators using real-time data were considered: a simple-to-implement regression-based method and a more computationally demanding k-nearest neighbour method. Our example scenario of UK data was taken from the M25 London orbital motorway during 2003 and the results compared in terms of the root-mean-square prediction error. It was found that where the variability was greatest (typically during the rush hours periods or periods of flow breakdowns) the regression and nearest neighbour estimators reduced the prediction error substantially compared with a naive estimator constructed from the historical mean journey time. Only as the lag between the decision time and the journey start time increased to beyond around 2 hours did the potential to improve upon the historical mean estimator diminish. Thus, there is considerable scope for prediction methods combined with access to real-time data to improve the accuracy in journey time estimates. In so doing, they reduce the uncertainty in estimating the generalized cost of travel. The regression-based prediction estimator has a particularly low computational overhead, in contrast to the nearest neighbour estimator, which makes it entirely suitable for an online implementation. Finally, the project demonstrates both the value of preserving historical archives of transport related datasets as well as provision of access to real-time measurements.

UCAM-CL-TR-677

Eiko Yoneki:

ECCO: Data centric asynchronous communication

December 2006, 210 pages, PDF
PhD thesis (Lucy Cavendish College, September 2006)

Abstract: This dissertation deals with data centric networking in distributed systems, which relies on content addressing instead of host addressing for participating nodes, thus providing network independence for applications. Publish/subscribe asynchronous group communication realises the vision of data centric networking that is particularly important for networks supporting mobile clients over heterogeneous wireless networks. In such networks, client applications prefer to receive specific data and require selective data dissemination. Underlying mechanisms such as asynchronous message passing, distributed message filtering and query/subscription management are essential. Furthermore, recent progress in wireless sensor networks brought a new dimension of data processing in ubiquitous computing, where the sensors are used to gather high volumes of different data types and to feed them as contexts to a wide range of applications.

Particular emphasis has been placed on fundamental design of event representation. Besides existing event attributes, event order, and continuous context information such as time or geographic location can be incorporated within an event description. Data representation of event and query will be even more important in future ubiquitous computing, where events flow over heterogeneous networks. This dissertation presents a multidimensional event representation (i.e., Hypercube structure in RTree) for efficient indexing, filtering, matching, and scalability in publish/subscribe systems. The hypercube event with a typed content-based publish/subscribe system for wide-area networks is demonstrated for improving the event filtering process.

As a primary focus, this dissertation investigates a structureless, asynchronous group communication over wireless ad hoc networks named 'ECCO Pervasive Publish/Subscribe' (ECCO-PPS). ECCO-PPS uses context-adaptive controlled flooding, which takes a cross-layer approach between middleware and network layers and provides a content-based publish/subscribe paradigm. Traditionally events have been payload data within network layer components; the network layer never touches the data contents. However, application data have more influence on data dissemination in ubiquitous computing scenarios.

The state information of the local node may be the event forwarding trigger. Thus, the model of publish/subscribe must become more symmetric, with events being disseminated based on rules and conditions defined by the events themselves. The event can thus choose the destinations instead of relying on the potential receivers' decision. The publish/subscribe system offers a data centric approach, where the destination address is not described with any explicit network address. The symmetric publish/subscribe paradigm brings another level to the data-centric paradigm, leading to a fundamental change in functionality at the network level of asynchronous group communication and membership maintenance.

To add an additional dimension of event process-

ing in global computing, It is important to understand event aggregation, filtering and correlation. Temporal ordering of events is essential for event correlation over distributed systems. This dissertation introduces generic composite event semantics with interval-based semantics for event detection. This precisely defines complex timing constraints among correlated event instances.

In conclusion, this dissertation provides advanced data-centric asynchronous communication, which provides efficiency, reliability, and robustness, while adapting to the underlying network environments.

UCAM-CL-TR-678

Andrew D. Twigg:

Compact forbidden-set routing

December 2006, 115 pages, PDF
PhD thesis (King's College, June 2006)

Abstract: We study the compact forbidden-set routing problem. We describe the first compact forbidden-set routing schemes that do not suffer from non-convergence problems often associated with Bellman-Ford iterative schemes such as the interdomain routing protocol, BGP. For degree-d n-node graphs of treewidth t, our schemes use space $O(t^2\ d\ \mathrm{polylog}(n))$ bits per node; a trivial scheme uses $O(n^2)$ and routing trees use $\Omega(n)$ per node (these results have since been improved and extended – see [Courcelle, Twigg, Compact forbidden-set routing, 24th Symposium on Theoretical Aspects of Computer Science, Aachen 2007]. We also show how to do forbidden-set routing on planar graphs between nodes whose distance is less than a parameter l. We prove a lower bound on the space requirements of forbidden-set routing for general graphs, and show that the problem is related to constructing an efficient distributed representation of all the separators of an undirected graph. Finally, we consider routing while taking into account path costs of intermediate nodes and show that this requires large routing labels. We also study a novel way of approximating forbidden-set routing using quotient graphs of low treewidth.

UCAM-CL-TR-679

Karen Spärck Jones:

Automatic summarising: a review and discussion of the state of the art

January 2007, 67 pages, PDF

Abstract: This paper reviews research on automatic summarising over the last decade. This period has seen a rapid growth of work in the area stimulated by technology and by several system evaluation programmes. The review makes use of several frameworks to organise the review, for summarising, for systems, for the task factors affecting summarising, and for evaluation design and practice.

The review considers the evaluation strategies that have been applied to summarising and the issues they raise, and the major summary evaluation programmes. It examines the input, purpose and output factors that have been investigated in summarising research in the last decade, and discusses the classes of strategy, both extractive and non-extractive, that have been explored, illustrating the range of systems that have been built. This analysis of strategies is amplified by accounts of specific exemplar systems.

The conclusions drawn from the review are that automatic summarisation research has made valuable progress in the last decade, with some practically useful approaches, better evaluation, and more understanding of the task. However as the review also makes clear, summarising systems are often poorly motivated in relation to the factors affecting summaries, and evaluation needs to be taken significantly further so as to engage with the purposes for which summaries are intended and the contexts in which they are used.

A reduced version of this report, entitled 'Automatic summarising: the state of the art' will appear in Information Processing and Management, 2007.

UCAM-CL-TR-680

Jing Su, James Scott, Pan Hui, Eben Upton, Meng How Lim, Christophe Diot, Jon Crowcroft, Ashvin Goel, Eyal de Lara:

Haggle: Clean-slate networking for mobile devices

January 2007, 30 pages, PDF

Abstract: Haggle is a layerless networking architecture for mobile devices. It is motivated by the infrastructure dependence of applications such as email and web browsing, even in situations where infrastructure is not necessary to accomplish the end user goal, e.g. when the destination is reachable by ad hoc neighbourhood communication. In this paper we present details of Haggle's architecture, and of the prototype implementation which allows existing email and web applications to become infrastructure-independent, as we show with an experimental evaluation.

UCAM-CL-TR-681

Piotr Zieliński:

Indirect channels: a bandwidth-saving technique for fault-tolerant protocols

April 2007, 24 pages, PDF

Abstract: Sending large messages known to the recipient is a waste of bandwidth. Nevertheless, many fault-tolerant agreement protocols send the same large message between each pair of participating processes. This practical problem has recently been addressed in the context of Atomic Broadcast by presenting a specialized algorithm.

This paper proposes a more general solution by providing virtual indirect channels that physically transmit message ids instead of full messages if possible. Indirect channels are transparent to the application; they can be used with any distributed algorithm, even with unreliable channels or malicious participants. At the same time, they provide rigorous theoretical properties.

Indirect channels are conservative: they do not allow manipulating message ids if full messages are not known. This paper also investigates the consequences of relaxing this assumption on the latency and correctness of Consensus and Atomic Broadcast implementations: new algorithms and lower bounds are shown.

UCAM-CL-TR-682

Juliano Iyoda:

Translating HOL functions to hardware

April 2007, 89 pages, PDF
PhD thesis (Hughes Hall, October 2006)

Abstract: Delivering error-free products is still a major challenge for hardware and software engineers. Due to the increasingly growing complexity of computing systems, there is a demand for higher levels of automation in formal verification.

This dissertation proposes an approach to generate formally verified circuits automatically. The main outcome of our project is a compiler implemented on top of the theorem prover HOL4 which translates a subset of higher-order logic to circuits. The subset of the logic is a first-order tail-recursive functional language. The compiler takes a function f as argument and automatically produces the theorem "$\vdash$ C implements f" where C is a circuit and "implements" is a correctness relation between a circuit and a function. We achieve full mechanisation of proofs by defining theorems which are composable. The correctness of a circuit can be mechanically determined by the correctness of its subcircuits. This technology allows the designer to focus on higher levels of abstraction instead of reasoning and verifying systems at the gate level.

A pretty-printer translates netlists described in higher-order logic to structural Verilog. Our compiler is integrated with Altera tools to run our circuits in FPGAs. Thus the theorem prover is used as an environment for supporting the development process from formal specification to implementation.

Our approach has been tested with fairly substantial case studies. We describe the design and the verification of a multiplier and a simple microcomputer which has shown us that the compiler supports small and medium-sized applications. Although this approach does not scale to industrial-sized applications yet, it is a first step towards the implementation of a new technology that can raise the level of mechanisation in formal verification.

UCAM-CL-TR-683

Martin Kleppmann:

Simulation of colliding constrained rigid bodies

April 2007, 65 pages, PDF

Abstract: I describe the development of a program to simulate the dynamic behaviour of interacting rigid bodies. Such a simulation may be used to generate animations of articulated characters in 3D graphics applications. Bodies may have an arbitrary shape, defined by a triangle mesh, and may be connected with a variety of different joints. Joints are represented by constraint functions which are solved at run-time using Lagrange multipliers. The simulation performs collision detection and prevents penetration of rigid bodies by applying impulses to colliding bodies and reaction forces to bodies in resting contact.

The simulation is shown to be physically accurate and is tested on several different scenes, including one of an articulated human character falling down a flight of stairs.

An appendix describes how to derive arbitrary constraint functions for the Lagrange multiplier method. Collisions and joints are both represented as constraints, which allows them to be handled with a unified algorithm. The report also includes some results relating to the use of quaternions in dynamic simulations.

UCAM-CL-TR-684

Pan Hui, Jon Crowcroft:

Bubble Rap: Forwarding in small world DTNs in ever decreasing circles

May 2007, 44 pages, PDF

Abstract: In this paper we seek to improve understanding of the structure of human mobility, and to use this in the design of forwarding algorithms for Delay Tolerant Networks for the dissemination of data amongst mobile users.

Cooperation binds but also divides human society into communities. Members of the same community interact with each other preferentially. There is structure in human society. Within society and its communities, individuals have varying popularity. Some people are more popular and interact with more people

than others; we may call them hubs. Popularity ranking is one facet of the population. In many physical networks, some nodes are more highly connected to each other than to the rest of the network. The set of such nodes are usually called clusters, communities, cohesive groups or modules. There is structure to social networking. Different metrics can be used such as information flow, Freeman betweenness, closeness and inference power, but for all of them, each node in the network can be assigned a global centrality value.

What can be inferred about individual popularity, and the structure of human society from measurements within a network? How can the local and global characteristics of the network be used practically for information dissemination? We present and evaluate a sequence of designs for forwarding algorithms for Pocket Switched Networks, culminating in Bubble, which exploit increasing levels of information about mobility and interaction.

UCAM-CL-TR-685

John Daugman, Cathryn Downing:

Effect of severe image compression on iris recognition performance

May 2007, 20 pages, PDF

Abstract: We investigate three schemes for severe compression of iris images, in order to assess what their impact would be on recognition performance of the algorithms deployed today for identifying persons by this biometric feature. Currently, standard iris images are 600 times larger than the IrisCode templates computed from them for database storage and search; but it is administratively desired that iris data should be stored, transmitted, and embedded in media in the form of images rather than as templates computed with proprietary algorithms. To reconcile that goal with its implications for bandwidth and storage, we present schemes that combine region-of-interest isolation with JPEG and JPEG2000 compression at severe levels, and we test them using a publicly available government database of iris images. We show that it is possible to compress iris images to as little as 2 KB with minimal impact on recognition performance. Only some 2% to 3% of the bits in the IrisCode templates are changed by such severe image compression. Standard performance metrics such as error trade-off curves document very good recognition performance despite this reduction in data size by a net factor of 150, approaching a convergence of image data size and template size.

UCAM-CL-TR-686

Andrew C. Rice:

Dependable systems for Sentient Computing

May 2007, 150 pages, PDF

Abstract: Computers and electronic devices are continuing to proliferate throughout our lives. Sentient Computing systems aim to reduce the time and effort required to interact with these devices by composing them into systems which fade into the background of the user's perception. Failures are a significant problem in this scenario because their occurrence will pull the system into the foreground as the user attempts to discover and understand the fault. However, attempting to exist and interact with users in a real, unpredictable, physical environment rather than a well-constrained virtual environment makes failures inevitable.

This dissertation describes a study of dependability. A dependable system permits applications to discover the extent of failures and to adapt accordingly such that their continued behaviour is intuitive to users of the system.

Cantag, a reliable marker-based machine-vision system, has been developed to aid the investigation of dependability. The description of Cantag includes specific contributions for marker tracking such as rotationally invariant coding schemes and reliable back-projection for circular tags. An analysis of Cantag's theoretical performance is presented and compared to its real-world behaviour. This analysis is used to develop optimised tag designs and performance metrics. The use of validation is proposed to permit runtime calculation of observable metrics and verification of system components. Formal proof methods are combined with a logical validation framework to show the validity of performance optimisations.

UCAM-CL-TR-687

Viktor Vafeiadis, Matthew Parkinson:

A marriage of rely/guarantee and separation logic

June 2007, 31 pages, PDF

Abstract: In the quest for tractable methods for reasoning about concurrent algorithms both rely/guarantee logic and separation logic have made great advances. They both seek to tame, or control, the complexity of concurrent interactions, but neither is the ultimate approach. Rely-guarantee copes naturally with interference, but its specifications are complex because they describe the entire state. Conversely separation logic has difficulty dealing with interference, but its specifications are simpler because they describe only the relevant state that the program accesses.

We propose a combined system which marries the two approaches. We can describe interference naturally (using a relation as in rely/guarantee), and where there is no interference, we can reason locally (as in separation logic). We demonstrate the advantages of the combined approach by verifying a lock-coupling list algorithm, which actually disposes/frees removed nodes.

UCAM-CL-TR-688

Sam Staton:

Name-passing process calculi: operational models and structural operational semantics

June 2007, 245 pages, PDF
PhD thesis (Girton College, December 2006)

Abstract: This thesis is about the formal semantics of name-passing process calculi. We study operational models by relating various different notions of model, and we analyse structural operational semantics by extracting a congruence rule format from a model theory. All aspects of structural operational semantics are addressed: behaviour, syntax, and rule-based inductive definitions.

A variety of models for name-passing behaviour are considered and developed. We relate classes of indexed labelled transition systems, proposed by Cattani and Sewell, with coalgebraic models proposed by Fiore and Turi. A general notion of structured coalgebra is introduced and developed, and a natural notion of structured bisimulation is related to Sangiorgi's open bisimulation for the π-calculus. At first the state spaces are organised as presheaves, but it is reasonable to constrain the models to sheaves in a category known as the Schanuel topos. This sheaf topos is exhibited as equivalent to a category of named-sets proposed by Montanari and Pistore for efficient verification of name-passing systems.

Syntax for name-passing calculi involves variable binding and substitution. Gabbay and Pitts proposed nominal sets as an elegant model for syntax with binding, and we develop a framework for substitution in this context. The category of nominal sets is equivalent to the Schanuel topos, and so syntax and behaviour can be studied within one universe.

An abstract account of structural operational semantics was developed by Turi and Plotkin. They explained the inductive specification of a system by rules in the GSOS format of Bloom et al., in terms of initial algebra recursion for lifting a monad of syntax to a category of behaviour. The congruence properties of bisimilarity can be observed at this level of generality. We study this theory in the general setting of structured coalgebras, and then for the specific case of name-passing systems, based on categories of nominal sets.

At the abstract level of category theory, classes of rules are understood as natural transformations. In the concrete domain, though, rules for name-passing systems are formulae in a suitable logical framework. By imposing a format on rules in Pitts's nominal logic, we characterise a subclass of rules in the abstract domain. Translating the abstract results, we conclude that, for a name-passing process calculus defined by rules in this format, a variant of open bisimilarity is a congruence.

UCAM-CL-TR-689

Ursula H. Augsdörfer, Neil A. Dodgson, Malcolm A. Sabin:

Removing polar rendering artifacts in subdivision surfaces

June 2007, 7 pages, PDF

Abstract: There is a belief that subdivision schemes require the subdominant eigenvalue, λ, to be the same around extraordinary vertices as in the regular regions of the mesh. This belief is owing to the polar rendering artifacts which occur around extraordinary points when λ is significantly larger than in the regular regions. By constraining the tuning of subdivision schemes to solutions which fulfill this condition we may prevent ourselves from finding the optimal limit surface. We show that the perceived problem is purely a rendering artifact and that it does not reflect the quality of the underlying limit surface. Using the bounded curvature Catmull-Clark scheme as an example, we describe three practical methods by which this rendering artifact can be removed, thereby allowing us to tune subdivision schemes using any appropriate values of λ.

UCAM-CL-TR-690

Russell Glen Ross:

Cluster storage for commodity computation

June 2007, 178 pages, PDF
PhD thesis (Wolfson College, December 2006)

Abstract: Standards in the computer industry have made basic components and entire architectures into commodities, and commodity hardware is increasingly being used for the heavy lifting formerly reserved for specialised platforms. Now software and services are following. Modern updates to virtualization technology make it practical to subdivide commodity servers and manage groups of heterogeneous services using commodity operating systems and tools, so services can be packaged and managed independent of the hardware on which they run. Computation as a commodity is soon to follow, moving beyond the specialised applications typical of today's utility computing.

In this dissertation, I argue for the adoption of service clusters—clusters of commodity machines under central control, but running services in virtual machines for arbitrary, untrusted clients—as the basic building block for an economy of flexible commodity computation. I outline the requirements this platform imposes on its storage system and argue that they are necessary for service clusters to be practical, but are not found in existing systems.

Next I introduce Envoy, a distributed file system for service clusters. In addition to meeting the needs of a new environment, Envoy introduces a novel file distribution scheme that organises metadata and cache management according to runtime demand. In effect, the file system is partitioned and control of each part given to the client that uses it the most; that client in turn acts as a server with caching for other clients that require concurrent access. Scalability is limited only by runtime contention, and clients share a perfectly consistent cache distributed across the cluster. As usage patterns change, the partition boundaries are updated dynamically, with urgent changes made quickly and more minor optimisations made over a longer period of time.

Experiments with the Envoy prototype demonstrate that service clusters can support cheap and rapid deployment of services, from isolated instances to groups of cooperating components with shared storage demands.

UCAM-CL-TR-691

Neil A. Dodgson, Malcolm A. Sabin, Richard Southern:

Preconditions on geometrically sensitive subdivision schemes

August 2007, 13 pages, PDF

Abstract: Our objective is to create subdivision schemes with limit surfaces which are surfaces useful in engineering (spheres, cylinders, cones etc.) without resorting to special cases. The basic idea explored by us previously in the curve case is that if the property that all vertices lie on an object of the required class can be preserved through the subdivision refinement, it will be preserved into the limit surface also. The next obvious step was to try a bivariate example. We therefore identified the simplest possible scheme and implemented it. However, this misbehaved quite dramatically. This report, by doing the limit analysis, identifies why the misbehaviour occurred, and draws conclusions about how the problems should be avoided.

UCAM-CL-TR-692

Alan F. Blackwell:

Toward an undergraduate programme in Interdisciplinary Design

July 2007, 13 pages, PDF

Abstract: This technical report describes an experimental syllabus proposal that was developed for the Cambridge Computer Science Tripos (the standard undergraduate degree programme in Computer Science at Cambridge). The motivation for the proposal was to create an innovative research-oriented taught course that would be compatible with the broader policy goals of the Crucible network for research in interdisciplinary design. As the course is not proceeding, the syllabus is published here for use by educators and educational researchers with interests in design teaching.

UCAM-CL-TR-693

Piotr Zieliński:

Automatic classification of eventual failure detectors

July 2007, 21 pages, PDF

Abstract: Eventual failure detectors, such as Ω or $\Diamond$P, can make arbitrarily many mistakes before they start providing correct information. This paper shows that any detector implementable in an purely asynchronous system can be implemented as a function of only the order of most-recently heard-from processes. The finiteness of this representation means that eventual failure detectors can be enumerated and their relative strengths tested automatically. The results for systems with two and three processes are presented.

Implementability can also be modelled as a game between Prover and Disprover. This approach not only speeds up automatic implementability testing, but also results in shorter and more intuitive proofs. I use this technique to identify the new weakest failure detector anti-Ω and prove its properties. Anti-Ω outputs process ids and, while not necessarily stabilizing, it ensures that some correct process is eventually never output.

UCAM-CL-TR-694

Piotr Zieliński:

Anti-Ω: the weakest failure detector for set agreement

July 2007, 24 pages, PDF

Abstract: In the set agreement problem, n processes have to decide on at most $n-1$ of the proposed values. This paper shows that the anti-Ω failure detector is both sufficient and necessary to implement set agreement in an asynchronous shared-memory system equipped with registers. Each query to anti-Ω returns a single process id; the specification ensures that there is a correct process whose id is returned only finitely many times.

UCAM-CL-TR-695

Karen Su, Inaki Berenguer, Ian J. Wassell, Xiaodong Wang:

Efficient maximum-likelihood decoding of spherical lattice codes

July 2007, 29 pages, PDF

Abstract: A new framework for efficient and exact Maximum-Likelihood (ML) decoding of spherical lattice codes is developed. It employs a double-tree structure: The first is that which underlies established tree-search decoders; the second plays the crucial role of guiding the primary search by specifying admissible candidates and is our focus in this report. Lattice codes have long been of interest due to their rich structure, leading to numerous decoding algorithms for unbounded lattices, as well as those with axis-aligned rectangular shaping regions. Recently, spherical Lattice Space-Time (LAST) codes were proposed to realize the optimal diversity-multiplexing tradeoff of MIMO channels. We address the so-called boundary control problem arising from the spherical shaping region defining these codes. This problem is complicated because of the varying number of candidates potentially under consideration at each search stage; it is not obvious how to address it effectively within the frameworks of existing schemes. Our proposed strategy is compatible with all sequential tree-search detectors, as well as auxiliary processing such as the MMSE-GDFE and lattice reduction. We demonstrate the superior performance and complexity profiles achieved when applying the proposed boundary control in conjunction with two current efficient ML detectors and show an improvement of 1dB over the state-of-the-art at a comparable complexity.

UCAM-CL-TR-696

Oliver J. Woodman:

An introduction to inertial navigation

August 2007, 37 pages, PDF

Abstract: Until recently the weight and size of inertial sensors has prohibited their use in domains such as human motion capture. Recent improvements in the performance of small and lightweight micromachined electromechanical systems (MEMS) inertial sensors have made the application of inertial techniques to such problems possible. This has resulted in an increased interest in the topic of inertial navigation, however current introductions to the subject fail to sufficiently describe the error characteristics of inertial systems.

We introduce inertial navigation, focusing on strapdown systems based on MEMS devices. A combination of measurement and simulation is used to explore the error characteristics of such systems. For a simple inertial navigation system (INS) based on the Xsens Mtx inertial measurement unit (IMU), we show that the average error in position grows to over 150 m after 60 seconds of operation. The propagation of orientation errors caused by noise perturbing gyroscope signals is identified as the critical cause of such drift. By simulation we examine the significance of individual noise processes perturbing the gyroscope signals, identifying white noise as the process which contributes most to the overall drift of the system.

Sensor fusion and domain specific constraints can be used to reduce drift in INSs. For an example INS we show that sensor fusion using magnetometers can reduce the average error in position obtained by the system after 60 seconds from over 150 m to around 5 m. We conclude that whilst MEMS IMU technology is rapidly improving, it is not yet possible to build a MEMS based INS which gives sub-meter position accuracy for more than one minute of operation.

UCAM-CL-TR-697

Chris J. Purcell:

Scaling Mount Concurrency: scalability and progress in concurrent algorithms

August 2007, 155 pages, PDF
PhD thesis (Trinity College, July 2007)

Abstract: As processor speeds plateau, chip manufacturers are turning to multi-processor and multi-core designs to increase performance. As the number of simultaneous threads grows, Amdahl's Law means the performance of programs becomes limited by the cost that does not scale: communication, via the memory subsystem. Algorithm design is critical in minimizing these costs.

In this dissertation, I first show that existing instruction set architectures must be extended to allow general scalable algorithms to be built. Since it is impractical to entirely abandon existing hardware, I then present a reasonably scalable implementation of a map built on the widely-available compare-and-swap primitive, which outperforms existing algorithms for a range of usages.

Thirdly, I introduce a new primitive operation, and show that it provides efficient and scalable solutions to several problems before proving that it satisfies strong theoretical properties. Finally, I outline possible hardware implementations of the primitive with different properties and costs, and present results from a hardware evaluation, demonstrating that the new primitive can provide good practical performance.

UCAM-CL-TR-698

Simon J. Hollis:

Pulse-based, on-chip interconnect

September 2007, 186 pages, PDF
PhD thesis (Queens' College, June 2007)

Abstract: This thesis describes the development of an on-chip point-to-point link, with particular emphasis on the reduction of its global metal area footprint.

To reduce its metal footprint, the interconnect uses a serial transmission approach. 8-bit data is sent using just two wires, through a pulse-based technique, inspired by the GasP interconnect from Sun Microsystems. Data and control signals are transmitted bi-directionally on a wire using this double-edged, pulse-based signalling protocol, and formatted using a variant of dual-rail encoding. These choices enable a reduction in the number of wires needed, an improvement in the acknowledgement overhead of the asynchronous protocol, and the ability to cross clock domains without synchronisation hazards.

New, stateful, repeaters are demonstrated, and results from spice simulations of the system show that data can be transferred at over 1Gbit/s, over 1mm of minimum-sized, minimally-spaced metal 5 wiring, on a 180nm (0.18um) technology. This reduces to only 926Mbit/s, when 10mm of wiring is considered, and represents a channel utilisation of a very attractive 45% of theoretical capacity at this length. Analysis of latencies, energy consumption, and area use are also provided.

The point-to-point link is then expanded with the invention and demonstration of a router and an arbitrated merge element, to produce a Network-on-Chip (NoC) design, called RasP. The full system is then evaluated, and peak throughput is shown to be 763Mbit/s for 1mm of wiring, reducing to 599Mbit/s for 10mm of the narrow metal 5 interconnect.

Finally, RasP is compared in performance with the Chain interconnect from the University of Manchester. Results for the metrics of throughput, latency, energy consumption and area footprint show that the two systems perform very similarly — the maximum absolute deviation is under 25% for throughput, latency and area; and the energy-efficiency of RasP is approximately twice that of Chain. Between the two systems, RasP has the smaller latency, energy and area requirements and is shown to be a viable alternative NoC design.

UCAM-CL-TR-699

Richard Southern, Neil A. Dodgson:

A smooth manifold based construction of approximating lofted surfaces

October 2007, 17 pages, PDF

Abstract: We present a new method for constructing a smooth manifold approximating a curve network or control mesh. In our two-step method, smooth vertex patches are initially defined by extrapolating and then blending a univariate or bivariate surface representation. Each face is then constructed by blending together the segments of each vertex patch corresponding to the face corners. By approximating the input curve network, rather than strictly interpolating it, we have greater flexibility in controlling surface behaviour and have local control. Additionally no initial control mesh fitting or fairing needs to be performed, and no derivative information is needed to ensure continuity at patch boundaries.

UCAM-CL-TR-700

Maja Vuković:

Context aware service composition

October 2007, 225 pages, PDF
PhD thesis (Newnham College, April 2006)

Abstract: Context aware applications respond and adapt to changes in the computing environment. For example, they may react when the location of the user or the capabilities of the device used change. Despite the increasing importance and popularity of such applications, advances in application models to support their development have not kept up. Legacy application design models, which embed contextual dependencies in the form of if-then rules specifying how applications should react to context changes, are still widely used. Such models are impractical to accommodate the large variety of possibly even unanticipated context types and their values.

This dissertation proposes a new application model for building context aware applications, considering them as dynamically composed sequences of calls to services, software components that perform well-defined computational operations and export open interfaces through which they can be invoked. This work employs goal-oriented inferencing from planning technologies for selecting the services and assembling the sequence of their execution, allowing different compositions to result from different context parameters such as resources available, time constraints, and user location. Contextual changes during the execution of the services may trigger further re-composition causing the application to evolve dynamically.

An important challenge in providing a context aware service composition facility is dealing with failures that may occur, for instance as a result of context changes or missing service descriptions. To handle composition failures, this dissertation introduces GoalMorph, a system which transforms failed composition requests into alternative ones that can be solved.

This dissertation describes the design and implementation of the proposed framework for context aware service composition. Experimental evaluation

of a realistic infotainment application demonstrates that the framework provides an effcient and scalable solution. Furthermore, it shows that GoalMorph transforms goals successfully, increasing the utility of achieved goals without imposing a prohibitive composition time overhead.

By developing the proposed framework for fault-tolerant, context aware service composition this work ultimately lowers the barrier for building extensible applications that automatically adapt to the user's context. This represents a step towards a new paradigm for developing adaptive software to accommodate the increasing dynamicity of computing environments.

UCAM-CL-TR-701

Jacques Jean-Alain Fournier:

Vector microprocessors for cryptography

October 2007, 174 pages, PDF
PhD thesis (Trinity Hall, April 2007)

Abstract: Embedded security devices like 'Trusted Platforms' require both scalability (of power, performance and area) and flexibility (of software and countermeasures). This thesis illustrates how data parallel techniques can be used to implement scalable architectures for cryptography. Vector processing is used to provide high performance, power efficient and scalable processors. A programmable vector 4-stage pipelined co-processor, controlled by a scalar MIPS compatible processor, is described. The instruction set of the co-processor is defined for cryptographic algorithms like AES and Montgomery modular multiplication for RSA and ECC. The instructions are assessed using an instruction set simulator based on the ArchC tool. This instruction set simulator is used to see the impact of varying the vector register depth (p) and the number of vector processing units (r). Simulations indicate that for vector versions of AES, RSA and ECC the performance improves in O(log(r)). A cycle-accurate synthesisable Verilog model of the system (VeMICry) is implemented in TSMC's 90nm technology and used to show that the best area/power/performance tradeoff is reached for r = (p/4). Also, this highly scalable design allows area/power/performance trade-offs to be made for a panorama of applications ranging from smartcards to servers. This thesis is, to my best knowledge, the first attempt to implement embedded cryptography using vector processing techniques.

UCAM-CL-TR-702

Alisdair Wren:

Relationships for object-oriented programming languages

November 2007, 153 pages, PDF
PhD thesis (Sidney Sussex College, March 2007)

Abstract: Object-oriented approaches to software design and implementation have gained enormous popularity over the past two decades. However, whilst models of software systems routinely allow software engineers to express relationships between objects, object-oriented programming languages lack this ability. Instead, relationships must be encoded using complex reference structures. When the model cannot be expressed directly in code, it becomes more difficult for programmers to see a correspondence between design and implementation – the model no longer faithfully documents the code. As a result, programmer intuition is lost, and error becomes more likely, particularly during maintenance of an unfamiliar software system.

This thesis explores extensions to object-oriented languages so that relationships may be expressed with the same ease as objects. Two languages with relationships are specified: RelJ, which offers relationships in a class-based language based on Java, and QSigma, which is an object calculus with heap query.

In RelJ, relationship declarations exist at the same level as class declarations: relationships are named, they may have fields and methods, they may inherit from one another and their instances may be referenced just like objects. Moving into the object-based world, QSigma is based on the sigma-calculi of Abadi and Cardelli, extended with the ability to query the heap. Heap query allows objects to determine how they are referenced by other objects, such that single references are sufficient for establishing an inter-object relationship observable by all participants. Both RelJ and QSigma are equipped with a formal type system and semantics to ensure type safety in the presence of these extensions.

By giving formal models of relationships in both class- and object-based settings, we can obtain general principles for relationships in programming languages and, therefore, establish a correspondence between implementation and design.

UCAM-CL-TR-703

Jon Crowcroft, Tim Deegan,
Christian Kreibich, Richard Mortier,
Nicholas Weaver:

Lazy Susan: dumb waiting as proof of work

November 2007, 23 pages, PDF

Abstract: The open nature of Internet services has been of great value to users, enabling dramatic innovation and evolution of services. However, this openness permits many abuses of open-access Internet services such as web, email, and DNS. To counteract such abuses, a number of so called proof-of-work schemes have been proposed. They aim to prevent or limit such abuses by demanding potential clients of the service to prove that they have carried out some amount of work before

they will be served. In this paper we show that existing resource-based schemes have several problems, and instead propose latency-based proof-of-work as a solution. We describe centralised and distributed variants, introducing the problem class of non-parallelisable shared secrets in the process. We also discuss application of this technique at the network layer as a way to prevent Internet distributed denial-of-service attacks.

UCAM-CL-TR-704

Paul William Hunter:

Complexity and infinite games on finite graphs

November 2007, 170 pages, PDF
PhD thesis (Hughes Hall, July 2007)

Abstract: This dissertation investigates the interplay between complexity, infinite games, and finite graphs. We present a general framework for considering two-player games on finite graphs which may have an infinite number of moves and we consider the computational complexity of important related problems. Such games are becoming increasingly important in the field of theoretical computer science, particularly as a tool for formal verification of non-terminating systems. The framework introduced enables us to simultaneously consider problems on many types of games easily, and this is demonstrated by establishing previously unknown complexity bounds on several types of games.

We also present a general framework which uses infinite games to define notions of structural complexity for directed graphs. Many important graph parameters, from both a graph theoretic and algorithmic perspective, can be defined in this system. By considering natural generalizations of these games to directed graphs, we obtain a novel feature of digraph complexity: directed connectivity. We show that directed connectivity is an algorithmically important measure of complexity by showing that when it is limited, many intractable problems can be efficiently solved. Whether it is structurally an important measure is yet to be seen, however this dissertation makes a preliminary investigation in this direction.

We conclude that infinite games on finite graphs play an important role in the area of complexity in theoretical computer science.

UCAM-CL-TR-705

Alan C. Lawrence:

Optimizing compilation with the Value State Dependence Graph

December 2007, 183 pages, PDF
PhD thesis (Churchill College, May 2007)

Abstract: Most modern compilers are based on variants of the Control Flow Graph. Developments on this representation—specifically, SSA form and the Program Dependence Graph (PDG)—have focused on adding and refining data dependence information, and these suggest the next step is to use a purely data-dependence-based representation such as the VDG (Ernst et al.) or VSDG (Johnson et al.).

This thesis studies such representations, identifying key differences in the information carried by the VSDG and several restricted forms of PDG, which relate to functional programming and continuations. We unify these representations in a new framework for specifying the sharing of resources across a computation.

We study the problems posed by using the VSDG, and argue that existing techniques have not solved the sequentialization problem of mapping VSDGs back to CFGs. We propose a new compiler architecture breaking sequentialization into several stages which focus on different characteristics of the input VSDG, and tend to be concerned with different properties of the output and target machine. The stages integrate a wide variety of important optimizations, exploit opportunities offered by the VSDG to address many common phase-order problems, and unify many operations previously considered distinct.

Focusing on branch-intensive code, we demonstrate how effective control flow—sometimes superior to that of the original source code, and comparable to the best CFG optimization techniques—can be reconstructed from just the dataflow information comprising the VSDG. Further, a wide variety of more invasive optimizations involving the duplication and specialization of program elements are eased because the VSDG relaxes the CFG's overspecification of instruction and branch ordering. Specifically we identify the optimization of nested branches as generalizing the problem of minimizing boolean expressions.

We conclude that it is now practical to discard the control flow information rather than maintain it in parallel as is done in many previous approaches (e.g. the PDG).

UCAM-CL-TR-706

Steven J. Murdoch:

Covert channel vulnerabilities in anonymity systems

December 2007, 140 pages, PDF
PhD thesis (Girton College, August 2007)

Abstract: The spread of wide-scale Internet surveillance has spurred interest in anonymity systems that protect users' privacy by restricting unauthorised access to their identity. This requirement can be considered as a flow control policy in the well established field of multi-level secure systems. I apply previous research on covert

channels (unintended means to communicate in violation of a security policy) to analyse several anonymity systems in an innovative way.

One application for anonymity systems is to prevent collusion in competitions. I show how covert channels may be exploited to violate these protections and construct defences against such attacks, drawing from previous covert channel research and collusion-resistant voting systems.

In the military context, for which multilevel secure systems were designed, covert channels are increasingly eliminated by physical separation of interconnected single-role computers. Prior work on the remaining network covert channels has been solely based on protocol specifications. I examine some protocol implementations and show how the use of several covert channels can be detected and how channels can be modified to resist detection.

I show how side channels (unintended information leakage) in anonymity networks may reveal the behaviour of users. While drawing on previous research on traffic analysis and covert channels, I avoid the traditional assumption of an omnipotent adversary. Rather, these attacks are feasible for an attacker with limited access to the network. The effectiveness of these techniques is demonstrated by experiments on a deployed anonymity network, Tor.

Finally, I introduce novel covert and side channels which exploit thermal effects. Changes in temperature can be remotely induced through CPU load and measured by their effects on crystal clock skew. Experiments show this to be an effective attack against Tor. This side channel may also be usable for geolocation and, as a covert channel, can cross supposedly infallible air-gap security boundaries.

This thesis demonstrates how theoretical models and generic methodologies relating to covert channels may be applied to find practical solutions to problems in real-world anonymity systems. These findings confirm the existing hypothesis that covert channel analysis, vulnerabilities and defences developed for multilevel secure systems apply equally well to anonymity systems.

UCAM-CL-TR-707

Ian Caulfield:

Complexity-effective superscalar embedded processors using instruction-level distributed processing

December 2007, 130 pages, PDF
PhD thesis (Queens' College, May 2007)

Abstract: Modern trends in mobile and embedded devices require ever increasing levels of performance, while maintaining low power consumption and silicon area usage. This thesis presents a new architecture for a high-performance embedded processor, based upon the instruction-level distributed processing (ILDP) methodology. A qualitative analysis of the complexity of an ILDP implementation as compared to both a typical scalar RISC CPU and a superscalar design is provided, which shows that the ILDP architecture eliminates or greatly reduces the size of a number of structures present in a superscalar architecture, allowing its complexity and power consumption to compare favourably with a simple scalar design.

The performance of an implementation of the ILDP architecture is compared to some typical processors used in high-performance embedded systems. The effect on performance of a number of the architectural parameters is analysed, showing that many of the parallel structures used within the processor can be scaled to provide less parallelism with little cost to the overall performance. In particular, the size of the register file can be greatly reduced with little average effect on performance – a size of 32 registers, with 16 visible in the instruction set, is shown to provide a good trade-off between area/power and performance.

Several novel developments to the ILDP architecture are then described and analysed. Firstly, a scheme to halve the number of processing elements and thus greatly reduce silicon area and power consumption is outlined but proves to result in a 12–14% drop in performance. Secondly, a method to reduce the area and power requirements of the memory logic in the architecture is presented which can achieve similar performance to the original architecture with a large reduction in area and power requirements or, at an increased area/power cost, can improve performance by approximately 24%. Finally, a new organisation for the register file is proposed, which reduces the silicon area used by the register file by approximately three-quarters and allows even greater power savings, especially in the case where processing elements are power gated.

Overall, it is shown that the ILDP methodology is a viable approach for future embedded system design, and several new variants on the architecture are contributed. Several areas of useful future research are highlighted, especially with respect to compiler design for the ILDP paradigm.

UCAM-CL-TR-708

Chi-Kin Chau, Jon Crowcroft, Kang-Won Lee, Starsky H.Y. Wong:

IDRM: Inter-Domain Routing Protocol for Mobile Ad Hoc Networks

January 2008, 24 pages, PDF

Abstract: Inter-domain routing is an important component to allow interoperation among heterogeneous network domains operated by different organizations. Although inter-domain routing has been extensively studied in the Internet, it remains relatively unexplored in the Mobile Ad Hoc Networks (MANETs) space. In MANETs, the inter-domain routing problem is challenged by: (1) dynamic network topology, and (2) diverse intra-domain ad hoc routing protocols. In this paper, we propose a networking protocol called IDRM (Inter-Domain Routing Protocol for MANETs) to enable interoperation among MANETs. IDRM can handle the dynamic nature of MANETs and support policy-based routing similarly to BGP. We first discuss the design challenges for inter-domain routing in MANETs, and then present the design of IDRM with illustrative examples. Finally, we present a simulation-based study to understand the operational effectiveness of inter-domain routing and show that the overhead of IDRM is moderate.

UCAM-CL-TR-709

Ford Long Wong:

Protocols and technologies for security in pervasive computing and communications

January 2008, 167 pages, PDF
PhD thesis (Girton College, August 2007)

Abstract: As the state-of-the-art edges towards Mark Weiser's vision of ubiquitous computing (ubicomp), we found that we have to revise some previous assumptions about security engineering for this domain. Ubicomp devices have to be networked together to be able to realize their promise. To communicate securely amongst themselves, they have to establish secret session keys, but this is a difficult problem when this is done primarily over radio in an ad-hoc scenario, i.e. without the aid of an infrastructure (such as a PKI), and when it is assumed that the devices are resource-constrained and cannot perform complex calculations. Secondly, when ubicomp devices are carried by users as personal items, their permanent identifiers inadvertently allow the users to be tracked, to the detriment of user privacy. Unless there are deliberate improvements in designing for location privacy, ubicomp devices can be trivially detected, and linked to individual users, with discomfiting echoes of a surveillance society. Our findings and contributions are thus as follow. In considering session key establishment, we learnt that asymmetric cryptography is not axiomatically infeasible, and may in fact be essential, to counter possible attackers, for some of the more computationally capable (and important) devices. We next found existing attacker models to be inadequate, along with existing models of bootstrapping security associations, in ubicomp. We address the inadequacies with a contribution which we call: 'multi-channel security protocols', by leveraging on multiple channels, with different properties, existing in the said environment. We gained an appreciation of the fact that location privacy is really a multi-layer problem, particularly so in ubicomp, where an attacker often may have access to different layers. Our contributions in this area are to advance the design for location privacy by introducing a MAC-layer proposal with stronger unlinkability, and a physical-layer proposal with stronger unobservability.

UCAM-CL-TR-710

Lucy G. Brace-Evans:

Event structures with persistence

February 2008, 113 pages, PDF
PhD thesis (St. John's College, October 2007)

Abstract: Increasingly, the style of computation is changing. Instead of one machine running a program sequentially, we have systems with many individual agents running in parallel. The need for mathematical models of such computations is therefore ever greater.

There are many models of concurrent computations. Such models can, for example, provide a semantics to process calculi and thereby suggest behavioural equivalences between processes. They are also key to the development of automated tools for reasoning about concurrent systems. In this thesis we explore some applications and generalisations of one particular model – event structures. We describe a variety of kinds of morphism between event structures. Each kind expresses a different sort of behavioural relationship. We demonstrate the way in which event structures can model both processes and types of processes by recalling a semantics for Affine HOPLA, a higher order process language. This is given in terms of asymmetric spans of event structures. We show that such spans support a trace construction. This allows the modelling of feedback and suggests a semantics for non-deterministic dataflow processes in terms of spans. The semantics given is shown to be consistent with Kahn's fixed point construction when we consider spans modelling deterministic processes.

A generalisation of event structures to include persistent events is proposed. Based on previously described morphisms between classical event structures, we define several categories of event structures with persistence. We show that, unlike for the corresponding categories of classical event structures, all are isomorphic to Kleisli categories of monads on the most restricted category. Amongst other things, this provides us with a way of understanding the asymmetric spans mentioned previously as symmetric spans where one morphism is modified by a monad. Thus we provide a general setting for future investigations involving event structures.

UCAM-CL-TR-711

Saar Drimer, Steven J. Murdoch, Ross Anderson:

Thinking inside the box: system-level failures of tamper proofing

February 2008, 37 pages, PDF

Abstract: PIN entry devices (PEDs) are critical security components in EMV smartcard payment systems as they receive a customer's card and PIN. Their approval is subject to an extensive suite of evaluation and certification procedures. In this paper, we demonstrate that the tamper proofing of PEDs is unsatisfactory, as is the certification process. We have implemented practical low-cost attacks on two certified, widely-deployed PEDs – the Ingenico i3300 and the Dione Xtreme. By tapping inadequately protected smartcard communications, an attacker with basic technical skills can expose card details and PINs, leaving cardholders open to fraud. We analyze the anti-tampering mechanisms of the two PEDs and show that, while the specific protection measures mostly work as intended, critical vulnerabilities arise because of the poor integration of cryptographic, physical and procedural protection. As these vulnerabilities illustrate a systematic failure in the design process, we propose a methodology for doing it better in the future. They also demonstrate a serious problem with the Common Criteria. We discuss the incentive structures of the certification process, and show how they can lead to problems of the kind we identified. Finally, we recommend changes to the Common Criteria framework in light of the lessons learned.

An abridged version of this paper is to appear at the IEEE Symposium on Security and Privacy, May 2008, Oakland, CA, US.

UCAM-CL-TR-712

Christian Richardt:

Flash-exposure high dynamic range imaging: virtual photography and depth-compensating flash

March 2008, 9 pages, PDF

Abstract: I present a revised approach to flash-exposure high dynamic range (HDR) imaging and demonstrate two applications of this image representation. The first application enables the creation of realistic 'virtual photographs' for arbitrary flash-exposure settings, based on a single flash-exposure HDR image. The second application is a novel tone mapping operator for flash-exposure HDR images based on the idea of an 'intelligent flash'. It compensates for the depth-related brightness fall-off occurring in flash photographs by taking the ambient illumination into account.

UCAM-CL-TR-713

Pan Hui:

People are the network: experimental design and evaluation of social-based forwarding algorithms

March 2008, 160 pages, PDF
PhD thesis (Churchill College, July 2007)

Abstract: Cooperation binds but also divides human society into communities. Members of the same community interact with each other preferentially. There is structure in human society. Within society and its communities, individuals have varying popularity. Some people are more popular and interact with more people than others; we may call them hubs. I develop methods to extract this kind of social information from experimental traces and use it to choose the next hop forwarders in Pocket Switched Networks (PSNs). I find that by incorporating social information, forwarding efficiency can be significantly improved. For practical reasons, I also develop distributed algorithms for inferring communities.

Forwarding in Delay Tolerant Networks (DTNs), or more particularly PSNs, is a challenging problem since human mobility is usually difficult to predict. In this thesis, I aim to tackle this problem using an experimental approach by studying real human mobility. I perform six mobility experiments in different environments. The resultant experimental datasets are valuable for the research community. By analysing the experimental data, I find out that the inter-contact time of humans follows a power-law distribution with coefficient smaller than 1 (over the range of 10 minutes to 1 day). I study the limits of "oblivious" forwarding in the experimental environment and also the impact of the power-law coefficient on message delivery.

In order to study social-based forwarding, I develop methods to infer human communities from the data and use these in the study of social-aware forwarding. I propose several social-aware forwarding schemes and evaluate them on different datasets. I find out that by combining community and centrality information, forwarding efficiency can be significantly improved, and I call this scheme BUBBLE forwarding with the analogy that each community is a BUBBLE with big bubbles containing smaller bubbles. For practical deployment of these algorithms, I propose distributed community detection schemes, and also propose methods to approximate node centrality in the system.

Besides the forwarding study, I also propose a layerless data-centric architecture for the PSN scenario to address the problem with the status quo in communication (e.g. an infrastructuredependent and synchronous API), which brings PSN one step closer to real-world deployment.

UCAM-CL-TR-714

Tim Moreton:

A wide-area file system for migrating virtual machines

March 2008, 163 pages, PDF
PhD thesis (King's College, February 2007)

Abstract: Improvements in processing power and core bandwidth set against fundamental constraints on wide-area latency increasingly emphasise the position in the network at which services are deployed. The XenoServer project is building a platform for distributed computing that facilitates the migration of services between hosts to minimise client latency and balance load in response to changing patterns of demand. Applications run inside whole-system virtual machines, allowing the secure multiplexing of host resources.

Since services are specified in terms of a complete root file system and kernel image, a key component of this architecture is a substrate that provides an abstraction akin to local disks for these virtual machines, whether they are running, migrating or suspended. However, the same combination of wide-area latency, constrained bandwidth and global scale that motivates the XenoServer platform itself impedes the location, management and rapid transfer of storage between deployment sites. This dissertation describes Xest, a novel wide-area file system that aims to address these challenges.

I examine Xest's design, centred on the abstraction of virtual disks, volumes that allow only a single writer yet are transparently available despite migration. Virtual disks support the creation of snapshots and may be rapidly forked into copies that can be modified independently. This encourages an architectural separation into node-local file system and global content distribution framework and reduces the dependence of local operations on wide-area interactions.

I then describe how Xest addresses the dual problem of latency and scale by managing, caching, advertising and retrieving storage on the basis of groups, sets of files that correspond to portions of inferred working sets of client applications. Coarsening the granularity of these interfaces further decouples local and global activity: fewer units can lead to fewer interactions and the maintenance of less addressing state. The precision of these interfaces is retained by clustering according to observed access patterns and, in response to evidence of poor clusterings, selectively degrading groups into their constituent elements.

I evaluate a real deployment of Xest over a wide-area testbed. Doing so entails developing new tools for capturing and replaying traces to simulate virtual machine workloads. My results demonstrate the practicality and high performance of my design and illustrate the trade-offs involved in modifying the granularity of established storage interfaces.

UCAM-CL-TR-715

Feng Hao:

On using fuzzy data in security mechanisms

April 2008, 69 pages, PDF
PhD thesis (Queens' College, April 2007)

Abstract: Under the microscope, every physical object has unique features. It is impossible to clone an object, reproducing exactly the same physical traits. This unclonability principle has been applied in many security applications. For example, the science of biometrics is about measuring unique personal features. It can authenticate individuals with a high level of assurance. Similarly, a paper document can be identified by measuring its unique physical properties, such as randomly-interleaving fiber structure.

Unfortunately, when physical measurements are involved, errors arise inevitably and the obtained data are fuzzy by nature. This causes two main problems: 1) fuzzy data cannot be used as a cryptographic key, as cryptography demands the key be precise; 2) fuzzy data cannot be sorted easily, which prevents efficient information retrieval. In addition, biometric measurements create a strong binding between a person and his unique features, which may conflict with personal privacy. In this dissertation, we study these problems in detail and propose solutions.

First, we propose a scheme to derive error-free keys from fuzzy data, such as iris codes. There are two types of errors within iris codes: background-noise errors and burst errors. Accordingly, we devise a two-layer error correction technique, which first corrects the background-noise errors using a Hadamard code, then the burst errors using a Reed-Solomon code. Based on a database of 700 iris images, we demonstrate that an error-free key of 140 bits can be reliably reproduced from genuine iris codes with a 99.5% success rate. In addition, despite the irrevocability of the underlying biometric data, the keys produced using our technique can be easily revoked or updated.

Second, we address the search problem for a large fuzzy database that stores iris codes or data with a similar structure. Currently, the algorithm used in all public deployments of iris recognition is to search exhaustively through a database of iris codes, looking for a match that is close enough. We propose a much more efficient search algorithm: Beacon Guided Search (BGS). BGS works by indexing iris codes, adopting a "multiple colliding segments principle" along with an early termination strategy to reduce the search range dramatically. We evaluate this algorithm using 632,500 real-world iris codes, showing a substantial speed-up over exhaustive search with a negligible loss of precision. In addition, we demonstrate that our empirical findings match theoretical analysis.

Finally, we study the anonymous-veto problem, which is more commonly known as the Dining Cryptographers problem: how to perform a secure multiparty computation of the boolean-OR function, while preserving the privacy of each input bit. The solution to this problem has general applications in security going way beyond biometrics. Even though there have been several solutions presented over the past 20 years, we propose a new solution called: Anonymous Veto Network (AV-net). Compared with past work, the AV-net protocol provides the strongest protection of each delegate's privacy against collusion; it requires only two rounds of broadcast, fewer than any other solutions; the computational load and bandwidth usage are the lowest among the available techniques; and our protocol does not require any private channels or third parties. Overall, it seems unlikely that, with the same underlying technology, there can be any other solutions significantly more efficient than ours.

UCAM-CL-TR-716

Gavin M. Bierman, Matthew J. Parkinson, James Noble:

UpgradeJ: Incremental typechecking for class upgrades

April 2008, 35 pages, PDF

Abstract: One of the problems facing developers is the constant evolution of components that are used to build applications. This evolution is typical of any multi-person or multi-site software project. How can we program in this environment? More precisely, how can language design address such evolution? In this paper we attack two significant issues that arise from constant component evolution: we propose language- level extensions that permit multiple, co-existing versions of classes and the ability to dynamically upgrade from one version of a class to another, whilst still maintaining type safety guarantees and requiring only lightweight extensions to the runtime infrastructure. We show how our extensions, whilst intuitive, provide a great deal of power by giving a number of examples. Given the subtlety of the problem, we formalize a core fragment of our language and prove a number of important safety properties.

UCAM-CL-TR-717

Stephen Julian Rymill:

Psychologically-based simulation of human behaviour

June 2008, 250 pages, PDF
PhD thesis (Jesus College, 2006)

Abstract: The simulation of human behaviour is a key area of computer graphics as there is currently a great demand for animations consisting of virtual human characters, ranging from film special effects to building design. Currently, animated characters can either be laboriously created by hand, or by using an automated system: however, results from the latter may still look artificial and require much further manual work.

The aim of this work is to improve the automated simulation of human behaviour by making use of ideas from psychology research; the ways in which this research has been used are made clear throughout this thesis. It has influenced all aspects of the design:

- Collision avoidance techniques are based on observed practices.
- Actors have simulated vision and attention.
- Actors can be given a variety of moods and emotions to affect their behaviour.

This thesis discusses the benefits of the simulation of attention; this technique recreates the eye movements of each actor, and allows each actor to build up its own mental model of its surroundings. It is this model that the actor then uses in its decisions on how to behave: techniques for collision prediction and collision avoidance are discussed. On top of this basic behaviour, variability is introduced by allowing all actors to have different sets of moods and emotions, which influence all aspects of their behaviour. The real-time 3D simulation created to demonstrate the actors' behaviour is also described.

This thesis demonstrates that the use of techniques based on psychology research leads to a qualitative and quantitative improvement in the simulation of human behaviour; this is shown through a variety of pictures and videos, and by results of numerical experiments and user testing. Results are compared with previous work in the field, and with real human behaviour.

UCAM-CL-TR-718

Tyler Moore:

Cooperative attack and defense in distributed networks

June 2008, 172 pages, PDF
PhD thesis (St. John's College, March 2008)

Abstract: The advance of computer networking has made cooperation essential to both attackers and defenders. Increased decentralization of network ownership requires devices to interact with entities beyond their own realm of control. The distribution of intelligence forces decisions to be taken at the edge. The exposure of devices makes multiple, simultaneous attacker-chosen compromise a credible threat. Motivation for this thesis derives from the observation that it is often easier for attackers to cooperate than for defenders to do so. I describe a number of attacks which exploit cooperation to devastating effect. I also propose

and evaluate defensive strategies which require cooperation.

I first investigate the security of decentralized, or 'ad-hoc', wireless networks. Many have proposed preloading symmetric keys onto devices. I describe two practical attacks on these schemes. First, attackers may compromise several devices and share the pre-loaded secrets to impersonate legitimate users. Second, whenever some keys are not pre-assigned but exchanged upon deployment, a revoked attacker can rejoin the network.

I next consider defensive strategies where devices collectively decide to remove a malicious device from the network. Existing voting-based protocols are made resilient to the attacks I have developed, and I propose alternative strategies that can be more efficient and secure. First, I describe a reelection protocol which relies on positive affirmation from peers to continue participation. Then I describe a more radical alternative called suicide: a good device removes a bad one unilaterally by declaring both devices dead. Suicide offers significant improvements in speed and efficiency compared to voting-based decision mechanisms. I then apply suicide and voting to revocation in vehicular networks.

Next, I empirically investigate attack and defense in another context: phishing attacks on the Internet. I have found evidence that one group responsible for half of all phishing, the rock-phish gang, cooperates by pooling hosting resources and by targeting many banks simultaneously. These cooperative attacks are shown to be far more effective.

I also study the behavior of defenders – banks and Internet service providers – who must cooperate to remove malicious sites. I find that phishing-website lifetimes follow a long-tailed lognormal distribution. While many sites are removed quickly, others remain much longer. I examine several feeds from professional 'take-down' companies and find that a lack of data sharing helps many phishing sites evade removal for long time periods.

One anti-phishing organization has relied on volunteers to submit and verify suspected phishing sites. I find its voting-based decision mechanism to be slower and less comprehensive than unilateral verification performed by companies. I also note that the distribution of user participation is highly skewed, leaving the scheme vulnerable to manipulation.

UCAM-CL-TR-719

William H. Billingsley:

The Intelligent Book: technologies for intelligent and adaptive textbooks, focussing on Discrete Mathematics

June 2008, 156 pages, PDF
PhD thesis (Wolfson College, April 2007)

Abstract: An "Intelligent Book" is a Web-based textbook that contains exercises that are backed by computer models or reasoning systems. Within the exercises, students work using appropriate graphical notations and diagrams for the subject matter, and comments and feedback from the book are related into the content model of the book. The content model can be extended by its readers. This dissertation examines the question of how to provide an Intelligent Book that can support undergraduate questions in Number Theory, and particularly questions that allow the student to write a proof as the answer. Number Theory questions pose a challenge not only because the student is working on an unfamiliar topic in an unfamiliar syntax, but also because there is no straightforward procedure for how to prove an arbitrary Number Theory problem.

The main contribution is a system for supporting student-written proof exercises, backed by the Isabelle/HOL automated proof assistant and a set of teaching scripts. Students write proofs using MathsTiles: a graphical notation consisting of composable tiles, each of which can contain an arbitrary piece of mathematics or logic written by the teacher. These tiles resemble parts of the proof as it might be written on paper, and are translated into Isabelle/HOL'ws Isar syntax on the server. Unlike traditional syntax-directed editors, MathsTiles allow students to freely sketch out parts of an answer and do not constrain the order in which an answer is written. They also allow details of the language to change between or even during questions.

A number of smaller contributions are also presented. By using the dynamic nature of MathsTiles, a type of proof exercise is developed where the student must search for the statements he or she wishes to use. This allows questions to be supported by informal modelling, making them much easier to write, but still ensures that the interface does not act as a prop for the answer. The concept of searching for statements is extended to develop "massively multiple choice" questions: a mid-point between the multiple choice and short answer formats. The question architecture that is presented is applicable across different notational forms and different answer analysis techniques. The content architecture uses an informal ontology that enables students and untrained users to add and adapt content within the book, including adding their own chapters, while ensuring the content can also be referred to by the models and systems that advise students during exercises.

UCAM-CL-TR-720

Lauri I.W. Pesonen:

A capability-based access control architecture for multi-domain publish/subscribe systems

June 2008, 175 pages, PDF
PhD thesis (Wolfson College, December 2007)

Abstract: Publish/subscribe is emerging as the favoured communication paradigm for large-scale, wide-area distributed systems. The publish/subscribe many-to-many interaction model together with asynchronous messaging provides an efficient transport for highly distributed systems in high latency environments with direct peer-to-peer interactions amongst the participants.

Decentralised publish/subscribe systems implement the event service as a network of event brokers. The broker network makes the system more resilient to failures and allows it to scale up efficiently as the number of event clients increases. In many cases such distributed systems will only be feasible when implemented over the Internet as a joint effort spanning multiple administrative domains. The participating members will benefit from the federated event broker networks both with respect to the size of the system as well as its fault-tolerance.

Large-scale, multi-domain environments require access control; users will have different privileges for sending and receiving instances of different event types. Therefore, we argue that access control is vital for decentralised publish/subscribe systems, consisting of multiple independent administrative domains, to ever be deployable in large scale.

This dissertation presents MAIA, an access control mechanism for decentralised, type-based publish/subscribe systems. While the work concentrates on type-based publish/subscribe the contributions are equally applicable to both topic and content-based publish/subscribe systems.

Access control in distributed publish/subscribe requires secure, distributed naming, and mechanisms for enforcing access control policies. The first contribution of this thesis is a mechanism for names to be referenced unambiguously from policy without risk of forgeries. The second contribution is a model describing how signed capabilities can be used to grant domains and their members' access rights to event types in a scalable and expressive manner. The third contribution is a model for enforcing access control in the decentralised event service by encrypting event content.

We illustrate the design and implementation of MAIA with a running example of the UK Police Information Technology Organisation and the UK police forces.

UCAM-CL-TR-721

Ben W. Medlock:

Investigating classification for natural language processing tasks

June 2008, 138 pages, PDF
PhD thesis (Fitzwilliam College, September 2007)

Abstract: This report investigates the application of classification techniques to four natural language processing (NLP) tasks. The classification paradigm falls within the family of statistical and machine learning (ML) methods and consists of a framework within which a mechanical 'learner' induces a functional mapping between elements drawn from a particular sample space and a set of designated target classes. It is applicable to a wide range of NLP problems and has met with a great deal of success due to its flexibility and firm theoretical foundations.

The first task we investigate, topic classification, is firmly established within the NLP/ML communities as a benchmark application for classification research. Our aim is to arrive at a deeper understanding of how class granularity affects classification accuracy and to assess the impact of representational issues on different classification models. Our second task, content-based spam filtering, is a highly topical application for classification techniques due to the ever-worsening problem of unsolicited email. We assemble a new corpus and formulate a state-of-the-art classifier based on structured language model components. Thirdly, we introduce the problem of anonymisation, which has received little attention to date within the NLP community. We define the task in terms of obfuscating potentially sensitive references to real world entities and present a new publicly-available benchmark corpus. We explore the implications of the subjective nature of the problem and present an interactive model for anonymising large quantities of data based on syntactic analysis and active learning. Finally, we investigate the task of hedge classification, a relatively new application which is currently of growing interest due to the expansion of research into the application of NLP techniques to scientific literature for information extraction. A high level of annotation agreement is obtained using new guidelines and a new benchmark corpus is made publicly available. As part of our investigation, we develop a probabilistic model for training data acquisition within a semi-supervised learning framework which is explored both theoretically and experimentally.

Throughout the report, many common themes of fundamental importance to classification for NLP are addressed, including sample representation, performance evaluation, learning model selection, linguistically-motivated feature engineering, corpus construction and real-world application.

UCAM-CL-TR-722

Mbou Eyole-Monono:

Energy-efficient sentient computing

July 2008, 138 pages, PDF
PhD thesis (Trinity College, January 2008)

Abstract: In a bid to improve the interaction between computers and humans, it is becoming necessary to make increasingly larger deployments of sensor networks. These clusters of small electronic devices can be embedded in our surroundings and can detect and

react to physical changes. They will make computers more proactive in general by gathering and interpreting useful information about the physical environment through a combination of measurements. Increasing the performance of these devices will mean more intelligence can be embedded within the sensor network. However, most conventional ways of increasing performance often come with the burden of increased power dissipation which is not an option for energy-constrained sensor networks. This thesis proposes, develops and tests a design methodology for performing greater amounts of processing within a sensor network while satisfying the requirement for low energy consumption. The crux of the thesis is that there is a great deal of concurrency present in sensor networks which when combined with a tightly-coupled group of small, fast, energy-conscious processors can result in a significantly more efficient network. The construction of a multiprocessor system aimed at sensor networks is described in detail. It is shown that a routine critical to sensor networks can be sped up with the addition of a small set of primitives. The need for a very fast inter-processor communication mechanism is highlighted, and the hardware scheduler developed as part of this effort forms the cornerstone of the new sentient computing framework by facilitating thread operations and minimising the time required for context-switching. The experimental results also show that end-to-end latency can be reduced in a flexible way through multiprocessing.

UCAM-CL-TR-723

Richard Southern:

Animation manifolds for representing topological alteration

July 2008, 131 pages, PDF
PhD thesis (Clare Hall, February 2008)

Abstract: An animation manifold encapsulates an animation sequence of surfaces contained within a higher dimensional manifold with one dimension being time. An iso–surface extracted from this structure is a frame of the animation sequence.

In this dissertation I make an argument for the use of animation manifolds as a representation of complex animation sequences. In particular animation manifolds can represent transitions between shapes with differing topological structure and polygonal density.

I introduce the animation manifold, and show how it can be constructed from a keyframe animation sequence and rendered using raytracing or graphics hardware. I then adapt three Laplacian editing frameworks to the higher dimensional context. I derive new boundary conditions for both primal and dual Laplacian methods, and present a technique to adaptively regularise the sampling of a deformed manifold after editing.

The animation manifold can be used to represent a morph sequence between surfaces of arbitrary topology. I present a novel framework for achieving this by connecting planar cross sections in a higher dimension with a new constrained Delaunay triangulation. Topological alteration is achieved by using the Voronoi skeleton, a novel structure which provides a fast medial axis approximation.

UCAM-CL-TR-724

Ulrich Paquet:

Bayesian inference for latent variable models

July 2008, 137 pages, PDF
PhD thesis (Wolfson College, March 2007)

Abstract: Bayes' theorem is the cornerstone of statistical inference. It provides the tools for dealing with knowledge in an uncertain world, allowing us to explain observed phenomena through the refinement of belief in model parameters. At the heart of this elegant framework lie intractable integrals, whether in computing an average over some posterior distribution, or in determining the normalizing constant of a distribution. This thesis examines both deterministic and stochastic methods in which these integrals can be treated. Of particular interest shall be parametric models where the parameter space can be extended with additional latent variables to get distributions that are easier to handle algorithmically.

Deterministic methods approximate the posterior distribution with a simpler distribution over which the required integrals become tractable. We derive and examine a new generic α-divergence message passing scheme for a multivariate mixture of Gaussians, a particular modeling problem requiring latent variables. This algorithm minimizes local α-divergences over a chosen posterior factorization, and includes variational Bayes and expectation propagation as special cases.

Stochastic (or Monte Carlo) methods rely on a sample from the posterior to simplify the integration tasks, giving exact estimates in the limit of an infinite sample. Parallel tempering and thermodynamic integration are introduced as 'gold standard' methods to sample from multimodal posterior distributions and determine normalizing constants. A parallel tempered approach to sampling from a mixture of Gaussians posterior through Gibbs sampling is derived, and novel methods are introduced to improve the numerical stability of thermodynamic integration.

A full comparison with parallel tempering and thermodynamic integration shows variational Bayes, expectation propagation, and message passing with the Hellinger distance $\alpha = 1/2$ to be perfectly suitable for model selection, and for approximating the predictive distribution with high accuracy.

Variational and stochastic methods are combined in a novel way to design Markov chain Monte Carlo

(MCMC) transition densities, giving a variational transition kernel, which lower bounds an exact transition kernel. We highlight the general need to mix variational methods with other MCMC moves, by proving that the variational kernel does not necessarily give a geometrically ergodic chain.

UCAM-CL-TR-725

Hamed Haddadi, Damien Fay, Almerima Jamakovic, Olaf Maennel, Andrew W. Moore, Richard Mortier, Miguel Rio, Steve Uhlig:

Beyond node degree: evaluating AS topology models

July 2008, 16 pages, PDF

Abstract: Many models have been proposed to generate Internet Autonomous System (AS) topologies, most of which make structural assumptions about the AS graph. In this paper we compare AS topology generation models with several observed AS topologies. In contrast to most previous works, we avoid making assumptions about which topological properties are important to characterize the AS topology. Our analysis shows that, although matching degree-based properties, the existing AS topology generation models fail to capture the complexity of the local interconnection structure between ASs. Furthermore, we use BGP data from multiple vantage points to show that additional measurement locations significantly affect local structure properties, such as clustering and node centrality. Degree-based properties, however, are not notably affected by additional measurements locations. These observations are particularly valid in the core. The shortcomings of AS topology generation models stems from an underestimation of the complexity of the connectivity in the core caused by inappropriate use of BGP data.

UCAM-CL-TR-726

Viktor Vafeiadis:

Modular fine-grained concurrency verification

July 2008, 148 pages, PDF
PhD thesis (Selwyn College, July 2007)

Abstract: Traditionally, concurrent data structures are protected by a single mutual exclusion lock so that only one thread may access the data structure at any time. This coarse-grained approach makes it relatively easy to reason about correctness, but it severely limits parallelism. More advanced algorithms instead perform synchronisation at a finer grain. They employ sophisticated synchronisation schemes (both blocking and non-blocking) and are usually written in low-level languages such as C.

This dissertation addresses the formal verification of such algorithms. It proposes techniques that are modular (and hence scalable), easy for programmers to use, and yet powerful enough to verify complex algorithms. In doing so, it makes two theoretical and two practical contributions to reasoning about fine-grained concurrency.

First, building on rely/guarantee reasoning and separation logic, it develops a new logic, RGSep, that subsumes these two logics and enables simple, modular proofs of fine-grained concurrent algorithms that use complex dynamically allocated data structures and may explicitly deallocate memory. RGSep allows for ownership-based reasoning and ownership transfer between threads, while maintaining the expressiveness of binary relations to describe inter-thread interference.

Second, it describes techniques for proving linearisability, the standard correctness condition for fine-grained concurrent algorithms. The main proof technique is to introduce auxiliary single-assignment variables to capture the linearisation point and to inline the abstract effect of the program at that point as auxiliary code.

Third, it demonstrates this approach by proving linearisability of a collection of concurrent list and stack algorithms, as well as providing the first correctness proofs of the RDCSS and MCAS implementations of Harris et al.

Finally, it describes a prototype safety checker, SmallfootRG, for fine-grained concurrent programs that is based on RGSep. SmallfootRG proves simple safety properties for a number of list and stack algorithms and verifies the absence of memory leaks.

UCAM-CL-TR-727

Ruoshui Liu, Ian J. Wassell:

A novel auto-calibration system for wireless sensor motes

September 2008, 65 pages, PDF

Abstract: In recent years, Wireless Sensor Networks (WSNs) research has undergone a quiet revolution, providing a new paradigm for sensing and disseminating information from various environments. In reality, the wireless propagation channel in many harsh environments has a significant impact on the coverage range and quality of the radio links between the wireless nodes (motes). Therefore, the use of diversity techniques (e.g., frequency diversity and spatial diversity) must be considered to ameliorate the notoriously variable and unpredictable point-to-point radio communication links. However, in order to determine the space and frequency diversity characteristics of the channel,

accurate measurements need to be made. The most representative and inexpensive solution is to use motes, however they suffer poor accuracy owing to their low-cost and compromised radio frequency (RF) performance.

In this report we present a novel automated calibration system for characterising mote RF performance. The proposed strategy provides us with good knowledge of the actual mote transmit power, RSSI characteristics and receive sensitivity by establishing calibration tables for transmitting and receiving mote pairs over their operating frequency range. The validated results show that our automated calibration system can achieve an increase of ± 1.5 dB in the RSSI accuracy. In addition, measurements of the mote transmit power show a significant difference from that claimed in the manufacturer's data sheet. The proposed calibration method can also be easily applied to wireless sensor motes from virtually any vendor, provided they have a RF connector.

UCAM-CL-TR-728

Peter J.C. Brown, Christopher T. Faigle:

A robust efficient algorithm for point location in triangulations

February 1997, 16 pages, PDF

Abstract: This paper presents a robust alternative to previous approaches to the problem of point location in triangulations represented using the quadedge data structure. We generalise the reasons for the failure of an earlier routine to terminate when applied to certain non-Delaunay triangulations. This leads to our new deterministic algorithm which we prove is guaranteed to terminate. We also present a novel heuristic for choosing a starting edge for point location queries and show that this greatly enhances the efficiency of point location for the general case.

UCAM-CL-TR-729

Damien Fay, Hamed Haddadi, Steve Uhlig, Andrew W. Moore, Richard Mortier, Almerima Jamakovic:

Weighted spectral distribution

September 2008, 13 pages, PDF

Abstract: Comparison of graph structures is a frequently encountered problem across a number of problem domains. Comparing graphs requires a metric to discriminate which features of the graphs are considered important. The spectrum of a graph is often claimed to contain all the information within a graph, but the raw spectrum contains too much information to be directly used as a useful metric. In this paper we introduce a metric, the weighted spectral distribution, that improves on the raw spectrum by discounting those eigenvalues believed to be unimportant and emphasizing the contribution of those believed to be important.

We use this metric to optimize the selection of parameter values for generating Internet topologies. Our metric leads to parameter choices that appear sensible given prior knowledge of the problem domain: the resulting choices are close to the default values of the topology generators and, in the case of the AB generator, fall within the expected region. This metric provides a means for meaningfully optimizing parameter selection when generating topologies intended to share structure with, but not match exactly, measured graphs.

UCAM-CL-TR-730

Robert J. Ennals:

Adaptive evaluation of non-strict programs

August 2008, 243 pages, PDF
PhD thesis (King's College, June 2004)

Abstract: Most popular programming languages are strict. In a strict language, the binding of a variable to an expression coincides with the evaluation of the expression.

Non-strict languages attempt to make life easier for programmers by decoupling expression binding and expression evaluation. In a non-strict language, a variable can be bound to an unevaluated expression, and such expressions can be passed around just like values in a strict language. This separation allows the programmer to declare a variable at the point that makes most logical sense, rather than at the point at which its value is known to be needed.

Non-strict languages are usually evaluated using a technique called Lazy Evaluation. Lazy Evaluation will only evaluate an expression when its value is known to be needed. While Lazy Evaluation minimises the total number of expressions evaluated, it imposes a considerable bookkeeping overhead, and has unpredictable space behaviour.

In this thesis, we present a new evaluation strategy which we call Optimistic Evaluation. Optimistic Evaluation blends lazy and eager evaluation under the guidance of an online profiler. The online profiler observes the running program and decides which expressions should be evaluated lazily, and which should be evaluated eagerly. We show that the worst case performance of Optimistic Evaluation relative to Lazy Evaluation can be bounded with an upper bound chosen by the user. Increasing this upper bound allows the profiler to take greater risks and potentially achieve better average performance.

This thesis describes both the theory and practice of Optimistic Evaluation. We start by giving an overview

of Optimistic Evaluation. We go on to present a formal model, which we use to justify our design. We then detail how we have implemented Optimistic Evaluation as part of an industrial-strength compiler. Finally, we provide experimental results to back up our claims.

UCAM-CL-TR-731

Matthew Johnson:

A new approach to Internet banking

September 2008, 113 pages, PDF
PhD thesis (Trinity Hall, July 2008)

Abstract: This thesis investigates the protection landscape surrounding online banking. First, electronic banking is analysed for vulnerabilities and a survey of current attacks is carried out. This is represented graphically as an attack tree describing the different ways in which online transactions can be attacked.

The discussion then moves on to various defences which have been developed, categorizing them and analyzing how successful they are at protecting against the attacks given in the first chapter. This covers everything from TLS encryption through phishing site detection to two-factor authentication.

Having declared all current schemes for protecting online banking lacking in some way, the key aspects of the problem are identified. This is followed by a proposal for a more robust defence system which uses a small security device to create a trusted path to the customer, rather than depend upon trusting the customer's computer. The protocol for this system is described along with all the other restrictions required for actual use. This is followed by a description of a demonstration implementation of the system.

Extensions to the system are then proposed, designed to afford extra protection for the consumer and also to support other types of device. There is then a discussion of ways of managing keys in a heterogeneous system, rather than one managed by a single entity.

The conclusion discusses the weaknesses of the proposed scheme and evaluates how successful it is likely to be in practice and what barriers there may be to adoption in the banking system.

UCAM-CL-TR-732

Alban Rrustemi:

Computing surfaces – a platform for scalable interactive displays

November 2008, 156 pages, PDF
PhD thesis (St Edmund's College, November 2008)

Abstract: Recent progress in electronic, display and sensing technologies makes possible a future with omnipresent, arbitrarily large interactive display surfaces. Nonetheless, current methods of designing display systems with multi-touch sensitivity do not scale. This thesis presents computing surfaces as a viable platform for resolving forthcoming scalability limitations.

Computing surfaces are composed of a homogeneous network of physically adjoined, small sensitive displays with local computation and communication capabilities. In this platform, inherent scalability is provided by a distributed architecture. The regular spatial distribution of resources presents new demands on the way surface input and output information is managed and processed.

Direct user input with touch based gestures needs to account for the distributed architecture of computing surfaces. A scalable middleware solution that conceals the tiled architecture is proposed for reasoning with touch-based gestures. The validity of this middleware is proven in a case study, where a fully distributed algorithm for online recognition of unistrokes – a particular class of touch-based gestures – is presented and evaluated.

Novel interaction techniques based around interactive display surfaces involve direct manipulation with displayed digital objects. In order to facilitate such interactions in computing surfaces, an efficient distributed algorithm to perform 2D image transformations is introduced and evaluated. The performance of these transformations is heavily influenced by the arbitration policies of the interconnection network. One approach for improving the performance of these transformations in conventional network architectures is proposed and evaluated.

More advanced applications in computing surfaces require the presence of some notion of time. An efficient algorithm for internal time synchronisation is presented and evaluated. A hardware solution is adopted to minimise the delay uncertainty of special timestamp messages. The proposed algorithm allows efficient, scalable time synchronisation among clusters of tiles. A hardware reference platform is constructed to demonstrate the basic principles and features of computing surfaces. This platform and a complementary simulation environment is used for extensive evaluation and analysis.

UCAM-CL-TR-733

Darren Edge:

Tangible user interfaces for peripheral interaction

December 2008, 237 pages, PDF
PhD thesis (Jesus College, January 2008)

Abstract: Since Mark Weiser's vision of ubiquitous computing in 1988, many research efforts have been made to move computation away from the workstation

and into the world. One such research area focuses on "Tangible" User Interfaces or TUIs – those that provide both physical representation and control of underlying digital information.

This dissertation describes how TUIs can support a "peripheral" style of interaction, in which users engage in short, dispersed episodes of low-attention interaction with digitally-augmented physical tokens. The application domain in which I develop this concept is the office context, where physical tokens can represent items of common interest to members of a team whose work is mutually interrelated, but predominantly performed independently by individuals at their desks.

An "analytic design process" is introduced as a way of developing TUI designs appropriate for their intended contexts of use. This process is then used to present the design of a bimanual desktop TUI that complements the existing workstation, and encourages peripheral interaction in parallel with workstation-intensive tasks. Implementation of a prototype TUI is then described, comprising "task" tokens for work-time management, "document" tokens for face-to-face sharing of collaborative documents, and "contact" tokens for awareness of other team members' status and workload. Finally, evaluation of this TUI is presented via description of its extended deployment in a real office context.

The main empirically-grounded results of this work are a categorisation of the different ways in which users can interact with physical tokens, and an identification of the qualities of peripheral interaction that differentiate it from other interaction styles. The foremost benefits of peripheral interaction were found to arise from the freedom with which tokens can be appropriated to create meaningful information structures of both cognitive and social significance, in the physical desktop environment and beyond.

UCAM-CL-TR-734

Philip Tuddenham:

Tabletop interfaces for remote collaboration

December 2008, 243 pages, PDF
PhD thesis (Gonville and Caius College, June 2008)

Abstract: Effective support for synchronous remote collaboration has long proved a desirable yet elusive goal for computer technology. Although video views showing the remote participants have recently improved, technologies providing a shared visual workspace of the task still lack support for the visual cues and work practices of co-located collaboration.

Researchers have recently demonstrated shared workspaces for remote collaboration using large horizontal interactive surfaces. These remote tabletop interfaces may afford the beneficial work practices associated with co-located collaboration around tables. However, there has been little investigation of remote tabletop interfaces beyond limited demonstrations. There is currently little theoretical basis for their design, and little empirical characterisation of their support for collaboration. The construction of remote tabletop applications also presents considerable technical challenges.

This dissertation addresses each of these areas. Firstly, a theory of workspace awareness is applied to consider the design of remote tabletop interfaces and the work practices that they may afford.

Secondly, two technical barriers to the rapid exploration of useful remote tabletop applications are identified: the low resolution of conventional tabletop displays; and the lack of support for existing user interface components. Techniques from multi-projector display walls are applied to address these problems. The resulting method is evaluated empirically and used to create a number of novel tabletop interfaces.

Thirdly, an empirical investigation compares remote and co-located tabletop interfaces. The findings show how the design of remote tabletop interfaces leads to collaborators having a high level of awareness of each other's actions in the workspace. This enables smooth transitions between individual and group work, together with anticipation and assistance, similar to co-located tabletop collaboration. However, remote tabletop collaborators use different coordination mechanisms from co-located collaborators. The results have implications for the design and future study of these interfaces.

UCAM-CL-TR-735

Diarmuid Ó Séaghdha:

Learning compound noun semantics

December 2008, 167 pages, PDF
PhD thesis (Corpus Christi College, July 2008)

Abstract: This thesis investigates computational approaches for analysing the semantic relations in compound nouns and other noun-noun constructions. Compound nouns in particular have received a great deal of attention in recent years due to the challenges they pose for natural language processing systems. One reason for this is that the semantic relation between the constituents of a compound is not explicitly expressed and must be retrieved from other sources of linguistic and world knowledge.

I present a new scheme for the semantic annotation of compounds, describing in detail the motivation for the scheme and the development process. This scheme is applied to create an annotated dataset for use in compound interpretation experiments. The results of a dual-annotator experiment indicate that good agreement can be obtained with this scheme relative to previously reported results and also provide insights into the challenging nature of the annotation task.

I describe two corpus-driven paradigms for comparing pairs of nouns: lexical similarity and relational similarity. Lexical similarity is based on comparing each

constituent of a noun pair to the corresponding constituent of another pair. Relational similarity is based on comparing the contexts in which both constituents of a noun pair occur together with the corresponding contexts of another pair. Using the flexible framework of kernel methods, I develop techniques for implementing both similarity paradigms.

A standard approach to lexical similarity represents words by their co-occurrence distributions. I describe a family of kernel functions that are designed for the classification of probability distributions. The appropriateness of these distributional kernels for semantic tasks is suggested by their close connection to proven measures of distributional lexical similarity. I demonstrate the effectiveness of the lexical similarity model by applying it to two classification tasks: compound noun interpretation and the 2007 SemEval task on classifying semantic relations between nominals.

To implement relational similarity I use kernels on strings and sets of strings. I show that distributional set kernels based on a multinomial probability model can be computed many times more efficiently than previously proposed kernels, while still achieving equal or better performance. Relational similarity does not perform as well as lexical similarity in my experiments. However, combining the two models brings an improvement over either model alone and achieves state-of-the-art results on both the compound noun and SemEval Task 4 datasets.

UCAM-CL-TR-736

Mike Dodds, Xinyu Feng,
Matthew Parkinson, Viktor Vafeiadis:

Deny-guarantee reasoning

January 2009, 82 pages, PDF

Abstract: Rely-guarantee is a well-established approach to reasoning about concurrent programs that use parallel composition. However, parallel composition is not how concurrency is structured in real systems. Instead, threads are started by 'fork' and collected with 'join' commands. This style of concurrency cannot be reasoned about using rely-guarantee, as the life-time of a thread can be scoped dynamically. With parallel composition the scope is static.

In this paper, we introduce deny-guarantee reasoning, a reformulation of rely-guarantee that enables reasoning about dynamically scoped concurrency. We build on ideas from separation logic to allow interference to be dynamically split and recombined, in a similar way that separation logic splits and joins heaps. To allow this splitting, we use deny and guarantee permissions: a deny permission specifies that the environment cannot do an action, and guarantee permission allow us to do an action. We illustrate the use of our proof system with examples, and show that it can encode all the original rely-guarantee proofs. We also present the semantics and soundness of the deny-guarantee method.

UCAM-CL-TR-737

Na Xu:

Static contract checking for Haskell

December 2008, 175 pages, PDF
PhD thesis (Churchill College, August 2008)

Abstract: Program errors are hard to detect and are costly, to both programmers who spend significant efforts in debugging, and for systems that are guarded by runtime checks. Static verification techniques have been applied to imperative and object-oriented languages, like Java and C#, for checking basic safety properties such as memory leaks. In a pure functional language, many of these basic properties are guaranteed by design, which suggests the opportunity for verifying more sophisticated program properties. Nevertheless, few automatic systems for doing so exist. In this thesis, we show the challenges and solutions to verifying advanced properties of a pure functional language, Haskell. We describe a sound and automatic static verification framework for Haskell, that is based on contracts and symbolic execution. Our approach gives precise blame assignments at compile-time in the presence of higher-order functions and laziness.

First, we give a formal definition of contract satisfaction which can be viewed as a denotational semantics for contracts. We then construct two contract checking wrappers, which are dual to each other, for checking the contract satisfaction. We prove the soundness and completeness of the construction of the contract checking wrappers with respect to the definition of the contract satisfaction. This part of my research shows that the two wrappers are projections with respect to a partial ordering crashes-more-often and furthermore, they form a projection pair and a closure pair. These properties give contract checking a strong theoretical foundation.

As the goal is to detect bugs during compile time, we symbolically execute the code constructed by the contract checking wrappers and prove the soundness of this approach. We also develop a technique named counter-example-guided (CEG) unrolling which only unroll function calls on demand. This technique speeds up the checking process.

Finally, our verification approach makes error tracing much easier compared with the existing set-based analysis. Thus equipped, we are able to tell programmers during compile-time which function to blame and why if there is a bug in their program. This is a breakthrough for lazy languages because it is known to be difficult to report such informative messages either at compile-time or run-time.

UCAM-CL-TR-738

Scott Fairbanks:

High precision timing using self-timed circuits

January 2009, 99 pages, PDF
PhD thesis (Gonville and Caius College, September 2004)

Abstract: Constraining the events that demarcate periods on a VLSI chip to precise instances of time is the task undertaken in this thesis. High speed sampling and clock distribution are two example applications. Foundational to my approach is the use of self-timed data control circuits.

Specially designed self-timed control circuits deliver high frequency timing signals with precise phase relationships. The frequency and the phase relationships are controlled by varying the number of self-timed control stages and the number of tokens they control.

The self-timed control circuits are constructed with simple digital logic gates. The digital logic gates respond to a range of analog values with a continuum of precise and controlled delays. The control circuits implement their functionality efficiently. This allows the gates to drive long wires and distribute the timing signals over a large area. Also gate delays are short and few, allowing for high frequencies.

The self-timed control circuits implement the functionality of a FIFO that is then closed into a ring. Timing tokens ripple through the rings. The FIFO stages use digital handshaking protocols to pass the timing tokens between the stages. The FIFO control stage detects the phase between the handshake signals on its inputs and produces a signal that is sent back to the producers with a delay that is a function of the phase relationship of the input signals.

The methods described are not bound to the same process and systematic skew limitations of existing methods. For a certain power budget, timing signals are generated and distributed with significantly less power with the approaches to be presented than with conventional methods.

UCAM-CL-TR-739

Salman Taherian:

State-based Publish/Subscribe for sensor systems

January 2009, 240 pages, PDF
PhD thesis (St John's College, June 2008)

Abstract: Recent technological advances have enabled the creation of networks of sensor devices. These devices are typically equipped with basic computational and communication capabilities. Systems based on these devices can deduce high-level, meaningful information about the environment that may be useful to applications. Due to their scale, distributed nature, and the limited resources available to sensor devices, these systems are inherently complex. Shielding applications from this complexity is a challenging problem.

To address this challenge, I present a middleware called SPS (State-based Publish/Subscribe). It is based on a combination of a State-Centric data model and a Publish/Subscribe (Pub/Sub) communication paradigm. I argue that a state-centric data model allows applications to specify environmental situations of interest in a more natural way than existing solutions. In addition, Pub/Sub enables scalable many-to-many communication between sensors, actuators, and applications.

This dissertation initially focuses on Resource-constrained Sensor Networks (RSNs) and proposes State Filters (SFs), which are lightweight, stateful, event filtering components. Their design is motivated by the redundancy and correlation observed in sensor readings produced close together in space and time. By performing context-based data processing, SFs increase Pub/Sub expressiveness and improve communication efficiency.

Secondly, I propose State Maintenance Components (SMCs) for capturing more expressive conditions in heterogeneous sensor networks containing more resourceful devices. SMCs extend SFs with data fusion and temporal and spatial data manipulation capabilities. They can also be composed together (in a DAG) to deduce higher level information. SMCs operate independently from each other and can therefore be decomposed for distributed processing within the network.

Finally, I present a Pub/Sub protocol called QPS (Quad-PubSub) for location-aware Wireless Sensor Networks (WSNs). QPS is central to the design of my framework as it facilitates messaging between state-based components, applications, sensors, and actuators. In contrast to existing data dissemination protocols, QPS has a layered architecture. This allows for the transparent operation of routing protocols that meet different Quality of Service (QoS) requirements.

UCAM-CL-TR-740

Tal Sobol-Shikler:

Analysis of affective expression in speech

January 2009, 163 pages, PDF
PhD thesis (Girton College, March 2007)

Abstract: This dissertation presents analysis of expressions in speech. It describes a novel framework for dynamic recognition of acted and naturally evoked expressions and its application to expression mapping and to multi-modal analysis of human-computer interactions.

The focus of this research is on analysis of a wide range of emotions and mental states from non-verbal expressions in speech. In particular, on inference of complex mental states, beyond the set of basic emotions, including naturally evoked subtle expressions and mixtures of expressions.

This dissertation describes a bottom-up computational model for processing of speech signals. It combines the application of signal processing, machine learning and voting methods with novel approaches to the design, implementation and validation. It is based on a comprehensive framework that includes all the development stages of a system. The model represents paralinguistic speech events using temporal abstractions borrowed from various disciplines such as musicology, engineering and linguistics. The model consists of a flexible and expandable architecture. The validation of the model extends its scope to different expressions, languages, backgrounds, contexts and applications.

The work adapts an approach that an utterance is not an isolated entity but rather a part of an interaction and should be analysed in this context. The analysis in context includes relations to events and other behavioural cues. Expressions of mental states are related not only in time but also by their meaning and content. This work demonstrates the relations between the lexical definitions of mental states, taxonomies and theoretical conceptualization of mental states and their vocal correlates. It examines taxonomies and theoretical conceptualisation of mental states in relation to their vocal characteristics. The results show that a very wide range of mental state concepts can be mapped, or described, using a high-level abstraction in the form of a small sub-set of concepts which are characterised by their vocal correlates.

This research is an important step towards comprehensive solutions that incorporate social intelligence cues for a wide variety of applications and for multidisciplinary research.

UCAM-CL-TR-741

David N. Cottingham:

Vehicular wireless communication

January 2009, 264 pages, PDF
PhD thesis (Churchill College, September 2008)

Abstract: Transportation is vital in everyday life. As a consequence, vehicles are increasingly equipped with onboard computing devices. Moreover, the demand for connectivity to vehicles is growing rapidly, both from business and consumers. Meanwhile, the number of wireless networks available in an average city in the developed world is several thousand. Whilst this theoretically provides near-ubiquitous coverage, the technology type is not homogeneous.

This dissertation discusses how the diversity in communication systems can be best used by vehicles. Focussing on road vehicles, it first details the technologies available, the difficulties inherent in the vehicular environment, and how intelligent handover algorithms could enable seamless connectivity. In particular, it identifies the need for a model of the coverage of wireless networks.

In order to construct such a model, the use of vehicular sensor networks is proposed. The Sentient Van, a platform for vehicular sensing, is introduced, and details are given of experiments carried out concerning the performance of IEEE 802.11x, specifically for vehicles. Using the Sentient Van, a corpus of 10 million signal strength readings was collected over three years. This data, and further traces, are used in the remainder of the work described, thus distinguishing it in using entirely real world data.

Algorithms are adapted from the field of 2-D shape simplification to the problem of processing thousands of signal strength readings. By applying these to the data collected, coverage maps are generated that contain extents. These represent how coverage varies between two locations on a given road. The algorithms are first proven fit for purpose using synthetic data, before being evaluated for accuracy of representation and compactness of output using real data.

The problem of how to select the optimal network to connect to is then addressed. The coverage map representation is converted into a multi-planar graph, where the coverage of all available wireless networks is included. This novel representation also includes the ability to hand over between networks, and the penalties so incurred. This allows the benefits of connecting to a given network to be traded off with the cost of handing over to it.

In order to use the multi-planar graph, shortest path routing is used. The theory underpinning multi-criteria routing is overviewed, and a family of routing metrics developed. These generate efficient solutions to the problem of calculating the sequence of networks that should be connected to over a given geographical route. The system is evaluated using real traces, finding that in 75% of the test cases proactive routing algorithms provide better QoS than a reactive algorithm. Moreover, the system can also be run to generate geographical routes that are QoS-aware.

This dissertation concludes by examining how coverage mapping can be applied to other types of data, and avenues for future research are proposed.

UCAM-CL-TR-742

Thomas Ridge, Michael Norrish, Peter Sewell:

TCP, UDP, and Sockets: Volume 3: The Service-level Specification

February 2009, 305 pages, PDF

Abstract: Despite more than 30 years of research on protocol specification, the major protocols deployed in the Internet, such as TCP, are described only in informal prose RFCs and executable code. In part this is because the scale and complexity of these protocols makes them challenging targets for formal descriptions, and because

techniques for mathematically rigorous (but appropriately loose) specification are not in common use.

In this work we show how these difficulties can be addressed. We develop a high-level specification for TCP and the Sockets API, describing the byte-stream service that TCP provides to users, expressed in the formalised mathematics of the HOL proof assistant. This complements our previous low-level specification of the protocol internals, and makes it possible for the first time to state what it means for TCP to be correct: that the protocol implements the service. We define a precise abstraction function between the models and validate it by testing, using verified testing infrastructure within HOL. Some errors may remain, of course, especially as our resources for testing were limited, but it would be straightforward to use the method on a larger scale. This is a pragmatic alternative to full proof, providing reasonable confidence at a relatively low entry cost.

Together with our previous validation of the low-level model, this shows how one can rigorously tie together concrete implementations, low-level protocol models, and specifications of the services they claim to provide, dealing with the complexity of real-world protocols throughout.

Similar techniques should be applicable, and even more valuable, in the design of new protocols (as we illustrated elsewhere, for a MAC protocol for the SWIFT optically switched network). For TCP and Sockets, our specifications had to capture the historical complexities, whereas for a new protocol design, such specification and testing can identify unintended complexities at an early point in the design.

UCAM-CL-TR-743

Rebecca F. Watson:

Optimising the speed and accuracy of a Statistical GLR Parser

March 2009, 145 pages, PDF
PhD thesis (Darwin College, September 2007)

Abstract: The focus of this thesis is to develop techniques that optimise both the speed and accuracy of a unification-based statistical GLR parser. However, we can apply these methods within a broad range of parsing frameworks. We first aim to optimise the level of tag ambiguity resolved during parsing, given that we employ a front-end PoS tagger. This work provides the first broad comparison of tag models as we consider both tagging and parsing performance. A dynamic model achieves the best accuracy and provides a means to overcome the trade-off between tag error rates in single tag per word input and the increase in parse ambiguity over multipletag per word input. The second line of research describes a novel modification to the inside-outside algorithm, whereby multiple inside and outside probabilities are assigned for elements within the packed parse forest data structure. This algorithm enables us to compute a set of 'weighted GRs' directly from this structure. Our experiments demonstrate substantial increases in parser accuracy and throughput for weighted GR output.

Finally, we describe a novel confidence-based training framework, that can, in principle, be applied to any statistical parser whose output is defined in terms of its consistency with a given level and type of annotation. We demonstrate that a semisupervised variant of this framework outperforms both Expectation-Maximisation (when both are constrained by unlabelled partial-bracketing) and the extant (fully supervised) method. These novel training methods utilise data automatically extracted from existing corpora. Consequently, they require no manual effort on behalf of the grammar writer, facilitating grammar development.

UCAM-CL-TR-744

Anna Ritchie:

Citation context analysis for information retrieval

March 2009, 119 pages, PDF
PhD thesis (New Hall, June 2008)

Abstract: This thesis investigates taking words from around citations to scientific papers in order to create an enhanced document representation for improved information retrieval. This method parallels how anchor text is commonly used in Web retrieval. In previous work, words from citing documents have been used as an alternative representation of the cited document but no previous experiment has combined them with a full-text document representation and measured effectiveness in a large scale evaluation.

The contributions of this thesis are twofold: firstly, we present a novel document representation, along with experiments to measure its effect on retrieval effectiveness, and, secondly, we document the construction of a new, realistic test collection of scientific research papers, with references (in the bibliography) and their associated citations (in the running text of the paper) automatically annotated. Our experiments show that the citation-enhanced document representation increases retrieval effectiveness across a range of standard retrieval models and evaluation measures.

In Chapter 2, we give the background to our work, discussing the various areas from which we draw together ideas: information retrieval, particularly link structure analysis and anchor text indexing, and bibliometrics, in particular citation analysis. We show that there is a close relatedness of ideas between these areas but that these ideas have not been fully explored experimentally. Chapter 3 discusses the test collection paradigm for evaluation of information retrieval systems and describes how and why we built our test collection. In Chapter 4 we introduce the ACL Anthology,

the archive of computational linguistics papers that our test collection is centred around. The archive contains the most prominent publications since the beginning of the field in the early 1960s, consisting of one journal plus conferences and workshops, resulting in over 10,000 papers. Chapter 5 describes how the PDF papers are prepared for our experiments, including identification of references and citations in the papers, once converted to plain text, and extraction of citation information to an XML database. Chapter 6 presents our experiments: we show that adding citation terms to the full-text of the papers improves retrieval effectiveness by up to 7.4%, that weighting citation terms higher relative to paper terms increases the improvement and that varying the context from which citation terms are taken has a significant effect on retrieval effectiveness. Our main hypothesis that citation terms enhance a full-text representation of scientific papers is thus proven.

There are some limitations to these experiments. The relevance judgements in our test collection are incomplete but we have experimentally verified that the test collection is, nevertheless, a useful evaluation tool. Using the Lemur toolkit constrained the method that we used to weight citation terms; we would like to experiment with a more realistic implementation of term weighting. Our experiments with different citation contexts did not conclude an optimal citation context; we would like to extend the scope of our investigation. Now that our test collection exists, we can address these issues in our experiments and leave the door open for more extensive experimentation.

UCAM-CL-TR-745

Scott Owens, Susmit Sarkar, Peter Sewell:

A better x86 memory model: x86-TSO (extended version)

March 2009, 52 pages, PDF

Abstract: Real multiprocessors do not provide the sequentially consistent memory that is assumed by most work on semantics and verification. Instead, they have relaxed memory models, typically described in ambiguous prose, which lead to widespread confusion. These are prime targets for mechanized formalization. In previous work we produced a rigorous x86-CC model, formalizing the Intel and AMD architecture specifications of the time, but those turned out to be unsound with respect to actual hardware, as well as arguably too weak to program above. We discuss these issues and present a new x86-TSO model that suffers from neither problem, formalized in HOL4. We believe it is sound with respect to real processors, reflects better the vendor's intentions, and is also better suited for programming. We give two equivalent definitions of x86-TSO: an intuitive operational model based on local write buffers, and an axiomatic total store ordering model, similar to that of the SPARCv8. Both are adapted to handle x86-specific features. We have implemented the axiomatic model in our memevents tool, which calculates the set of all valid executions of test programs, and, for greater confidence, verify the witnesses of such executions directly, with code extracted from a third, more algorithmic, equivalent version of the definition.

UCAM-CL-TR-746

Shishir Nagaraja, Ross Anderson:

The snooping dragon: social-malware surveillance of the Tibetan movement

March 2009, 12 pages, PDF

Abstract: In this note we document a case of malware-based electronic surveillance of a political organisation by the agents of a nation state. While malware attacks are not new, two aspects of this case make it worth serious study. First, it was a targeted surveillance attack designed to collect actionable intelligence for use by the police and security services of a repressive state, with potentially fatal consequences for those exposed. Second, the modus operandi combined social phishing with high-grade malware. This combination of well-written malware with well-designed email lures, which we call social malware, is devastatingly effective. Few organisations outside the defence and intelligence sector could withstand such an attack, and although this particular case involved the agents of a major power, the attack could in fact have been mounted by a capable motivated individual. This report is therefore of importance not just to companies who may attract the attention of government agencies, but to all organisations. As social-malware attacks spread, they are bound to target people such as accounts-payable and payroll staff who use computers to make payments. Prevention will be hard. The traditional defence against social malware in government agencies involves expensive and intrusive measures that range from mandatory access controls to tiresome operational security procedures. These will not be sustainable in the economy as a whole. Evolving practical low-cost defences against social-malware attacks will be a real challenge.

UCAM-CL-TR-747

Fei Song, Hongke Zhang, Sidong Zhang, Fernando Ramos, Jon Crowcroft:

An estimator of forward and backward delay for multipath transport

March 2009, 16 pages, PDF

Abstract: Multipath transport protocols require awareness of the capability of different paths being used for transmission. It is well known that round trip time (RTT) can be used to estimate retransmission timeout with reasonable accuracy. However, using RTT to evaluate the delay of forward or backward paths is not always suitable. In fact, these paths are usually dissimilar, and therefore the packet delay can be significantly different in each direction.

We propose a forward and backward delay estimator that aims to solve this problem. Based on the results of the estimator, a new retransmission heuristic mechanism for multipath transport is proposed. With this same technique we also build two other heuristics: A bottleneck bandwidth estimator and a shared congestion detector. These help the sender to choose the high bandwidth path in retransmission and ensure TCP-friendliness in multipath transport, respectively.

UCAM-CL-TR-748

Marco Canini, Wei Li, Andrew W. Moore:

GTVS: boosting the collection of application traffic ground truth

April 2009, 20 pages, PDF

Abstract: Interesting research in the areas of traffic classification, network monitoring, and application-orient analysis can not proceed with real trace data labeled with actual application information. However, hand-labeled traces are an extremely valuable but scarce resource in the traffic monitoring and analysis community, as a result of both privacy concerns and technical difficulties: hardly any possibility exists for payloaded data to be released to the public, while the intensive labor required for getting the ground-truth application information from the data severely constrains the feasibility of releasing anonymized versions of hand-labeled payloaded data.

The usual way to obtain the ground truth is fragile, inefficient and not directly comparable from one's work to another. This chapter proposes and details a methodology that significantly boosts the efficiency in compiling the application traffic ground truth. In contrast with other existing work, our approach maintains the high certainty as in hand-verification, while striving to save time and labor required for that. Further, it is implemented as an easy hands-on tool suite which is now freely available to the public.

In this paper we present a case study using a 30 minute real data trace to guide the readers through our ground-truth classification process. We also present a method, which is an extension of GTVS that efficiently classifies HTTP traffic by its purpose.

UCAM-CL-TR-749

Pan Hui, Eiko Yoneki, Jon Crowcroft, Shu-Yan Chan:

Identifying social communities in complex communications for network efficiency

May 2009, 14 pages, PDF

Abstract: Complex communication networks, more particular Mobile Ad Hoc Networks (MANET) and Pocket Switched Networks (PSN), rely on short range radio and device mobility to transfer data across the network. These kind of mobile networks contain duality in nature: they are radio networks at the same time also human networks, and hence knowledge from social networks can be also applicable here. In this paper, we demonstrate how identifying social communities can significantly improve the forwarding efficiencies in term of delivery ratio and delivery cost. We verify our hypothesis using data from five human mobility experiments and test on two application scenarios, asynchronous messaging and publish/subscribe service.

UCAM-CL-TR-750

Marco Canini, Wei Li, Martin Zadnik, Andrew W. Moore:

AtoZ: an automatic traffic organizer using NetFPGA

May 2009, 27 pages, PDF

Abstract: This paper introduces AtoZ, an automatic traffic organizer that provides end-users with control of how their applications use network resources. Such an approach contrasts with the moves of many ISPs towards network-wide application throttling and provider-centric control of an application's network-usage. AtoZ provides seamless per-application traffic-organizing on gigabit links, with minimal packet-delays and no unintended packet drops.

The AtoZ combines the high-speed packet processing of the NetFPGA with an efficient flow-behavior identification method. Currently users can enable AtoZ control over network resources by prohibiting certain applications and controlling the priority of others. We discuss deployment experience and use real traffic to illustrate how such an architecture enables several distinct features: high accuracy, high throughput, minimal delay, and efficient packet labeling – all in a low-cost, robust configuration that works alongside the home or enterprise access-router.

UCAM-CL-TR-751

David C. Turner:

Nominal domain theory for concurrency

July 2009, 185 pages, PDF
PhD thesis (Clare College, December 2008)

Abstract: Domain theory provides a powerful mathematical framework for describing sequential computation, but the traditional tools of domain theory are inapplicable to concurrent computation. Without a general mathematical framework it is hard to compare developments and approaches from different areas of study, leading to time and effort wasted in rediscovering old ideas in new situations.

A possible remedy to this situation is to build a denotational semantics based directly on computation paths, where a process denotes the set of paths that it may follow. This has been shown to be a remarkably powerful idea, but it lacks certain computational features. Notably, it is not possible to express the idea of names and name-generation within this simple path semantics.

Nominal set theory is a non-standard mathematical foundation that captures the notion of names in a general way. Building a mathematical development on top of nominal set theory has the effect of incorporating names into its fabric at a low level. Importantly, nominal set theory is sufficiently close to conventional foundations that it is often straightforward to transfer intuitions into the nominal setting.

Here the original path-based domain theory for concurrency is developed within nominal set theory, which has the effect of systematically adjoining name-generation to the model. This gives rise to an expressive metalanguage, Nominal HOPLA, which supports a notion of name-generation. Its denotational semantics is given entirely in terms of universal constructions on domains. An operational semantics is also presented, and relationships between the denotational and operational descriptions are explored.

The generality of this approach to including name generation into a simple semantic model indicates that it will be possible to apply the same techniques to more powerful domain theories for concurrency, such as those based on presheaves.

UCAM-CL-TR-752

Gerhard P. Hancke:

Security of proximity identification systems

July 2009, 161 pages, PDF
PhD thesis (Wolfson College, February 2008)

Abstract: RFID technology is the prevalent method for implementing proximity identification in a number of security sensitive applications. The perceived proximity of a token serves as a measure of trust and is often used as a basis for granting certain privileges or services. Ensuring that a token is located within a specified distance of the reader is therefore an important security requirement. In the case of high-frequency RFID systems the limited operational range of the near-field communication channel is accepted as implicit proof that a token is in close proximity to a reader. In some instances, it is also presumed that this limitation can provide further security services.

The first part of this dissertation presents attacks against current proximity identification systems. It documents how eavesdropping, skimming and relay attacks can be implemented against HF RFID systems. Experimental setups and practical results are provided for eavesdropping and skimming attacks performed against RFID systems adhering to the ISO 14443 and ISO 15693 standards. These attacks illustrate that the limited operational range cannot prevent unauthorised access to stored information on the token, or ensure that transmitted data remains confidential. The practical implementation of passive and active relay attacks against an ISO 14443 RFID system is also described. The relay attack illustrates that proximity identification should not rely solely on the physical characteristics of the communication channel, even if it could be shown to be location-limited. As a result, it is proposed that additional security measures, such as distance-bounding protocols, should be incorporated to verify proximity claims. A new method, using cover noise, is also proposed to make the backward communication channel more resistant to eavesdropping attacks.

The second part of this dissertation discusses distance-bounding protocols. These protocols determine an upper bound for the physical distance between two parties. A detailed survey of current proposals, investigating their respective merits and weaknesses, identifies general principles governing secure distance-bounding implementations. It is practically shown that an attacker can circumvent the distance bound by implementing attacks at the packet and physical layer of conventional communication channels. For this reason the security of a distance bound depends not only on the cryptographic protocol, but also on the time measurement provided by the underlying communication. Distance-bounding protocols therefore require special channels. Finally, a new distance-bounding protocol and a practical implementation of a suitable distance-bounding channel for HF RFID systems are proposed.

UCAM-CL-TR-753

John L. Miller, Jon Crowcroft:

Carbon: trusted auditing for P2P distributed virtual environments

August 2009, 20 pages, PDF

Abstract: Many Peer-to-Peer Distributed Virtual Environments (P2P DVE's) have been proposed, but none are widely deployed. One significant barrier to deployment is lack of security. This paper presents Carbon, a trusted auditing system for P2P DVE's which provides reasonable security with low per-client overhead. DVE's using Carbon perform offline auditing to evaluate DVE client correctness. Carbon audits can be used to catch DVE clients which break DVE rules – cheaters – so the DVE can punish them. We analyze the impact of applying Carbon to a peer-to-peer game with attributes similar to World of Warcraft. We show that 99.9% of cheaters – of a certain profile – can be caught with guided auditing and 2.3% bandwidth overhead, or 100% of cheaters can be caught with exhaustive auditing and 27% bandwidth overhead. The surprisingly low overhead for exhaustive auditing is the result of the small payload in most DVE packet updates, compared to the larger aggregate payloads in audit messages. Finally, we compare Carbon to PeerReview, and show that for DVE scenarios Carbon consumes significantly less resources – in typical cases by an order of magnitude – while sacrificing little protection.

UCAM-CL-TR-754

Frank Stajano, Paul Wilson:

Understanding scam victims: seven principles for systems security

August 2009, 22 pages, PDF
An updated, abridged and peer-reviewed version of this report appeared in Communications of the ACM 54(3):70-75, March 2011 [doi:10.1145/1897852.1897872]. Please cite the refereed CACM version in any related work.

Abstract: The success of many attacks on computer systems can be traced back to the security engineers not understanding the psychology of the system users they meant to protect. We examine a variety of scams and "short cons" that were investigated, documented and recreated for the BBC TV programme The Real Hustle and we extract from them some general principles about the recurring behavioural patterns of victims that hustlers have learnt to exploit.

We argue that an understanding of these inherent "human factors" vulnerabilities, and the necessity to take them into account during design rather than naïvely shifting the blame onto the "gullible users", is a fundamental paradigm shift for the security engineer which, if adopted, will lead to stronger and more resilient systems security.

UCAM-CL-TR-755

Yujian Gao, Aimin Hao, Qinping Zhao, Neil A. Dodgson:

Skin-detached surface for interactive large mesh editing

September 2009, 18 pages, PDF

Abstract: We propose a method for interactive deformation of large detailed meshes. Our method allows the users to manipulate the mesh directly using freely-selected handles on the mesh. To best preserve surface details, we introduce a new surface representation, the skin-detached surface. It represents a detailed surface model as a peeled "skin" added over a simplified surface model. The "skin" contains the details of the surface while the simplified mesh maintains the basic shape. The deformation process consists of three steps: At the mesh loading stage, the "skin" is precomputed according to the detailed mesh and detached from the simplified mesh. Then we deform the simplified mesh following the nonlinear gradient domain mesh editing approach to satisfy the handle position constraints. Finally the detailed "skin" is remapped onto the simplified mesh, resulting in a deformed detailed mesh. We investigate the advantages as well as the limitations of our method by implementing a prototype system and applying it to several examples.

UCAM-CL-TR-756

Hamed Haddadi, Damien Fay, Steve Uhlig, Andrew W. Moore, Richard Mortier, Almerima Jamakovic:

Analysis of the Internet's structural evolution

September 2009, 13 pages, PDF

Abstract: In this paper we study the structural evolution of the AS topology as inferred from two different datasets over a period of seven years. We use a variety of topological metrics to analyze the structural differences revealed in the AS topologies inferred from the two different datasets. In particular, to focus on the evolution of the relationship between the core and the periphery, we make use of the weighted spectral distribution.

We find that the traceroute dataset has increasing difficulty in sampling the periphery of the AS topology, largely due to limitations inherent to active probing. Such a dataset has too limited a view to properly observe topological changes at the AS-level compared to a dataset largely based on BGP data. We also highlight limitations in current measurements that require a better sampling of particular topological properties of the Internet. Our results indicate that the Internet is changing from a core-centered, strongly customer-provider oriented, disassortative network, to a soft-hierarchical, peering-oriented, assortative network.

UCAM-CL-TR-757

Mark Adcock:

Improving cache performance by runtime data movement

July 2009, 174 pages, PDF
PhD thesis (Christ's College, June 2009)

Abstract: The performance of a recursive data structure (RDS) increasingly depends on good data cache behaviour, which may be improved by software/hardware prefetching or by ensuring that the RDS has a good data layout. The latter is harder but more effective, and requires solving two separate problems: firstly ensuring that new RDS nodes are allocated in a good location in memory, and secondly preventing a degradation in layout when the RDS changes shape due to pointer updates.

The first problem has been studied in detail, but only two major classes of solutions to the second exist. Layout degradation may be side-stepped by using a 'cache-aware' RDS, one designed to have inherently good cache behaviour (e.g. using a B-Tree in place of a binary search tree), but such structures are difficult to devise and implement. A more automatic solution in some languages is to use a 'layout-improving' garbage collector, which attempt to improve heap data layout during collection using online profiling of data access patterns. This may carry large performance, memory and latency overheads.

In this thesis we investigate the insertion of code into a program which attempts to move RDS nodes at runtime to prevent or reduce layout degradation. Such code affects only the performance of a program not its semantics. The body of this thesis is a thorough and systematic evaluation of three different forms of data movement. The first method adapts existing work on static RDS data layout, performing ad-hoc single node movements at a program's pointer-update sites, which is simple to apply and effective in practice, but the performance gain may be hard to predict. The second method performs infrequent movement of larger groups of nodes, borrowing techniques from garbage collection but also embedding data movement in existing traversals of the RDS; the benefit of performing additional data movement to compact the heap is also demonstrated. The third method restores a pre-chosen layout after each RDS pointer update, which is a complex but effective technique, and may be viewed both as an optimisation and as a way of synthesising new cache-aware RDSs.

Concentrating on both maximising performance while minimising latency and extra memory usage, two fundamental RDSs are used for the investigation, representative of two common data access patterns (linear and branching). The methods of this thesis compare favourably to upper bounds on performance and to the canonical cache-aware solutions. This thesis shows the value of runtime data movement, and as well as producing optimisation useful in their own right may be used to guide the design of future cache-aware RDSs and layout-improving garbage collectors.

UCAM-CL-TR-758

Alexey Gotsman:

Logics and analyses for concurrent heap-manipulating programs

October 2009, 160 pages, PDF
PhD thesis (Churchill College, September 2009)

Abstract: Reasoning about concurrent programs is difficult because of the need to consider all possible interactions between concurrently executing threads. The problem is especially acute for programs that manipulate shared heap-allocated data structures, since heap-manipulation provides more ways for threads to interact. Modular reasoning techniques sidestep this difficulty by considering every thread in isolation under some assumptions on its environment.

In this dissertation we develop modular program logics and program analyses for the verification of concurrent heap-manipulating programs. Our approach is to exploit reasoning principles provided by program logics to construct modular program analyses and to use this process to obtain further insights into the logics. In particular, we build on concurrent separation logic—a Hoare-style logic that allows modular manual reasoning about concurrent programs written in a simple heap-manipulating programming language.

Our first contribution is to show the soundness of concurrent separation logic without the conjunction rule and the restriction that resource invariants be precise, and to construct an analysis for concurrent heap-manipulating programs that exploits this modified reasoning principle to achieve modularity. The analysis can be used to automatically verify a number of safety properties, including memory safety, data-structure integrity, data-race freedom, the absence of memory leaks, and the absence of assertion violations. We show that we can view the analysis as generating proofs in our variant of the logic, which enables the use of its results in proof-carrying code or theorem proving systems.

Reasoning principles expressed by program logics are most often formulated for only idealised programming constructs. Our second contribution is to develop logics and analyses for modular reasoning about features present in modern languages and libraries for concurrent programming: storable locks (i.e., locks dynamically created and destroyed in the heap), first-order procedures, and dynamically-created threads.

UCAM-CL-TR-759

Eric Koskinen, Matthew Parkinson,
Maurice Herlihy:

Coarse-grained transactions (extended version)

August 2011, 34 pages, PDF

Abstract: Traditional transactional memory systems suffer from overly conservative conflict detection, yielding so-called false conflicts, because they are based on fine-grained, low-level read/write conflicts. In response, the recent trend has been toward integrating various abstract data-type libraries using ad-hoc methods of high-level conflict detection. These proposals have led to improved performance but a lack of a unified theory has led to confusion in the literature.

We clarify these recent proposals by defining a generalization of transactional memory in which a transaction consists of course-grained (abstract data-type) operations rather than simply memory read/write operations. We provide semantics for both pessimistic (e.g. transactional boosting) and optimistic (e.g. traditional TMs and recent alternatives) execution. We show that both are included in the standard atomic semantics, yet find that the choice imposes different requirements on the coarse-grained operations: pessimistic requires operations be left-movers, optimistic requires right-movers. Finally, we discuss how the semantics applies to numerous TM implementation details discussed widely in the literature.

UCAM-CL-TR-760

Alan F. Blackwell, Lee Wilson, Alice Street,
Charles Boulton, John Knell:

Radical innovation: crossing knowledge boundaries with interdisciplinary teams

November 2009, 124 pages, PDF

Abstract: Interdisciplinary innovation arises from the positive effects that result when stepping across the social boundaries that we structure knowledge by. Those boundaries include academic disciplines, government departments, companies' internal functions, companies and sectors, and the boundaries between these domains. In the knowledge economy, it is often the case that the right knowledge to solve a problem is in a different place to the problem itself, so interdisciplinary innovation is an essential tool for the future. There are also many problems today that need more than one kind of knowledge to solve them, so interdisciplinary innovation is also an essential tool for the challenging problems of today.

This report presents the results of an in-depth study into successful interdisciplinary innovation, focusing on the personal experiences of the people who achieve it. It is complementary to organisational research, and to research on the economic impact of innovation, but has primarily adopted perspectives and methods from other disciplines. Instead, this report has been developed by a team that is itself interdisciplinary, with a particular focus on anthropology, design research, and strategic policy. It also draws on reports from expert witnesses and invited commentators in many other fields.

UCAM-CL-TR-761

Jonathan J. Davies:

Programming networks of vehicles

November 2009, 292 pages, PDF
PhD thesis (Churchill College, September 2008)

Abstract: As computers become smaller in size and advances in communications technology are made, we hypothesise that a new range of applications involving computing in road vehicles will emerge. These applications may be enabled by the future arrival of general-purpose computing platforms in vehicles. Many of these applications will involve the collection, processing and distribution of data sampled by sensors on large numbers of vehicles. This dissertation is primarily concerned with addressing how these applications can be designed and implemented by programmers.

We explore how a vehicular sensor platform may be built and how data from a variety of sensors can be sampled and stored. Applications exploiting such platforms will infer higher-level information from the raw sensor data collected. We present the design and implementation of one such application which involves processing vehicles' location histories into an up-to-date road map.

Our experience shows that there is a problem with programming this kind of application: the number of vehicles and the nature of computational infrastructure available are not known until the application is executed. By comparison, existing approaches to programming applications in wireless sensor networks tend to assume that the nature of the network architecture is known at design-time. This is not an appropriate assumption to make in vehicular sensor networks. Instead, this dissertation proposes that the functionality of applications is designed and implemented at a higher level and the problem of deciding how and where its components are to be executed is left to a compiler. We call this 'late physical binding'.

This approach brings the benefit that applications can be automatically adapted and optimised for execution in a wide range of environments. We describe a suite of transformations which can change the order in which components of the program are executed whilst preserving its semantic integrity. These transformations

may affect several of the application's characteristics such as its execution time or energy consumption.

The practical utility of this approach is demonstrated through a novel programming language based on Java. Two examples of diverse applications are presented which demonstrate that the language and compiler can be used to create non-trivial applications. Performance measurements show that the compiler can introduce parallelism to make more efficient use of resources and reduce an application's execution time. One of the applications belongs to a class of distributed systems beyond merely processing vehicular sensor data, suggesting that the late physical binding paradigm has broader application to other areas of distributed computing.

UCAM-CL-TR-762

Evangelia Kalyvianaki:

Resource provisioning for virtualized server applications

November 2009, 161 pages, PDF
PhD thesis (Lucy Cavendish College, August 2008)

Abstract: Data centre virtualization creates an agile environment for application deployment. Applications run within one or more virtual machines and are hosted on various servers throughout the data centre. One key mechanism provided by modern virtualization technologies is dynamic resource allocation. Using this technique virtual machines can be allocated resources as required and therefore, occupy only the necessary resources for their hosted application. In fact, two of the main challenges faced by contemporary data centres, server consolidation and power saving, can be tackled efficiently by capitalising on this mechanism.

This dissertation shows how to dynamically adjust the CPU resources allocated to virtualized server applications in the presence of workload fluctuations. In particular it employs a reactive approach to resource provisioning based on feedback control and introduces five novel controllers. All five controllers adjust the application allocations based on past utilisation observations.

A subset of the controllers integrate the Kalman filtering technique to track the utilisations and based on which they predict the allocations for the next interval. This approach is particularly attractive for the resource management problem since the Kalman filter uses the evolution of past utilisations to adjust the allocations. In addition, the adaptive Kalman controller which adjusts its parameters online and dynamically estimates the utilisation dynamics, is able to differentiate substantial workload changes from small fluctuations for unknown workloads.

In addition, this dissertation captures, models, and builds controllers based on the CPU resource coupling of application components. In the case of multi-tier applications, these controllers collectively allocate resources to all application tiers detecting saturation points across components. This results in them acting faster to workload variations than their single-tier counterparts.

All controllers are evaluated against the Rubis benchmark application deployed on a prototype virtualized cluster built for this purpose.

UCAM-CL-TR-763

Saar Drimer:

Security for volatile FPGAs

November 2009, 169 pages, PDF
PhD thesis (Darwin College, August 2009)

Abstract: With reconfigurable devices fast becoming complete systems in their own right, interest in their security properties has increased. While research on "FPGA security" has been active since the early 2000s, few have treated the field as a whole, or framed its challenges in the context of the unique FPGA usage model and application space. This dissertation sets out to examine the role of FPGAs within a security system and how solutions to security challenges can be provided. I offer the following contributions:

I motivate authenticating configurations as an additional capability to FPGA configuration logic, and then describe a flexible security protocol for remote reconfiguration of FPGA-based systems over insecure networks. Non-volatile memory devices are used for persistent storage when required, and complement the lack of features in some FPGAs with tamper proofing in order to maintain specified security properties. A unique advantage of the protocol is that it can be implemented on some existing FPGAs (i.e., it does not require FPGA vendors to add functionality to their devices). Also proposed is a solution to the "IP distribution problem" where designs from multiple sources are integrated into a single bitstream, yet must maintain their confidentiality.

I discuss the difficulty of reproducing and comparing FPGA implementation results reported in the academic literature. Concentrating on cryptographic implementations, problems are demonstrated through designing three architecture-optimized variants of the AES block cipher and analyzing the results to show that single figures of merit, namely "throughput" or "throughput per slice", are often meaningless without the context of an application. To set a precedent for reproducibility in our field, the HDL source code, simulation testbenches and compilation instructions are made publicly available for scrutiny and reuse.

Finally, I examine payment systems as ubiquitous embedded devices, and evaluate their security vulnerabilities as they interact in a multi-chip environment. Using FPGAs as an adversarial tool, a man-in-the-middle attack against these devices is demonstrated. An FPGA-based defense is also demonstrated: the first secure wired "distance bounding" protocol implementation.

This is then put in the context of securing reconfigurable systems.

UCAM-CL-TR-764

Caroline V. Gasperin:

Statistical anaphora resolution in biomedical texts

December 2009, 124 pages, PDF
PhD thesis (Clare Hall, August 2008)

Abstract: This thesis presents a study of anaphora in biomedical scientific literature and focuses on tackling the problem of anaphora resolution in this domain. Biomedical literature has been the focus of many information extraction projects; there are, however, very few works on anaphora resolution in biomedical scientific full-text articles. Resolving anaphora is an important step in the identification of mentions of biomedical entities about which information could be extracted.

We have identified coreferent and associative anaphoric relations in biomedical texts. Among associative relations we were able to distinguish 3 main types: biotype, homolog and set-member relations. We have created a corpus of biomedical articles that are annotated with anaphoric links between noun phrases referring to biomedical entities of interest. Such noun phrases are typed according to a scheme that we have developed based on the Sequence Ontology; it distinguishes 7 types of entities: gene, part of gene, product of gene, part of product, subtype of gene, supertype of gene and gene variant.

We propose a probabilistic model for the resolution of anaphora in biomedical texts. The model seeks to find the antecedents of anaphoric expressions, both coreferent and associative, and also to identify discourse-new expressions. The model secures good performance despite being trained on a small corpus: it achieves 55-73% precision and 57-63% recall on coreferent cases, and reasonable performance on different classes of associative cases. We compare the performance of the model with a rule-based baseline system that we have also developed, a naive Bayes system and a decision trees system, showing that the ours outperforms the others.

We have experimented with active learning in order to select training samples to improve the performance of our probabilistic model. It was not, however, more successful than random sampling.

UCAM-CL-TR-765

Magnus O. Myreen:

Formal verification of machine-code programs

December 2009, 109 pages, PDF
PhD thesis (Trinity College, December 2008)

Abstract: Formal program verification provides mathematical means of increasing assurance for the correctness of software. Most approaches to program verification are either fully automatic and prove only weak properties, or alternatively are manual and labour intensive to apply; few target realistically modelled machine code. The work presented in this dissertation aims to ease the effort required in proving properties of programs on top of detailed models of machine code. The contributions are novel approaches for both verification of existing programs and methods for automatically constructing correct code.

For program verification, this thesis presents a new approach based on translation: the problem of proving properties of programs is reduced, via fully-automatic deduction, to a problem of proving properties of recursive functions. The translation from programs to recursive functions is shown to be implementable in a theorem prover both for simple while-programs as well as real machine code. This verification-after-translation approach has several advantages over established approaches of verification condition generation. In particular, the new approach does not require annotating the program with assertions. More importantly, the proposed approach separates the verification proof from the underlying model so that specific resource names, some instruction orderings and certain control-flow structures become irrelevant. As a result proof reuse is enabled to a greater extent than in currently used methods. The scalability of this new approach is illustrated through the verification of ARM, x86 and PowerPC implementations of a copying garbage collector.

For construction of correct code this thesis presents a new compiler which maps functions from logic, via proof, down to multiple carefully modelled commercial machine languages. Unlike previously published work on compilation from higher-order logic, this compiler allows input functions to be partially specified and supports a broad range of user-defined extensions. These features enabled the production of formally verified machine-code implementations of a LISP interpreter, as a case study.

The automation and proofs have been implemented in the HOL4 theorem prover, using a new machine-code Hoare triple instantiated to detailed specifications of ARM, x86 and PowerPC instruction set architectures.

UCAM-CL-TR-766

Julian M. Smith:

Towards robust inexact geometric computation

December 2009, 186 pages, PDF
PhD thesis (St. Edmund's College, July 2009)

Abstract: Geometric algorithms implemented using rounded arithmetic are prone to robustness problems.

Geometric algorithms are often a mix of arithmetic and combinatorial computations, arising from the need to create geometric data structures that are themselves a complex mix of numerical and combinatorial data. Decisions that influence the topology of a geometric structure are made on the basis of certain arithmetic calculations, but the inexactness of these calculations may lead to inconsistent decisions, causing the algorithm to produce a topologically invalid result or to fail catastrophically. The research reported here investigates ways to produce robust algorithms with inexact computation.

I present two algorithms for operations on piecewise linear (polygonal/polyhedral) shapes. Both algorithms are topologically robust, meaning that they are guaranteed to generate a topologically valid result from a topologically valid input, irrespective of numerical errors in the computations. The first algorithm performs the Boolean operation in 3D, and also in 2D. The main part of this algorithm is a series of interdependent operations. The relationship between these operations ensures a consistency in these operations, which, I prove, guarantees the generation of a shape representation with valid topology. The basic algorithm may generate geometric artifacts such as gaps and slivers, which generally can be removed by a data-smoothing post-process. The second algorithm presented performs simplification in 2D, converting a geometrically invalid (but topologically valid) shape representation into one that is fully valid. This algorithm is based on a variant of the Bentley-Ottmann sweep line algorithm, but with additional rules to handle situations not possible under an exact implementation.

Both algorithms are presented in the context of what is required of an algorithm in order for it to be classed as robust in some sense. I explain why the formulaic approach used for the Boolean algorithm cannot readily be used for the simplification process. I also give essential code details for a C++ implementation of the 2D simplification algorithm, and discuss the results of extreme tests designed to show up any problems. Finally, I discuss floating-point arithmetic, present error analysis for the floating-point computation of the intersection point between two segments in 2D, and discuss how such errors affect both the simplification algorithm and the basic Boolean algorithm in 2D.

UCAM-CL-TR-767

Massimo Ostilli, Eiko Yoneki,
Ian X. Y. Leung, Jose F. F. Mendes,
Pietro Lió, Jon Crowcroft:

Ising model of rumour spreading in interacting communities

January 2010, 24 pages, PDF

Abstract: We report a preliminary investigation on interactions between communities in a complex network using the Ising model to analyse the spread of information among real communities. The inner opinion of a given community is forced to change through the introduction of a unique external source and we analyse how the other communities react to this change. We model two conceptual external sources: namely, "Strong-belief", and "propaganda", by an infinitely strong inhomogeneous external field and a finite uniform external field, respectively. In the former case, the community changes independently from other communities while in the latter case according also to interactions with the other communities. We apply our model to synthetic networks as well as various real world data ranging from human physical contact networks to online social networks. The experimental results using real world data clearly demonstrate two distinct scenarios of phase transitions characterised by the presence of strong memory effects when the graph and coupling parameters are above a critical threshold.

UCAM-CL-TR-768

Cecily Morrison, Adona Iosif, Miklos Danka:

Report on existing open-source electronic medical records

February 2010, 12 pages, PDF

Abstract: In this report we provide an overview of existing open-source electronic medical records and assess them against the criteria established by the EViDence group.

UCAM-CL-TR-769

Sriram Srinivasan:

Kilim: A server framework with lightweight actors, isolation types and zero-copy messaging

February 2010, 127 pages, PDF
PhD thesis (King's College, February 2010)

Abstract: Internet services are implemented as hierarchical aggregates of communicating components: networks of data centers, networks of clusters in a data center, connected servers in a cluster, and multiple virtual machines on a server machine, each containing several operating systems processes. This dissertation argues for extending this structure to the intra-process level, with networks of communicating actors. An actor is a single-threaded state machine with a private heap and a thread of its own. It communicates with other actors using well-defined and explicit messaging protocols. Actors must be light enough to comfortably match the inherent concurrency in the problem space, and to exploit all available parallelism. Our aims are two-fold: (a) to treat SMP systems as they really are:

distributed systems with eventual consistency, and (b) recognize from the outset that a server is always part of a larger collection of communicating components, thus eliminating the mindset mismatch between concurrent programming and distributed programming.

Although the actor paradigm is by no means new, our design points are informed by drawing parallels between the macro and micro levels. As with components in a distributed system, we expect that actors must be isolatable in a number of ways: memory isolation, fault isolation, upgrade isolation, and execution isolation. The application should be able to have a say in actor placement and scheduling, and actors must be easily monitorable.

Our primary contribution is in showing that these requirements can be satisfied in a language and environment such as Java, without changes to the source language or to the virtual machine, and without leaving much of the idiomatic ambit of Java, with its mindset of pointers and mutable state. In other words, one does not have to move to a concurrency-oriented language or to an entirely immutable object paradigm.

We demonstrate an open-source toolkit called Kilim that provides (a) ultra-lightweight actors (faster and lighter than extant environments such as Erlang), (b) a type system that guarantees memory isolation between threads by separating internal objects from exportable messages and by enforcing ownership and structural constraints on the latter (linearity and tree-structure, respectively) and, (c) a library with I/O support and customizable synchronization constructs and schedulers.

We show that this solution is simpler to program than extant solutions, yet statically guaranteed to be free of low-level data races. It is also faster, more scalable and more stable (in increasing scale) in two industrial strength evaluations: interactive web services (comparing Kilim Web Server to Jetty) and databases (comparing Berkeley DB to a Kilim variant of it).

UCAM-CL-TR-770

Jatinder Singh:

Controlling the dissemination and disclosure of healthcare events

February 2010, 193 pages, PDF
PhD thesis (St. John's College, September 2009)

Abstract: Information is central to healthcare: for proper care, information must be shared. Modern healthcare is highly collaborative, involving interactions between users from a range of institutions, including primary and secondary care providers, researchers, government and private organisations. Each has specific data requirements relating to the service they provide, and must be informed of relevant information as it occurs.

Personal health information is highly sensitive. Those who collect/hold data as part of the care process are responsible for protecting its confidentiality, in line with patient consent, codes of practice and legislation. Ideally, one should receive only that information necessary for the tasks they perform—on a need-to-know basis.

Healthcare requires mechanisms to strictly control information dissemination. Many solutions fail to account for the scale and heterogeneity of the environment. Centrally managed data services impede the local autonomy of health institutions, impacting security by diminishing accountability and increasing the risks/impacts of incorrect disclosures. Direct, synchronous (request-response) communication requires an enumeration of every potential information source/sink. This is impractical when considering health services at a national level. Healthcare presents a data-driven environment highly amenable to an event-based infrastructure, which can inform, update and alert relevant parties of incidents as they occur. Event-based data dissemination paradigms, while efficient and scalable, generally lack the rigorous access control mechanisms required for health infrastructure.

This dissertation describes how publish/subscribe, an asynchronous, push-based, many-to-many middleware communication paradigm, is extended to include mechanisms for actively controlling information disclosure. We present Interaction Control: a data-control layer above a publish/subscribe service allowing the definition of context-aware policy rules to authorise information channels, transform information and restrict data propagation according to the circumstances. As dissemination policy is defined at the broker-level and enforced by the middleware, client compliance is ensured. Although policy enforcement involves extra processing, we show that in some cases the control mechanisms can actually improve performance over a general publish/subscribe implementation. We build Interaction Control mechanisms into integrated database-brokers to provide a rich representation of state; while facilitating audit, which is essential for accountability.

Healthcare requires the sharing of sensitive information across federated domains of administrative control. Interaction Control provides the means for balancing the competing concerns of information sharing and protection. It enables those responsible for information to meet their data management obligations, through specification of fine-grained disclosure policy.

UCAM-CL-TR-771

Cecily Morrison:

Bodies-in-Space: investigating technology usage in co-present group interaction

March 2010, 147 pages, PDF
PhD thesis (Darwin College, August 2009)

Abstract: With mobile phones in people's pockets, digital devices in people's homes, and information systems in group meetings at work, technology is frequently present when people interact with each other. Unlike devices used by a single person at a desk, people, rather than machines, are the main focus in social settings. An important difference then between these two scenarios, individual and group, is the role of the body. Although non-verbal behaviour is not part of human-computer interaction, it is very much part of human-human interaction. This dissertation explores bodies-in-space — people's use of spatial and postural positioning of their bodies to maintain a social interaction when technology is supporting the social interaction of a co-present group.

I begin this dissertation with a review of literature, looking at how and when bodies-in-space have been accounted for in research and design processes of technology for co-present groups. I include examples from both human-computer interaction, as well the social sciences more generally. Building on this base, the following four chapters provide examples and discussion of methods to: (1) see (analytically), (2) notate, (3) adjust (choreograph), and (4) research in the laboratory, bodies-in-space. I conclude with reflections on the value of capturing bodies-in-space in the process of designing technology for co-present groups and emphasise a trend towards end-user involvement and its consequences for the scope of human-computer interaction research.

All of the research in this dissertation derives from, and relates to, the real-world context of an intensive care unit of a hospital and was part of assessing the deployment of an electronic patient record.

UCAM-CL-TR-772

Matthew R. Lakin:

An executable meta-language for inductive definitions with binders

March 2010, 171 pages, PDF
PhD thesis (Queens' College, March 2010)

Abstract: A testable prototype can be invaluable for identifying bugs during the early stages of language development. For such a system to be useful in practice it should be quick and simple to generate prototypes from the language specification.

This dissertation describes the design and development of a new programming language called alphaML, which extends traditional functional programming languages with specific features for producing correct, executable prototypes. The most important new features of alphaML are for the handling of names and binding structures in user-defined languages. To this end, alphaML uses the techniques of nominal sets (due to Pitts and Gabbay) to represent names explicitly and handle binding correctly up to alpha-renaming. The language also provides built-in support for constraint solving and non-deterministic search.

We begin by presenting a generalised notion of systems defined by a set of schematic inference rules. This is our model for the kind of languages that might be implemented using alphaML. We then present the syntax, type system and operational semantics of the alphaML language and proceed to define a sound and complete embedding of schematic inference rules. We turn to program equivalence and define a standard notion of operational equivalence between alphaML expressions and use this to prove correctness results about the representation of data terms involving binding and about schematic formulae and inductive definitions.

The fact that binding can be represented correctly in alphaML is interesting for technical reasons, because the language dispenses with the notion of globally distinct names present in most systems based on nominal methods. These results, along with the encoding of inference rules, constitute the main technical payload of the dissertation. However, our approach complicates the solving of constraints between terms. Therefore, we develop a novel algorithm for solving equality and freshness constraints between nominal terms which does not rely on standard devices such as swappings and suspended permutations. Finally, we discuss an implementation of alphaML, and conclude with a summary of the work and a discussion of possible future extensions.

UCAM-CL-TR-773

Thomas J. Cashman:

NURBS-compatible subdivision surfaces

March 2010, 99 pages, PDF
PhD thesis (Queens' College, January 2010)

Abstract: Two main technologies are available to design and represent freeform surfaces: Non-Uniform Rational B-Splines (NURBS) and subdivision surfaces. Both representations are built on uniform B-splines, but they extend this foundation in incompatible ways, and different industries have therefore established a preference for one representation over the other. NURBS are the dominant standard for Computer-Aided Design, while subdivision surfaces are popular for applications in animation and entertainment. However there are benefits of subdivision surfaces (arbitrary topology) which would be useful within Computer-Aided Design, and features of NURBS (arbitrary degree and non-uniform parametrisations) which would make good additions to current subdivision surfaces.

I present NURBS-compatible subdivision surfaces, which combine topological freedom with the ability to represent any existing NURBS surface exactly. Subdivision schemes that extend either non-uniform or general-degree B-spline surfaces have appeared before, but this dissertation presents the first surfaces able to handle both challenges simultaneously. To achieve this

I develop a novel factorisation of knot insertion rules for non-uniform, general-degree B-splines.

Many subdivision surfaces have poor second-order behaviour near singularities. I show that it is possible to bound the curvatures of the general-degree subdivision surfaces created using my factorisation. Bounded-curvature surfaces have previously been created by 'tuning' uniform low-degree subdivision schemes; this dissertation shows that general-degree schemes can be tuned in a similar way. As a result, I present the first general-degree subdivision schemes with bounded curvature at singularities.

Previous subdivision schemes, both uniform and non-uniform, have inserted knots indiscriminately, but the factorised knot insertion algorithm I describe in this dissertation grants the flexibility to insert knots selectively. I exploit this flexibility to preserve convexity in highly non-uniform configurations, and to create locally uniform regions in place of non-uniform knot intervals. When coupled with bounded-curvature modifications, these techniques give the first non-uniform subdivision schemes with bounded curvature.

I conclude by combining these results to present NURBS-compatible subdivision surfaces: arbitrary-topology, non-uniform and general-degree surfaces which guarantee high-quality second-order surface properties.

UCAM-CL-TR-774

John Wickerson, Mike Dodds, Matthew Parkinson:

Explicit stabilisation for modular rely-guarantee reasoning

March 2010, 29 pages, PDF

Abstract: We propose a new formalisation of stability for Rely-Guarantee, in which an assertion's stability is encoded into its syntactic form. This allows two advances in modular reasoning. Firstly, it enables Rely-Guarantee, for the first time, to verify concurrent libraries independently of their clients' environments. Secondly, in a sequential setting, it allows a module's internal interference to be hidden while verifying its clients. We demonstrate our approach by verifying, using RGSep, the Version 7 Unix memory manager, uncovering a twenty-year-old bug in the process.

UCAM-CL-TR-775

Anil Madhavapeddy:

Creating high-performance, statically type-safe network applications

March 2010, 169 pages, PDF
PhD thesis (Robinson College, April 2006)

Abstract: A typical Internet server finds itself in the middle of a virtual battleground, under constant threat from worms, viruses and other malware seeking to subvert the original intentions of the programmer. In particular, critical Internet servers such as OpenSSH, BIND and Sendmail have had numerous security issues ranging from low-level buffer overflows to subtle protocol logic errors. These problems have cost billions of dollars as the growth of the Internet exposes increasing numbers of computers to electronic malware. Despite the decades of research on techniques such as model-checking, type-safety and other forms of formal analysis, the vast majority of server implementations continue to be written unsafely and informally in C/C++.

In this dissertation we propose an architecture for constructing new implementations of standard Internet protocols which integrates mature formal methods not currently used in deployed servers: (i) static type systems from the ML family of functional languages; (ii) model checking to verify safety properties exhaustively about aspects of the servers; and (iii) generative meta-programming to express high-level constraints for the domain-specific tasks of packet parsing and constructing non-deterministic state machines. Our architecture—dubbed MELANGE—is based on Objective Caml and contributes two domain-specific languages: (i) the Meta Packet Language (MPL), a data description language used to describe the wire format of a protocol and output statically type-safe code to handle network traffic using high-level functional data structures; and (ii) the Statecall Policy Language (SPL) for constructing non-deterministic finite state automata which are embedded into applications and dynamically enforced, or translated into PROMELA and statically model-checked.

Our research emphasises the importance of delivering efficient, portable code which is feasible to deploy across the Internet. We implemented two complex protocols—SSH and DNS—to verify our claims, and our evaluation shows that they perform faster than their standard counterparts OpenSSH and BIND, in addition to providing static guarantees against some classes of errors that are currently a major source of security problems.

UCAM-CL-TR-776

Kathryn E. Gray, Alan Mycroft:

System tests from unit tests

March 2010, 27 pages, PDF

Abstract: Large programs have bugs; software engineering practices reduce the number of bugs in deployed systems by relying on a combination of unit tests, to filter out bugs in individual procedures, and system tests, to identify bugs in an integrated system.

Our previous work showed how Floyd-Hoare triples, {P}C{Q}, could also be seen as unit tests, i.e.

formed a link between verification and test-based validation. A transactional-style implementation allows test post-conditions to refer to values of data structures both before and after test execution. Here we argue that this style of specifications, with a transactional implementation, provide a novel source of system tests.

Given a set of unit tests for a system, we can run programs in test mode on real data. Based on an analysis of the unit tests, we intersperse the program's execution with the pre- and post-conditions from the test suite to expose bugs or incompletenesses in either the program or the test suite itself. We use the results of these tests, as well as branch-trace coverage information, to identify and report anomalies in the running program.

UCAM-CL-TR-777

Thomas Dinsdale-Young, Mike Dodds, Philippa Gardner, Matthew Parkinson, Viktor Vafeiadis:

Concurrent Abstract Predicates

April 2010, 43 pages, PDF

Abstract: Abstraction is key to understanding and reasoning about large computer systems. Abstraction is simple to achieve if the relevant data structures are disjoint, but rather difficult when they are partially shared, as is the case for concurrent modules. We present a program logic for reasoning abstractly about data structures, that provides a fiction of disjointness and permits compositional reasoning. The internal details of a module are completely hidden from the client by concurrent abstract predicates. We reason about a module's implementation using separation logic with permissions, and provide abstract specifications for use by client programs using concurrent abstract predicates. We illustrate our abstract reasoning by building two implementations of a lock module on top of hardware instructions, and two implementations of a concurrent set module on top of the lock module.

UCAM-CL-TR-778

Viktor Vafeiadis:

Automatically proving linearizability

September 2016, 19 pages, PDF

Abstract: This technical report presents a practical automatic verification procedure for proving linearizability (i.e., atomicity and functional correctness) of concurrent data structure implementations. The procedure uses a novel instrumentation to verify logically pure executions, and is evaluated on a number of standard concurrent stack, queue and set algorithms.

UCAM-CL-TR-779

Eric K. Henderson:

A text representation language for contextual and distributional processing

April 2010, 207 pages, PDF
PhD thesis (Fitzwilliam College, 2009)

Abstract: This thesis examines distributional and contextual aspects of linguistic processing in relation to traditional symbolic approaches. Distributional processing is more commonly associated with statistical methods, while an integrated representation of context spanning document and syntactic structure is lacking in current linguistic representations. This thesis addresses both issues through a novel symbolic text representation language.

The text representation language encodes information from all levels of linguistic analysis in a semantically motivated form. Using object-oriented constructs in a recursive structure that can be derived from the syntactic parse, the language provides a common interface for symbolic and distributional processing. A key feature of the language is a recursive treatment of context at all levels of representation. The thesis gives a detailed account of the form and syntax of the language, as well as a treatment of several important constructions. Comparisons are made with other linguistic and semantic representations, and several of the distinguishing features are demonstrated through experiments.

The treatment of context in the representation language is discussed at length. The recursive structure employed in the representation is explained and motivated by issues involving document structure. Applications of the contextual representation in symbolic processing are demonstrated through several experiments.

Distributional processing is introduced using traditional statistical techniques to measure semantic similarity. Several extant similarity metrics are evaluated using a novel evaluation metric involving adjective antonyms. The results provide several insights into the nature of distributional processing, and this motivates a new approach based on characteristic adjectives.

Characteristic adjectives are distributionally derived and semantically differentiated vectors associated with a node in a semantic taxonomy. They are significantly lower-dimensioned then their undifferentiated source vectors, while retaining a strong correlation to their position in the semantic space. Their properties and derivation are described in detail and an experimental evaluation of their semantic content is presented.

Finally, the distributional techniques to derive characteristic adjectives are extended to encompass symbolic processing. Rules involving several types of symbolic patterns are distributionally derived from a source corpus, and applied to the text representation language.

Polysemy is addressed in the derivation by limiting distributional information to monosemous words. The derived rules show a significant improvement at disambiguating nouns in a test corpus.

UCAM-CL-TR-780

Michael Roe:

Cryptography and evidence

May 2010, 75 pages, PDF
PhD thesis (Clare College, April 1997)

Abstract: The invention of public-key cryptography led to the notion that cryptographically protected messages could be used as evidence to convince an impartial adjudicator that a disputed event had in fact occurred. Information stored in a computer is easily modified, and so records can be falsified or retrospectively modified. Cryptographic protection prevents modification, and it is hoped that this will make cryptographically protected data acceptable as evidence. This usage of cryptography to render an event undeniable has become known as non-repudiation. This dissertation is an enquiry into the fundamental limitations of this application of cryptography, and the disadvantages of the techniques which are currently in use. In the course of this investigation I consider the converse problem, of ensuring that an instance of communication between computer systems leaves behind no unequivocal evidence of its having taken place. Features of communications protocols that were seen as defects from the standpoint of non-repudiation can be seen as benefits from the standpoint of this converse problem, which I call "plausible deniability".

UCAM-CL-TR-781

Minor E. Gordon:

Stage scheduling for CPU-intensive servers

June 2010, 119 pages, PDF
PhD thesis (Jesus College, December 2009)

Abstract: The increasing prevalence of multicore, multiprocessor commodity hardware calls for server software architectures that are cycle-efficient on individual cores and can maximize concurrency across an entire machine. In order to achieve both ends this dissertation advocates stage architectures that put software concurrency foremost and aggressive CPU scheduling that exploits the common structure and runtime behavior of CPU-intensive servers. For these servers user-level scheduling policies that multiplex one kernel thread per physical core can outperform those that utilize pools of worker threads per stage on CPU-intensive workloads. Boosting the hardware efficiency of servers in userspace means a single machine can handle more users without tuning, operating system modifications, or better hardware.

UCAM-CL-TR-782

Jonathan M. Hayman:

Petri net semantics

June 2010, 252 pages, PDF
PhD thesis (Darwin College, January 2009)

Abstract: Petri nets are a widely-used model for concurrency. By modelling the effect of events on local components of state, they reveal how the events of a process interact with each other, and whether they can occur independently of each other by operating on disjoint regions of state.

Despite their popularity, we are lacking systematic syntax-driven techniques for defining the semantics of programming languages inside Petri nets in an analogous way that Plotkin's Structural Operational Semantics defines a transition system semantics. The first part of this thesis studies a generally-applicable framework for the definition of the net semantics of a programming language.

The net semantics is used to study concurrent separation logic, a Hoare-style logic used to prove partial correctness of pointer-manipulating concurrent programs. At the core of the logic is the notion of separation of ownership of state, allowing us to infer that proven parallel processes operate on the disjoint regions of the state that they are seen to own. In this thesis, a notion of validity of the judgements capturing the subtle notion of ownership is given and soundness of the logic with respect to this model is shown. The model is then used to study the independence of processes arising from the separation of ownership. Following from this, a form of refinement is given which is capable of changing the granularity assumed of the program's atomic actions.

Amongst the many different models for concurrency, there are several forms of Petri net. Category theory has been used in the past to establish connections between them via adjunctions, often coreflections, yielding common constructions across the models and relating concepts such as bisimulation. The most general forms of Petri net have, however, fallen outside this framework. Essentially, this is due to the most general forms of net having an implicit symmetry in their behaviour that other forms of net cannot directly represent.

The final part of this thesis shows how an abstract framework for defining symmetry in models can be applied to obtain categories of Petri net with symmetry. This is shown to recover, up to symmetry, the universal characterization of unfolding operations on general Petri nets, allowing coreflections up to symmetry between the category of general Petri nets and other categories of net.

UCAM-CL-TR-783

Dan O'Keeffe:

Distributed complex event detection for pervasive computing

July 2010, 170 pages, PDF
PhD thesis (St. John's College, December 2009)

Abstract: Pervasive computing is a model of information processing that augments computers with sensing capabilities and distributes them into the environment. Many pervasive computing applications are reactive in nature, in that they perform actions in response to events (i.e. changes in state of the environment). However, these applications are typically interested in high-level complex events, in contrast to the low-level primitive events produced by sensors. The goal of this thesis is to support the detection of complex events by filtering, aggregating, and combining primitive events.

Supporting complex event detection in pervasive computing environments is a challenging problem. Sensors may have limited processing, storage, and communication capabilities. In addition, battery powered sensing devices have limited energy resources. Since they are embedded in the environment, recharging may be difficult or impossible. To prolong the lifetime of the system, it is vital that these energy resources are used efficiently. Further complications arise due to the distributed nature of pervasive computing systems. The lack of a global clock can make it impossible to order events from different sources. Events may be delayed or lost en route to their destination, making it difficult to perform timely and accurate complex event detection. Finally, pervasive computing systems may be large, both geographically and in terms of the number of sensors. Architectures to support pervasive computing applications should therefore be highly scalable.

We make several contributions in this dissertation. Firstly, we present a flexible language for specifying complex event patterns. The language provides developers with a variety of parameters to control the detection process, and is designed for use in an open distributed environment. Secondly, we provide the ability for applications to specify a variety of detection policies. These policies allow the system to determine the best way of handling lost and delayed events. Of particular interest is our 'no false-positive' detection policy. This allows a reduction in detection latency while ensuring that only correct events are generated for applications sensitive to false positives. Finally, we show how complex event detector placement can be optimized over a federated event-based middleware. In many cases, detector distribution can reduce unnecessary communication with resource constrained sensors.

UCAM-CL-TR-784

Ripduman Sohan, Andrew Rice, Andrew W. Moore, Kieran Mansley:

Characterizing 10 Gbps network interface energy consumption

July 2010, 10 pages, PDF

Abstract: Understanding server energy consumption is fast becoming an area of interest given the increase in the per-machine energy footprint of modern servers and the increasing number of servers required to satisfy demand. In this paper we (i) quantify the energy overhead of the network subsystem in modern servers by measuring, reporting and analyzing power consumption in six 10 Gbps and four 1 Gbps interconnects at a fine-grained level; (ii) introduce two metrics for calculating the energy efficiency of a network interface from the perspective of network throughput and host CPU usage; (iii) compare the efficiency of multiport 1 Gbps interconnects as an alternative to 10 Gbps interconnects; and (iv) conclude by offering recommendations for improving network energy efficiency for system deployment and network interface designers.

UCAM-CL-TR-785

Aaron R. Coble:

Anonymity, information, and machine-assisted proof

July 2010, 171 pages, PDF
PhD thesis (King's College, January 2010)

Abstract: This report demonstrates a technique for proving the anonymity guarantees of communication systems, using a mechanised theorem-prover. The approach is based on Shannon's theory of information and can be used to analyse probabilistic programs. The information-theoretic metrics that are used for anonymity provide quantitative results, even in the case of partial anonymity. Many of the developments in this text are applicable to information leakage in general, rather than solely to privacy properties. By developing the framework within a mechanised theorem-prover, all proofs are guaranteed to be logically and mathematically consistent with respect to a given model. Moreover, the specification of a system can be parameterised and desirable properties of the system can quantify over those parameters; as a result, properties can be proved about the system in general, rather than specific instances.

In order to develop the analysis framework described in this text, the underlying theories of information, probability, and measure had to be formalised in the theorem-prover; those formalisations are explained in detail. That foundational work is of general interest

and not limited to the applications illustrated here. The meticulous, extensional approach that has been taken ensures that mathematical consistency is maintained.

A series of examples illustrate how formalised information theory can be used to analyse and prove the information leakage of programs modelled in the theorem-prover. Those examples consider a number of different threat models and show how they can be characterised in the framework proposed.

Finally, the tools developed are used to prove the anonymity of the dining cryptographers (DC) protocol, thereby demonstrating the use of the framework and its applicability to proving privacy properties; the DC protocol is a standard benchmark for new methods of analysing anonymity systems. This work includes the first machine-assisted proof of anonymity of the DC protocol for an unbounded number of cryptographers.

UCAM-CL-TR-786

Arnab Banerjee:

Communication flows in power-efficient Networks-on-Chips

August 2010, 107 pages, PDF
PhD thesis (Girton College, March 2009)

Abstract: Networks-on-Chips (NoCs) represent a scalable wiring solution for future chips, with dynamic allocation-based networks able to provide good utilisation of the scarce available resources. This thesis develops power-efficient, dynamic, packet-switched NoCs which can support on-chip communication flows.

Given the severe power constraint already present in VLSI, a power efficient NoC design direction is first developed. To accurately explore the impact of various design parameters on NoC power dissipation, 4 different router designs are synthesised, placed and routed in a 90nm process. This demonstrates that the power demands are dominated by the data-path and not the control-path, leading to the key finding that, from the energy perspective, it is justifiable to use more computation to optimise communication.

A review of existing research shows the near-ubiquitous nature of stream-like communication flows in future computing systems, making support for flows within NoCs critically important. It is shown that in several situations, current NoCs make highly inefficient use of network resources in the presence of communication flows. To resolve this problem, a scalable mechanism is developed to enable the identification of flows, with a flow defined as all packets going to the same destination. The number of virtual-channels that can be used by a single flow is then limited to the minimum required, ensuring efficient resource utilisation.

The issue of fair resource allocation between flows is next investigated. The locally fair, packet-based allocation strategies of current NoCs are shown not to provide fairness between flows. The mechanism already developed to identify flows by their destination nodes is extended to enable flows to be identified by source-destination address pairs. Finally, a modification to the link scheduling mechanism is proposed to achieve max-min fairness between flows.

UCAM-CL-TR-787

Daniel Bernhardt:

Emotion inference from human body motion

October 2010, 227 pages, PDF
PhD thesis (Selwyn College, January 2010)

Abstract: The human body has evolved to perform sophisticated tasks from locomotion to the use of tools. At the same time our body movements can carry information indicative of our intentions, inter-personal attitudes and emotional states. Because our body is specialised to perform a variety of everyday tasks, in most situations emotional effects are only visible through subtle changes in the qualities of movements and actions. This dissertation focuses on the automatic analysis of emotional effects in everyday actions.

In the past most efforts to recognise emotions from the human body have focused on expressive gestures which are archetypal and exaggerated expressions of emotions. While these are easier to recognise by humans and computational pattern recognisers they very rarely occur in natural scenarios. The principal contribution of this dissertation is hence the inference of emotional states from everyday actions such as walking, knocking and throwing. The implementation of the system draws inspiration from a variety of disciplines including psychology, character animation and speech recognition. Complex actions are modelled using Hidden Markov Models and motion primitives. The manifestation of emotions in everyday actions is very subtle and even humans are far from perfect at picking up and interpreting the relevant cues because emotional influences are usually minor compared to constraints arising from the action context or differences between individuals.

This dissertation describes a holistic approach which models emotional, action and personal influences in order to maximise the discriminability of different emotion classes. A pipeline is developed which incrementally removes the biases introduced by different action contexts and individual differences. The resulting signal is described in terms of posture and dynamic features and classified into one of several emotion classes using statistically trained Support Vector Machines. The system also goes beyond isolated expressions and is able to classify natural action sequences. I use Level Building to segment action sequences and combine component classifications using an incremental voting scheme which is suitable for online applications. The system is comprehensively evaluated along a number of dimensions using a corpus of

motion-captured actions. For isolated actions I evaluate the generalisation performance to new subjects. For action sequences I study the effects of reusing models trained on the isolated cases vs adapting models to connected samples. The dissertation also evaluates the role of modelling the influence of individual user differences. I develop and evaluate a regression-based adaptation scheme. The results bring us an important step closer to recognising emotions from body movements, embracing the complexity of body movements in natural scenarios.

UCAM-CL-TR-788

Byron Cook, Eric Koskinen, Moshe Vardi:

Branching-time reasoning for programs (extended version)

July 2011, 38 pages, PDF

Abstract: We describe a reduction from temporal property verification to a program analysis problem. Our reduction is an encoding which, with the use of procedures and nondeterminism, enables existing interprocedural program analysis tools to naturally perform the reasoning necessary for proving temporal properties (eg. backtracking, eventuality checking, tree counterexamples for branching-time properties, abstraction refinement, etc.). Our reduction is state-based in nature but also forms the basis of an efficient algorithm for verifying trace-based properties, when combined with an iterative symbolic determinization technique, due to Cook and Koskinen.

In this extended version, we formalize our encoding as a guarded transition system G, parameterized by a finite set of ranking functions and the temporal logic property. We establish soundness between a safety property of G and the validity of a branching time temporal logic property ∀CTL. ∀CTL is a sufficient logic for proving properties written in the trace-based Linear Temporal Logic via the iterative algorithm.

Finally using examples drawn from the PostgreSQL database server, Apache web server, and Windows OS kernel, we demonstrate the practical viability of our work.

UCAM-CL-TR-789

Byron Cook, Eric Koskinen:

Making prophecies with decision predicates

November 2010, 29 pages, PDF

Abstract: We describe a new algorithm for proving temporal properties expressed in LTL of infinite-state programs. Our approach takes advantage of the fact that LTL properties can often be proved more efficiently using techniques usually associated with the branching-time logic CTL than they can with native LTL algorithms. The caveat is that, in certain instances, nondeterminism in the system's transition relation can cause CTL methods to report counterexamples that are spurious with respect to the original LTL formula. To address this problem we describe an algorithm that, as it attempts to apply CTL proof methods, finds and then removes problematic nondeterminism via an analysis on the potentially spurious counterexamples. Problematic nondeterminism is characterized using decision predicates, and removed using a partial, symbolic determinization procedure which introduces new prophecy variables to predict the future outcome of these choices. We demonstrate—using examples taken from the PostgreSQL database server, Apache web server, and Windows OS kernel—that our method can yield enormous performance improvements in comparison to known tools, allowing us to automatically prove properties of programs where we could not prove them before.

UCAM-CL-TR-790

Ted Briscoe, Ben Medlock, Øistein Andersen:

Automated assessment of ESOL free text examinations

November 2010, 31 pages, PDF

Abstract: In this report, we consider the task of automated assessment of English as a Second Language (ESOL) examination scripts written in response to prompts eliciting free text answers. We review and critically evaluate previous work on automated assessment for essays, especially when applied to ESOL text. We formally define the task as discriminative preference ranking and develop a new system trained and tested on a corpus of manually-graded scripts. We show experimentally that our best performing system is very close to the upper bound for the task, as defined by the agreement between human examiners on the same corpus. Finally we argue that our approach, unlike extant solutions, is relatively prompt-insensitive and resistant to subversion, even when its operating principles are in the public domain. These properties make our approach significantly more viable for high-stakes assessment.

UCAM-CL-TR-791

Andreas Vlachos:

Semi-supervised learning for biomedical information extraction

November 2010, 113 pages, PDF
PhD thesis (Peterhouse College, December 2009)

Abstract: This thesis explores the application of semi-supervised learning to biomedical information extraction. The latter has emerged in recent years as a challenging application domain for natural language processing techniques. The challenge stems partly from the lack of appropriate resources that can be used as labeled training data. Therefore, we choose to focus on semi-supervised learning techniques which enable us to take advantage of human supervision combined with unlabeled data.

We begin with a short introduction to biomedical information extraction and semi-supervised learning in Chapter 1. Chapter 2 focuses on the task of biomedical named entity recognition. Using raw abstracts and a dictionary of gene names we develop two systems for this task. Furthermore, we discuss annotation issues and demonstrate how the performance can be improved using user feedback in realistic conditions. In Chapter 3 we develop two biomedical event extraction systems: a rule-based one and a machine learning based one. The former needs only an annotated dictionary and syntactic parsing as input, while the latter requires partial event annotation additionally. Both systems achieve performances comparable to systems utilizing fully annotated training data. Chapter 4 discusses the task of lexical-semantic clustering using Dirichlet process mixture models. We review the unsupervised learning method used, which allows the number of clusters discovered to be determined by the data. Furthermore, we introduce a new clustering evaluation measure that addresses some shortcomings of the existing measures. Chapter 5 introduces a method of guiding the clustering solution using pairwise links between instances. Furthermore, we present a method of selecting these pairwise links actively in order to decrease the amount of supervision required. Finally, Chapter 6 assesses the contributions of this thesis and highlights directions for future work.

UCAM-CL-TR-792

James P. Bridge:

Machine learning and automated theorem proving

November 2010, 180 pages, PDF
PhD thesis (Corpus Christi College, October 2010)

Abstract: Computer programs to find formal proofs of theorems have a history going back nearly half a century. Originally designed as tools for mathematicians, modern applications of automated theorem provers and proof assistants are much more diverse. In particular they are used in formal methods to verify software and hardware designs to prevent costly, or life threatening, errors being introduced into systems from microchips to controllers for medical equipment or space rockets.

Despite this, the high level of human expertise required in their use means that theorem proving tools are not widely used by non specialists, in contrast to computer algebra packages which also deal with the manipulation of symbolic mathematics. The work described in this dissertation addresses one aspect of this problem, that of heuristic selection in automated theorem provers. In theory such theorem provers should be automatic and therefore easy to use; in practice the heuristics used in the proof search are not universally optimal for all problems so human expertise is required to determine heuristic choice and to set parameter values.

Modern machine learning has been applied to the automation of heuristic selection in a first order logic theorem prover. One objective was to find if there are any features of a proof problem that are both easy to measure and provide useful information for determining heuristic choice. Another was to determine and demonstrate a practical approach to making theorem provers truly automatic.

In the experimental work, heuristic selection based on features of the conjecture to be proved and the associated axioms is shown to do better than any single heuristic. Additionally a comparison has been made between static features, measured prior to the proof search process, and dynamic features that measure changes arising in the early stages of proof search. Further work was done on determining which features are important, demonstrating that good results are obtained with only a few features required.

UCAM-CL-TR-793

Shazia Afzal:

Affect inference in learning environments: a functional view of facial affect analysis using naturalistic data

December 2010, 146 pages, PDF
PhD thesis (Murray Edwards College, May 2010)

Abstract: This research takes an application-oriented stance on affective computing and addresses the problem of automatic affect inference within learning technologies. It draws from the growing understanding of the centrality of emotion in the learning process and the fact that, as yet, this crucial link is not addressed in the design of learning technologies. This dissertation specifically focuses on examining the utility of facial affect analysis to model the affective state of a learner in a one-on-one learning setting.

Although facial affect analysis using posed or acted data has been studied in great detail for a couple of decades now, research using naturalistic data is still a challenging problem. The challenges are derived from the complexity in conceptualising affect, the methodological and technical difficulties in measuring it, and the emergent ethical concerns in realising automatic

affect inference by computers. However, as the context of this research is derived from, and relates to, a real-world application environment, it is based entirely on naturalistic data. The whole pipeline – of identifying the requirements, to collection of data, to the development of an annotation protocol, to labelling of data, and the eventual analyses – both quantitative and qualitative; is described in this dissertation. In effect, a framework for conducting research using natural data is set out and the challenges encountered at each stage identified.

Apart from the challenges associated with the perception and measurement of affect, this research emphasises that there are additional issues that require due consideration by virtue of the application context. As such, in light of the discussed observations and results, this research concludes that we need to understand the nature and expression of emotion in the context of technology use, and pursue creative exploration of what is perhaps a qualitatively different form of emotion expression and communication.

UCAM-CL-TR-794

Øistein E. Andersen:

Grammatical error prediction

January 2011, 163 pages, PDF
PhD thesis (Girton College, 2010)

Abstract: In this thesis, we investigate methods for automatic detection, and to some extent correction, of grammatical errors. The evaluation is based on manual error annotation in the Cambridge Learner Corpus (CLC), and automatic or semi-automatic annotation of error corpora is one possible application, but the methods are also applicable in other settings, for instance to give learners feedback on their writing or in a proofreading tool used to prepare texts for publication.

Apart from the CLC, we use the British National Corpus (BNC) to get a better model of correct usage, WordNet for semantic relations, other machine-readable dictionaries for orthography/morphology, and the Robust Accurate Statistical Parsing (RASP) system to parse both the CLC and the BNC and thereby identify syntactic relations within the sentence. An ancillary outcome of this is a syntactically annotated version of the BNC, which we have made publicly available.

We present a tool called GenERRate, which can be used to introduce errors into a corpus of correct text, and evaluate to what extent the resulting synthetic error corpus can complement or replace a real error corpus.

Different methods for detection and correction are investigated, including: sentence-level binary classification based on machine learning over n-grams of words, n-grams of part-of-speech tags and grammatical relations; automatic identification of features which are highly indicative of individual errors; and development of classifiers aimed more specifically at given error types, for instance concord errors based on syntactic structure and collocation errors based on co-occurrence statistics from the BNC, using clustering to deal with data sparseness. We show that such techniques can detect, and sometimes even correct, at least certain error types as well as or better than human annotators.

We finally present an annotation experiment in which a human annotator corrects and supplements the automatic annotation, which confirms the high detection/correction accuracy of our system and furthermore shows that such a hybrid set-up gives higher-quality annotation with considerably less time and effort expended compared to fully manual annotation.

UCAM-CL-TR-795

Aurelie Herbelot:

Underspecified quantification

February 2011, 163 pages, PDF
PhD thesis (Trinity Hall, 2010)

Abstract: Many noun phrases in text are ambiguously quantified: syntax doesn't explicitly tell us whether they refer to a single entity or to several and, in main clauses, what portion of the set denoted by the subject Nbar actually takes part in the event expressed by the verb. For instance, when we utter the sentence 'Cats are mammals', it is only world knowledge that allows our hearer to infer that we mean 'All cats are mammals', and not 'Some cats are mammals'. This ambiguity effect is interesting at several levels. Theoretically, it raises cognitive and linguistic questions. To what extent does syntax help humans resolve the ambiguity? What problem-solving skills come into play when syntax is insufficient for full resolution? How does ambiguous quantification relate to the phenomenon of genericity, as described by the linguistic literature? From an engineering point of view, the resolution of quantificational ambiguity is essential to the accuracy of some Natural Language Processing tasks.

We argue that the quantification ambiguity phenomenon can be described in terms of underspecification and propose a formalisation for what we call 'underquantified' subject noun phrases. Our formalisation is motivated by inference requirements and covers all cases of genericity.

Our approach is then empirically validated by human annotation experiments. We propose an annotation scheme that follows our theoretical claims with regard to underquantification. Our annotation results strengthen our claim that all noun phrases can be analysed in terms of quantification. The produced corpus allows us to derive a gold standard for quantification resolution experiments and is, as far as we are aware, the first attempt to analyse the distribution of null quantifiers in English.

We then create a baseline system for automatic quantification resolution, using syntax to provide discriminating features for our classification. We show

that results are rather poor for certain classes and argue that some level of pragmatics is needed, in combination with syntax, to perform accurate resolution. We explore the use of memory-based learning as a way to approximate the problem-solving skills available to humans at the level of pragmatic understanding.

UCAM-CL-TR-796

Jonathan Mak:

Facilitating program parallelisation: a profiling-based approach

March 2011, 120 pages, PDF
PhD thesis (St. John's College, November 2010)

Abstract: The advance of multi-core architectures signals the end of universal speed-up of software over time. To continue exploiting hardware developments, effort must be invested in producing software that can be split up to run on multiple cores or processors. Many solutions have been proposed to address this issue, ranging from explicit to implicit parallelism, but consensus has yet to be reached on the best way to tackle such a problem.

In this thesis we propose a profiling-based interactive approach to program parallelisation. Profilers gather dependence information on a program, which is then used to automatically parallelise the program at source-level. The programmer can then examine the resulting parallel program, and using critical path information from the profiler, identify and refactor parallelism bottlenecks to enable further parallelism. We argue that this is an efficient and effective method of parallelising general sequential programs.

Our first contribution is a comprehensive analysis of limits of parallelism in several benchmark programs, performed by constructing Dynamic Dependence Graphs (DDGs) from execution traces. We show that average available parallelism is often high, but realising it would require various changes in compilation, language or computation models. As an example, we show how using a spaghetti stack structure can lead to a doubling of potential parallelism.

The rest of our thesis demonstrates how some of this potential parallelism can be realised under the popular fork-join parallelism model used by Cilk, TBB, OpenMP and others. We present a tool-chain with two main components: Embla 2, which uses DDGs from profiled dependences to estimate the amount of task-level parallelism in programs; and Woolifier, a source-to-source transformer that uses Embla 2's output to parallelise the programs. Using several case studies, we demonstrate how this tool-chain greatly facilitates program parallelisation by performing an automatic best-effort parallelisation and presenting critical paths in a concise graphical form so that the programmer can quickly locate parallelism bottlenecks, which when refactored can lead to even greater potential parallelism and significant actual speed-ups (up to around 25 on a 32-effective-core machine).

UCAM-CL-TR-797

Boris Feigin:

Interpretational overhead in system software

April 2011, 116 pages, PDF
PhD thesis (Homerton College, September 2010)

Abstract: Interpreting a program carries a runtime penalty: the interpretational overhead. Traditionally, a compiler removes interpretational overhead by sacrificing inessential details of program execution. However, a broad class of system software is based on non-standard interpretation of machine code or a higher-level language. For example, virtual machine monitors emulate privileged instructions; program instrumentation is used to build dynamic call graphs by intercepting function calls and returns; and dynamic software updating technology allows program code to be altered at runtime. Many of these frameworks are performance-sensitive and several efficiency requirements—both formal and informal—have been put forward over the last four decades. Largely independently, the concept of interpretational overhead received much attention in the partial evaluation ("program specialization") literature. This dissertation contributes a unifying understanding of efficiency and interpretational overhead in system software.

Starting from the observation that a virtual machine monitor is a self-interpreter for machine code, our first contribution is to reconcile the definition of efficient virtualization due to Popek and Goldberg with Jones optimality, a measure of the strength of program specializers. We also present a rational reconstruction of hardware virtualization support ("trap-and-emulate") from context-threaded interpretation, a technique for implementing fast interpreters due to Berndl et al.

As a form of augmented execution, virtualization shares many similarities with program instrumentation. Although several low-overhead instrumentation frameworks are available on today's hardware, there has been no formal understanding of what it means for instrumentation to be efficient. Our second contribution is a definition of efficiency for program instrumentation in the spirit of Popek and Goldberg's work. Instrumentation also incurs an implicit overhead because instrumentation code needs access to intermediate execution states and this is antagonistic to optimization. The third contribution is to use partial equivalence relations (PERs) to express the dependence of instrumentation on execution state, enabling an instrumentation/optimization trade-off. Since program instrumentation, applied at runtime, constitutes a kind of dynamic software update, we can similarly restrict allowable future updates to be consistent with existing optimizations. Finally, treating "old" and "new" code in

a dynamically-updatable program as being written in different languages permits a semantic explanation of a safety rule that was originally introduced as a syntactic check.

UCAM-CL-TR-798

Periklis Akritidis:

Practical memory safety for C

June 2011, 136 pages, PDF
PhD thesis (Wolfson College, May 2010)

Abstract: Copious amounts of high-performance and low-level systems code are written in memory-unsafe languages such as C and C++. Unfortunately, the lack of memory safety undermines security and reliability; for example, memory-corruption bugs in programs can breach security, and faults in kernel extensions can bring down the entire operating system. Memory-safe languages, however, are unlikely to displace C and C++ in the near future; thus, solutions for future and existing C and C++ code are needed.

Despite considerable prior research, memory-safety problems in C and C++ programs persist because the existing proposals that are practical enough for production use cannot offer adequate protection, while comprehensive proposals are either too slow for practical use, or break backwards compatibility by requiring significant porting or generating binary-incompatible code.

To enable practical protection against memory-corruption attacks and operating system crashes, I designed new integrity properties preventing dangerous memory corruption at low cost instead of enforcing strict memory safety to catch every memory error at high cost. Then, at the implementation level, I aggressively optimised for the common case, and streamlined execution by modifying memory layouts as far as allowed without breaking binary compatibility.

I developed three compiler-based tools for analysing and instrumenting unmodified source code to automatically generate binaries hardened against memory errors: BBC and WIT to harden user-space C programs, and BGI to harden and to isolate Microsoft Windows kernel extensions. The generated code incurs low performance overhead and is binary-compatible with uninstrumented code. BBC offers strong protection with lower overhead than previously possible for its level of protection; WIT further lowers overhead while offering stronger protection than previous solutions of similar performance; and BGI improves backwards compatibility and performance over previous proposals, making kernel extension isolation practical for commodity systems.

UCAM-CL-TR-799

Thomas Tuerk:

A separation logic framework for HOL

June 2011, 271 pages, PDF
PhD thesis (Downing College, December 2010)

Abstract: Separation logic is an extension of Hoare logic due to O'Hearn and Reynolds. It was designed for reasoning about mutable data structures. Because separation logic supports local reasoning, it scales better than classical Hoare logic and can easily be used to reason about concurrency. There are automated separation logic tools as well as several formalisations in interactive theorem provers. Typically, the automated separation logic tools are able to reason about shallow properties of large programs. They usually consider just the shape of data structures, not their data-content. The formalisations inside theorem provers can be used to prove interesting, deep properties. However, they typically lack automation. Another shortcoming is that there are a lot of slightly different separation logics. For each programming language and each interesting property a new kind of separation logic seems to be invented.

In this thesis, a general framework for separation logic is developed inside the HOL4 theorem prover. This framework is based on Abstract Separation Logic, an abstract, high level variant of separation logic. Abstract Separation Logic is a general separation logic such that many other separation logics can be based on it. This framework is instantiated in a first step to support a stack with read and write permissions following ideas of Parkinson, Bornat and Calcagno. Finally, the framework is further instantiated to build a separation logic tool called Holfoot. It is similar to the tool Smallfoot, but extends it from reasoning about shape properties to fully functional specifications.

To my knowledge this work presents the first formalisation of Abstract Separation Logic inside a theorem prover. By building Holfoot on top of this formalisation, I could demonstrate that Abstract Separation Logic can be used as a basis for realistic separation logic tools. Moreover, this work demonstrates that it is feasable to implement such separation logic tools inside a theorem prover. Holfoot is highly automated. It can verify Smallfoot examples automatically inside HOL4. Moreover, Holfoot can use the full power of HOL4. This allows Holfoot to verify fully functional specifications. Simple fully functional specifications can be handled automatically using HOL4's tools and libraries or external SMT solvers. More complicated ones can be handled using interactive proofs inside HOL4. In contrast, most other separation logic tools can reason just about the shape of data structures. Others reason only about data properties that can be solved using SMT solvers.

UCAM-CL-TR-800

James R. Srinivasan:

Improving cache utilisation

June 2011, 184 pages, PDF
PhD thesis (Jesus College, April 2011)

Abstract: Microprocessors have long employed caches to help hide the increasing latency of accessing main memory. The vast majority of previous research has focussed on increasing cache hit rates to improve cache performance, while lately decreasing power consumption has become an equally important issue. This thesis examines the lifetime of cache lines in the memory hierarchy, considering whether they are live (will be referenced again before eviction) or dead (will not be referenced again before eviction). Using these two states, the cache utilisation (proportion of the cache which will be referenced again) can be calculated.

This thesis demonstrates that cache utilisation is relatively poor over a wide range of benchmarks and cache configurations. By focussing on techniques to improve cache utilisation, cache hit rates are increased while overall power consumption may also be decreased.

Key to improving cache utilisation is an accurate predictor of the state of a cache line. This thesis presents a variety of such predictors, mostly based upon the mature field of branch prediction, and compares them against previously proposed predictors. The most appropriate predictors are then demonstrated in two applications: Improving victim cache performance through filtering, and reducing cache pollution during aggressive prefetching

These applications are primarily concerned with improving cache performance and are analysed using a detailed microprocessor simulator. Related applications, including decreasing power consumption, are also discussed, as are the applicability of these techniques to multiprogrammed and multiprocessor systems.

UCAM-CL-TR-801

Amitabha Roy:

Software lock elision for x86 machine code

July 2011, 154 pages, PDF
PhD thesis (Emmanuel College, April 2011)

Abstract: More than a decade after becoming a topic of intense research there is no transactional memory hardware nor any examples of software transactional memory use outside the research community. Using software transactional memory in large pieces of software needs copious source code annotations and often means that standard compilers and debuggers can no longer be used. At the same time, overheads associated with software transactional memory fail to motivate programmers to expend the needed effort to use software transactional memory. The only way around the overheads in the case of general unmanaged code is the anticipated availability of hardware support. On the other hand, architects are unwilling to devote power and area budgets in mainstream microprocessors to hardware transactional memory, pointing to transactional memory being a "niche" programming construct. A deadlock has thus ensued that is blocking transactional memory use and experimentation in the mainstream.

This dissertation covers the design and construction of a software transactional memory runtime system called SLE_x86 that can potentially break this deadlock by decoupling transactional memory from programs using it. Unlike most other STM designs, the core design principle is transparency rather than performance. SLE_x86 operates at the level of x86 machine code, thereby becoming immediately applicable to binaries for the popular x86 architecture. The only requirement is that the binary synchronise using known locking constructs or calls such as those in Pthreads or OpenMP libraries. SLE_x86 provides speculative lock elision (SLE) entirely in software, executing critical sections in the binary using transactional memory. Optionally, the critical sections can also be executed without using transactions by acquiring the protecting lock.

The dissertation makes a careful analysis of the impact on performance due to the demands of the x86 memory consistency model and the need to transparently instrument x86 machine code. It shows that both of these problems can be overcome to reach a reasonable level of performance, where transparent software transactional memory can perform better than a lock. SLE_x86 can ensure that programs are ready for transactional memory in any form, without being explicitly written for it.

UCAM-CL-TR-802

Johanna Geiß:

Latent semantic sentence clustering for multi-document summarization

July 2011, 156 pages, PDF
PhD thesis (St. Edmund's College, April 2011)

Abstract: This thesis investigates the applicability of Latent Semantic Analysis (LSA) to sentence clustering for Multi-Document Summarization (MDS). In contrast to more shallow approaches like measuring similarity of sentences by word overlap in a traditional vector space model, LSA takes word usage patterns into account. So far LSA has been successfully applied to different Information Retrieval (IR) tasks like information filtering and document classification (Dumais, 2004). In the course of this research, different parameters essential to sentence clustering using a hierarchical agglomerative

clustering algorithm (HAC) in general and in combination with LSA in particular are investigated. These parameters include, inter alia, information about the type of vocabulary, the size of the semantic space and the optimal numbers of dimensions to be used in LSA. These parameters have not previously been studied and evaluated in combination with sentence clustering (chapter 4).

This thesis also presents the first gold standard for sentence clustering in MDS. To be able to evaluate sentence clusterings directly and classify the influence of the different parameters on the quality of sentence clustering, an evaluation strategy is developed that includes gold standard comparison using different evaluation measures (chapter 5). Therefore the first compound gold standard for sentence clustering was created. Several human annotators were asked to group similar sentences into clusters following guidelines created for this purpose (section 5.4). The evaluation of the human generated clusterings revealed that the human annotators agreed on clustering sentences above chance. Analysis of the strategies adopted by the human annotators revealed two groups – hunters and gatherers – who differ clearly in the structure and size of the clusters they created (chapter 6).

On the basis of the evaluation strategy the parameters for sentence clustering and LSA are optimized (chapter 7). A final experiment in which the performance of LSA in sentence clustering for MDS is compared to the simple word matching approach of the traditional Vector Space Model (VSM) revealed that LSA produces better quality sentence clusters for MDS than VSM.

UCAM-CL-TR-803

Ekaterina V. Shutova:

Computational approaches to figurative language

August 2011, 219 pages, PDF
PhD thesis (Pembroke College, March 2011)

Abstract: The use of figurative language is ubiquitous in natural language text and it is a serious bottleneck in automatic text understanding. A system capable of interpreting figurative language would be extremely beneficial to a wide range of practical NLP applications. The main focus of this thesis is on the phenomenon of metaphor. I adopt a statistical data-driven approach to its modelling, and create the first open-domain system for metaphor identification and interpretation in unrestricted text. In order to verify that similar methods can be applied to modelling other types of figurative language, I then extend this work to the task of interpretation of logical metonymy.

The metaphor interpretation system is capable of discovering literal meanings of metaphorical expressions in text. For the metaphors in the examples "All of this stirred an unfathomable excitement in her" or "a carelessly leaked report" the system produces interpretations "All of this provoked an unfathomable excitement in her" and "a carelessly disclosed report" respectively. It runs on unrestricted text and to my knowledge is the only existing robust metaphor paraphrasing system. It does not employ any hand-coded knowledge, but instead derives metaphorical interpretations from a large text corpus using statistical pattern-processing. The system was evaluated with the aid of human judges and it operates with the accuracy of 81%.

The metaphor identification system automatically traces the analogies involved in the production of a particular metaphorical expression in a minimally supervised way. The system generalises over the analogies by means of verb and noun clustering, i.e. identification of groups of similar concepts. This generalisation makes it capable of recognising previously unseen metaphorical expressions in text, e.g. having once seen a metaphor 'stir excitement' the system concludes that 'swallow anger' is also used metaphorically. The system identifies metaphorical expressions with a high precision of 79%.

The logical metonymy processing system produces a list of metonymic interpretations disambiguated with respect to their word sense. It then automatically organises them into a novel class-based model of logical metonymy inspired by both empirical evidence and linguistic theory. This model provides more accurate and generalised information about possible interpretations of metonymic phrases than previous approaches.

UCAM-CL-TR-804

Sean B. Holden:

The HasGP user manual

September 2011, 18 pages, PDF

Abstract: HasGP is an experimental library implementing methods for supervised learning using Gaussian process (GP) inference, in both the regression and classification settings. It has been developed in the functional language Haskell as an investigation into whether the well-known advantages of the functional paradigm can be exploited in the field of machine learning, which traditionally has been dominated by the procedural/object-oriented approach, particularly involving C/C++ and Matlab. HasGP is open-source software released under the GPL3 license. This manual provides a short introduction on how install the library, and how to apply it to supervised learning problems. It also provides some more in-depth information on the implementation of the library, which is aimed at developers. In the latter, we also show how some of the specific functional features of Haskell, in particular the ability to treat functions as first-class objects, and the use of typeclasses and monads, have informed the design of the library. This manual applies to HasGP version 0.1, which is the initial release of the library.

UCAM-CL-TR-805

Simon Hay:

A model personal energy meter

September 2011, 207 pages, PDF
PhD thesis (Girton College, August 2011)

Abstract: Every day each of us consumes a significant amount of energy, both directly through transport, heating and use of appliances, and indirectly from our needs for the production of food, manufacture of goods and provision of services.

This dissertation investigates a personal energy meter which can record and apportion an individual's energy usage in order to supply baseline information and incentives for reducing our environmental impact.

If the energy costs of large shared resources are split evenly without regard for individual consumption each person minimises his own losses by taking advantage of others. Context awareness offers the potential to change this balance and apportion energy costs to those who cause them to be incurred. This dissertation explores how sensor systems installed in many buildings today can be used to apportion energy consumption between users, including an evaluation of a range of strategies in a case study and elaboration of the overriding principles that are generally applicable. It also shows how second-order estimators combined with location data can provide a proxy for fine-grained sensing.

A key ingredient for apportionment mechanisms is data on energy usage. This may come from metering devices or buildings directly, or from profiling devices and using secondary indicators to infer their power state. A mechanism for profiling devices to determine the energy costs of specific activities, particularly applicable to shared programmable devices is presented which can make this process simpler and more accurate. By combining crowd-sourced building-inventory information and a simple building energy model it is possible to estimate an individual's energy use disaggregated by device class with very little direct sensing.

Contextual information provides crucial cues for apportioning the use and energy costs of resources, and one of the most valuable sources from which to infer context is location. A key ingredient for a personal energy meter is a low cost, low infrastructure location system that can be deployed on a truly global scale. This dissertation presents a description and evaluation of the new concept of inquiry-free Bluetooth tracking that has the potential to offer indoor location information with significantly less infrastructure and calibration than other systems.

Finally, a suitable architecture for a personal energy meter on a global scale is demonstrated using a mobile phone application to aggregate energy feeds based on the case studies and technologies developed.

UCAM-CL-TR-806

Damien Fay, Jérôme Kunegis, Eiko Yoneki:

On joint diagonalisation for dynamic network analysis

October 2011, 12 pages, PDF

Abstract: Joint diagonalisation (JD) is a technique used to estimate an average eigenspace of a set of matrices. Whilst it has been used successfully in many areas to track the evolution of systems via their eigenvectors; its application in network analysis is novel. The key focus in this paper is the use of JD on matrices of spanning trees of a network. This is especially useful in the case of real-world contact networks in which a single underlying static graph does not exist. The average eigenspace may be used to construct a graph which represents the 'average spanning tree' of the network or a representation of the most common propagation paths. We then examine the distribution of deviations from the average and find that this distribution in real-world contact networks is multi-modal; thus indicating several modes in the underlying network. These modes are identified and are found to correspond to particular times. Thus JD may be used to decompose the behaviour, in time, of contact networks and produce average static graphs for each time. This may be viewed as a mixture between a dynamic and static graph approach to contact network analysis.

UCAM-CL-TR-807

Ola Mahmoud:

Second-order algebraic theories

October 2011, 133 pages, PDF
PhD thesis (Clare Hall, March 2011)

Abstract: Second-order universal algebra and second-order equational logic respectively provide a model theory and a formal deductive system for languages with variable binding and parameterised metavariables. This dissertation completes the algebraic foundations of second-order languages from the viewpoint of categorical algebra.

In particular, the dissertation introduces the notion of second-order algebraic theory. A main role in the definition is played by the second-order theory of equality M, representing the most elementary operators and equations present in every second-order language. We show that M can be described abstractly via the universal property of being the free cartesian category on an exponentiable object. Thereby, in the tradition of categorical algebra, a second-order algebraic theory consists of a cartesian category TH and a strict cartesian identity-on-objects functor from M to TH that preserves the universal exponentiable object of M.

At the syntactic level, we establish the correctness of our definition by showing a categorical equivalence between second-order equational presentations and second-order algebraic theories. This equivalence, referred to as the Second-Order Syntactic Categorical Type Theory Correspondence, involves distilling a notion of syntactic translation between second-order equational presentations that corresponds to the canonical notion of morphism between second-order algebraic theories. Syntactic translations provide a mathematical formalisation of notions such as encodings and transforms for second-order languages.

On top of the aforementioned syntactic correspondence, we furthermore establish the Second-Order Semantic Categorical Type Theory Correspondence. This involves generalising Lawvere's notion of functorial model of algebraic theories to the second-order setting. By this semantic correspondence, second-order functorial semantics is shown to correspond to the model theory of second-order universal algebra.

We finally show that the core of the theory surrounding Lawvere theories generalises to the second order as well. Instances of this development are the existence of algebraic functors and monad morphisms in the second-order universe. Moreover, we define a notion of translation homomorphism that allows us to establish a 2-categorical type theory correspondence.

UCAM-CL-TR-808

Matko Botinčan, Mike Dodds, Suresh Jagannathan:

Resource-sensitive synchronisation inference by abduction

January 2012, 57 pages, PDF

Abstract: We present an analysis which takes as its input a sequential program, augmented with annotations indicating potential parallelization opportunities, and a sequential proof, written in separation logic, and produces a correctly-synchronized parallelized program and proof of that program. Unlike previous work, ours is not an independence analysis; we insert synchronization constructs to preserve relevant dependencies found in the sequential program that may otherwise be violated by a naïve translation. Separation logic allows us to parallelize fine-grained patterns of resource-usage, moving beyond straightforward points-to analysis.

Our analysis works by using the sequential proof to discover dependencies between different parts of the program. It leverages these discovered dependencies to guide the insertion of synchronization primitives into the parallelized program, and ensure that the resulting parallelized program satisfies the same specification as the original sequential program. Our analysis is built using frame inference and abduction, two techniques supported by an increasing number of separation logic tools.

UCAM-CL-TR-809

John L. Miller:

Distributed virtual environment scalability and security

October 2011, 98 pages, PDF
PhD thesis (Hughes Hall, October 2011)

Abstract: Distributed virtual environments (DVEs) have been an active area of research and engineering for more than 20 years. The most widely deployed DVEs are network games such as Quake, Halo, and World of Warcraft (WoW), with millions of users and billions of dollars in annual revenue. Deployed DVEs remain expensive centralized implementations despite significant research outlining ways to distribute DVE workloads.

This dissertation shows previous DVE research evaluations are inconsistent with deployed DVE needs. Assumptions about avatar movement and proximity – fundamental scale factors – do not match WoW's workload, and likely the workload of other deployed DVEs. Alternate workload models are explored and preliminary conclusions presented. Using realistic workloads it is shown that a fully decentralized DVE cannot be deployed to today's consumers, regardless of its overhead.

Residential broadband speeds are improving, and this limitation will eventually disappear. When it does, appropriate security mechanisms will be a fundamental requirement for technology adoption.

A trusted auditing system ("Carbon") is presented which has good security, scalability, and resource characteristics for decentralized DVEs. When performing exhaustive auditing, Carbon adds 27% network overhead to a decentralized DVE with a WoW-like workload. This resource consumption can be reduced significantly, depending upon the DVE's risk tolerance. Finally, the Pairwise Random Protocol (PRP) is described. PRP enables adversaries to fairly resolve probabilistic activities, an ability missing from most decentralized DVE security proposals.

Thus, this dissertation's contribution is to address two of the obstacles for deploying research on decentralized DVE architectures. First, lack of evidence that research results apply to existing DVEs. Second, the lack of security systems combining appropriate security guarantees with acceptable overhead.

UCAM-CL-TR-810

Nick Barrow-Williams:

Proximity Coherence for chip-multiprocessors

November 2011, 164 pages, PDF
PhD thesis (Trinity Hall, January 2011)

Abstract: Many-core architectures provide an efficient way of harnessing the growing numbers of transistors available in modern fabrication processes; however, the parallel programs run on these platforms are increasingly limited by the energy and latency costs of communication. Existing designs provide a functional communication layer but do not necessarily implement the most efficient solution for chip-multiprocessors, placing limits on the performance of these complex systems. In an era of increasingly power limited silicon design, efficiency is now a primary concern that motivates designers to look again at the challenge of cache coherence.

The first step in the design process is to analyse the communication behaviour of parallel benchmark suites such as Parsec and SPLASH-2. This thesis presents work detailing the sharing patterns observed when running the full benchmarks on a simulated 32-core x86 machine. The results reveal considerable locality of shared data accesses between threads with consecutive operating system assigned thread IDs. This pattern, although of little consequence in a multi-node system, corresponds to strong physical locality of shared data between adjacent cores on a chip-multiprocessor platform.

Traditional cache coherence protocols, although often used in chip-multiprocessor designs, have been developed in the context of older multi-node systems. By redesigning coherence protocols to exploit new patterns such as the physical locality of shared data, improving the efficiency of communication, specifically in chip-multiprocessors, is possible. This thesis explores such a design – Proximity Coherence – a novel scheme in which L1 load misses are optimistically forwarded to nearby caches via new dedicated links rather than always being indirected via a directory structure.

UCAM-CL-TR-811

A. Theodore Markettos:

Active electromagnetic attacks on secure hardware

December 2011, 217 pages, PDF
PhD thesis (Clare Hall, March 2010)

Abstract: The field of side-channel attacks on cryptographic hardware has been extensively studied. In many cases it is easier to derive the secret key from these attacks than to break the cryptography itself. One such sidechannel attack is the electromagnetic side-channel attack, giving rise to electromagnetic analysis (EMA).

EMA, when otherwise known as 'TEMPEST' or 'compromising emanations', has a long history in the military context over almost the whole of the twentieth century. The US military also mention three related attacks, believed to be: HIJACK (modulation of secret data onto conducted signals), NONSTOP (modulation of secret data onto radiated signals) and TEAPOT (intentional malicious emissions).

In this thesis I perform a fusion of TEAPOT and HIJACK/NONSTOP techniques on secure integrated circuits. An attacker is able to introduce one or more frequencies into a cryptographic system with the intention of forcing it to misbehave or to radiate secrets.

I demonstrate two approaches to this attack:

To perform the reception, I assess a variety of electromagnetic sensors to perform EMA. I choose an inductive hard drive head and a metal foil electric field sensor to measure near-field EM emissions.

The first approach, named the re-emission attack, injects frequencies into the power supply of a device to cause it to modulate up baseband signals. In this way I detect data-dependent timing from a 'secure' microcontroller. Such up-conversion enables a more compact and more distant receiving antenna.

The second approach involves injecting one or more frequencies into the power supply of a random number generator that uses jitter of ring oscillators as its random number source. I am able to force injection locking of the oscillators, greatly diminishing the entropy available.

I demonstrate this with the random number generators on two commercial devices. I cause a 2004 EMV banking smartcard to fail statistical test suites by generating a periodicity. For a secure 8-bit microcontroller that has been used in banking ATMs, I am able to reduce the random number entropy from 2^{32} to 225. This enables a 50% probability of a successful attack on cash withdrawal in 15 attempts.

UCAM-CL-TR-812

Pedro Brandão:

Abstracting information on body area networks

January 2012, 144 pages, PDF
PhD thesis (Magdalene College, July 2011)

Abstract: Healthcare is changing, correction, healthcare is in need of change. The population ageing, the increase in chronic and heart diseases and just the increase in population size will overwhelm the current hospital-centric healthcare.

There is a growing interest by individuals to monitor their own physiology. Not only for sport activities, but also to control their own diseases. They are changing from the passive healthcare receiver to a proactive self-healthcare taker. The focus is shifting from hospital centred treatment to a patient-centric healthcare monitoring.

Continuous, everyday, wearable monitoring and actuating is part of this change. In this setting, sensors that monitor the heart, blood pressure, movement, brain activity, dopamine levels, and actuators that pump insulin, 'pump' the heart, deliver drugs to specific organs, stimulate the brain are needed as pervasive components in and on the body. They will tend for

people's need of self-monitoring and facilitate healthcare delivery.

These components around a human body that communicate to sense and act in a coordinated fashion make a Body Area Network (BAN). In most cases, and in our view, a central, more powerful component will act as the coordinator of this network. These networks aim to augment the power to monitor the human body and react to problems discovered with this observation. One key advantage of this system is their overarching view of the whole network. That is, the central component can have an understanding of all the monitored signals and correlate them to better evaluate and react to problems. This is the focus of our thesis.

In this document we argue that this multi-parameter correlation of the heterogeneous sensed information is not being handled in BANs. The current view depends exclusively on the application that is using the network and its understanding of the parameters. This means that every application will oversee the BAN's heterogeneous resources managing them directly without taking into consideration other applications, their needs and knowledge.

There are several physiological correlations already known by the medical field. Correlating blood pressure and cross sectional area of blood vessels to calculate blood velocity, estimating oxygen delivery from cardiac output and oxygen saturation, are such examples. This knowledge should be available in a BAN and shared by the several applications that make use of the network. This architecture implies a central component that manages the knowledge and the resources. And this is, in our view, missing in BANs.

Our proposal is a middleware layer that abstracts the underlying BAN's resources to the application, providing instead an information model to be queried. The model describes the correlations for producing new information that the middleware knows about. Naturally, the raw sensed data is also part of the model. The middleware hides the specificities of the nodes that constitute the BAN, by making available their sensed production. Applications are able to query for information attaching requirements to these requests. The middleware is then responsible for satisfying the requests while optimising the resource usage of the BAN.

Our architecture proposal is divided in two corresponding layers, one that abstracts the nodes' hardware (hiding node's particularities) and the information layer that describes information available and how it is correlated. A prototype implementation of the architecture was done to illustrate the concept.

UCAM-CL-TR-813

Andrew B. Lewis:

Reconstructing compressed photo and video data

February 2012, 148 pages, PDF
PhD thesis (Trinity College, June 2011)

Abstract: Forensic investigators sometimes need to verify the integrity and processing history of digital photos and videos. The multitude of storage formats and devices they need to access also presents a challenge for evidence recovery. This thesis explores how visual data files can be recovered and analysed in scenarios where they have been stored in the JPEG or H.264 (MPEG-4 AVC) compression formats.

My techniques make use of low-level details of lossy compression algorithms in order to tell whether a file under consideration might have been tampered with. I also show that limitations of entropy coding sometimes allow us to recover intact files from storage devices, even in the absence of filesystem and container metadata.

I first show that it is possible to embed an imperceptible message within a uniform region of a JPEG image such that the message becomes clearly visible when the image is recompressed at a particular quality factor, providing a visual warning that recompression has taken place.

I then use a precise model of the computations involved in JPEG decompression to build a specialised compressor, designed to invert the computations of the decompressor. This recompressor recovers the compressed bitstreams that produce a given decompression result, and, as a side-effect, indicates any regions of the input which are inconsistent with JPEG decompression. I demonstrate the algorithm on a large database of images, and show that it can detect modifications to decompressed image regions.

Finally, I show how to rebuild fragmented compressed bitstreams, given a syntax description that includes information about syntax errors, and demonstrate its applicability to H.264/AVC Baseline profile video data in memory dumps with randomly shuffled blocks.

UCAM-CL-TR-814

Arjuna Sathiaseelan, Jon Crowcroft:

The free Internet: a distant mirage or near reality?

February 2012, 10 pages, PDF

Abstract: Through this short position paper, we hope to convey our thoughts on the need for free Internet access and describe possible ways of achieving this – hoping this stimulates a useful discussion.

UCAM-CL-TR-815

Christian Richardt:

Colour videos with depth: acquisition, processing and evaluation

March 2012, 132 pages, PDF
PhD thesis (Gonville & Caius College, November 2011)

Abstract: The human visual system lets us perceive the world around us in three dimensions by integrating evidence from depth cues into a coherent visual model of the world. The equivalent in computer vision and computer graphics are geometric models, which provide a wealth of information about represented objects, such as depth and surface normals. Videos do not contain this information, but only provide per-pixel colour information. In this dissertation, I hence investigate a combination of videos and geometric models: videos with per-pixel depth (also known as RGBZ videos). I consider the full life cycle of these videos: from their acquisition, via filtering and processing, to stereoscopic display.

UCAM-CL-TR-816

Jean E. Martina:

Verification of security protocols based on multicast communication

March 2012, 150 pages, PDF
PhD thesis (Clare College, February 2011)

Abstract: Over an insecure network, agents need means to communicate securely. These means are often called security protocols. Security protocols, although constructed through the arrangement of simple security blocks, normally yield complex goals. They seem simple at a first glance, but hide subtleties that allow them to be exploited.

One way of trying to systematically capture such subtleties is through the use of formal methods. The maturity of some methods for protocol verification is a fact today. But these methods are still not able to capture the whole set of security protocols being designed. With the convergence to an online world, new security goals are proposed and new protocols need to be designed. The evolution of formal verification methods becomes a necessity to keep pace with this ongoing development.

This thesis covers the Inductive Method and its extensions. The Inductive Method is a formalism to specify and verify security protocols based on structural induction and higher-order logic proofs. This account of our extensions enables the Inductive Method to reason about non-Unicast communication and threshold cryptography.

We developed a new set of theories capable of representing the entire set of known message casting frameworks. Our theories enable the Inductive Method to reason about a whole new set of protocols. We also specified a basic abstraction of threshold cryptography as a way of proving the extensibility of the method to new cryptographic primitives. We showed the feasibility of our specifications by revisiting a classic protocol, now verified under our framework. Secrecy verification under a mixed environment of Multicast and Unicast was also done for a Byzantine security protocol.

UCAM-CL-TR-817

Joseph Bonneau, Cormac Herley, Paul C. van Oorschot, Frank Stajano:

The quest to replace passwords: a framework for comparative evaluation of Web authentication schemes

March 2012, 32 pages, PDF

Abstract: We evaluate two decades of proposals to replace text passwords for general-purpose user authentication on the web using a broad set of twenty-five usability, deployability and security benefits that an ideal scheme might provide. The scope of proposals we survey is also extensive, including password management software, federated login protocols, graphical password schemes, cognitive authentication schemes, one-time passwords, hardware tokens, phone-aided schemes and biometrics. Our comprehensive approach leads to key insights about the difficulty of replacing passwords. Not only does no known scheme come close to providing all desired benefits: none even retains the full set of benefits which legacy passwords already provide. In particular, there is a wide range between schemes offering minor security benefits beyond legacy passwords, to those offering significant security benefits in return for being more costly to deploy or difficult to use. We conclude that many academic proposals have failed to gain traction because researchers rarely consider a sufficiently wide range of real-world constraints. Beyond our analysis of current schemes, our framework provides an evaluation methodology and benchmark for future web authentication proposals.

This report is an extended version of the peer-reviewed paper by the same name. In about twice as many pages it gives full ratings for 35 authentication schemes rather than just 9.

UCAM-CL-TR-818

Robert N. M. Watson:

New approaches to operating system security extensibility

April 2012, 184 pages, PDF
PhD thesis (Wolfson College, October 2010)

Abstract: This dissertation proposes new approaches to commodity computer operating system (OS) access control extensibility that address historic problems

with concurrency and technology transfer. Access control extensibility addresses a lack of consensus on operating system policy model at a time when security requirements are in flux: OS vendors, anti-virus companies, firewall manufacturers, smart phone developers, and application writers require new tools to express policies tailored to their needs. By proposing principled approaches to access control extensibility, this work allows OS security to be "designed in" yet remain flexible in the face of diverse and changing requirements.

I begin by analysing system call interposition, a popular extension technology used in security research and products, and reveal fundamental and readily exploited concurrency vulnerabilities. Motivated by these failures, I propose two security extension models: the TrustedBSD Mandatory Access Control (MAC) Framework, a flexible kernel access control extension framework for the FreeBSD kernel, and Capsicum, practical capabilities for UNIX.

The MAC Framework, a research project I began before starting my PhD, allows policy modules to dynamically extend the kernel access control policy. The framework allows policies to integrate tightly with kernel synchronisation, avoiding race conditions inherent to system call interposition, as well as offering reduced development and technology transfer costs for new security policies. Over two chapters, I explore the framework itself, and its transfer to and use in several products: the open source FreeBSD operating system, nCircle's enforcement appliances, and Apple's Mac OS X and iOS operating systems.

Capsicum is a new application-centric capability security model extending POSIX. Capsicum targets application writers rather than system designers, reflecting a trend towards security-aware applications such as Google's Chromium web browser, that map distributed security policies into often inadequate local primitives. I compare Capsicum with other sandboxing techniques, demonstrating improved performance, programmability, and security.

This dissertation makes original contributions to challenging research problems in security and operating system design. Portions of this research have already had a significant impact on industry practice.

UCAM-CL-TR-819

Joseph Bonneau:

Guessing human-chosen secrets

May 2012, 161 pages, PDF
PhD thesis (Churchill College, May 2012)

Abstract: Authenticating humans to computers remains a notable weak point in computer security despite decades of effort. Although the security research community has explored dozens of proposals for replacing or strengthening passwords, they appear likely to remain entrenched as the standard mechanism of human-computer authentication on the Internet for years to come. Even in the optimistic scenario of eliminating passwords from most of today's authentication protocols using trusted hardware devices or trusted servers to perform federated authentication, passwords will persist as a means of "last-mile" authentication between humans and these trusted single sign-on deputies.

This dissertation studies the difficulty of guessing human-chosen secrets, introducing a sound mathematical framework modeling human choice as a skewed probability distribution. We introduce a new metric, alpha-guesswork, which can accurately model the resistance of a distribution against all possible guessing attacks. We also study the statistical challenges of estimating this metric using empirical data sets which can be modeled as a large random sample from the underlying probability distribution.

This framework is then used to evaluate several representative data sets from the most important categories of human-chosen secrets to provide reliable estimates of security against guessing attacks. This includes collecting the largest-ever corpus of user-chosen passwords, with nearly 70 million, the largest list of human names ever assembled for research, the largest data sets of real answers to personal knowledge questions and the first data published about human choice of banking PINs. This data provides reliable numbers for designing security systems and highlights universal limitations of human-chosen secrets.

UCAM-CL-TR-820

Eiko Yoneki, Amitabha Roy:

A unified graph query layer for multiple databases

August 2012, 22 pages, PDF

Abstract: There is increasing demand to store and query data with an inherent graph structure. Examples of such data include those from online social networks, the semantic web and from navigational queries on spatial data such as maps. Unfortunately, traditional relational databases have fallen short where such graph structured data is concerned. This has led to the development of specialised graph databases such as Neo4j. However, traditional databases continue to have a wide usage base and have desirable properties such as the capacity to support a high volume of transactions while offering ACID semantics. In this paper we argue that it is in fact possible to unify different database paradigms together in the case of graph structured data through the use of a common query language and data loader that we have named Crackle (a wordplay on Gra[ph]QL). Crackle provides an expressive and powerful query library in Clojure (a functional LISP dialect for JVMs). It also provides a data loader that is capable of interfacing transparently with various data sources such as PostgreSQL databases and the Redis key-value store. Crackle shields programmers from the

backend database by allowing them to write queries in Clojure. Additionally, its graph-focused prefetchers are capable of closing the hitherto large gap between a PostgreSQL database and a specialised graph database such as Neo4j from as much 326x (with a SQL query) to as low as 6x (when using Crackle). We also include a detailed performance analysis that identifies ways to further reduce this gap with Crackle. This brings into question the performance argument for specialised graph databases such as Neo4j by providing comparable performance on supposedly legacy data sources.

UCAM-CL-TR-821

Charles Reams:

Modelling energy efficiency for computation

October 2012, 135 pages, PDF
PhD thesis (Clare College, October 2012)

Abstract: In the last decade, efficient use of energy has become a topic of global significance, touching almost every area of modern life, including computing. From mobile to desktop to server, energy efficiency concerns are now ubiquitous. However, approaches to the energy problem are often piecemeal and focus on only one area for improvement.

I argue that the strands of the energy problem are inextricably entangled and cannot be solved in isolation. I offer a high-level view of the problem and, building from it, explore a selection of subproblems within the field. I approach these with various levels of formality, and demonstrate techniques to make improvements on all levels. The original contributions are as follows:

Chapter 3 frames the energy problem as one of optimisation with constraints, and explores the impact of this perspective for current commodity products. This includes considerations of the hardware, software and operating system. I summarise the current situation in these respects and propose directions in which they could be improved to better support energy management.

Chapter 4 presents mathematical techniques to compute energy-optimal schedules for long-running computations. This work reflects the server-domain concern with energy cost, producing schedules that exploit fluctuations in power cost over time to minimise expenditure rather than raw energy. This assumes certain idealised models of power, performance, cost, and workload, and draws precise formal conclusions from them.

Chapter 5 considers techniques to implement energy-efficient real-time streaming. Two classes of problem are considered: first, hard real-time streaming with fixed, predictable frame characteristics; second, soft real-time streaming with a quality-of-service guarantee and probabilistic descriptions of per-frame workload. Efficient algorithms are developed for scheduling frame execution in an energy-efficient way while still guaranteeing hard real-time deadlines. These schedules determine appropriate values for power-relevant parameters, such as dynamic voltage–frequency scaling.

A key challenge for future work will be unifying these diverse approaches into one "Theory of Energy" for computing. The progress towards this is summarised in Chapter 6. The thesis concludes by sketching future work towards this Theory of Energy.

UCAM-CL-TR-822

Richard A. Russell:

Planning with preferences using maximum satisfiability

October 2012, 160 pages, PDF
PhD thesis (Gonville and Caius College, September 2011)

Abstract: The objective of automated planning is to synthesise a plan that achieves a set of goals specified by the user. When achieving every goal is not feasible, the planning system must decide which ones to plan for and find the lowest cost plan. The system should take as input a description of the user's preferences and the costs incurred through executing actions. Goal utility dependencies arise when the utility of achieving a goal depends on the other goals that are achieved with it. This complicates the planning procedure because achieving a new goal can alter the utilities of all the other goals currently achieved.

In this dissertation we present methods for solving planning problems with goal utility dependencies by compiling them to a variant of satisfiability known as weighted partial maximum satisfiability (WPMax-SAT). An optimal solution to the encoding is found using a general-purpose solver. The encoding is constructed such that its optimal solution can be used to construct a plan that is most preferred amongst other plans of length that fit within a prespecified horizon. We evaluate this approach against an integer programming based system using benchmark problems taken from past international planning competitions.

We study how a WPMax-SAT solver might benefit from incorporating a procedure known as survey propagation. This is a message passing algorithm that estimates the probability that a variable is constrained to be a particular value in a randomly selected satisfying assignment. These estimates are used to influence variable/value decisions during search for a solution. Survey propagation is usually presented with respect to the satisfiability problem, and its generalisation, SP(y), with respect to the maximum satisfiability problem. We extend the argument that underpins these two algorithms to derive a new set of message passing equations for application to WPMax-SAT problems. We evaluate the success of this method by applying it to our encodings of planning problems with goal utility dependencies.

Our results indicate that planning with preferences using WPMax-SAT is competitive and sometimes more successful than an integer programming approach – solving two to three times more subproblems in some domains, while being outperformed by a smaller margin in others. In some domains, we also find that using information provided by survey propagation in a WPMax-SAT solver to select variable/value pairs for the earliest decisions can, on average, direct search to lower cost solutions than a uniform sampling strategy combined with a popular heuristic.

UCAM-CL-TR-823

Amitabha Roy, Karthik Nilakant,
Valentin Dalibard, Eiko Yoneki:

Mitigating I/O latency in SSD-based graph traversal

November 2012, 27 pages, PDF

Abstract: Mining large graphs has now become an important aspect of many applications. Recent interest in low cost graph traversal on single machines has lead to the construction of systems that use solid state drives (SSDs) to store the graph. An SSD can be accessed with far lower latency than magnetic media, while remaining cheaper than main memory. Unfortunately SSDs are slower than main memory and algorithms running on such systems are hampered by large IO latencies when accessing the SSD. In this paper we present two novel techniques to reduce the impact of SSD IO latency on semi-external memory graph traversal. We introduce a variant of the Compressed Sparse Row (CSR) format that we call Compressed Enumerated Encoded Sparse Offset Row (CEESOR). CEESOR is particularly efficient for graphs with hierarchical structure and can reduce the space required to represent connectivity information by amounts varying from 5% to as much as 76%. CEESOR allows a larger number of edges to be moved for each unit of IO transfer from the SSD to main memory and more effective use of operating system caches. Our second contribution is a runtime prefetching technique that exploits the ability of solid state drives to service multiple random access requests in parallel. We present a novel Run Along SSD Prefetcher (RASP). RASP is capable of hiding the effect of IO latency in single threaded graph traversal in breadth-first and shorted path order to the extent that it improves iteration time for large graphs by amounts varying from 2.6X-6X.

UCAM-CL-TR-824

Simon Frankau:

Hardware synthesis from a stream-processing functional language

November 2012, 202 pages, PDF
PhD thesis (St. John's College, July 2004)

Abstract: As hardware designs grow exponentially larger, there is an increasing challenge to use transistor budgets effectively. Without higher-level synthesis tools, so much effort may be spent on low-level details that it becomes impractical to effectively design circuits of the size that can be fabricated. This possibility of a design gap has been documented for some time now.

One solution is the use of domain-specific languages. This thesis covers the use of software-like languages to describe algorithms that are to be implemented in hardware. Hardware engineers can use the tools to improve their productivity and effectiveness in this particular domain. Software engineers can also use this approach to benefit from the parallelism available in modern hardware (such as reconfigurable systems and FPGAs), while retaining the convenience of a software description.

In this thesis a statically-allocated pure functional language, SASL, is introduced. Static allocation makes the language suited to implementation in fixed hardware resources. The I/O model is based on streams (linear lazy lists), and implicit parallelism is used in order to maintain a software-like approach. The thesis contributes constraints which allow the language to be statically-allocated, and synthesis techniques for SASL targeting both basic CSP and a graph-based target that may be compiled to a register-transfer level (RTL) description.

Further chapters examine the optimisation of the language, including the use of lenient evaluation to increase parallelism, the introduction of closures and general lazy evaluation, and the use of non-determinism in the language. The extensions are examined in terms of the restrictions required to ensure static allocation, and the techniques required to synthesise them.

UCAM-CL-TR-825

Jonathan Anderson:

Privacy engineering for social networks

December 2012, 255 pages, PDF
PhD thesis (Trinity College, July 2012)

Abstract: In this dissertation, I enumerate several privacy problems in online social networks (OSNs) and describe a system called Footlights that addresses them. Footlights is a platform for distributed social applications that allows users to control the sharing of private information. It is designed to compete with the performance of todays centralised OSNs, but it does not trust centralised infrastructure to enforce security properties.

Based on several socio-technical scenarios, I extract concrete technical problems to be solved and show how the existing research literature does not solve them.

Addressing these problems fully would fundamentally change users interactions with OSNs, providing real control over online sharing.

I also demonstrate that todays OSNs do not provide this control: both user data and the social graph are vulnerable to practical privacy attacks.

Footlights storage substrate provides private, scalable, sharable storage using untrusted servers. Under realistic assumptions, the direct cost of operating this storage system is less than one US dollar per user-year. It is the foundation for a practical shared filesystem, a perfectly unobservable communications channel and a distributed application platform.

The Footlights application platform allows third-party developers to write social applications without direct access to users private data. Applications run in a conned environment with a private-by-default security model: applications can only access user information with explicit user consent. I demonstrate that practical applications can be written on this platform.

The security of Footlights user data is based on public-key cryptography, but users are able to log in to the system without carrying a private key on a hardware token. Instead, users authenticate to a set of authentication agents using a weak secret such as a user-chosen password or randomly-assigned 4-digit number. The protocol is designed to be secure even in the face of malicious authentication agents.

UCAM-CL-TR-826

Fernando M. V. Ramos:

GREEN IPTV: a resource and energy efficient network for IPTV

December 2012, 152 pages, PDF
PhD thesis (Clare Hall, November 2012)

Abstract: The distribution of television is currently dominated by three technologies: over-the-air broadcast, cable, and satellite. The advent of IP networks and the increased availability of broadband access created a new vehicle for the distribution of TV services. The distribution of digital TV services over IP networks, or IPTV, offers carriers flexibility and added value in the form of additional services. It causes therefore no surprise the rapid roll-out of IPTV services by operators worldwide in the past few years.

IPTV distribution imposes stringent requirements on both performance and reliability. It is therefore challenging for an IPTV operator to guarantee the quality of experience expected by its users, and doing so in an efficient manner. In this dissertation I investigate some of the challenges faced by IPTV distribution network operators, and I propose novel techniques to address these challenges.

First, I address one of the major concerns of IPTV network deployment: channel change delay. This is the latency experienced by users when switching between TV channels. Synchronisation and buffering of video streams can cause channel change delays of several seconds. I perform an empirical analysis of a particular solution to the channel change delay problem, namely, predictive pre-joining of TV channels. In this scheme each Set Top Box simultaneously joins additional multicast groups (TV channels) along with the one requested by the user. If the user switches to any of these channels next, switching latency is virtually eliminated, and user experience is improved. The results show that it is possible to eliminate zapping delay for a significant percentage of channel switching requests with little impact in access network bandwidth cost.

Second, I propose a technique to increase the resource and energy efficiency of IPTV networks. This technique is based on a simple paradigm: avoiding waste. To reduce the inefficiencies of current static multicast distribution schemes, I propose a semi-dynamic scheme where only a selection of TV multicast groups is distributed in the network, instead of all. I perform an empirical evaluation of this method and conclude that its use results in significant bandwidth reductions without compromising service performance. I also demonstrate that these reductions may translate into significant energy savings in the future.

Third, to increase energy efficiency further I propose a novel energy and resource friendly protocol for core optical IPTV networks. The idea is for popular IPTV traffic to optically bypass the network nodes, avoiding electronic processing. I evaluate this proposal empirically and conclude that the introduction of optical switching techniques results in a significant increase in the energy efficiency of IPTV networks.

All the schemes I present in this dissertation are evaluated by means of trace-driven analyses using a dataset from an operational IPTV service provider. Such thorough and realistic evaluation enables the assessment of the proposed techniques with an increased level of confidence, and is therefore a strength of this dissertation.

UCAM-CL-TR-827

Omar S. Choudary:

The smart card detective: a hand-held EMV interceptor

December 2012, 55 pages, PDF

Abstract: Several vulnerabilities have been found in the EMV system (also known as Chip and PIN). Saar Drimer and Steven Murdoch have successfully implemented a relay attack against EMV using a fake terminal. Recently the same authors have found a method to successfully complete PIN transactions without actually entering the correct PIN. The press has published this vulnerability but they reported such a scenario as being hard to execute in practice because it requires specialized and complex hardware.

As proposed by Ross Anderson and Mike Bond in 2006, I decided to create a miniature man-in-the-middle device to defend smartcard users against relay attacks.

As a result of my MPhil project work I created a hand-held device, called Smart Card Defender (SCD), which intercepts the communication between smartcard and terminal. The device has been built using a low cost ATMEL AT90USB1287 microcontroller and other readily available electronic components. The total cost of the SCD has been around £100, but an industrial version could be produced for less than £20.

I implemented several applications using the SCD, including the defense against the relay attack as well as the recently discovered vulnerability to complete a transaction without using the correct PIN.

All the applications have been successfully tested on CAP readers and live terminals. Furthermore, I have performed real tests using the SCD at several shops in town.

From the experiments using the SCD, I have noticed some particularities of the CAP protocol compared to the EMV standard. I have also discovered that the smartcard does not follow the physical transport protocol exactly. Such findings are presented in detail, along with a discussion of the results.

UCAM-CL-TR-828

Rosemary M. Francis:

Exploring networks-on-chip for FPGAs

January 2013, 121 pages, PDF
PhD thesis (Darwin College, July 2009)

Abstract: Developments in fabrication processes have shifted the cost ratio between wires and transistors to allow new trade-offs between computation and communication. Rising clock speeds have lead to multi-cycle cross-chip communication and pipelined buses. It is then a small step from pipelining to switching and the development of multi-core networked systems-on-chip. Modern FPGAs are also now home to complex systems-on-chip. A change in the way we structure the computation demands a change in the way we structure the communication on-chip.

This thesis looks at Network-on-Chip design for FPGAs beyond the trade-offs between hard (silicon) and soft (configurable) designs. FPGAs are capable of extremely flexible, statically routed bit-based wiring, but this flexibility comes at a high area, latency and power cost. Soft NoCs are able to maintain this flexibility, but do not necessarily make good use of the computation-communication trade-off. Hard NoCs are more efficient when used, but are forced to operate below capacity by the soft IP cores. It is also difficult to design hard NoCs with the flexibility needed without wasting silicon when the network is not used.

In the first part of this thesis I explore the capability of Time-Division Multiplexed (TDM) wiring to bridge the gap between the fine-grain static FPGA wiring and the bus-based dynamic routing of a NoC. By replacing some of the static FPGA wiring with TDM wiring I am able to time division multiplex hard routers and make better use of the non-configurable area. The cost of a hard network is reduced by moving some of the area cost from the routers into reusable TDM wiring components. The TDM wiring improves the interface between the hard routers and soft IP blocks which leads to higher logic density overall. I show that TDM wiring makes hard routers a flexible and efficient alternative to soft interconnect.

The second part of this thesis looks at the feasibility of replacing all static wiring on the FPGA with TDM wiring. The aim was to increase the routing capacity of the FPGA whilst decreasing the area used to implement it. An ECAD flow was developed to explore the extent to which the amount of wiring can be reduced. The results were then used to design the TDM circuitry.

My results show that an 80% reduction in the amount of wiring is possible though time-division multiplexing. This reduction is sufficient to increase the routing capacity of the FPGA whilst maintaining similar or better logic density. This TDM wiring can be used to implement area and power-efficient hard networks-on-chip with good flexibility, as well as improving the performance of other hard IP blocks.

UCAM-CL-TR-829

Philip Christopher Paul:

Microelectronic security measures

February 2013, 177 pages, PDF
PhD thesis (Pembroke College, January 2009)

Abstract: In this dissertation I propose the concept of tamper protection grids for microelectronic security devices made from organic electronic materials. As security devices have become ubiquitous in recent years, they are becoming targets for criminal activity. One general attack route to breach the security is to carry out physical attack after depackaging a device. Commercial security devices use a metal wire mesh within the chip to protect against these attacks. However, as a microchip is physically robust, the mesh is not affected by depackaging.

As a better way of protecting security devices against attacks requiring the chip package to be removed, I investigate a protection grid that is vulnerable to damage if the packaging is tampered with. The protection grid is connected directly to standard bond pads on the microchip, to allow direct electronic measurements, saving the need for complex sensor structures. That way, a security device can monitor the package for integrity, and initiate countermeasures if required.

The feasibility of organic tamper protection grids was evaluated. To establish the viability of the concept,

a fabrication method for these devices was developed, the sensitivity to depackaging was assessed, and practical implementation issues were evolved. Inkjet printing was chosen as fabrication route, as devices can be produced at low cost while preserving flexibility of layout. A solution to the problem of adverse surface interaction was found to ensure good print quality on the hydrophobic chip surface. Standard contacts between chip and grid are non-linear and degrade between measurements, however it was shown that stable ohmic contacts are possible using a silver buffer layer. The sensitivity of the grid to reported depackaging methods was tested, and improvements to the structure were found to maximise damage to the grid upon tampering with the package. Practical issues such as measurement stability with temperature and age were evaluated, as well as a first prototype to assess the achievable measurement accuracy. The evaluation of these practical issues shows directions for future work that can develop organic protection grids beyond the proof of concept.

Apart from the previously mentioned invasive attacks, there is a second category of attacks, non-invasive attacks, that do not require the removal of the chip packaging. The most prominent non-invasive attack is power analysis in which the power consumption of a device is used as oracle to reveal the secret key of a security device. Logic gates were designed and fabricated with data-independent power consumption in each clock cycle. However, it is shown that this is not sufficient to protect the secret key. Despite balancing the discharged capacitances in each clock cycle, the power consumed still depends on the data input. While the overall charge consumed in each clock cycle matches to a few percent, differences within a clock cycle can easily be measured. It was shown that the dominant cause for this imbalance is early propagation, which can be mitigated by ensuring that evaluation in a gate only takes place after all inputs are present. The second major source of imbalance are mismatched discharge paths in logic gates, which result in data-dependent evaluation times of a gate. This source of imbalance is not as trivial to remove, as it conflicts with balancing the discharged capacitances in each clock cycle.

UCAM-CL-TR-830

Paul J. Fox:

Massively parallel neural computation

March 2013, 105 pages, PDF
PhD thesis (Jesus College, October 2012)

Abstract: Reverse-engineering the brain is one of the US National Academy of Engineering's 'Grand Challenges'. The structure of the brain can be examined at many different levels, spanning many disciplines from low-level biology through psychology and computer science. This thesis focusses on real-time computation of large neural networks using the Izhikevich spiking neuron model.

Neural computation has been described as 'embarrassingly parallel' as each neuron can be thought of as an independent system, with behaviour described by a mathematical model. However, the real challenge lies in modelling neural communication. While the connectivity of neurons has some parallels with that of electrical systems, its high fan-out results in massive data processing and communication requirements when modelling neural communication, particularly for real-time computations.

It is shown that memory bandwidth is the most significant constraint to the scale of real-time neural computation, followed by communication bandwidth, which leads to a decision to implement a neural computation system on a platform based on a network of Field Programmable Gate Arrays (FPGAs), using commercial off-the-shelf components with some custom supporting infrastructure. This brings implementation challenges, particularly lack of on-chip memory, but also many advantages, particularly high-speed transceivers. An algorithm to model neural communication that makes efficient use of memory and communication resources is developed and then used to implement a neural computation system on the multi-FPGA platform.

Finding suitable benchmark neural networks for a massively parallel neural computation system proves to be a challenge. A synthetic benchmark that has biologically-plausible fan-out, spike frequency and spike volume is proposed and used to evaluate the system. It is shown to be capable of computing the activity of a network of 256k Izhikevich spiking neurons with a fan-out of 1k in real-time using a network of 4 FPGA boards. This compares favourably with previous work, with the added advantage of scalability to larger neural networks using more FPGAs.

It is concluded that communication must be considered as a first-class design constraint when implementing massively parallel neural computation systems.

UCAM-CL-TR-831

Meredydd Luff:

Communication for programmability and performance on multi-core processors

April 2013, 89 pages, PDF
PhD thesis (Gonville & Caius College, November 2012)

Abstract: The transition to multi-core processors has yielded a fundamentally new sort of computer. Software can no longer benefit passively from improvements in processor technology, but must perform its computations in parallel if it is to take advantage of the continued increase in processing power. Software

development has yet to catch up, and with good reason: parallel programming is hard, error-prone and often unrewarding.

In this dissertation, I consider the programmability challenges of the multi-core era, and examine three angles of attack.

I begin by reviewing alternative programming paradigms which aim to address these changes, and investigate two popular alternatives with a controlled pilot experiment. The results are inconclusive, and subsequent studies in that field have suffered from similar weakness. This leads me to conclude that empirical user studies are poor tools for designing parallel programming systems.

I then consider one such alternative paradigm, transactional memory, which has promising usability characteristics but suffers performance overheads so severe that they mask its benefits. By modelling an ideal inter-core communication mechanism, I propose using our embarrassment of parallel riches to mitigate these overheads. By pairing "helper" processors with application threads, I offload the overheads of software transactional memory, thereby greatly mitigating the problem of serial overhead.

Finally, I address the mechanics of inter-core communication. Due to the use of cache coherence to preserve the programming model of previous processors, explicitly communicating between the cores of any modern multi-core processor is painfully slow. The schemes proposed so far to alleviate this problem are complex, insufficiently general, and often introduce new resources which cannot be virtualised transparently by a time-sharing operating system. I propose and describe an asynchronous remote store instruction, which is issued by one core and completed asynchronously by another into its own local cache. I evaluate several patterns of parallel communication, and determine that the use of remote stores greatly increases the performance of common synchronisation kernels. I quantify the benefit to the feasibility of fine-grained parallelism. To finish, I use this mechanism to implement my parallel STM scheme, and demonstrate that it performs well, reducing overheads significantly.

UCAM-CL-TR-832

Gregory A. Chadwick:

Communication centric, multi-core, fine-grained processor architecture

April 2013, 165 pages, PDF
PhD thesis (Fitzwilliam College, September 2012)

Abstract: With multi-core architectures now firmly entrenched in many application areas both computer architects and programmers now face new challenges. Computer architects must increase core count to increase explicit parallelism available to the programmer in order to provide better performance whilst leaving the programming model presented tractable. The programmer must find ways to exploit this explicit parallelism provided that scales well with increasing core and thread availability.

A fine-grained computation model allows the programmer to expose a large amount of explicit parallelism and the greater the level of parallelism exposed the better increasing core counts can be utilised. However a fine-grained approach implies many interworking threads and the overhead of synchronising and scheduling these threads can eradicate any scalability advantages a fine-grained program may have.

Communication is also a key issue in multi-core architecture. Wires do not scale as well as gates, making communication relatively more expensive compared to computation so optimising communication between cores on chip becomes important.

This dissertation presents an architecture designed to enable scalable fine-grained computation that is communication aware (allowing a programmer to optimise for communication). By combining a tagged memory, where each word is augmented with a presence bit signifying whether or not data is present in that word, with a hardware based scheduler, which allows a thread to wait upon a word becoming present with low overhead. A flexible and scalable architecture well suited to fine-grained computation can be created, one which enables this without needing the introduction of many new architectural features or instructions. Communication is made explicit by enforcing that accesses to a given area of memory will always go to the same cache, removing the need for a cache coherency protocol.

The dissertation begins by reviewing the need for multi-core architecture and discusses the major issues faced in their construction. It moves on to look at fine-grained computation in particular. The proposed architecture, known as Mamba, is then presented in detail with several software techniques suitable for use with it introduced. An FPGA implementation of Mamba is then evaluated against a similar architecture that lacks the extensions Mamba has for assisting in fine-grained computation (namely a memory tagged with presence bits and a hardware scheduler). Microbenchmarks examining the performance of FIFO based communication, MCS locks (an efficient spin-lock implementation based around queues) and barriers demonstrate Mamba's scalability and insensitivity to thread count. A SAT solver implementation demonstrates that these benefits have a real impact on an actual application.

UCAM-CL-TR-833

Alan F. Blackwell, Ignatios Charalampidis:

Practice-led design and evaluation of a live visual constraint language

May 2013, 16 pages, PDF

Abstract: We report an experimental evaluation of Palimpsest, a novel purely-visual programming language. A working prototype of Palimpsest had been developed following a practice-led process, in order to assess whether tools for use in the visual arts can usefully be created by adopting development processes that emulate arts practice. This initial prototype was received more positively by users who have high self-efficacy in both visual arts and computer use. A number of potential usability improvements are identified, structured according to the Cognitive Dimensions of Notations framework.

UCAM-CL-TR-834

John Wickerson:

Concurrent verification for sequential programs

May 2013, 149 pages, PDF
PhD thesis (Churchill College, December 2012)

Abstract: This dissertation makes two contributions to the field of software verification. The first explains how verification techniques originally developed for concurrency can be usefully applied to sequential programs. The second describes how sequential programs can be verified using diagrams that have a parallel nature.

The first contribution involves a new treatment of stability in verification methods based on rely-guarantee. When an assertion made in one thread of a concurrent system cannot be invalidated by the actions of other threads, that assertion is said to be 'stable'. Stability is normally enforced through side-conditions on rely-guarantee proof rules. This dissertation instead proposes to encode stability information into the syntactic form of the assertion. This approach, which we call explicit stabilisation, brings several benefits. First, we empower rely-guarantee with the ability to reason about library code for the first time. Second, when the rely-guarantee method is redeployed in a sequential setting, explicit stabilisation allows more details of a module's implementation to be hidden when verifying clients. Third, explicit stabilisation brings a more nuanced understanding of the important issue of stability in concurrent and sequential verification; such an understanding grows ever more important as verification techniques grow ever more complex.

The second contribution is a new method of presenting program proofs conducted in separation logic. Building on work by Jules Bean, the ribbon proof is a diagrammatic alternative to the standard 'proof outline'. By emphasising the structure of a proof, ribbon proofs are intelligible and hence pedagogically useful. Because they contain less redundancy than proof outlines, and allow each proof step to be checked locally, they are highly scalable; this we illustrate with a ribbon proof of the Version 7 Unix memory manager. Where proof outlines are cumbersome to modify, ribbon proofs can be visually manoeuvred to yield proofs of variant programs. We describe the ribbon proof system, prove its soundness and completeness, and outline a prototype tool for mechanically checking the diagrams it produces.

UCAM-CL-TR-835

Maximilian C. Bolingbroke:

Call-by-need supercompilation

May 2013, 230 pages, PDF
PhD thesis (Robinson College, April 2013)

Abstract: This thesis shows how supercompilation, a powerful technique for transformation and analysis of functional programs, can be effectively applied to a call-by-need language. Our setting will be core calculi suitable for use as intermediate languages when compiling higher-order, lazy functional programming languages such as Haskell.

We describe a new formulation of supercompilation which is more closely connected to operational semantics than the standard presentation. As a result of this connection, we are able to exploit a standard Sestoft-style operational semantics to build a supercompiler which, for the first time, is able to supercompile a call-by-need language with unrestricted recursive let bindings.

We give complete descriptions of all of the (surprisingly tricky) components of the resulting supercompiler, showing in detail how standard formulations of supercompilation have to be adapted for the call-by-need setting.

We show how the standard technique of generalisation can be extended to the call-by-need setting. We also describe a novel generalisation scheme which is simpler to implement than standard generalisation techniques, and describe a completely new form of generalisation which can be used when supercompiling a typed language to ameliorate the phenomenon of supercompilers overspecialising functions on their type arguments.

We also demonstrate a number of non-generalisation-based techniques that can be used to improve the quality of the code generated by the supercompiler. Firstly, we show how let-speculation can be used to ameliorate the effects of the work-duplication checks that are inherent to call-by-need supercompilation. Secondly, we demonstrate how the standard idea of 'rollback' in supercompilation can be adapted to our presentation of the supercompilation algorithm.

We have implemented our supercompiler as an optimisation pass in the Glasgow Haskell Compiler. We perform a comprehensive evaluation of our implementation on a suite of standard call-by-need benchmarks. We improve the runtime of the benchmarks in our suite by a geometric mean of 42%, and reduce the amount of memory which the benchmarks allocate by a geometric mean of 34%.

UCAM-CL-TR-836

Janina Voigt, Alan Mycroft:

Aliasing contracts: a dynamic approach to alias protection

June 2013, 27 pages, PDF

Abstract: Object-oriented programming languages allow multiple variables to refer to the same object, a situation known as aliasing. Aliasing is a powerful tool which enables sharing of objects across a system. However, it can cause serious encapsulation breaches if not controlled properly; through aliasing, internal parts of aggregate objects can be exposed and potentially modified by any part of the system.

A number of schemes for controlling aliasing have been proposed, including Clarke et al.'s ownership types and Boyland et al.'s capabilities. However, many existing systems lack flexibility and expressiveness, making it difficult in practice to program common idioms or patterns which rely on sharing, such as iterators.

We introduce aliasing contracts, a dynamic alias protection scheme which is highly flexible and expressive. Aliasing contracts allow developers to express assumptions about which parts of a system can access particular objects. Aliasing contracts attempt to be a universal approach to alias protection; they can be used to encode various existing schemes.

UCAM-CL-TR-837

Hamed Haddadi, Richard Mortier, Derek McAuley, Jon Crowcroft:

Human-data interaction

June 2013, 9 pages, PDF

Abstract: The time has come to recognise the emerging topic of Human-Data Interaction (HDI). It arises from the need, both ethical and practical, to engage users to a much greater degree with the collection, analysis, and trade of their personal data, in addition to providing them with an intuitive feedback mechanism. HDI is inherently inter-disciplinary, encapsulating elements not only of traditional computer science ranging across data processing, systems design, visualisation and interaction design, but also of law, psychology, behavioural economics, and sociology. In this short paper we elaborate the motivation for studying the nature and dynamics of HDI, and we give some thought to challenges and opportunities in developing approaches to this novel discipline.

UCAM-CL-TR-838

Silvia Breu:

Mining and tracking in evolving software

June 2013, 104 pages, PDF
PhD thesis (Newnham College, April 2011)

Abstract: Every large program contains a small fraction of functionality that resists clean encapsulation. For example, code for debugging or locking is hard to keep hidden using object-oriented mechanisms alone. This problem gave rise to aspect-oriented programming: such cross-cutting functionality is factored out into so-called aspects and these are woven back into mainline code during compilation. However, for existing software systems to benefit from AOP, the cross-cutting concerns must be identified first (aspect mining) before the system can be re-factored into an aspect-oriented design.

This thesis on mining and tracking cross-cutting concerns makes three contributions: firstly, it presents aspect mining as both a theoretical idea and a practical and scalable application. By analysing where developers add code to a program, our history-based aspect mining (HAM) identifies and ranks cross-cutting concerns. Its effectiveness and high precision was evaluated using industrial-sized open-source projects such as ECLIPSE.

Secondly, the thesis takes the work on software evolution one step further. Knowledge about a concern's implementation can become invalid as the system evolves. We address this problem by defining structural and textual patterns among the elements identified as relevant to a concern's implementation. The inferred patterns are documented as rules that describe a concern in a formal (intensional) rather than a merely textual (extensional) manner. These rules can then be used to track an evolving concern's implementation in conjunction with the development history.

Finally, we implemented this technique for Java in an Eclipse plug-in called ISIS4J and evaluated it using a number of concerns. For that we again used the development history of an open-source project. The evaluation shows not only the effectiveness of our approach, but also to what extent our approach supports the tracking of a concern's implementation despite, for example, program code extensions or refactorings.

UCAM-CL-TR-839

Colin Kelly:

Automatic extraction of property norm-like data from large text corpora

September 2013, 154 pages, PDF
PhD thesis (Trinity Hall, September 2012)

Abstract: Traditional methods for deriving property-based representations of concepts from text have focused on extracting unspecified relationships (e.g., "car — petrol") or only a sub-set of possible relation types, such as hyponymy/hypernymy (e.g., "car is-a vehicle") or meronymy/metonymy (e.g., "car has wheels").

We propose a number of varied approaches towards the extremely challenging task of automatic, large-scale acquisition of unconstrained, human-like property norms (in the form "concept relation feature", e.g., "elephant has trunk", "scissors used for cutting", "banana is yellow") from large text corpora. We present four distinct extraction systems for our task. In our first two experiments we manually develop syntactic and lexical rules designed to extract property norm-like information from corpus text. We explore the impact of corpus choice, investigate the efficacy of reweighting our output through WordNet-derived semantic clusters, introduce a novel entropy calculation specific to our task, and test the usefulness of other classical word-association metrics.

In our third experiment we employ semi-supervised learning to generalise from our findings thus far, viewing our task as one of relation classification in which we train a support vector machine on a known set of property norms. Our feature extraction performance is encouraging; however the generated relations are restricted to those found in our training set. Therefore in our fourth and final experiment we use an improved version of our semi-supervised system to initially extract only features for concepts. We then use the concepts and extracted features to anchor an unconstrained relation extraction stage, introducing a novel backing-off technique which assigns relations to concept/feature pairs using probabilistic information.

We also develop and implement an array of evaluations for our task. In addition to the previously employed ESSLLI gold standard, we offer five new evaluation techniques: fMRI activation prediction, EEG activation prediction, a conceptual structure statistics evaluation, a human-generated semantic similarity evaluation and a WordNet semantic similarity comparison. We also comprehensively evaluate our three best systems using human annotators.

Throughout our experiments, our various systems' output is promising but our final system is by far the best-performing. When evaluated against the ESSLLI gold standard it achieves a precision of 44.1%, compared to the 23.9% precision of the current state of the art. Furthermore, our final system's Pearson correlation with human- generated semantic similarity measurements is strong at 0.742, and human judges marked 71.4% of its output as correct/plausible.

UCAM-CL-TR-840

Marek Rei:

Minimally supervised dependency-based methods for natural language processing

September 2013, 169 pages, PDF
PhD thesis (Churchill College, December 2012)

Abstract: This work investigates minimally-supervised methods for solving NLP tasks, without requiring explicit annotation or training data. Our motivation is to create systems that require substantially reduced effort from domain and/or NLP experts, compared to annotating a corresponding dataset, and also offer easier domain adaptation and better generalisation properties.

We apply these principles to four separate language processing tasks and analyse their performance compared to supervised alternatives. First, we investigate the task of detecting the scope of speculative language, and develop a system that applies manually-defined rules over dependency graphs. Next, we experiment with distributional similarity measures for detecting and generating hyponyms, and describe a new measure that achieves the highest performance on hyponym generation. We also extend the distributional hypothesis to larger structures and propose the task of detecting entailment relations between dependency graph fragments of various types and sizes. Our system achieves relatively high accuracy by combining distributional and lexical similarity scores. Finally, we describe a self-learning framework for improving the accuracy of an unlexicalised parser, by calculating relation probabilities using its own dependency output. The method requires only a large in-domain text corpus and can therefore be easily applied to different domains and genres.

While fully supervised approaches generally achieve the highest results, our experiments found minimally supervised methods to be remarkably competitive. By moving away from explicit supervision, we aim to better understand the underlying patterns in the data, and to create systems that are not tied to any specific domains, tasks or resources.

UCAM-CL-TR-841

Arjuna Sathiaseelan, Dirk Trossen, Ioannis Komnios, Joerg Ott, Jon Crowcroft:

Information centric delay tolerant networking: an internet architecture for the challenged

September 2013, 11 pages, PDF

Abstract: Enabling universal Internet access is one of the key issues that is currently being addressed globally. However the existing Internet architecture is seriously challenged to ensure universal service provisioning. This technical report puts forth our vision to make the Internet more accessible by architecting a universal communication architectural framework combining two emerging architecture and connectivity approaches: Information Centric Networking (ICN) and

Delay/Disruption Tolerant Networking (DTN). Such an unified architecture will aggressively seek to widen the connectivity options and provide flexible service models beyond what is currently pursued in the field of universal service provisioning.

UCAM-CL-TR-842

Helen Yannakoudakis:

Automated assessment of English-learner writing

October 2013, 151 pages, PDF
PhD thesis (Wolfson College, December 2012)

Abstract: In this thesis, we investigate automated assessment (AA) systems of free text that automatically analyse and score the quality of writing of learners of English as a second (or other) language. Previous research has employed techniques that measure, in addition to writing competence, the semantic relevance of a text written in response to a given prompt. We argue that an approach which does not rely on task-dependent components or data, and directly assesses learner English, can produce results as good as prompt-specific models. Furthermore, it has the advantage that it may not require re-training or tuning for new prompts or assessment tasks. We evaluate the performance of our models against human scores, manually annotated in the Cambridge Learner Corpus, a subset of which we have released in the public domain to facilitate further research on the task.

We address AA as a supervised discriminative machine learning problem, investigate methods for assessing different aspects of writing prose, examine their generalisation to different corpora, and present state-of-the-art models. We focus on scoring general linguistic competence and discourse coherence and cohesion, and report experiments on detailed analysis of appropriate techniques and feature types derived automatically from generic text processing tools, on their relative importance and contribution to performance, and on comparison with different discriminative models, whilst also experimentally motivating novel feature types for the task. Using outlier texts, we examine and address validity issues of AA systems and, more specifically, their robustness to subversion by writers who understand something of their workings. Finally, we present a user interface that visualises and uncovers the 'marking criteria' represented in AA models, that is, textual features identified as highly predictive of a learner's level of attainment. We demonstrate how the tool can support their linguistic interpretation and enhance hypothesis formation about learner grammars, in addition to informing the development of AA systems and further improving their performance.

UCAM-CL-TR-843

Robin Message:

Programming for humans: a new paradigm for domain-specific languages

November 2013, 140 pages, PDF
PhD thesis (Robinson College, March 2013)

Abstract: Programming is a difficult, specialist skill. Despite much research in software engineering, programmers still work like craftsmen or artists, not engineers. As a result, programs cannot easily be modified, joined together or customised. However, unlike a craft product, once a programmer has created their program, it can replicated infinitely and perfectly. This means that programs are often not a good fit for their end-users because this infinite duplication gives their creators an incentive to create very general programs.

My thesis is that we can create better paradigms, languages and data structuring techniques to enable end-users to create their own programs.

The first contribution is a new paradigm for programming languages which explicitly separates control and data flow. For example, in a web application, the control level would handle user clicks and database writes, while the data level would handle form inputs and database reads. The language is strongly typed, with type reconstruction. We believe this paradigm is particularly suited to end-user programming of interactive applications.

The second contribution is an implementation of this paradigm in a specialised visual programming language for novice programmers to develop web applications. We describe our programming environment, which has a novel layout algorithm that maps control and data flow onto separate dimensions. We show that experienced programmers are more productive in this system than the alternatives.

The third contribution is a novel data structuring technique which infers fuzzy types from example data. This inference is theoretically founded on Bayesian statistics. Our inference aids programmers in moving from semi-structured data to typed programs. We discuss how this data structuring technique could be visualised and integrated with our visual programming environment.

UCAM-CL-TR-844

Wei Ming Khoo:

Decompilation as search

November 2013, 119 pages, PDF
PhD thesis (Hughes Hall, August 2013)

Abstract: Decompilation is the process of converting programs in a low-level representation, such as machine code, into high-level programs that are human readable, compilable and semantically equivalent. The current de facto approach to decompilation is largely modelled on compiler theory and only focusses on one or two of these desirable goals at a time.

This thesis makes the case that decompilation is more effectively accomplished through search. It is observed that software development is seldom a clean slate process and much software is available in public repositories. To back this claim, evidence is presented from three categories of software development: corporate software development, open source projects and malware creation. Evidence strongly suggests that code reuse is prevalent in all categories.

Two approaches to search-based decompilation are proposed. The first approach borrow inspiration from information retrieval, and constitutes the first contribution of this thesis. It uses instruction mnemonics, control-flow sub-graphs and data constants, which can be quickly extracted from a disassembly, and relies on the popular text search engine CLucene. The time taken to analyse a function is small enough to be practical and the technique achieves an F2 measure of above 83.0% for two benchmarks.

The second approach and contribution of this thesis is perturbation analysis, which is able to differentiate between algorithms implementing the same functionality, e.g. bubblesort versus quicksort, and between different implementations of the same algorithm, e.g. quicksort from Wikipedia versus quicksort from Rosetta code. Test-based indexing (TBI) uses random testing to characterise the input-output behaviour of a function; perturbation-based indexing (PBI) is TBI with additional input-output behaviour obtained through perturbation analysis. TBI/PBI achieves an F2 measure of 88.4% on five benchmarks involving different compilers and compiler options.

To perform perturbation analysis, function prototyping is needed, the standard way comprising liveness and reaching-definitions analysis. However, it is observed that in practice actual prototypes fall into one of a few possible categories, enabling the type system to be simplified considerably. The third and final contribution is an approach to prototype recovery that follows the principle of conformant execution, in the form of inlined data source tracking, to infer arrays, pointer-to-pointers and recursive data structures.

UCAM-CL-TR-845

Stephen Kell:

Black-box composition of mismatched software components

December 2013, 251 pages, PDF
PhD thesis (Christ's College, December 2010)

Abstract: Software is expensive to develop. Much of that expense can be blamed on difficulties in combining, integrating or re-using separate pieces of software, and in maintaining such compositions. Conventional development tools approach composition in an inherently narrow way. Specifically, they insist on modules that are plug-compatible, meaning that they must fit together down to a very fine level of detail, and that are homogeneous, meaning that they must be written according to the same conventions and (usually) in the same programming language. In summary, modules must have matched interfaces to compose. These inflexibilities, in turn, motivate more software creation and concomitant expense: they make programming approaches based on integration and re-use unduly expensive. This means that reimplementation from scratch is often chosen in preference to adaptation of existing implementations.

This dissertation presents several contributions towards lessening this problem. It centres on the design of a new special-purpose programming language, called Cake. This language is specialised to the task of describing how components having mismatched interfaces (i.e., not plug-compatible, and perhaps not homogeneous) may be adapted so that they compose as required. It is a language significantly more effective at capturing relationships between mismatched interfaces than general-purpose programming languages. Firstly, we outline the language's design, which centres on reconciling interface differences in the form high-level correspondence rules which relate different interfaces. Secondly, since Cake is designed to be a practical tool which can be a convenient and easily-integrated tool under existing development practices, we describe an implementation of Cake in detail and explain how it achieves this integration. Thirdly, we evaluate Cake on real tasks: by applying it to integration tasks which have already been performed under conventional approaches, we draw meaningful comparisons demonstrating a smaller (quantitative) size of required code and lesser (qualitative) complexity of the code that is required. Finally, Cake applies to a wide range of input components; we sketch extensions to Cake which render it capable of composing components that are heterogeneous with respect to a carefully identified set of stylistic concerns which we describe in detail.

UCAM-CL-TR-846

Daniel Bates:

Exploiting tightly-coupled cores

January 2014, 162 pages, PDF
PhD thesis (Robinson College, July 2013)

Abstract: As we move steadily through the multicore era, and the number of processing cores on each chip continues to rise, parallel computation becomes increasingly important. However, parallelising an application is often difficult because of dependencies between different regions of code which require cores

to communicate. Communication is usually slow compared to computation, and so restricts the opportunities for profitable parallelisation. In this work, I explore the opportunities provided when communication between cores has a very low latency and low energy cost. I observe that there are many different ways in which multiple cores can be used to execute a program, allowing more parallelism to be exploited in more situations, and also providing energy savings in some cases. Individual cores can be made very simple and efficient because they do not need to exploit parallelism internally. The communication patterns between cores can be updated frequently to reflect the parallelism available at the time, allowing better utilisation than specialised hardware which is used infrequently.

In this dissertation I introduce Loki: a homogeneous, tiled architecture made up of many simple, tightly-coupled cores. I demonstrate the benefits in both performance and energy consumption which can be achieved with this arrangement and observe that it is also likely to have lower design and validation costs and be easier to optimise. I then determine exactly where the performance bottlenecks of the design are, and where the energy is consumed, and look into some more-advanced optimisations which can make parallelism even more profitable.

UCAM-CL-TR-847

Jatinder Singh, Jean Bacon:

SBUS: a generic policy-enforcing middleware for open pervasive systems

February 2014, 20 pages, PDF

Abstract: Currently, application components tend to be bespoke and closed, running in vertical silos (single applications/systems). To realise the potential of pervasive systems, and emerging distributed systems more generally, it must be possible to use components system-wide, perhaps in ways and for purposes not envisaged by their designers. It follows that while the infrastructure and resources underlying applications still require management, so too do the applications themselves, in terms of how and when they (inter)operate. To achieve such context-dependent, personalised operation we believe that the application logic embodied in components should be separated from the policy that coordinates them, specifying where and how they should be used.

SBUS is an open, decentralised, application-independent policy-enforcing middleware, developed towards this aim. To enable the flexible and complex interactions required by pervasive systems, it supports a wide range of interaction patterns, including event driven operation, request-response, and data (message) streaming, and features a flexible security model. Crucially, SBUS is dynamically reconfigurable, allowing components to be managed from outside application logic, by authorised third-parties. This paves the way for policy-driven systems, where policy can operate across infrastructure and applications to realise both traditional and new functionality.

This report details the SBUS middleware and the role of policy enforcement in enabling pervasive, distributed systems.

UCAM-CL-TR-848

James G. Jardine:

Automatically generating reading lists

February 2014, 164 pages, PDF
PhD thesis (Robinson College, August 2013)

Abstract: This thesis addresses the task of automatically generating reading lists for novices in a scientific field. Reading lists help novices to get up to speed in a new field by providing an expert-directed list of papers to read. Without reading lists, novices must resort to ad-hoc exploratory scientific search, which is an inefficient use of time and poses a danger that they might use biased or incorrect material as the foundation for their early learning.

The contributions of this thesis are fourfold. The first contribution is the ThemedPageRank (TPR) algorithm for automatically generating reading lists. It combines Latent Topic Models with Personalised PageRank and Age Adjustment in a novel way to generate reading lists that are of better quality than those generated by state-of-the-art search engines. TPR is also used in this thesis to reconstruct the bibliography for scientific papers. Although not designed specifically for this task, TPR significantly outperforms a state-of-the-art system purpose-built for the task. The second contribution is a gold-standard collection of reading lists against which TPR is evaluated, and against which future algorithms can be evaluated. The eight reading lists in the gold-standard were produced by experts recruited from two universities in the United Kingdom. The third contribution is the Citation Substitution Coefficient (CSC), an evaluation metric for evaluating the quality of reading lists. CSC is better suited to this task than standard IR metrics such as precision, recall, F-score and mean average precision because it gives partial credit to recommended papers that are close to gold-standard papers in the citation graph. This partial credit results in scores that have more granularity than those of the standard IR metrics, allowing the subtle differences in the performance of recommendation algorithms to be detected. The final contribution is a light-weight algorithm for Automatic Term Recognition (ATR). As will be seen, technical terms play an important role in the TPR algorithm. This light-weight algorithm extracts technical terms from the titles of documents without the need for the complex apparatus required by most state-of-the-art ATR algorithms. It is also capable of extracting very long technical terms, unlike many other ATR algorithms.

Four experiments are presented in this thesis. The first experiment evaluates TPR against state-of-the-art search engines in the task of automatically generating reading lists that are comparable to expert-generated gold-standards. The second experiment compares the performance of TPR against a purpose-built state-of-the-art system in the task of automatically reconstructing the reference lists of scientific papers. The third experiment involves a user study to explore the ability of novices to build their own reading lists using two fundamental components of TPR: automatic technical term recognition and topic modelling. A system exposing only these components is compared against a state-of-the-art scientific search engine. The final experiment is a user study that evaluates the technical terms discovered by the ATR algorithm and the latent topics generated by TPR. The study enlists thousands of users of Qiqqa, research management software independently written by the author of this thesis.

UCAM-CL-TR-849

Marcelo Bagnulo Braun, Jon Crowcroft:

SNA: Sourceless Network Architecture

March 2014, 12 pages, PDF

Abstract: Why are there source addresses in datagrams? What alternative architecture can one conceive to provide all of the current, and some new functionality, currently dependant on a conflicting set of uses for this field. We illustrate how this can be achieved by reinterpreting the 32-bit field in IPv4 headers to help the Internet solve a range of current and future problems.

UCAM-CL-TR-850

Robert N.M. Watson, Peter G. Neumann, Jonathan Woodruff, Jonathan Anderson, David Chisnall, Brooks Davis, Ben Laurie, Simon W. Moore, Steven J. Murdoch, Michael Roe:

Capability Hardware Enhanced RISC Instructions: CHERI Instruction-set architecture

April 2014, 131 pages, PDF

Abstract: This document describes the rapidly maturing design for the Capability Hardware Enhanced RISC Instructions (CHERI) Instruction-Set Architecture (ISA), which is being developed by SRI International and the University of Cambridge. The document is intended to capture our evolving architecture, as it is being refined, tested, and formally analyzed. We have now reached 70% of the time for our research and development cycle.

CHERI is a hybrid capability-system architecture that combines new processor primitives with the commodity 64-bit RISC ISA enabling software to efficiently implement fine-grained memory protection and a hardware-software object-capability security model. These extensions support incrementally adoptable, high-performance, formally based, programmer-friendly underpinnings for fine-grained software decomposition and compartmentalization, motivated by and capable of enforcing the principle of least privilege. The CHERI system architecture purposefully addresses known performance and robustness gaps in commodity ISAs that hinder the adoption of more secure programming models centered around the principle of least privilege. To this end, CHERI blends traditional paged virtual memory with a per-address-space capability model that includes capability registers, capability instructions, and tagged memory that have been added to the 64-bit MIPS ISA via a new capability coprocessor.

CHERI's hybrid approach, inspired by the Capsicum security model, allows incremental adoption of capability-oriented software design: software implementations that are more robust and resilient can be deployed where they are most needed, while leaving less critical software largely unmodified, but nevertheless suitably constrained to be incapable of having adverse effects. For example, we are focusing conversion efforts on low-level TCB components of the system: separation kernels, hypervisors, operating system kernels, language runtimes, and userspace TCBs such as web browsers. Likewise, we see early-use scenarios (such as data compression, image processing, and video processing) that relate to particularly high-risk software libraries, which are concentrations of both complex and historically vulnerability-prone code combined with untrustworthy data sources, while leaving containing applications unchanged.

This report describes the CHERI architecture and design, and provides reference documentation for the CHERI instruction-set architecture (ISA) and potential memory models, along with their requirements. It also documents our current thinking on integration of programming languages and operating systems. Our ongoing research includes two prototype processors employing the CHERI ISA, each implemented as an FPGA soft core specified in the Bluespec hardware description language (HDL), for which we have integrated the application of formal methods to the Bluespec specifications and the hardware-software implementation.

UCAM-CL-TR-851

Robert N.M. Watson, David Chisnall, Brooks Davis, Wojciech Koszek, Simon W. Moore, Steven J. Murdoch, Peter G. Neumann, Jonathan Woodruff:

Capability Hardware Enhanced RISC Instructions: CHERI User's guide

April 2014, 26 pages, PDF

Abstract: The CHERI User's Guide documents the software environment for the Capability Hardware Enhanced RISC Instructions (CHERI) prototype developed by SRI International and the University of Cambridge. The User's Guide is targeted at hardware and software developers working with capability-enhanced software. It describes the CheriBSD operating system, a version of the FreeBSD operating system that has been adapted to support userspace capability systems via the CHERI ISA, and the CHERI Clang/LLVM compiler suite. It also describes the earlier Deimos demonstration microkernel.

UCAM-CL-TR-852

Robert N.M. Watson, Jonathan Woodruff, David Chisnall, Brooks Davis, Wojciech Koszek, A. Theodore Markettos, Simon W. Moore, Steven J. Murdoch, Peter G. Neumann, Robert Norton, Michael Roe:

Bluespec Extensible RISC Implementation: BERI Hardware reference

April 2014, 76 pages, PDF

Abstract: The BERI Hardware Reference documents the Bluespec Extensible RISC Implementation (BERI) developed by SRI International and the University of Cambridge. The reference is targeted at hardware and software developers working with the BERI1 and BERI2 processor prototypes in simulation and synthesized to FPGA targets. We describe how to use the BERI1 and BERI2 processors in simulation, the BERI1 debug unit, the BERI unit-test suite, how to use BERI with Altera FPGAs and Terasic DE4 boards, the 64-bit MIPS and CHERI ISAs implemented by the prototypes, the BERI1 and BERI2 processor implementations themselves, and the BERI Programmable Interrupt Controller (PIC).

UCAM-CL-TR-853

Robert N.M. Watson, David Chisnall, Brooks Davis, Wojciech Koszek, Simon W. Moore, Steven J. Murdoch, Peter G. Neumann, Jonathan Woodruff:

Bluespec Extensible RISC Implementation: BERI Software reference

April 2014, 34 pages, PDF

Abstract: The BERI Software Reference documents how to build and use FreeBSD on the Bluespec Extensible RISC Implementation (BERI) developed by SRI International and the University of Cambridge. The reference is targeted at hardware and software programmers who will work with BERI or BERI-derived systems.

UCAM-CL-TR-854

Dominic Orchard:

Programming contextual computations

May 2014, 223 pages, PDF
PhD thesis (Jesus College, January 2013)

Abstract: Modern computer programs are executed in a variety of different contexts: on servers, handheld devices, graphics cards, and across distributed environments, to name a few. Understanding a program's contextual requirements is therefore vital for its correct execution. This dissertation studies contextual computations, ranging from application-level notions of context to lower-level notions of context prevalent in common programming tasks. It makes contributions in three areas: mathematically structuring contextual computations, analysing contextual program properties, and designing languages to facilitate contextual programming.

Firstly, existing work which mathematically structures contextual computations using comonads (in programming and semantics) is analysed and extended. Comonads are shown to exhibit a shape preservation property which restricts their applicability to a subset of contextual computations. Subsequently, novel generalisations of comonads are developed, including the notion of an indexed comonad, relaxing shape-preservation restrictions.

Secondly, a general class of static analyses called coeffect systems is introduced to describe the propagation of contextual requirements throughout a program. Indexed comonads, with some additional structure, are shown to provide a semantics for languages whose contextual properties are captured by a coeffect analysis.

Finally, language constructs are presented to ease the programming of contextual computations. The benefits of these language features, the mathematical structuring, and coeffect systems are demonstrated by a language for container programming which guarantees optimisations and safety invariants.

UCAM-CL-TR-855

Patrick K.A. Wollner, Isak Herman,
Haikal Pribadi, Leonardo Impett,
Alan F. Blackwell:

Mephistophone

June 2014, 8 pages, PDF

Abstract: The scope of this project is the creation of a controller for composition, performance and interaction with sound. Interactions can be classified to one of three types: (i) end-user triggering, controlling, editing, and manipulation of sounds with varying temporal dimensions; (ii) inclusion of multi-sensor feedback mechanisms including end-user biological monitoring; and (iii) integration of sensed, semi-random, environmental factors as control parameters to the output of the system.

The development of the device has been completed in two stages: (i) conceptual scoping has defined the interaction space for the development of this machine; (ii) prototype development has resulted in the creation of a functioning prototype and culminated in a series of live performances. The final stage presupposes a custom interaction design for each artistic partner, reinforcing the conceptual role of the device as a novel mechanism for personalized, visualizable, tangible interaction with sound.

UCAM-CL-TR-856

Awais Athar:

Sentiment analysis of scientific citations

June 2014, 114 pages, PDF
PhD thesis (Girton College, April 2014)

Abstract: While there has been growing interest in the field of sentiment analysis for different text genres in the past few years, relatively less emphasis has been placed on extraction of opinions from scientific literature, more specifically, citations. Citation sentiment detection is an attractive task as it can help researchers in identifying shortcomings and detecting problems in a particular approach, determining the quality of a paper for ranking in citation indexes by including negative citations in the weighting scheme, and recognising issues that have not been addressed as well as possible gaps in current research approaches.

Current approaches assume that the sentiment present in the citation sentence represents the true sentiment of the author towards the cited paper and do not take further informal mentions of the citations elsewhere in the article into account. There have also been no attempts to evaluate citation sentiment on a large corpus.

This dissertation focuses on the detection of sentiment towards the citations in a scientific article. The detection is performed using the textual information from the article. I address three sub-tasks and present new large corpora for each of the tasks.

Firstly, I explore different feature sets for detection of sentiment in explicit citations. For this task, I present a new annotated corpus of more than 8,700 citation sentences which have been labelled as positive, negative or objective towards the cited paper. Experimenting with different feature sets, I show the best result of micro-F score 0.760 is obtained using n-grams of length and dependency relations.

Secondly, I show that the assumption that sentiment is limited only to the explicit citation is incorrect. I present a citation context corpus where more than 200,000 sentences from 1,034 paper—reference pairs have been annotated for sentiment. These sentences contain 1,741 citations towards 20 cited papers. I show that including the citation context in the analysis increases the subjective sentiment by almost 185%. I propose new features which help in extracting the citation context and examine their effect on sentiment analysis.

Thirdly, I tackle the task of identifying significant citations. I propose features which help discriminate these from citations in passing, and show that they provide statistically significant improvements over a rule-based baseline.

UCAM-CL-TR-857

Heidi Howard:

ARC: Analysis of Raft Consensus

July 2014, 69 pages, PDF
BA dissertation (Pembroke College, May 2014)

Abstract: The Paxos algorithm, despite being synonymous with distributed consensus for a decade, is famously difficult to reason about and implement due to its non-intuitive approach and underspecification. In response, this project implemented and evaluated a framework for constructing fault-tolerant applications, utilising the recently proposed Raft algorithm for distributed consensus. Constructing a simulation framework for our implementation enabled us to evaluate the protocol on everything from understandability and efficiency to correctness and performance in diverse network environments. We propose a range of optimisations to the protocol and released to the community a testbed for developing further optimisations and investigating optimal protocol parameters for real-world deployments.

UCAM-CL-TR-858

Jonathan D. Woodruff:

CHERI: A RISC capability machine for practical memory safety

July 2014, 112 pages, PDF
PhD thesis (Clare Hall, March 2014)

Abstract: This work presents CHERI, a practical extension of the 64-bit MIPS instruction set to support capabilities for fine-grained memory protection.

Traditional paged memory protection has proved inadequate in the face of escalating security threats and proposed solutions include fine-grained protection tables (Mondrian Memory Protection) and hardware fat-pointer protection (Hardbound). These have emphasised transparent protection for C executables but have lacked flexibility and practicality. Intel's recent memory protection extensions (iMPX) attempt to adopt some of these ideas and are flexible and optional but lack the strict correctness of these proposals.

Capability addressing has been the classical solution to efficient and strong memory protection but it has been thought to be incompatible with common instruction sets and also with modern program structure which uses a flat memory space with global pointers.

CHERI is a fusion of capabilities with a paged flat memory producing a program-managed fat pointer capability model. This protection mechanism scales from application sandboxing to efficient byte-level memory safety with per-pointer permissions. I present an extension to the 64-bit MIPS architecture on FPGA that runs standard FreeBSD and supports self-segmenting applications in user space.

Unlike other recent proposals, the CHERI implementation is open-source and of sufficient quality to support software development as well as community extension of this work. I compare with published memory safety mechanisms and demonstrate competitive performance while providing assurance and greater flexibility with simpler hardware requirements.

UCAM-CL-TR-859

Lucian Carata, Oliver Chick, James Snee, Ripduman Sohan, Andrew Rice, Andy Hopper:

Resourceful: fine-grained resource accounting for explaining service variability

September 2014, 12 pages, PDF

Abstract: Increasing server utilization in modern datacenters also increases the likelihood of contention on physical resources and unexpected behavior due to side-effects from interfering applications. Existing resource accounting mechanisms are too coarse-grained for allowing services to track the causes of such variations in their execution. We make the case for measuring resource consumption at system-call level and outline the design of Resourceful, a system that offers applications the ability of querying this data at runtime with low overhead, accounting for costs incurred both synchronously and asynchronously after a given call.

UCAM-CL-TR-860

Steffen Loesch:

Program equivalence in functional metaprogramming via nominal Scott domains

October 2014, 164 pages, PDF
PhD thesis (Trinity College, May 2014)

Abstract: A prominent feature of metaprogramming is to write algorithms in one programming language (the meta-language) over structures that represent the programs of another programming language (the object-language). Whenever the object-language has binding constructs (and most programming languages do), we run into tedious issues concerning the semantically correct manipulation of binders.

In this thesis we study a semantic framework in which these issues can be dealt with automatically by the meta-language. Our framework takes the user-friendly 'nominal' approach to metaprogramming in which bound objects are named.

Specifically, we develop mathematical tools for giving logical proofs that two metaprograms (of our framework) are equivalent. We consider two programs to be equivalent if they always give the same observable results when they are run as part of any larger codebase. This notion of program equivalence, called contextual equivalence, is examined for an extension of Plotkin's archetypal functional programming language PCF with nominal constructs for metaprogramming, called PNA. Historically, PCF and its denotational semantics based on Scott domains were hugely influential in the study of contextual equivalence. We mirror Plotkin's classical results with PNA and a denotational semantics based on a variant of Scott domains that is modelled within the logic of nominal sets. In particular, we prove the following full abstraction result: two PNA programs are contextually equivalent if and only if they denote equal elements of the nominal Scott domain model. This is the first full abstraction result we know of for languages combining higher-order functions with some form of locally scoped names, which uses a domain theory based on ordinary extensional functions, rather

than using the more intensional approach of game semantics.

To obtain full abstraction, we need to add two new programming language constructs to PNA, one for existential quantification over names and one for 'definite description' over names. Adding only one of them is insufficient, as we give proofs that full abstraction fails if either is left out.

UCAM-CL-TR-861

Tadas Baltrusaitis:

Automatic facial expression analysis

October 2014, 218 pages, PDF
PhD thesis (Fitzwilliam College, March 2014)

Abstract: Humans spend a large amount of their time interacting with computers of one type or another. However, computers are emotionally blind and indifferent to the affective states of their users. Human-computer interaction which does not consider emotions, ignores a whole channel of available information.

Faces contain a large portion of our emotionally expressive behaviour. We use facial expressions to display our emotional states and to manage our interactions. Furthermore, we express and read emotions in faces effortlessly. However, automatic understanding of facial expressions is a very difficult task computationally, especially in the presence of highly variable pose, expression and illumination. My work furthers the field of automatic facial expression tracking by tackling these issues, bringing emotionally aware computing closer to reality.

Firstly, I present an in-depth analysis of the Constrained Local Model (CLM) for facial expression and head pose tracking. I propose a number of extensions that make location of facial features more accurate.

Secondly, I introduce a 3D Constrained Local Model (CLM-Z) which takes full advantage of depth information available from various range scanners. CLM-Z is robust to changes in illumination and shows better facial tracking performance.

Thirdly, I present the Constrained Local Neural Field (CLNF), a novel instance of CLM that deals with the issues of facial tracking in complex scenes. It achieves this through the use of a novel landmark detector and a novel CLM fitting algorithm. CLNF outperforms state-of-the-art models for facial tracking in presence of difficult illumination and varying pose.

Lastly, I demonstrate how tracked facial expressions can be used for emotion inference from videos. I also show how the tools developed for facial tracking can be applied to emotion inference in music.

UCAM-CL-TR-862

Henrik Lieng:

Surface modelling for 2D imagery

October 2014, 177 pages, PDF
PhD thesis (Fitzwilliam College, June 2014)

Abstract: Vector graphics provides powerful tools for drawing scalable 2D imagery. With the rise of mobile computers, of different types of displays and image resolutions, vector graphics is receiving an increasing amount of attention. However, vector graphics is not the leading framework for creating and manipulating 2D imagery. The reason for this reluctance of employing vector graphical frameworks is that it is difficult to handle complex behaviour of colour across the 2D domain.

A challenging problem within vector graphics is to define smooth colour functions across the image. In previous work, two approaches exist. The first approach, known as diffusion curves, diffuses colours from a set of input curves and points. The second approach, known as gradient meshes, defines smooth colour functions from control meshes. These two approaches are incompatible: diffusion curves do not support the local behaviour provided by gradient meshes and gradient meshes do not support freeform curves as input. My research aims to narrow the gap between diffusion curves and gradient meshes.

With this aim in mind, I propose solutions to create control meshes from freeform curves. I demonstrate that these control meshes can be used to render a vector primitive similar to diffusion curves using subdivision surfaces. With the use of subdivision surfaces, instead of a diffusion process, colour gradients can be locally controlled using colour gradient curves associated with the input curves.

The advantage of local control is further explored in the setting of vector-centric image processing. I demonstrate that a certain contrast enhancement profile, known as the Cornsweet profile, can be modelled via surfaces in images. This approach does not produce saturation artefacts related with previous filter-based methods. Additionally, I demonstrate various approaches to artistic filtering, where the artist locally models given artistic effects.

Gradient meshes are restricted to rectangular topology of the control meshes. I argue that this restriction hinders the applicability of the approach and its potential to be used with control meshes extracted from freeform curves. To this end, I propose a mesh-based vector primitive that supports arbitrary manifold topology of the control mesh.

UCAM-CL-TR-863

Jatinder Singh, Jean Bacon, Jon Crowcroft, Anil Madhavapeddy, Thomas Pasquier, W. Kuan Hon, Christopher Millard:

Regional clouds: technical considerations

November 2014, 18 pages, PDF

Abstract: The emergence and rapid uptake of cloud computing services raise a number of legal challenges. Recently, there have been calls for regional clouds; where policy makers from various states have proposed cloud computing services that are restricted to serving (only) their particular geographic region. At a technical level, such rhetoric is rooted in the means for control.

This paper explores the technical considerations underpinning a regional cloud, including the current state of cloud provisioning, what can be achieved using existing technologies, and the potential of ongoing research. Our discussion covers technology at various system levels, including network-centric controls, cloud platform management, and governance mechanisms (including encryption and information flow control) for cloud providers, applications, tenants, and end-users.

UCAM-CL-TR-864

Robert N. M. Watson, Peter G. Neumann, Jonathan Woodruff, Jonathan Anderson, David Chisnall, Brooks Davis, Ben Laurie, Simon W. Moore, Steven J. Murdoch, Michael Roe:

Capability Hardware Enhanced RISC Instructions: CHERI Instruction-set architecture

December 2014, 142 pages, PDF

Abstract: This technical report describes CHERI ISAv3, the third version of the Capability Hardware Enhanced RISC Instructions (CHERI) Instruction-Set Architecture (ISA). CHERI is being developed by SRI International and the University of Cambridge. This design captures four years of research, development, refinement, formal analysis, and testing, and is a substantial enhancement to the ISA version described in UCAM-CL-TR-850. Key improvements lie in tighter C-language integration, and more mature support for software object-capability models; these changes result from experience gained in adapting substantial software stacks to run on prototype hardware.

The CHERI instruction set is based on a hybrid capability-system architecture that adds new capability-system primitives to a commodity 64-bit RISC ISA enabling software to efficiently implement fine-grained memory protection and a hardware-software object-capability security model. These extensions support incrementally adoptable, high-performance, formally based, programmer-friendly underpinnings for fine-grained software decomposition and compartmentalization, motivated by and capable of enforcing the principle of least privilege.

The CHERI system architecture purposefully addresses known performance and robustness gaps in commodity ISAs that hinder the adoption of more secure programming models centered around the principle of least privilege. To this end, CHERI blends traditional paged virtual memory with a per-address-space capability model that includes capability registers, capability instructions, and tagged memory that have been added to the 64-bit MIPS ISA via a new capability coprocessor. CHERI also learns from the C-language fat-pointer literature: CHERI capabilities can describe not only regions of memory, but can also capture C pointer semantics allowing capabilities to be substituted for pointers in generated code.

CHERI's hybrid system approach, inspired by the Capsicum security model, allows incremental adoption of capability-oriented software design: software implementations that are more robust and resilient can be deployed where they are most needed, while leaving less critical software largely unmodified, but nevertheless suitably constrained to be incapable of having adverse effects. For example, we are focusing conversion efforts on low-level TCB components of the system: separation kernels, hypervisors, operating system kernels, language runtimes, and userspace TCBs such as web browsers. Likewise, we see early-use scenarios (such as data compression, protocol parsing, image processing, and video processing) that relate to particularly high-risk software libraries, which are concentrations of both complex and historically vulnerability-prone code combined with untrustworthy data sources, while leaving containing applications unchanged.

This report describes the CHERI Instruction-Set Architecture (ISA) and design, and provides reference documentation and potential memory models, along with their requirements. It also briefly addresses the CHERI system hardware-software architecture, documenting our current thinking on integrating programming languages and operating systems with the CHERI hardware.

UCAM-CL-TR-865

Christopher S.F. Smowton:

I/O Optimisation and elimination via partial evaluation

December 2014, 129 pages, PDF
PhD thesis (Churchill College, November 2014)

Abstract: Computer programs commonly repeat work. Short programs go through the same initialisation sequence each time they are run, and long-running servers may be given a sequence of similar or identical requests. In both cases, there is an opportunity to save time by reusing previously computed results; however, programmers often do not exploit that opportunity because to do so would cost development time and increase code complexity.

Partial evaluation is a semi-automatic technique for specialising programs or parts thereof to perform better in particular circumstances, and can reduce repeated work by generating a program variant that is specialised for use in frequently-occurring circumstances. However, existing partial evaluators are limited in both their depth of analysis and their support for real-world programs, making them ineffective at specialising practical software.

In this dissertation, I present a new, more accurate partial evaluation system that can specialise programs, written in low-level languages including C and C++, that interact with the operating system to read external data. It is capable of specialising programs that are challenging to analyse, including those which use arbitrarily deep pointer indirection, unsafe type casts, and multi-threading. I use this partial evaluator to specialise programs with respect to files that they read from disk, or data they consume from the network, producing specialised variants that perform better when that external data is as expected, but which continue to function like the original program when it is not. To demonstrate the system's practical utility, I evaluate the system specialising real-world software, and show that it can achieve significant runtime improvements with little manual assistance.

UCAM-CL-TR-866

Ilias Giechaskiel, George Panagopoulos, Eiko Yoneki:

PDTL: Parallel and distributed triangle listing for massive graphs

April 2015, 14 pages, PDF

Abstract: This paper presents the first distributed triangle listing algorithm with provable CPU, I/O, Memory, and Network bounds. Finding all triangles (3-cliques) in a graph has numerous applications for density and connectivity metrics. The majority of existing algorithms for massive graphs are sequential processing and distributed versions of algorithms do not guarantee their CPU, I/O, Memory or Network requirements. Our Parallel and Distributed Triangle Listing (PDTL) framework focuses on efficient external-memory access in distributed environments instead of fitting subgraphs into memory. It works by performing efficient orientation and load-balancing steps, and replicating graphs across machines by using an extended version of Hu et al.'s Massive Graph Triangulation algorithm. As a result, PDTL suits a variety of computational environments, from single-core machines to high-end clusters. PDTL computes the exact triangle count on graphs of over 6 billion edges and 1 billion vertices (e.g. Yahoo graphs), outperforming and using fewer resources than the state-of-the-art systems PowerGraph, OPT, and PATRIC by 2 times to 4 times. Our approach highlights the importance of I/O considerations in a distributed environment, which has received less attention in the graph processing literature.

UCAM-CL-TR-867

Nikolai Sultana:

Higher-order proof translation

April 2015, 158 pages, PDF
PhD thesis (Trinity College, April 2014)

Abstract: The case for interfacing logic tools together has been made countless times in the literature, but it is still an important research question. There are various logics and respective tools for carrying out formal developments, but practitioners still lament the difficulty of reliably exchanging mathematical data between tools.

Writing proof-translation tools is hard. The problem has both a theoretical side (to ensure that the translation is adequate) and a practical side (to ensure that the translation is feasible and usable). Moreover, the source and target proof formats might be less documented than desired (or even necessary), and this adds a dash of reverse-engineering to what should be a system integration task.

This dissertation studies proof translation for higher-order logic. We will look at the qualitative benefits of locating the translation close to the source (where the proof is generated), the target (where the proof is consumed), and in between (as an independent tool from the proof producer and consumer).

Two ideas are proposed to alleviate the difficulty of building proof translation tools. The first is a proof translation framework that is structured as a compiler. Its target is specified as an abstract machine, which captures the essential features of its implementations. This framework is designed to be performant and extensible. Second, we study proof transformations that convert refutation proofs from a broad class of consistency-preserving calculi (such as those used by many proof-finding tools) into proofs in validity-preserving calculi (the kind used by many proof-checking tools). The basic method is very simple, and involves applying a single transformation uniformly to all of the source calculi's inferences, rather than applying ad hoc (rule specific) inference interpretations.

UCAM-CL-TR-868

Robert N. M. Watson, Jonathan Woodruff, David Chisnall, Brooks Davis, Wojciech Koszek, A. Theodore Markettos, Simon W. Moore, Steven J. Murdoch, Peter G. Neumann, Robert Norton, Michael Roe:

Bluespec Extensible RISC Implementation: BERI Hardware reference

April 2015, 82 pages, PDF

Abstract: The BERI Hardware Reference describes the Bluespec Extensible RISC Implementation (BERI) prototype developed by SRI International and the University of Cambridge. The reference is targeted at hardware and software developers working with the BERI1 and BERI2 processor prototypes in simulation and synthesized to FPGA targets. We describe how to use the BERI1 and BERI2 processors in simulation, the BERI1 debug unit, the BERI unit-test suite; how to use BERI with Altera FPGAs and Terasic DE4 boards, the 64-bit MIPS and CHERI ISAs implemented by the prototypes, the BERI1 and BERI2 processor implementations themselves, and the BERI Programmable Interrupt Controller (PIC).

UCAM-CL-TR-869

Robert N. M. Watson, David Chisnall, Brooks Davis, Wojciech Koszek, Simon W. Moore, Steven J. Murdoch, Peter G. Neumann, Jonathan Woodruff:

Bluespec Extensible RISC Implementation: BERI Software reference

April 2015, 27 pages, PDF

Abstract: The BERI Software Reference documents how to build and use the FreeBSD operating system on the Bluespec Extensible RISC Implementation (BERI) developed by SRI International and the University of Cambridge. The reference is targeted at hardware and software programmers who will work with BERI or BERI-derived systems.

UCAM-CL-TR-870

Ali Mustafa Zaidi:

Accelerating control-flow intensive code in spatial hardware

May 2015, 170 pages, PDF
PhD thesis (St. Edmund's College, February 2014)

Abstract: Designers are increasingly utilizing spatial (e.g. custom and reconfigurable) architectures to improve both efficiency and performance in increasingly heterogeneous systems-on-chip. Unfortunately, while such architectures can provide orders of magnitude better efficiency and performance on numeric applications, they exhibit poor performance when implementing sequential, control-flow intensive code. This thesis studies the problem of improving sequential code performance in spatial hardware without sacrificing its inherent efficiency advantage.

I propose (a) switching from a statically scheduled to a dynamically scheduled, dataflow execution model, and (b) utilizing a newly developed compiler intermediate representation (IR) designed to expose ILP in spatial hardware, even in the presence of complex control flow. I describe this new IR – the Value State Flow Graph (VSFG) – and how it statically exposes ILP from control-flow intensive code by enabling control-dependence analysis, execution along multiple flows of control, as well as aggressive control-flow speculation. I also present a High-Level Synthesis (HLS) toolchain, that compiles unmodified high-level language code to dataflow custom hardware, via the LLVM compiler infrastructure.

I show that for control-flow intensive code, VSFG-based custom hardware performance approaches, or even exceeds the performance of a complex superscalar processor, while consuming only 1/4x the energy of an efficient in-order processor, and 1/8x that of a complex out-of-order processor. I also present a discussion of compile-time optimizations that may be attempted to further improve both efficiency and performance for VSFG-based hardware, including using alias analysis to statically partition and parallelize memory operations.

This work demonstrates that it is possible to use custom and/or reconfigurable hardware in heterogeneous systems to improve the efficiency of frequently executed sequential code, without compromising performance relative to an energy inefficient out-of-order superscalar processor.

UCAM-CL-TR-871

Peter R. Calvert:

Architecture-neutral parallelism via the Join Calculus

July 2015, 150 pages, PDF
PhD thesis (Trinity College, September 2014)

Abstract: Ever since the UNCOL efforts in the 1960s, compilers have sought to use both source-language-neutral and architecture-neutral intermediate representations. The advent of web applets led to the JVM where such a representation was also used for distribution. This trend has continued and now many mainstream applications are distributed using the JVM or .NET formats. These languages can be efficiently run on a target architecture (e.g. using JIT techniques). However, such intermediate languages have been predominantly sequential, supporting only rudimentary concurrency primitives such as threads. This thesis proposes a parallel intermediate representation with analogous goals. The specific contributions made in this

work are based around a join calculus abstract machine (JCAM). These can be broadly categorised into three sections.

The first contribution is the definition of the abstract machine itself. The standard join calculus is modified to prevent implicit sharing of data as this is undesirable in non-shared memory architectures. It is then condensed into three primitive operations that can be targeted by optimisations and analyses. I claim that these three simple operations capture all the common styles of concurrency and parallelism used in current programming languages.

The work goes on to show how the JCAM intermediate representation can be implemented on shared-memory multi-core machines with acceptable overheads. This process illustrates some key program properties that can be exploited to give significant benefits in certain scenarios.

Finally, conventional control-flow analyses are adapted to the join calculus to allow the properties required for optimising compilation to be inferred. Along with the prototype compiler, this illustrates the JCAM's capabilities as a universal representation for parallelism.

UCAM-CL-TR-872

Jia Meng:

The integration of higher order interactive proof with first order automatic theorem proving

July 2015, 144 pages, PDF
PhD thesis (Churchill College, April 2005)

Abstract: Interactive and automatic theorem proving are the two most widely used computer-assisted theorem proving methods. Interactive proof tools such as HOL, Isabelle and PVS have been highly successful. They support expressive formalisms and have been used for verifying hardware, software, protocols, and so forth. Unfortunately interactive proof requires much effort from a skilled user. Many other tools are completely automatic, such as Vampire, SPASS and Otter. However, they cannot be used to verify large systems because their logic is inexpressive. This dissertation focuses on how to combine these two types of theorem proving to obtain the advantages of each of them. This research is carried out by investigating the integration of Isabelle with Vampire and SPASS.

Isabelle is an interactive theorem prover and it supports a multiplicity of logics, such as ZF and HOL. Vampire and SPASS are first order untyped resolution provers. The objective of this research is to design an effective method to support higher order interactive proof with any first order resolution prover. This integration can simplify the formal verification procedure by reducing the user interaction required during interactive proofs: many goals will be proved by automatic provers.

For such an integration to be effective, we must bridge the many differences between a typical interactive theorem prover and a resolution theorem prover. Examples of the differences are higher order versus first order; typed versus untyped.

Through experiments, we have designed and implemented a practical method to convert Isabelle's formalisms (ZF and HOL) into untyped first-order clauses. Isabelle/ZF's formulae that are not first-order need to be reformulated to first-order formulae before clause normal form transformation. For Isabelle/HOL, a sound modelling of its type system is designed first before translating its formulae into first-order clauses with its type information encoded. This method of formalization makes it possible to have Isabelle integrated with resolutions.

A large set of axioms is usually required to support interactive proofs but can easily overwhelm an automatic prover. We have experimented with various methods to solve this problem, including using different settings of an automatic prover and automatically eliminating irrelevant axioms.

The invocation of background automatic provers should be invisible to users, hence we have also designed and implemented an automatic calling procedure, which extracts all necessary information and sends it to an automatic prover at an appropriate point during an Isabelle proof.

Finally, the results and knowledge gained from this research are generally applicable and can be applied to future integration of any other interactive and automatic theorem provers.

UCAM-CL-TR-873

Khilan Gudka, Robert N.M. Watson, Jonathan Anderson, David Chisnall, Brooks Davis, Ben Laurie, Ilias Marinos, Peter G. Neumann, Alex Richardson:

Clean application compartmentalization with SOAAP (extended version)

August 2015, 35 pages, PDF

Abstract: Application compartmentalization, a vulnerability mitigation technique employed in programs such as OpenSSH and the Chrome web browser, decomposes software into sandboxed components to limit privileges leaked or otherwise available to attackers. However, compartmentalizing applications – and maintaining that compartmentalization – is hindered by ad hoc methodologies and significantly increased programming effort. In practice, programmers stumble

through (rather than overtly reason about) compartmentalization spaces of possible decompositions, unknowingly trading off correctness, security, complexity, and performance.

We present a new conceptual framework embodied in an LLVM-based tool: the Security-Oriented Analysis of Application Programs (SOAAP) that allows programmers to reason about compartmentalization using source-code annotations (compartmentalization hypotheses). We demonstrate considerable benefit when creating new compartmentalizations for complex applications, and analyze existing compartmentalized applications to discover design faults and maintenance issues arising from application evolution.

This technical report is an extended version of the similarly named conference paper presented at the 2015 ACM Conference on Computer and Communications Security (CCS).

UCAM-CL-TR-874

Zhen Bai:

Augmented Reality interfaces for symbolic play in early childhood

September 2015, 292 pages, PDF
PhD thesis (Jesus College, September 2014)

Abstract: Augmented Reality (AR) is an umbrella term for technologies that superimpose virtual contents onto the physical world. There is an emerging research focus on AR applications to improve quality of life for special user groups with diverse levels of age, skill, disabilities and knowledge.

Symbolic play is an early childhood activity that requires children to interpret elements of the real world in a non-literal way. Much research effort has been spent to enhance symbolic play due to its close link with critical development such as symbolic thought, creativity and social understanding.

In this thesis, I identified an analogy between the dual representational characteristics of AR and symbolic play. This led me to explore to what extent AR can promote cognitive and social development in symbolic play for young children with and without autism spectrum condition (ASC). To address this research goal, I developed a progressive AR design approach that requires progressive levels of mental effort to conceive symbolic thought during play. I investigated the usability of AR displays with the magic mirror metaphor to support physical object manipulation. Based on the progressive AR design approach and preparatory usability investigation, I designed an AR system to enhance solitary symbolic play, and another AR system to enhance social symbolic play. The effectiveness of each system was rigorously evaluated with reference to psychology literature.

Empirical results show that children with ASC 4-7 years old produced more solitary symbolic play with higher theme relevance using the first AR system as compared with an equivalent non-AR natural play setting. Typically developing children aged 4-6, using the second AR system, demonstrated improved social symbolic play in terms of comprehending emotional states of pretend roles, and constructing joint pretense on symbolic transformations using specially designed AR scaffoldings.

UCAM-CL-TR-875

Theodosia Togia:

The language of collaborative tagging

September 2015, 203 pages, PDF
PhD thesis (Trinity Hall, September 2015)

Abstract: Collaborative tagging is the process whereby people attach keywords, known as tags, to digital resources, such as text and images, in order to render them retrievable in the future. This thesis investigates how tags submitted by users in collaborative tagging systems function as descriptors of a resource's perceived content. Using computational and theoretical tools, I compare collaborative tagging with natural language description in order to determine whether or to what extent the former behaves as the latter.

I start the investigation by collecting a corpus of tagged images and exploring the relationship between a resource and a tag using theories from different disciplines, such as Library Science, Semiotics and Information Retrieval. Then, I study the lexical characteristics of individual tags, suggesting that tags behave as natural language words. The next step is to examine how tags combine when annotating a resource. It will be shown that similar combinatory constraints hold for tags assigned to a resource and for words as used in coherent text. This realisation will lead to the question of whether the similar combinatory patterns between tags and words are due to implicit semantic relations between the tags. To provide an answer, I conduct an experiment asking humans to submit both tags and textual descriptions for a set of images, constructing a parallel a corpus of more than one thousand tags-text annotations. Analysis of this parallel corpus provides evidence that a large number of tag pairs are connected via implicit semantic relations, whose nature is described. Finally, I investigate whether it is possible to automatically identify the semantically related tag pairs and make explicit their relationship, even in the absence of supporting image-specific text. I construct and evaluate a proof-of-concept system to demonstrate that this task is attainable.

UCAM-CL-TR-876

Robert N. M. Watson, Peter G. Neumann, Jonathan Woodruff, Michael Roe, Jonathan Anderson, David Chisnall,

Brooks Davis, Alexandre Joannou,
Ben Laurie, Simon W. Moore,
Steven J. Murdoch, Robert Norton,
Stacey Son:

Capability Hardware Enhanced RISC Instructions: CHERI Instruction-Set Architecture

September 2015, 198 pages, PDF

Abstract: This technical report describes CHERI ISAv4, the fourth version of the Capability Hardware Enhanced RISC Instructions (CHERI) Instruction-Set Architecture (ISA). CHERI is being developed by SRI International and the University of Cambridge. This design captures four years of research, development, refinement, formal analysis, and testing, and is a substantial enhancement to the ISA version described in UCAM-CL-TR-850. Key improvements lie in tighter C-language integration, and more mature support for software object-capability models; these changes result from experience gained in adapting substantial software stacks to run on prototype hardware.

The CHERI instruction set is based on a hybrid capability-system architecture that adds new capability-system primitives to a commodity 64-bit RISC ISA enabling software to efficiently implement fine-grained memory protection and a hardware-software object-capability security model. These extensions support incrementally adoptable, high-performance, formally based, programmer-friendly underpinnings for fine-grained software decomposition and compartmentalization, motivated by and capable of enforcing the principle of least privilege.

The CHERI system architecture purposefully addresses known performance and robustness gaps in commodity ISAs that hinder the adoption of more secure programming models centered around the principle of least privilege. To this end, CHERI blends traditional paged virtual memory with a per-address-space capability model that includes capability registers, capability instructions, and tagged memory that have been added to the 64-bit MIPS ISA via a new capability coprocessor. CHERI also learns from the C-language fat-pointer literature: CHERI capabilities can describe not only regions of memory, but can also capture C pointer semantics allowing capabilities to be substituted for pointers in generated code.

CHERI's hybrid system approach, inspired by the Capsicum security model, allows incremental adoption of capability-oriented software design: software implementations that are more robust and resilient can be deployed where they are most needed, while leaving less critical software largely unmodified, but nevertheless suitably constrained to be incapable of having adverse effects. For example, we are focusing conversion efforts on low-level TCB components of the system: separation kernels, hypervisors, operating system kernels, language runtimes, and userspace TCBs such as web browsers. Likewise, we see early-use scenarios (such as data compression, protocol parsing, image processing, and video processing) that relate to particularly high-risk software libraries, which are concentrations of both complex and historically vulnerability-prone code combined with untrustworthy data sources, while leaving containing applications unchanged.

This report describes the CHERI Instruction-Set Architecture (ISA) and design, and provides reference documentation and potential memory models, along with their requirements. It also briefly addresses the CHERI system hardware-software architecture, documenting our current thinking on integrating programming languages and operating systems with the CHERI hardware.

UCAM-CL-TR-877

Robert N. M. Watson, David Chisnall,
Brooks Davis, Wojciech Koszek,
Simon W. Moore, Steven J. Murdoch,
Peter G. Neumann, Jonathan Woodruff:

Capability Hardware Enhanced RISC Instructions: CHERI Programmer's Guide

September 2015, 58 pages, PDF

Abstract: This work presents CHERI, a practical extension of the 64-bit MIPS instruction set to support capabilities for fine-grained memory protection.

Traditional paged memory protection has proved inadequate in the face of escalating security threats and proposed solutions include fine-grained protection tables (Mondrian Memory Protection) and hardware fat-pointer protection (Hardbound). These have emphasised transparent protection for C executables but have lacked flexibility and practicality. Intel's recent memory protection extensions (iMPX) attempt to adopt some of these ideas and are flexible and optional but lack the strict correctness of these proposals.

Capability addressing has been the classical solution to efficient and strong memory protection but it has been thought to be incompatible with common instruction sets and also with modern program structure which uses a flat memory space with global pointers.

CHERI is a fusion of capabilities with a paged flat memory producing a program-managed fat pointer capability model. This protection mechanism scales from application sandboxing to efficient byte-level memory safety with per-pointer permissions. I present an extension to the 64-bit MIPS architecture on FPGA that runs standard FreeBSD and supports self-segmenting applications in user space.

Unlike other recent proposals, the CHERI implementation is open-source and of sufficient quality to support software development as well as community

extension of this work. I compare with published memory safety mechanisms and demonstrate competitive performance while providing assurance and greater flexibility with simpler hardware requirements.

UCAM-CL-TR-878

Marios O. Choudary:

Efficient multivariate statistical techniques for extracting secrets from electronic devices

September 2015, 164 pages, PDF
PhD thesis (Darwin College, July 2014)

Abstract: In 2002, Suresh Chari, Rao Josyula and Pankaj Rohatgi presented a very powerful method, known as the 'Template Attack', to infer secret values processed by a microcontroller, by analysing its power-supply current, generally known as its 'side-channel leakage'. This attack uses a profiling step to compute the parameters of a multivariate normal distribution from the leakage of a training device, and an attack step in which these parameters are used to infer a secret value (e.g. cryptographic key) from the leakage of a target device. This has important implications for many industries, such as pay-TV or banking, that use a microcontroller executing a cryptographic algorithm to authenticate their customers.

In this thesis, I describe efficient implementations of this template attack, that can push its limits further, by using efficient multivariate statistical analysis techniques. Firstly, I show that, using a linear discriminant score, we can avoid some numerical obstacles, and use a large number of leakage samples to improve the attack, while also drastically decreasing its computation time. I evaluate my implementations on an 8-bit microcontroller, using different compression methods, including Principal Component Analysis (PCA) and Fisher's Linear Discriminant Analysis (LDA), and I provide guidance for the choice of attack algorithm. My results show that we can determine almost perfectly an 8-bit target value, even when this value is manipulated by a single LOAD instruction.

Secondly, I show that variability caused by the use of either different devices or different acquisition campaigns can have a strong impact on the performance of these attacks. Using four different Atmel XMEGA 256 A3U 8-bit devices, I explore several variants of the template attack to compensate for this variability, and I show that, by adapting PCA and LDA to this context, we can reduce the entropy of an unknown 8-bit value to below 1.5 bits, even when using one device for profiling and another one for the attack.

Then, using factor analysis, I identify the main factors that contribute to the correlation between leakage samples, and analyse the influence of this correlation on template attacks. I show that, in some cases, by estimating the covariance matrix only from these main factors, we can improve the template attack. Furthermore, I show how to use factor analysis in order to generate arbitrary correlation matrices for the simulation of leakage traces that are similar to the real leakage.

Finally, I show how to implement PCA and LDA efficiently with the stochastic model presented by Schindler et al. in 2005, resulting in the most effective kind of profiled attack. Using these implementations, I demonstrate a profiled attack on a 16-bit target.

UCAM-CL-TR-879

Ramana Kumar:

Self-compilation and self-verification

February 2016, 148 pages, PDF
PhD thesis (Peterhouse College, May 2015)

Abstract: This dissertation presents two pieces of work, one building on the other, that advance the state of the art of formal verification. The focus, in both cases, is on proving end-to-end correctness for realistic implementations of computer software. The first piece is a verified compiler for a stateful higher-order functional programming language, CakeML, which is packaged into a verified read-eval-print loop (REPL). The second piece is a verified theorem-prover kernel for higher-order logic (HOL), designed to run in the verified REPL.

Self-compilation is the key idea behind the verification of the CakeML REPL, in particular, the new technique of proof-grounded bootstrapping of a verified compiler. The verified compiler is bootstrapped within the theorem prover used for its verification, and then packaged into a REPL. The result is an implementation of the REPL in machine code, verified against machine-code semantics. All the end-to-end correctness theorems linking this low-level implementation to its desired semantics are proved within the logic of the theorem prover. Therefore the trusted computing base (TCB) of the final implementation is smaller than for any previously verified compiler.

Just as self-compilation is a benchmark by which to judge a compiler, I propose self-verification as a benchmark for theorem provers, and present a method by which a theorem prover could verify its own implementation. By applying proof-grounded compilation (i.e., proof-grounded bootstrapping applied to something other than a compiler) to an implementation of a theorem prover, we obtain a theorem prover whose machine-code implementation is verified. To connect this result back to the semantics of the logic of the theorem prover, we need to formalise that semantics and prove a refinement theorem. I present some advances in the techniques that can be used to formalise HOL within itself, as well as demonstrating that the theorem prover, and its correctness proof, can be pushed through the verified compiler.

My thesis is that verification of a realistic implementation can be produced mostly automatically from a verified high-level implementation, via the use of verified compilation. I present a verified compiler and explain how it was bootstrapped to achieve a small TCB, and then explain how verified compilation can be used on a larger application, in particular, a theorem prover for higher-order logic. The application has two parts, one domain-specific and the other generic. For the domain-specific part, I formalise the semantics of higher-order logic and prove its inference system sound. For the generic part, I apply proof-grounded compilation to produce the verified implementation.

UCAM-CL-TR-880

Janina Voigt:

Access contracts: a dynamic approach to object-oriented access protection

February 2016, 171 pages, PDF
PhD thesis (Trinity College, May 2014)

Abstract: In object-oriented (OO) programming, variables do not contain objects directly but addresses of objects on the heap. Thus, several variables can point to the same object; we call this aliasing.

Aliasing is a central feature of OO programming that enables efficient sharing of objects across a system. This is essential for the implementation of many programming idioms, such as iterators. On the other hand, aliasing reduces modularity and encapsulation, making programs difficult to understand, debug and maintain.

Much research has been done on controlling aliasing. Alias protection schemes (such as Clarke et al.'s influential ownership types) limit which references can exist, thus guaranteeing the protection of encapsulated objects. Unfortunately, existing schemes are significantly restrictive and consequently have not been widely adopted by software developers.

This thesis makes three contributions to the area of alias protection. Firstly, it proposes aliasing contracts, a novel, dynamically-checked alias protection scheme for object-oriented programming languages. Aliasing contracts are highly flexible and expressive, addressing the limitations of existing work. We show that they can be used to model many existing alias protection schemes, providing a unifying approach to alias protection.

Secondly, we develop a prototype implementation of aliasing contracts in Java and use it to quantify the run-time performance of aliasing contracts. Since aliasing contracts are checked dynamically, they incur run-time performance overheads; however, our performance evaluation shows that using aliasing contracts for testing and debugging is nevertheless feasible.

Thirdly, we propose a static analysis which can verify simple aliasing contracts at compile time, including those contracts which model ownership types. Contracts which can be verified in this way can subsequently be removed from the program before it is executed. We show that such a combination of static and dynamic checking significantly improves the run-time performance of aliasing contracts.

UCAM-CL-TR-881

Juan M. Tirado, Ovidiu Serban, Qiang Guo, Eiko Yoneki:

Web data knowledge extraction

March 2016, 60 pages, PDF

Abstract: A constantly growing amount of information is available through the web. Unfortunately, extracting useful content from this massive amount of data still remains an open issue. The lack of standard data models and structures forces developers to create ad-hoc solutions from scratch. The advice of an expert is still needed in many situations where developers do not have the correct background knowledge. This forces developers to spend time acquiring the necessary background from the expert. In other directions, there are promising solutions employing machine learning techniques. However, increasing accuracy requires an increase in system complexity that cannot be endured in many projects.

In this work, we approach the web knowledge extraction problem using an expert centric methodology. This methodology defines a set of configurable, extendible and independent components that permit the reutilisation of large pieces of code among projects. Our methodology differs from similar solutions in its expert-driven design. This design makes it possible for a subject specialist to drive the knowledge extraction for a given set of documents. Additionally, we propose the utilization of machine assisted solutions that guide the expert during this process. To demonstrate the capabilities of our methodology, we present a real use case scenario in which public procurement data is extracted from the web-based repositories of several public institutions across Europe. We provide insightful details about the challenges we had to deal with in this case and additional discussions about how to apply our methodology.

UCAM-CL-TR-882

Niall Murphy:

Discovering and exploiting parallelism in DOACROSS loops

March 2016, 129 pages, PDF
PhD thesis (Darwin College, September 2015)

Abstract: Although multicore processors have been the norm for a decade, programmers still struggle to write parallel general-purpose applications, resulting in underutilised on-chip resources. Automatic parallelisation is a promising approach to improving the performance of such applications without burdening the programmer. I explore various techniques for automatically extracting parallelism which span the spectrum from conservative parallelisation, where transformations must be proven correct by the compiler, to optimistic parallelisation, where speculative execution allows the compiler to generate potentially unsafe code, with runtime supports to ensure correct execution.

Firstly I present a limit study of conservative parallelisation. I use a novel runtime profiling technique to find all data dependences which occur during execution and build an oracle data dependence graph. This oracle is used as input to the HELIX compiler which automatically generates parallelised code. In this way, the compiler is no longer limited by the inadequacies of compile-time dependence analysis and the performance of the generated code represents the upper limit of what can be achieved by HELIX-style parallelisation. I show that, despite shortcomings in the compile-time analysis, the oracle-parallelised code is no better than that ordinarily produced by HELIX.

Secondly I present a limit study of optimistic parallelisation. I implement a dataflow timing model that allows each instruction to execute as early as possible in an idealistic, zero-overhead machine, thus giving a theoretical limit to the parallelism which can be exploited. The study shows that considerable extra parallelism is available which, due to its dynamic nature, cannot be exploited by HELIX, even with the oracle dependence analysis.

Finally I demonstrate the design of a practical parallelisation system which combines the best aspects of both the conservative and optimistic parallelisation styles. I use runtime profiling to detect which HELIX-identified dependences cause frequent conflicts and synchronise these while allowing other code to run speculatively. This judicious speculation model achieves superior performance to HELIX and approaches the theoretical limit in many cases.

UCAM-CL-TR-883

Richard Russell, Sean B. Holden:

Survey propagation applied to weighted partial maximum satisfiability

March 2016, 15 pages, PDF

Abstract: We adapt the survey propagation method for application to weighted partial maximum satisfiability (WPMax-SAT) problems consisting of a mixture of hard and soft clauses. The aim is to find the truth assignment that minimises the total cost of unsatisfied soft clauses while satisfying all hard clauses. We use fixed points of the new set of message passing equations in a decimation procedure to reduce the size of WPMax-SAT problems. We evaluate this strategy on randomly generated WPMax-SAT problems and find that message passing frequently converges to nontrivial fixed points, and in many cases decimation results in simplified problems for which local search solvers are able to find truth assignments of comparable cost faster than without decimation.

UCAM-CL-TR-884

Zongyan Huang:

Machine learning and computer algebra

April 2016, 113 pages, PDF
PhD thesis (Lucy Cavendish College, August 2015)

Abstract: Computer algebra is a fundamental research field of computer science, and computer algebra systems are used in a wide range of problems, often leading to significant savings of both time and effort. However, many of these systems offer different heuristics, decision procedures, and parameter settings to tackle any given problem, and users need to manually select them in order to use the system.

In this context, the algorithm selection problem is the problem of selecting the most efficient setting of the computer algebra system when faced with a particular problem. These choices can dramatically affect the efficiency, or even the feasibility of finding a solution. Often we have to rely on human expertise to pick a suitable choice as there are no fixed rules that determine the best approach, and even for experts, the relationship between the problem at hand and the choice of an algorithm is far from obvious.

Machine learning techniques have been widely applied in fields where decisions are to be made without the presence of a human expert, such as in web search, text categorization, or recommender systems. My hypothesis is that machine learning can also be applied to help solve the algorithm selection problem for computer algebra systems.

In this thesis, we perform several experiments to determine the effectiveness of machine learning (specifically using support vector machines) for the problem of algorithm selection in three instances of computer algebra applications over real closed fields. Our three applications are: (i) choosing decision procedures and time limits in MetiTarski; (ii) choosing a heuristic for CAD variable ordering; (iii) predicting the usefulness of Gröbner basis preconditioning. The results show that machine learning can effectively be applied to these applications, with the machine learned choices being superior to both choosing a single fixed individual algorithm, as well as to random choice.

UCAM-CL-TR-885

Sean B. Holden:

HasGP: A Haskell library for Gaussian process inference

April 2016, 6 pages, PDF

Abstract: HasGP is a library providing supervised learning algorithms for Gaussian process (GP) regression and classification. While only one of many GP libraries available, it differs in that it represents an ongoing exploration of how machine learning research and deployment might benefit by moving away from the imperative/object-oriented style of implementation and instead employing the functional programming (FP) paradigm. HasGP is implemented in Haskell and is available under the GPL3 open source license.

UCAM-CL-TR-886

Ekaterina Kochmar:

Error detection in content word combinations

May 2016, 170 pages, PDF
PhD thesis (St. John's College, December 2014)

Abstract: This thesis addresses the task of error detection in the choice of content words focusing on adjective–noun and verb–object combinations. We show that error detection in content words is an underexplored area in research on learner language since (i) most previous approaches to error detection and correction have focused on other error types, and (ii) the approaches that have previously addressed errors in content words have not performed error detection proper. We show why this task is challenging for the existing algorithms and propose a novel approach to error detection in content words.

We note that since content words express meaning, an error detection algorithm should take the semantic properties of the words into account. We use a compositional distributional semantic framework in which we represent content words using their distributions in native English, while the meaning of the combinations is represented using models of compositional semantics. We present a number of measures that describe different properties of the modelled representations and can reliably distinguish between the representations of the correct and incorrect content word combinations. Finally, we cast the task of error detection as a binary classification problem and implement a machine learning classifier that uses the output of the semantic measures as features.

The results of our experiments confirm that an error detection algorithm that uses semantically motivated features achieves good accuracy and precision and outperforms the state-of-the-art approaches. We conclude that the features derived from the semantic representations encode important properties of the combinations that help distinguish the correct combinations from the incorrect ones.

The approach presented in this work can naturally be extended to other types of content word combinations. Future research should also investigate how the error correction component for content word combinations could be implemented.

UCAM-CL-TR-887

Robert M. Norton:

Hardware support for compartmentalisation

May 2016, 86 pages, PDF
PhD thesis (Clare Hall, September 2015)

Abstract: Compartmentalisation is a technique to reduce the impact of security bugs by enforcing the 'principle of least privilege' within applications. Splitting programs into separate components that each operate with minimal access to resources means that a vulnerability in one part is prevented from affecting the whole. However, the performance costs and development effort of doing this have so far prevented widespread deployment of compartmentalisation, despite the increasingly apparent need for better computer security. A major obstacle to deployment is that existing compartmentalisation techniques rely either on virtual memory hardware or pure software to enforce separation, both of which have severe performance implications and complicate the task of developing compartmentalised applications.

CHERI (Capability Hardware Enhanced RISC Instructions) is a research project which aims to improve computer security by allowing software to precisely express its memory access requirements using hardware support for bounded, unforgeable pointers known as capabilities. One consequence of this approach is that a single virtual address space can be divided into many independent compartments, with very efficient transitions and data sharing between them.

This dissertation analyses the compartmentalisation features of the CHERI Instruction Set Architecture (ISA). It includes: a summary of the CHERI ISA, particularly its compartmentalisation features; a description of a multithreaded CHERI CPU which runs the FreeBSD Operating System; the results of benchmarks that compare the characteristics of hardware supported compartmentalisation with traditional techniques; and an evaluation of proposed optimisations to the CHERI ISA to further improve domain crossing efficiency.

I find that the CHERI ISA provides extremely efficient, practical support for compartmentalisation and that there are opportunities for further optimisation if even lower overhead is required in the future.

UCAM-CL-TR-888

Sherif Akoush, Ripduman Sohan,
Andy Hopper:

Recomputation-based data reliability for MapReduce using lineage

May 2016, 19 pages, PDF

Abstract: Ensuring block-level reliability of MapReduce datasets is expensive due to the spatial overheads of replicating or erasure coding data. As the amount of data processed with MapReduce continues to increase, this cost will increase proportionally. In this paper we introduce Recomputation-Based Reliability in MapReduce (RMR), a system for mitigating the cost of maintaining reliable MapReduce datasets. RMR leverages record-level lineage of the relationships between input and output records in the job for the purposes of supporting block-level recovery. We show that collecting this lineage imposes low temporal overhead. We further show that the collected lineage is a fraction of the size of the output dataset for many MapReduce jobs. Finally, we show that lineage can be used to deterministically reproduce any block in the output. We quantitatively demonstrate that, by ensuring the reliability of the lineage rather than the output, we can achieve data reliability guarantees with a small storage requirement.

UCAM-CL-TR-889

Sherif Akoush, Ripduman Sohan,
Andrew Rice, Andy Hopper:

Evaluating the viability of remote renewable energy in datacentre computing

May 2016, 26 pages, PDF

Abstract: We investigate the feasibility of loosely-coupled distributed datacentre architectures colocated with and powered by renewable energy sources and interconnected using high-bandwidth low-latency data links. The design of these architectures advocates (i) variable machine availability: the number of machines accessible at a particular site at any given time is proportional to the amount of renewable energy available and (ii) workload deferment and migration: if there is insufficient site capacity, workloads are either halted or shifted to transient energy-rich locations.

While these architectures are attractive from an environmental perspective, their feasibility depends on (i) the requirement for additional hardware, (ii) the associated service interruptions, and (iii) the data and energy overhead of workload migration.

In this work we attempt to broadly quantify these overheads. We define a model of the basic design of the architecture incorporating the energy consumption of machines in the datacentre and the network. We further correlate this energy consumption with renewable energy available at different locations around the globe. Given this model we present two simulation-driven case studies based on data from Google and Facebook production clusters.

Generally we provide insights on the trade-offs associated with this off-grid architecture. For example, we show that an optimised configuration consisting of ten distributed datacentres results in a 2% increase in job completion time at the cost of a 50% increase in the number of machines required.

UCAM-CL-TR-890

Alistair G. Stead:

Using multiple representations to develop notational expertise in programming

June 2016, 301 pages, PDF
PhD thesis (Girton College, May 2015)

Abstract: The development of expertise with notations is an important skill for both creators and users of new technology in the future. In particular, the development of Notational Expertise (NE) is becoming valued in schools, where changes in curricula recommend that students of all ages use a range of programming representations to develop competency that can be used to ultimately create and manipulate abstract text notation. Educational programming environments making use of multiple external representations (MERs) that replicate the transitions occurring in schools within a single system present a promising area of research. They can provide support for scaffolding knowledge using low-abstraction representations, knowledge transfer, and addressing barriers faced during representation transition.

This thesis identifies analogies between the use of notation in mathematics education and programming to construct the Modes of Representational Abstraction (MoRA) framework, which supports the identification of representation transition strategies. These strategies were used to develop an educational programming environment, DrawBridge. Studies of the usage of DrawBridge highlighted the need for assessment mechanisms to measure NE. Empirical evaluation in a range of classrooms found that the MoRA framework provided useful insights and that low-abstraction representations provided a great source of motivation. Comparisons of assessments found that a novel assessment type, Adapted Parsons Problems, produced high student participation while still correlating with code-writing scores. Results strongly suggest that game-like features can encourage the use of abstract notation and increase students' acquisition of NE. Finally, findings show that students in higher year groups benefitted

more from using DrawBridge than students in lower year groups.

UCAM-CL-TR-891

Robert N. M. Watson, Peter G. Neumann, Jonathan Woodruff, Michael Roe, Jonathan Anderson, David Chisnall, Brooks Davis, Alexandre Joannou, Ben Laurie, Simon W. Moore, Steven J. Murdoch, Robert Norton, Stacey Son, Hongyan Xia:

Capability Hardware Enhanced RISC Instructions: CHERI Instruction-Set Architecture (Version 5)

June 2016, 242 pages, PDF

Abstract: This technical report describes CHERI ISAv5, the fifth version of the Capability Hardware Enhanced RISC Instructions (CHERI) Instruction-Set Architecture (ISA) being developed by SRI International and the University of Cambridge. This design captures six years of research, development, refinement, formal analysis, and testing, and is a substantial enhancement to the ISA versions described in earlier technical reports. This version introduces the CHERI-128 "compressed" capability format, adds further capability instructions to improve code generation, and rationalizes a number of ISA design choices (such as system permissions) as we have come to better understand mappings from C programming-language and MMU-based operating-system models into CHERI. It also contains improvements to descriptions, explanations, and rationale.

The CHERI instruction set is based on a hybrid capability-system architecture that adds new capability-system primitives to a commodity 64-bit RISC ISA enabling software to efficiently implement fine-grained memory protection and a hardware-software object-capability security model. These extensions support incrementally adoptable, high-performance, formally based, programmer-friendly underpinnings for fine-grained software decomposition and compartmentalization, motivated by and capable of enforcing the principle of least privilege. Fine-grained memory protection also provides direct mitigation of many widely deployed exploit techniques.

The CHERI system architecture purposefully addresses known performance and robustness gaps in commodity ISAs that hinder the adoption of more secure programming models centered around the principle of least privilege. To this end, CHERI blends traditional paged virtual memory with a per-address-space capability model that includes capability registers, capability instructions, and tagged memory that have been added to the 64-bit MIPS ISA. CHERI learns from the C-language fat-pointer literature: its capabilities describe fine-grained regions of memory and can be substituted for data or code pointers in generated code, protecting data and also providing Control-Flow Integrity (CFI). Strong monotonicity properties allow the CHERI capability model to express a variety of protection properties, from valid C-language pointer provenance and C-language bounds checking to implementing the isolation and controlled communication structures required for compartmentalization.

CHERI's hybrid system approach, inspired by the Capsicum security model, allows incremental adoption of capability-oriented software design: software implementations that are more robust and resilient can be deployed where they are most needed, while leaving less critical software largely unmodified, but nevertheless suitably constrained to be incapable of having adverse effects. For example, we are focusing conversion efforts on low-level TCB components of the system: separation kernels, hypervisors, operating-system kernels, language runtimes, and userspace TCBs such as web browsers. Likewise, we see early-use scenarios (such as data compression, protocol parsing, image processing, and video processing) that relate to particularly high-risk software libraries, which are concentrations of both complex and historically vulnerability-prone code combined with untrustworthy data sources, while leaving containing applications unchanged.

UCAM-CL-TR-892

A. Daniel Hall:

Pipelined image processing for pattern recognition

July 2016, 121 pages, PDF
PhD thesis (Queen's College, October 1991)

Abstract: Image processing for pattern recognition is both computationally intensive and algorithmically complex. The objective of the research presented here was to produce a fast inexpensive image processor for pattern recognition. This objective has been achieved by separating the computationally intensive pixel processing tasks from the algorithmically complex feature processing tasks.

The context for this work is explored in terms of image processor architecture, intermediate-level image processing tasks and pattern recognition.

A new language to describe pipelined neighbourhood operations on binary images ('PiNOLa': Pipelined Neighbourhood Operator Language) is presented. PiNOLa was implemented in Modula-2 to provide an interactive simulation system ('PiNOSim': Pipelined Neighbourhood Operator Simulator). Experiments using PiNOSim were conducted and a design for a topological feature extractor was produced.

A novel algorithm for connected component labelling in hardware is presented. This algorithm was

included in the PiNOSim program to enable the component labelling of features extracted using the language. The component labelling algorithm can be used with the topological feature extractor mentioned above. The result is a method of converting a binary raster scan into a stream of topological features grouped by connected object.

To test the potential performance of a system based on these ideas, some hardware ('GRIPPR': Generic Real-time Image Processor for Pattern Recognition) was designed. This machine was implemented using standard components linked to a PC based transputer board. To demonstrate an application of GRIPPR an Optical Character Recognition (OCR) system is presented. Finally, results demonstrating a continuous throughput of 1500 characters/second are given.

UCAM-CL-TR-893

Thomas F. J.-M. Pasquier:

Towards practical information flow control and audit

July 2016, 153 pages, PDF
PhD thesis (Jesus College, January 2016)

Abstract: In recent years, pressure from the general public and from policy makers has been for more and better control over personal data in cloud computing environments. Regulations put responsibilities on cloud tenants to ensure that proper measures are effected by their cloud provider. But there is currently no satisfactory mechanism to achieve this, leaving tenants open to potentially costly lawsuits.

Decentralised Information Flow Control (IFC) at system level is a data-centric Mandatory Access Control scheme that guarantees non-interference across security contexts, based on lattices defined by secrecy and integrity properties. Every data flow is continuously monitored to guarantee the enforcement of decentrally specified policies. Applications running above IFC enforcement need not be trusted and can interact. IFC constraints can be used to ensure that proper workflows are followed, as defined by regulations or contracts. For example, to ensure that end users' personal data are anonymised before being disclosed to third parties.

Information captured during IFC enforcement allows a directed graph representing whole-system data exchange to be generated. The coupling of policy enforcement and audit data capture allows system "noise" to be removed from audit data, and only information relevant to the policies in place to be recorded. It is possible to query these graphs to demonstrate that the system behaved according to regulation. For example, to demonstrate from run-time data that there is no path without anonymisation between an end-user and a third party.

UCAM-CL-TR-894

Christopher Bryant, Mariano Felice:

Issues in preprocessing current datasets for grammatical error correction

September 2016, 15 pages, PDF

Abstract: In this report, we describe some of the issues encountered when preprocessing two of the largest datasets for Grammatical Error Correction (GEC); namely the public FCE corpus and NUCLE (along with associated CoNLL test sets). In particular, we show that it is not straightforward to convert character level annotations to token level annotations and that sentence segmentation is more complex when annotations change sentence boundaries. These become even more complicated when multiple annotators are involved. We subsequently describe how we handle such cases and consider the pros and cons of different methods.

UCAM-CL-TR-895

Mariano Felice:

Artificial error generation for translation-based grammatical error correction

October 2016, 155 pages, PDF
PhD thesis (Hughes Hall, October 2016)

Abstract: Automated grammatical error correction for language learners has attracted a lot of attention in recent years, especially after a number of shared tasks that have encouraged research in the area. Treating the problem as a translation task from 'incorrect' into 'correct' English using statistical machine translation has emerged as a state-of-the-art approach but it requires vast amounts of corrected parallel data to produce useful results. Because manual annotation of incorrect text is laborious and expensive, we can generate artificial error-annotated data by injecting errors deliberately into correct text and thus produce larger amounts of parallel data with much less effort.

In this work, we review previous work on artificial error generation and investigate new approaches using random and probabilistic methods for constrained and general error correction. Our methods use error statistics from a reference corpus of learner writing to generate errors in native text that look realistic and plausible in context. We investigate a number of aspects that can play a part in the error generation process, such as the origin of the native texts, the amount of context used to find suitable insertion points, the type of information encoded by the error patterns and the output error distribution. In addition, we explore the use of linguistic

information for characterising errors and train systems using different combinations of real and artificial data.

Results of our experiments show that the use of artificial errors can improve system performance when they are used in combination with real learner errors, in line with previous research. These improvements are observed for both constrained and general correction, for which probabilistic methods produce the best results. We also demonstrate that systems trained on a combination of real and artificial errors can beat other highly-engineered systems and be more robust, showing that performance can be improved by focusing on the data rather than tuning system parameters.

Part of our work is also devoted to the proposal of the I-measure, a new evaluation scheme that scores corrections in terms of improvement on the original text and solves known issues with existing evaluation measures.

UCAM-CL-TR-896

Kumar Sharad:

Learning to de-anonymize social networks

December 2016, 158 pages, PDF
PhD thesis (Churchill College, November 2016)

Abstract: Releasing anonymized social network data for analysis has been a popular idea among data providers. Despite evidence to the contrary the belief that anonymization will solve the privacy problem in practice refuses to die. This dissertation contributes to the field of social graph de-anonymization by demonstrating that even automated models can be quite successful in breaching the privacy of such datasets. We propose novel machine-learning based techniques to learn the identities of nodes in social graphs, thereby automating manual, heuristic-based attacks. Our work extends the vast literature of social graph de-anonymization attacks by systematizing them.

We present a random-forests based classifier which uses structural node features based on neighborhood degree distribution to predict their similarity. Using these simple and efficient features we design versatile and expressive learning models which can learn the de-anonymization task just from a few examples. Our evaluation establishes their efficacy in transforming de-anonymization to a learning problem. The learning is transferable in that the model can be trained to attack one graph when trained on another.

Moving on, we demonstrate the versatility and greater applicability of the proposed model by using it to solve the long-standing problem of benchmarking social graph anonymization schemes. Our framework bridges a fundamental research gap by making cheap, quick and automated analysis of anonymization schemes possible, without even requiring their full description. The benchmark is based on comparison of structural information leakage vs. utility preservation. We study the trade-off of anonymity vs. utility for six popular anonymization schemes including those promising k-anonymity. Our analysis shows that none of the schemes are fit for the purpose.

Finally, we present an end-to-end social graph de-anonymization attack which uses the proposed machine learning techniques to recover node mappings across intersecting graphs. Our attack enhances the state of art in graph de-anonymization by demonstrating better performance than all the other attacks including those that use seed knowledge. The attack is seedless and heuristic free, which demonstrates the superiority of machine learning techniques as compared to hand-selected parametric attacks.

UCAM-CL-TR-897

Sheharbano Khattak:

Characterization of Internet censorship from multiple perspectives

January 2017, 170 pages, PDF
PhD thesis (Robinson College, January 2017)

Abstract: Internet censorship is rampant, both under the support of nation states and private actors, with important socio-economic and policy implications. Yet many issues around Internet censorship remain poorly understood because of the lack of adequate approaches to measure the phenomenon at scale. This thesis aims to help fill this gap by developing three methodologies to derive censorship ground truths, that are then applied to real-world datasets to study the effects of Internet censorship. These measurements are given foundation in a comprehensive taxonomy that captures the mechanics, scope, and dynamics of Internet censorship, complemented by a framework that is employed to systematize over 70 censorship resistance systems.

The first part of this dissertation analyzes "user-side censorship", where a device near the user, such as the local ISP or the national backbone, blocks the user's online communication. This study provides quantified insights into how censorship affects users, content providers, and Internet Service Providers (ISPs); as seen through the lens of traffic datasets captured at an ISP in Pakistan over a period of three years, beginning in 2011.

The second part of this dissertation moves to "publisher-side censorship". This is a new kind of blocking where the user's request arrives at the Web publisher, but the publisher (or something working on its behalf) refuses to respond based on some property of the user. Publisher-side censorship is explored in two contexts. The first is in the context of an anonymity network, Tor, involving a systematic enumeration and characterization of websites that treat Tor users differently from other users.

Continuing on the topic of publisher-side blocking, the second case study examines the Web's differential

treatment of users of adblocking software. The rising popularity of adblockers in recent years poses a serious threat to the online advertising industry, prompting publishers to actively detect users of adblockers and subsequently block them or otherwise coerce them to disable the adblocker. This study presents a first characterization of such practices across the Alexa Top 5,000 websites.

This dissertation demonstrates how the censor's blocking choices can leave behind a detectable pattern in network communications, that can be leveraged to establish exact mechanisms of censorship. This knowledge facilitates the characterization of censorship from different perspectives; uncovering entities involved in censorship, its targets, and the effects of such practices on stakeholders. More broadly, this study complements efforts to illuminate the nature, scale, and effects of opaque filtering practices; equipping policy-makers with the knowledge necessary to systematically and effectively respond to Internet censorship.

UCAM-CL-TR-898

Dongting Yu:

Access control for network management

January 2017, 108 pages, PDF
PhD thesis (Robinson College, May 2016)

Abstract: Network management inherently involves human input. From expressing business logic and network policy to the low level commands to networking devices, at least some tasks are done manually by operators. These tasks are a source of error whose consequences can be severe, since operators have high levels of access, and can bring down a whole network if they configure it improperly.

Software-Defined Networking (SDN) is a new network technology that brings even more flexibility (and risk) to network operations. Operators can now easily get third-party apps to run in their networks, or even host tenants and give them some control over their portion of the network. However security has not caught up, and it is easy for these third parties to access network resources without permission.

Access control is a mature concept; it has been studied for decades. In this dissertation I examine how access control can be used to solve the above network management problems. I look at the Border Gateway Protocol (BGP) from an operations perspective and propose a mandatory access control model using role-based policies to separate long-term invariants from day-to-day configurations. As a result the former would not be accidentally violated when configuring the latter, as this is a significant source of BGP misconfigurations today. Then, for SDN, I propose to add access control to controllers so that virtual controllers and applications that run within them cannot have unlimited access to the network infrastructure, as they do currently. Adding attribute-based access control makes the system much less fragile while it still retains the essential flexibility provided by SDN. Lastly, I propose a hierarchical architecture which, with SDN, can isolate security compromises even when some devices are physically compromised. This is achieved by using access control to both enable network access and deny unexpected connections.

UCAM-CL-TR-899

Douwe Kiela:

Deep embodiment: grounding semantics in perceptual modalities

February 2017, 128 pages, PDF
PhD thesis (Darwin College, July 2016)

Abstract: Multi-modal distributional semantic models address the fact that text-based semantic models, which represent word meanings as a distribution over other words, suffer from the grounding problem. This thesis advances the field of multi-modal semantics in two directions. First, it shows that transferred convolutional neural network representations outperform the traditional bag of visual words method for obtaining visual features. It is then shown that these representations may be applied successfully to various natural language processing tasks. Second, it performs the first ever experiments with grounding in the non-visual modalities of auditory and olfactory perception using raw data. Deep learning, a natural fit for deriving grounded representations, is used to obtain the highest-quality representations compared to more traditional approaches. Multi-modal representation learning leads to improvements over language-only models in a variety of tasks. If we want to move towards human-level artificial intelligence, we will need to build multi-modal models that represent the full complexity of human meaning, including its grounding in our various perceptual modalities.

UCAM-CL-TR-900

Valentin Dalibard:

A framework to build bespoke auto-tuners with structured Bayesian optimisation

February 2017, 182 pages, PDF
PhD thesis (St. John's College, January 2017)

Abstract: Due to their complexity, modern computer systems expose many configuration parameters which users must manually tune to maximise the performance of their applications. To relieve users of this burden, auto-tuning has emerged as an alternative in which a black-box optimiser iteratively evaluates configurations

to find efficient ones. A popular auto-tuning technique is Bayesian optimisation, which uses the results to incrementally build a probabilistic model of the impact of the parameters on performance. This allows the optimisation to quickly focus on efficient regions of the configuration space. Unfortunately, for many computer systems, either the configuration space is too large to develop a good model, or the time to evaluate performance is too long to be executed many times.

In this dissertation, I argue that by extracting a small amount of domain specific knowledge about a system, it is possible to build a bespoke auto-tuner with significantly better performance than its off-the-shelf counterparts. This could be performed, for example, by a system engineer who has a good understanding of the underlying system behaviour and wants to provide performance portability. This dissertation presents BOAT, a framework to build BespOke Auto-Tuners. BOAT offers a novel set of abstractions designed to make the exposition of domain knowledge easy and intuitive.

First, I introduce Structured Bayesian Optimisation (SBO), an extension of the Bayesian optimisation algorithm. SBO can leverage a bespoke probabilistic model of the system's behaviour, provided by the system engineer, to rapidly converge to high performance configurations. The model can benefit from observing many runtime measurements per evaluation of the system, akin to the use of profilers.

Second, I present Probabilistic-C++ a lightweight, high performance probabilistic programming library. It allows users to declare a probabilistic models of their system's behaviour and expose it to an SBO. Probabilistic programming is a recent tool from the Machine Learning community making the declaration of structured probabilistic models intuitive.

Third, I present a new optimisation scheduling abstraction which offers a structured way to express optimisations which themselves execute other optimisations. For example, this is useful to express Bayesian optimisations, which each iteration execute a numerical optimisation. Furthermore, it allows users to easily implement decompositions which exploit the loose coupling of subsets of the configuration parameters to optimise them almost independently.

UCAM-CL-TR-901

Chao Gao:

Signal maps for smartphone localisation

February 2017, 127 pages, PDF
PhD thesis (King's College, August 2016)

Abstract: Indoor positioning has been an active research area for 20 years. Systems based on dedicated infrastructure such as ultrasound or ultra-wideband (UWB) radio can provide centimetre-accuracy. But they are generally prohibitively expensive to build, deploy and maintain. Today, signal fingerprinting-based indoor positioning techniques, which use existing wireless infrastructure, are arguably the most prevalent. The fingerprinting-based positioning system matches the current signal observations (fingerprints) at a device to position it on a pre-defined fingerprint map. The map is created via some form of survey. However, a major deterrent of these systems is the initial creation and subsequent maintenance of the signal maps. The commonly used map building method is the so-called manual survey, during which a surveyor visits each point on a regular grid and measures the signal fingerprints there. This traditional method is laborious and not considered scalable. An emerging alternative to manual survey is the path survey, in which a surveyor moves continuously through the environment and signal measurements are taken by the surveying device along the path. A path survey is intuitively better than a manual survey, at least in terms of speed. But, path surveys have not been well-studied yet.

This thesis assessed the path survey quantitatively and rigorously, demonstrated that path survey can approach the manual survey in terms of accuracy if certain guidelines are followed. Automated survey systems have been proposed and a commodity smart-phone is the only survey device required. The proposed systems achieve sub-metre accuracy in recovering the survey trajectory both with and without environmental information (floor plans), and have been found to outperform the state-of-the-art in terms of robustness and scalability.

This thesis concludes that path survey can be streamlined by the proposed systems to replace the laborious manual survey. The proposed systems can promote the deployment of indoor positioning system in large-scale and complicated environments, especially in dynamic environments where frequent re-survey is needed.

UCAM-CL-TR-902

Raoul-Gabriel Urma:

Programming language evolution

February 2017, 129 pages, PDF
PhD thesis (Hughes Hall, September 2015)

Abstract: Programming languages are the way developers communicate with computers—just like natural languages let us communicate with one another. In both cases multiple languages have evolved. However, the programming language ecosystem changes at a much higher rate compared to natural languages. In fact, programming languages need to constantly evolve in response to user needs, hardware advances and research developments or they are otherwise displaced by newcomers. As a result, existing code written by developers may stop working with a newer language version. Consequently, developers need to "search" (analyse) and

"replace" (refactor) code in their code bases to support new language versions.

Traditionally, tools have focused on the replace aspect (refactoring) to support developers evolving their code bases. This dissertation argues that developers also need machine support focused on the search aspect.

This dissertation starts by characterising factors driving programming language evolution based on external versus internal forces. Next, it introduces a classification of changes introduced by language designers that affect developers. It then contributes three techniques to support developers in analysing their code bases.

First, we show a source code query system based on graphs that express code queries at a mixture of syntax-tree, type, control-flow graph and data-flow levels. We demonstrate how this system can support developers and language designers in locating various code patterns relevant in evolution.

Second, we design an optional run-time type system for Python, that lets developers manually specify contracts to identify semantic incompatibilities between Python 2 and Python 3.

Third, recognising that existing codebases do not have such contracts, we describe a dynamic analysis to automatically generate them.

UCAM-CL-TR-903

Jannis Bulian:

Parameterized complexity of distances to sparse graph classes

February 2017, 133 pages, PDF
PhD thesis (Clare College, February 2016)

Abstract: This dissertation offers contributions to the area of parameterized complexity theory, which studies the complexity of computational problems in a multivariate framework. We demonstrate that for graph problems, many popular parameters can be understood as distances that measure how far a graph is from belonging to a class of sparse graphs. We introduce the term distance parameter for such parameters and demonstrate their value in several ways.

The parameter tree-depth is uncovered as a distance parameter, and we establish several fixed-parameter tractability and hardness results for problems parameterized by tree-depth.

We introduce a new distance parameter elimination distance to a class C. The parameter measures distance in a way that naturally generalises the notion of vertex deletion by allowing deletions from every component in each deletion step. We show that tree-depth is a special case of this new parameter, introduce several characterisations of elimination distance, and discuss applications. In particular we demonstrate that Graph Isomorphism is FPT parameterized by elimination distance to bounded degree, extending known results. We also show that checking whether a given graph has elimination distance k to a minor-closed class C is FPT, and, moreover, if we are given a finite set of minors characterising C, that there is an explicit FPT algorithm for this.

We establish an algorithmic meta-theorem for several distance parameters, by showing that all graph problems that are both slicewise nowhere dense and slicewise first-order definable are FPT.

Lastly, several fixed-parameter tractability results are established for two graph problems that have natural distance parameterizations.

UCAM-CL-TR-904

Zheng Yuan:

Grammatical error correction in non-native English

March 2017, 145 pages, PDF
PhD thesis (St. Edmund's College, September 2016)

Abstract: Grammatical error correction (GEC) is the task of automatically correcting grammatical errors in written text. Previous research has mainly focussed on individual error types and current commercial proofreading tools only target limited error types. As sentences produced by learners may contain multiple errors of different types, a practical error correction system should be able to detect and correct all errors.

In this thesis, we investigate GEC for learners of English as a Second Language (ESL). Specifically, we treat GEC as a translation task from incorrect into correct English, explore new models for developing end-to-end GEC systems for all error types, study system performance for each error type, and examine model generalisation to different corpora. First, we apply Statistical Machine Translation (SMT) to GEC and prove that it can form the basis of a competitive all-errors GEC system. We implement an SMT-based GEC system which contributes to our winning system submitted to a shared task in 2014. Next, we propose a ranking model to re-rank correction candidates generated by an SMT-based GEC system. This model introduces new linguistic information and we show that it improves correction quality. Finally, we present the first study using Neural Machine Translation (NMT) for GEC. We demonstrate that NMT can be successfully applied to GEC and help capture new errors missed by an SMT-based GEC system.

While we focus on GEC for English, the methods presented in this thesis can be easily applied to any language.

UCAM-CL-TR-905

William Sonnex:

Fixed point promotion: taking the induction out of automated induction

March 2017, 170 pages, PDF
PhD thesis (Trinity College, September 2015)

Abstract: This thesis describes the implementation of Elea: the first automated theorem prover for properties of observational approximation between terms in a functional programming language. A term approximates another if no program context can tell them apart, except that the approximating term can be less defined than the approximated term. Elea can prove terms equivalent by proving approximation in both directions.

The language Elea proves approximations over is pure, call-by-name, and with non-strict data-types, a subset of the Haskell language. In a test set of established properties from the literature, Elea performs comparably well to the current state of the art in automated equivalence proof, but where these provers assume all terms are total, Elea does not.

Elea uses a novel method to prove approximations, fixed-point promotion, a method which does not suffer from some of the weaknesses of cyclic proof techniques such as induction or coinduction. Fixed-point promotion utilises an unfold-fold style term rewriting system, similar to supercompilation, which uses multiple novel term rewriting steps in order to simulate the lemma discovery techniques of automated cyclic provers.

Elea is proven sound using denotational semantics, including a novel method for verifying an unfold-fold style rewrite, a method I have called truncation fusion.

UCAM-CL-TR-906

Tomas Petricek:

Context-aware programming languages

March 2017, 218 pages, PDF
PhD thesis (Clare Hall, March 2017)

Abstract: The development of programming languages needs to reflect important changes in the way programs execute. In recent years, this has included the development of parallel programming models (in reaction to the multi-core revolution) or improvements in data access technologies. This thesis is a response to another such revolution – the diversification of devices and systems where programs run.

The key point made by this thesis is the realization that an execution environment or a context is fundamental for writing modern applications and that programming languages should provide abstractions for programming with context and verifying how it is accessed.

We identify a number of program properties that were not connected before, but model some notion of context. Our examples include tracking different execution platforms (and their versions) in cross-platform development, resources available in different execution environments (e.g. GPS sensor on a phone and database on the server), but also more traditional notions such as variable usage (e.g. in liveness analysis and linear logics) or past values in stream-based dataflow programming. Our first contribution is the discovery of the connection between the above examples and their novel presentation in the form of calculi (coeffect systems). The presented type systems and formal semantics highlight the relationship between different notions of context.

Our second contribution is the definition of two unified coeffect calculi that capture the common structure of the examples. In particular, our flat coeffect calculus models languages with contextual properties of the execution environment and our structural coeffect calculus models languages where the contextual properties are attached to the variable usage. We define the semantics of the calculi in terms of category theoretical structure of an indexed comonad (based on dualisation of the well-known monad structure), use it to define operational semantics and prove type safety of the calculi.

Our third contribution is a novel presentation of our work in the form of web-based interactive essay. This provides a simple implementation of three context-aware programming languages and lets the reader write and run simple context-aware programs, but also explore the theory behind the implementation including the typing derivation and semantics.

UCAM-CL-TR-907

Robert N. M. Watson, Peter G. Neumann, Jonathan Woodruff, Michael Roe, Jonathan Anderson, John Baldwin, David Chisnall, Brooks Davis, Alexandre Joannou, Ben Laurie, Simon W. Moore, Steven J. Murdoch, Robert Norton, Stacey Son, Hongyan Xia:

Capability Hardware Enhanced RISC Instructions: CHERI Instruction-Set Architecture (Version 6)

April 2017, 307 pages, PDF

Abstract: This technical report describes CHERI ISAv6, the sixth version of the Capability Hardware Enhanced RISC Instructions (CHERI) Instruction-Set Architecture (ISA) being developed by SRI International and the University of Cambridge. This design captures seven years of research, development, experimentation, refinement, formal analysis, and validation through hardware and software implementation. CHERI ISAv6 is a substantial enhancement to prior ISA versions: it introduces support for kernel-mode compartmentalization, jump-based rather than exception-based domain transition, architecture-abstracted and efficient tag restoration, and more efficient generated code. A new chapter addresses potential applications of the CHERI model to the RISC-V and x86-64 ISAs, previously described relative only to the 64-bit MIPS ISA. CHERI ISAv6 better explains our design rationale and research methodology.

CHERI is a hybrid capability-system architecture that adds new capability-system primitives to a commodity 64-bit RISC ISA enabling software to efficiently implement fine-grained memory protection and scalable software compartmentalization. Design goals have included incremental adoptability within current ISAs and software stacks, low performance overhead for memory protection, significant performance improvements for software compartmentalization, formal grounding, and programmer-friendly underpinnings. Throughout, we have focused on providing strong and efficient architectural foundations for the principles of least privilege and intentional use in the execution of software at multiple levels of abstraction, preventing and mitigating vulnerabilities.

The CHERI system architecture purposefully addresses known performance and robustness gaps in commodity ISAs that hinder the adoption of more secure programming models centered around the principle of least privilege. To this end, CHERI blends traditional paged virtual memory with an in-address-space capability model that includes capability registers, capability instructions, and tagged memory. CHERI builds on C-language fat-pointer literature: its capabilities describe fine-grained regions of memory and can be substituted for data or code pointers in generated code, protecting data and also improving control-flow robustness. Strong capability integrity and monotonicity properties allow the CHERI model to express a variety of protection properties, from enforcing valid C-language pointer provenance and bounds checking to implementing the isolation and controlled communication structures required for software compartmentalization.

CHERI's hybrid capability-system approach, inspired by the Capsicum security model, allows incremental adoption of capability-oriented design: software implementations that are more robust and resilient can be deployed where they are most needed, while leaving less critical software largely unmodified, but nevertheless suitably constrained to be incapable of having adverse effects. Potential deployment scenarios include low-level software Trusted Computing Bases (TCBs) such as separation kernels, hypervisors, and operating-system kernels, as well as userspace TCBs such as language runtimes and web browsers. Likewise, we see early-use scenarios (such as data compression, protocol parsing, and image processing) that relate to particularly high-risk software libraries, which are concentrations of both complex and historically vulnerability-prone code exposed to untrustworthy data sources, while leaving containing applications unchanged.

UCAM-CL-TR-908

Raphaël L. Proust:

ASAP: As Static As Possible memory management

July 2017, 147 pages, PDF
PhD thesis (Magdalene College, July 2016)

Abstract: Today, there are various ways to manage the memory of computer programs: garbage collectors of all kinds, reference counters, regions, linear types – each with benefits and drawbacks, each fit for specific settings, each appropriate to different problems, each with their own trade-offs.

Despite the plethora of techniques available, system programming (device drivers, networking libraries, cryptography applications, etc.) is still mostly done in C, even though memory management in C is notoriously unsafe. As a result, serious bugs are continuously discovered in system software.

In this dissertation, we study memory management strategies with a view to fitness for system programming.

First, we establish a framework to study memory management strategies. Often perceived as distinct categories, we argue that memory management approaches are actually part of a single design space. To this end, we establish a precise and powerful lexicon to describe memory management strategies of any kind. Using our newly established vocabulary, we further argue that this design space has not been exhaustively explored. We argue that one of the unexplored portion of this space, the static-automatic gap, contributes to the persistence of C in system programming.

Second, we develop ASAP: a new memory management technique that fits in the static-automatic gap. ASAP is fully automatic (not even annotations are required) and makes heavy use of static analysis. At compile time it inserts, in the original program, code that deallocates memory blocks as they becomes useless. We then show how ASAP interacts with various, advanced language features. Specifically, we extend ASAP to support polymorphism and mutability.

Third, we compare ASAP with existing approaches. One of the points of comparison we use is the behavioural suitability to system programming. We also explore how the ideas from ASAP can be combined

with other memory management strategies. We then show how ASAP handles programs satisfying the linear or region constraints. Finally, we explore the insights gained whilst developing and studying ASAP.

UCAM-CL-TR-909

Laurent Simon:

Exploring new attack vectors for the exploitation of smartphones

July 2017, 167 pages, PDF
PhD thesis (Homerton College, April 2016)

Abstract: Smartphones have evolved from simple candy-bar devices into powerful miniature computing platforms. Today's smartphones are complex multi-tenant platforms: users, OS providers, manufacturers, carriers, and app developers have to co-exist on a single device. As with other computing platforms, smartphone security research has dedicated a lot of effort, first, into the detection and prevention of ill-intentioned software; second, into the detection and mitigation of operating system vulnerabilities; and third, into the detection and mitigation of vulnerabilities in applications.

In this thesis, I take a different approach. I explore and study attack vectors that are specific to smartphones; that is, attack vectors that do not exist on other computing platforms because they are the result of these phones' intrinsic characteristics. One such characteristic is the sheer number of sensors and peripherals, such as an accelerometer, a gyroscope and a built-in camera. Their number keeps increasing with new usage scenarios, e.g. for health or navigation. So I show how to abuse the camera and microphone to infer a smartphone's motion during user input. I then correlate motion characteristics to the keyboard digits touched by a user so as to infer PINs. This can work even if the input is protected through a Trusted Execution Environment (TEE), the industry's preferred answer to the trusted path problem.

Another characteristic is their form factor, such as their small touch screen. New input methods have been devised to make user input easier, such as "gesture typing". So I study a new side channel that exploits hardware and software interrupt counters to infer what users type using this widely adopted input method.

Another inherent trait is that users carry smartphones everywhere. This increases the risk of theft or loss. In fact, in 2013 alone, 3.1M devices were stolen in the USA, and 120,000 in London. So I study the effectiveness of anti-theft software for the Android platform, and demonstrate a wide variety of vulnerabilities.

Yet another characteristic of the smartphone ecosystem if the pace at which new devices are released: users tend to replace their phone about every 2 years, compared to 4.5 years for their personal computers. For already 60% of users today, the purchase of a new smartphone is partly funded by selling the previous one. This can have privacy implications if the previous owner's personal data is not properly erased. So I study the effectiveness of the built-in sanitisation features in Android smartphones, lifting the curtains on their problems and their root causes.

UCAM-CL-TR-910

William Denman:

Automated verification of continuous and hybrid dynamical systems

July 2017, 178 pages, PDF
PhD thesis (Clare College, September 2014)

Abstract: The standard method used for verifying the behaviour of a dynamical system is simulation. But simulation can check only a finite number of operating conditions and system parameters, leading to a potentially incomplete verification result. This dissertation presents several automated theorem proving based methods that can, in contrast to simulation, completely guarantee the safety of a dynamical system model.

To completely verify a purely continuous dynamical system requires proving a universally quantified first order conjecture, which represents all possible trajectories of the system. Such a closed form solution may contain transcendental functions, rendering the problem undecidable in the general case. The automated theorem prover MetiTarski can be used to solve such a problem by reducing it to one over the real closed fields. The main issue is the doubly exponential complexity of the back-end decision procedures that it depends on. This dissertation proposes several techniques that make the required conjectures easier for MetiTarski to prove. The techniques are shown to be effective at reducing the amount of time required to prove high dimensional problems and are further demonstrated on a flight collision avoidance case study.

For hybrid systems, which contain both continuous and discrete behaviours, a different approach is proposed. In this case, qualitative reasoning is used to abstract the continuous state space into a set of discrete cells. Then, standard discrete state formal verification tools are used to verify properties over the abstraction. MetiTarski is employed to determine the feasibility of the cells and the transitions between them. As these checks are reduced to proving inequalities over the theory of the reals, it facilitates the analysis of non-polynomial hybrid systems that contain transcendental functions in their vector fields, invariants and guards. This qualitative abstraction framework has been implemented in the QUANTUM tool and is demonstrated on several hybrid system benchmark problems.

UCAM-CL-TR-911

Dominic Orchard, Mistral Contrastin,
Matthew Danish, Andrew Rice:

Proofs for 'Verifying Spatial Properties of Array Computations'

September 2017, 8 pages, PDF

Abstract: This technical report provides accompanying proofs for the paper: Verifying Spatial Properties of Array Computations. We show three properties of the lattice model of the stencil specification language. 1) that it is sound with respect to the equational theory of region specifications; 2) that is it sound with respect to the theory of region approximation; 3) that the inference algorithm is sound. We further derive useful identities on the specification language and properties of Union Normal Form—the data structure used to implement the model. Core definitions from the paper are restated here to make the overall report more self contained.

UCAM-CL-TR-912

Ang Kun Joo Michael, Emma Valla,
Natinael Solomon Neggatu,
Andrew W. Moore:

Network traffic classification via neural networks

September 2017, 25 pages, PDF

Abstract: The importance of network traffic classification has grown over the last decade. Coupled with advances in software and theory, the range of classification techniques has also increased. Network operators can predict demands in future traffic to high accuracy and better identify anomalous behavior. Multiple machine learning tools have been developed in this field and each have had varying degrees of success. In this paper we use supervised machine learning within a frequentist neural network to develop a model capable of achieving high classification accuracy and maintaining low system throughput. We will compare our model to previous work on Bayesian neural networks and other standard classification techniques in the context of real-time classification. The spatial and temporal stabilities of the different models will also be compared. Finally, we investigate the relationship between the convergence times of each model and the size of training dataset. Emphasis will be placed on experimental design and methodology to adequately justify and contextualize our analysis, as well as clarify the limitations of our results. Challenges in the field and areas for further work will also be discussed.

UCAM-CL-TR-913

Matic Horvat:

Hierarchical statistical semantic translation and realization

October 2017, 215 pages, PDF
PhD thesis (Churchill College, March 2017)

Abstract: Statistical machine translation (SMT) approaches extract translation knowledge automatically from parallel corpora. They additionally take advantage of monolingual text for target-side language modelling. Syntax-based SMT approaches also incorporate knowledge of source and/or target syntax by taking advantage of monolingual grammars induced from treebanks, and semantics-based SMT approaches use knowledge of source and/or target semantics in various forms. However, there has been very little research on incorporating the considerable monolingual knowledge encoded in deep, hand-built grammars into statistical machine translation. Since deep grammars can produce semantic representations, such an approach could be used for realization as well as MT.

In this thesis I present a hybrid approach combining some of the knowledge in a deep hand-built grammar, the English Resource Grammar (ERG), with a statistical machine translation approach. The ERG is used to parse the source sentences to obtain Dependency Minimal Recursion Semantics (DMRS) representations. DMRS representations are subsequently transformed to a form more appropriate for SMT, giving a parallel corpus with transformed DMRS on the source side and aligned strings on the target side. The SMT approach is based on hierarchical phrase-based translation (Hiero). I adapt the Hiero synchronous context-free grammar (SCFG) to comprise graph-to-string rules. DMRS graph-to-string SCFG is extracted from the parallel corpus and used in decoding to transform an input DMRS graph into a target string either for machine translation or for realization.

I demonstrate the potential of the approach for large-scale machine translation by evaluating it on the WMT15 English-German translation task. Although the approach does not improve on a state-of-the-art Hiero implementation, a manual investigation reveals some strengths and future directions for improvement. In addition to machine translation, I apply the approach to the MRS realization task. The approach produces realizations of high quality, but its main strength lies in its robustness. Unlike the established MRS realization approach using the ERG, the approach proposed in this thesis is able to realize representations that do not correspond perfectly to ERG semantic output, which will naturally occur in practical realization tasks. I demonstrate this in three contexts, by realizing representations derived from sentence compression, from robust parsing, and from the transfer-phase of an existing MT system.

In summary, the main contributions of this thesis are a novel architecture combining a statistical machine translation approach with a deep hand-built grammar and a demonstration of its practical usefulness as a large-scale machine translation system and a robust realization alternative to the established MRS realization approach.

UCAM-CL-TR-914

Diana Andreea Popescu, Noa Zilberman, Andrew W. Moore:

Characterizing the impact of network latency on cloud-based applications' performance

November 2017, 20 pages, PDF

Abstract: Businesses and individuals run increasing numbers of applications in the cloud. The performance of an application running in the cloud depends on the data center conditions and upon the resources committed to an application. Small network delays may lead to a significant performance degradation, which affects both the user's cost and the service provider's resource usage, power consumption and data center efficiency. In this work, we quantify the effect of network latency on several typical cloud workloads, varying in complexity and use cases. Our results show that different applications are affected by network latency to differing amounts. These insights into the effect of network latency on different applications have ramifications for workload placement and physical host sharing when trying to reach performance targets.

UCAM-CL-TR-915

Andrew Caines, Diane Nicholls, Paula Buttery:

Annotating errors and disfluencies in transcriptions of speech

December 2017, 10 pages, PDF

Abstract: This document presents our guidelines for the annotation of errors and disfluencies in transcriptions of speech. There is a well-established precedent for annotating errors in written texts but the same is not true of speech transcriptions. We describe our coding scheme, discuss examples and difficult cases, and introduce new codes to deal with features characteristic of speech.

UCAM-CL-TR-916

Robert N. M. Watson, Jonathan Woodruff, Michael Roe, Simon W. Moore, Peter G. Neumann:

Capability Hardware Enhanced RISC Instructions (CHERI): Notes on the Meltdown and Spectre Attacks

February 2018, 16 pages, PDF

Abstract: In this report, we consider the potential impact of recently announced Meltdown and Spectre microarchitectural side-channel attacks arising out of superscalar (out-of-order) execution on Capability Hardware Enhanced RISC Instructions (CHERI) computer architecture. We observe that CHERI's in-hardware permissions and bounds checking may be an effective form of mitigation for one variant of these attacks, in which speculated instructions can bypass software bounds checking. As with MMU-based techniques, CHERI remains vulnerable to side-channel leakage arising from speculative execution across compartment boundaries, leading us to propose a software-managed compartment ID to mitigate these vulnerabilities for other variants as well.

UCAM-CL-TR-917

Matko Botinčan:

Formal verification-driven parallelisation synthesis

March 2018, 163 pages, PDF
PhD thesis (Trinity College, September 2013)

Abstract: Concurrency is often an optimisation, rather than intrinsic to the functional behaviour of a program, i.e., a concurrent program is often intended to achieve the same effect of a simpler sequential counterpart, just faster. Error-free concurrent programming remains a tricky problem, beyond the capabilities of most programmers. Consequently, an attractive alternative to manually developing a concurrent program is to automatically synthesise one.

This dissertation presents two novel formal verification-based methods for safely transforming a sequential program into a concurrent one. The first method — an instance of proof-directed synthesis — takes as the input a sequential program and its safety proof, as well as annotations on where to parallelise, and produces a correctly-synchronised parallelised program, along with a proof of that program. The method uses the sequential proof to guide the insertion of synchronisation barriers to ensure that the parallelised program has the same behaviour as the original sequential version. The sequential proof, written in separation

logic, need only represent shape properties, meaning we can parallelise complex heap-manipulating programs without verifying every aspect of their behaviour.

The second method proposes specification-directed synthesis: given a sequential program, we extract a rich, stateful specification compactly summarising program behaviour, and use that specification for parallelisation. At the heart of the method is a learning algorithm which combines dynamic and static analysis. In particular, dynamic symbolic execution and the computational learning technique grammar induction are used to conjecture input-output specifications, and counterexample-guided abstraction refinement to confirm or refute the equivalence between the conjectured specification and the original program. Once equivalence checking succeeds, from the inferred specifications we synthesise code that executes speculatively in parallel — enabling automated parallelisation of irregular loops that are not necessary polyhedral, disjoint or with a static pipeline structure.

UCAM-CL-TR-918

Gregory Y. Tsipenyuk:

Evaluation of decentralized email architecture and social network analysis based on email attachment sharing

March 2018, 153 pages, PDF
PhD thesis (Sidney Sussex College, August 2017)

Abstract: Present day e-mail is provided by centralized services running in the cloud. The services transparently connect users behind middleboxes and provide backup, redundancy, and high availability at the expense of user privacy. In present day mobile environments, users can access and modify e-mail from multiple devices with updates reconciled on the central server. Prioritizing updates is difficult and may be undesirable. Moreover, legacy email protocols do not provide optimal e-mail synchronization and access. Recent phenomena of the Internet of Things (IoT) will see the number of interconnected devices grow to 27 billion by 2021. In the first part of my dissertation I am proposing a decentralized email architecture which takes advantage of user's a IoT devices to maintain a complete email history. This addresses the e-mail reconciliation issue and places data under user control. I replace legacy email protocols with a synchronization protocol to achieve eventual consistency of email and optimize bandwidth and energy usage. The architecture is evaluated on a Raspberry Pi computer.

There is an extensive body of research on Social Network Analysis (SNA) based on email archives. Typically, the analyzed network reflects either communication between users or a relationship between the e-mail and the information found in the e-mail's header and the body. This approach discards either all or some email attachments that cannot be converted to text; for instance, images. Yet attachments may use up to 90% of an e-mail archive size. In the second part of my dissertation I suggest extracting the network from e-mail attachments shared between users. I hypothesize that the network extracted from shared e-mail attachments might provide more insight into the social structure of the email archive. I evaluate communication and shared e-mail attachments networks by analyzing common centrality measures and classification and clustering algorithms. I further demonstrate how the analysis of the shared attachments network can be used to optimize the proposed decentralized e-mail architecture.

UCAM-CL-TR-919

Nada Amin, François-René Rideau:

Proceedings of the 2017 Scheme and Functional Programming Workshop

March 2018, 50 pages, PDF
Oxford, UK – 3 September 2017

Abstract: This report aggregates the papers presented at the eigtheenth annual Scheme and Functional Programming Workshop, hosted on September 3rd, 2017 in Oxford, UK and co-located with the twenty-second International Conference on Functional Programming.

The Scheme and Functional Programming Workshop is held every year to provide an opportunity for researchers and practitioners using Scheme and related functional programming languages like Racket, Clojure, and Lisp, to share research findings and discuss the future of the Scheme programming language.

Two full papers and three lightning talks were submitted to the workshop, and each submission was reviewed by three members of the program committee. After deliberation, all submissions were accepted to the workshop.

UCAM-CL-TR-920

Advait Sarkar:

Interactive analytical modelling

May 2018, 142 pages, PDF
PhD thesis (Emmanuel College, December 2016)

Abstract: This dissertation addresses the problem of designing tools for non-expert end-users performing analytical modelling, the activity of data analytics through machine learning. The research questions are, firstly, can tools be built for analytical modelling? Secondly, can these tools be made useful for non-experts?

Two case studies are made of building and applying machine learning models, firstly to analyse time series data, and secondly to analyse data in spreadsheets. Respectively, two prototypes are presented, Gatherminer

and BrainCel. It is shown how it is possible to visualise general tasks (e.g., prediction, validation, explanation) in these contexts, illustrating how these prototypes embody the research hypotheses, and confirmed through experimental evaluation that the prototypes do indeed have the desirable properties of facilitating analytical modelling and being accessible to non-experts.

These prototypes and their evaluations exemplify three theoretical contributions: (a) visual analytics can be viewed as an instance of end-user programming, (b) interactive machine learning can be viewed as dialogue, and (c) analytical modelling can be viewed as constructivist learning. Four principles for the design of analytical modelling systems are derived: begin the abstraction gradient at zero, abstract complex processes through heuristic automation, build expertise through iteration on multiple representations, and support dialogue through metamodels.

UCAM-CL-TR-921

Toby Moncaster:

Optimising data centre operation by removing the transport bottleneck

June 2018, 130 pages, PDF
PhD thesis (Wolfson College, February 2018)

Abstract: Data centres lie at the heart of almost every service on the Internet. Data centres are used to provide search results, to power social media, to store and index email, to host "cloud" applications, for online retail and to provide a myriad of other web services. Consequently the more efficient they can be made the better for all of us. The power of modern data centres is in combining commodity off-the-shelf server hardware and network equipment to provide what Google's Barrosso and Hölzle describe as "warehouse scale" computers.

Data centres rely on TCP, a transport protocol that was originally designed for use in the Internet. Like other such protocols, TCP has been optimised to maximise throughput, usually by filling up queues at the bottleneck. However, for most applications within a data centre network latency is more critical than throughput. Consequently the choice of transport protocol becomes a bottleneck for performance. My thesis is that the solution to this is to move away from the use of one-size-fits-all transport protocols towards ones that have been designed to reduce latency across the data centre and which can dynamically respond to the needs of the applications.

This dissertation focuses on optimising the transport layer in data centre networks. In particular I address the question of whether any single transport mechanism can be flexible enough to cater to the needs of all data centre traffic. I show that one leading protocol (DCTCP) has been heavily optimised for certain network conditions. I then explore approaches that seek to minimise latency for applications that care about it while still allowing throughput-intensive applications to receive a good level of service. My key contributions to this are Silo and Trevi.

Trevi is a novel transport system for storage traffic that utilises fountain coding to maximise throughput and minimise latency while being agnostic to drop, thus allowing storage traffic to be pushed out of the way when latency sensitive traffic is present in the network. Silo is an admission control system that is designed to give tenants of a multi-tenant data centre guaranteed low latency network performance. Both of these were developed in collaboration with others.

UCAM-CL-TR-922

Frank Stajano, Graham Rymer, Michelle Houghton:

Raising a new generation of cyber defenders

June 2018, 307 pages, PDF

Abstract: To address the skills gap in cyber security, in the 2015–16 academic year we launched two competitions for university students: the national (UK-wide) Inter-ACE and, in collaboration with MIT, the international Cambridge2Cambridge. After running the competitions for three years and growing them in several dimensions (including duration, budget, gender balance, number of participants, number of universities and number of countries), we distill our experience into a write-up of what we achieved and specific items of advice for our successors, discussing problems encountered and possible solutions in a variety of areas and suggesting future directions for further growth.

UCAM-CL-TR-923

Sam Ainsworth:

Prefetching for complex memory access patterns

July 2018, 146 pages, PDF
PhD thesis (Churchill College, February 2018)

Abstract: Modern-day workloads, particularly those in big data, are heavily memory-latency bound. This is because of both irregular memory accesses, which have no discernible pattern in their memory addresses, and large data sets that cannot fit in any cache. However, this need not be a barrier to high performance. With some data structure knowledge it is typically possible to bring data into the fast on-chip memory caches early, so that it is already available by the time it needs to be accessed. This thesis makes three contributions. I first contribute an automated software prefetching compiler technique to insert high-performance prefetches into

program code to bring data into the cache early, achieving 1.3x geometric mean speedup on the most complex processors, and 2.7x on the simplest. I also provide an analysis of when and why this is likely to be successful, which data structures to target, and how to schedule software prefetches well.

Then I introduce a hardware solution, the configurable graph prefetcher. This uses the example of breadth-first search on graph workloads to motivate how a hardware prefetcher armed with data-structure knowledge can avoid the instruction overheads, inflexibility and limited latency tolerance of software prefetching. The configurable graph prefetcher sits at the L1 cache and observes memory accesses, which can be configured by a programmer to be aware of a limited number of different data access patterns, achieving 2.3x geometric mean speedup on graph workloads on an out-of-order core. My final contribution extends the hardware used for the configurable graph prefetcher to make an event-triggered programmable prefetcher, using a set of a set of very small micro-controller-sized programmable prefetch units (PPUs) to cover a wide set of workloads. I do this by developing a highly parallel programming model that can be used to issue prefetches, thus allowing high-throughput prefetching with low power and area overheads of only around 3%, and a 3x geometric mean speedup for a variety of memory-bound applications. To facilitate its use, I then develop compiler techniques to help automate the process of targeting the programmable prefetcher. These provide a variety of tradeoffs from easiest to use to best performance.

UCAM-CL-TR-924

George Neville-Neil, Jonathan Anderson, Graeme Jenkinson, Brian Kidney, Domagoj Stolfa, Arun Thomas, Robert N. M. Watson:

OpenDTrace Specification version 1.0

August 2018, 235 pages, PDF

Abstract: OpenDTrace is a dynamic tracing facility offering full-system instrumentation, a high degree of flexibility, and portable semantics across a range of operating systems. Originally designed and implemented by Sun Microsystems (now Oracle), user-facing aspects of OpenDTrace, such as the D language and command-line tools, are well defined and documented. However, OpenDTraces internal formats the DTrace Intermediate Format (DIF), DTrace Object Format (DOF) and Compact C Trace Format (CTF) have primarily been documented through source-code comments rather than a structured specification. This technical report specifies these formats in order to better support the development of more comprehensive tests, new underlying execution substrates (such as just-in-time compilation), and future extensions. We not only cover the data structures present in OpenDTrace but also include a complete reference of all the low level instructions that are used by the byte code interpreter, all the built in global variables and subroutines. Our goal with this report is to provide not only a list of what is present in the code at any point in time, the what, but also explanations of how the system works as a whole, the how, and motivations for various design decisions that have been made along the way, the why. Throughout this report we use the name OpenDTrace to refer to the open-source project but retain the name DTrace when referring to data structures such as the DTrace Intermediate Format. OpenDTrace builds upon the foundations of the original DTrace code but provides new features, which were not present in the original. This document acts as a single source of truth for the current state of OpenDTrace as it is currently implemented and deployed.

UCAM-CL-TR-925

Ranjan Pal, Jon Crowcroft, Abhishek Kumar, Pan Hui, Hamed Haddadi, Swades De, Irene Ng, Sasu Tarkoma, Richard Mortier:

Privacy markets in the Apps and IoT age

September 2018, 45 pages, PDF

Abstract: In the era of the mobile apps and IoT, huge quantities of data about individuals and their activities offer a wave of opportunities for economic and societal value creation. However, the current personal data ecosystem is fragmented and inefficient. On one hand, end-users are not able to control access (either technologically, by policy, or psychologically) to their personal data which results in issues related to privacy, personal data ownership, transparency, and value distribution. On the other hand, this puts the burden of managing and protecting user data on apps and ad-driven entities (e.g., an ad-network) at a cost of trust and regulatory accountability. In such a context, data holders (e.g., apps) may take advantage of the individuals' inability to fully comprehend and anticipate the potential uses of their private information with detrimental effects for aggregate social welfare. In this paper, we investigate the problem of the existence and design of efficient ecosystems (modeled as markets in this paper) that aim to achieve a maximum social welfare state among competing data holders by preserving the heterogeneous privacy preservation constraints up to certain compromise levels, induced by their clients, and at the same time satisfying requirements of agencies (e.g., advertisers) that collect and trade client data for the purpose of targeted advertising, assuming the potential practical inevitability of some amount inappropriate data leakage on behalf of the data holders. Using concepts from supply-function economics, we propose the first mathematically rigorous and provably optimal

privacy market design paradigm that always results in unique equilibrium (i.e, stable) market states that can be either economically efficient or inefficient, depending on whether privacy trading markets are monopolistic or oligopolistic in nature. Subsequently, we characterize in closed form, the efficiency gap (if any) at market equilibrium.

UCAM-CL-TR-926

Ranjan Pal, Konstantinos Psounis, Abhishek Kumar, Jon Crowcroft, Pan Hui, Leana Golubchik, John Kelly, Aritra Chatterjee, Sasu Tarkoma:

Are cyber-blackouts in service networks likely?: implications for cyber risk management

October 2018, 32 pages, PDF

Abstract: Service liability interconnections among networked IT and IoT driven service organizations create potential channels for cascading service disruptions due to modern cybercrimes such as DDoS, APT, and ransomware attacks. The very recent Mirai DDoS and WannaCry ransomware attacks serve as famous examples of cyber-incidents that have caused catastrophic service disruptions worth billions of dollars across organizations around the globe. A natural question that arises in this context is "what is the likelihood of a cyber-blackout?", where the latter term is defined as: "the probability that all (or a major subset of) organizations in a service chain become dysfunctional in a certain manner due to a cyber-attack at some or all points in the chain".

The answer to this question has major implications to risk management businesses such as cyber-insurance when it comes to designing policies by risk-averse insurers for providing coverage to clients in the aftermath of such catastrophic network events. In this paper, we investigate this question in general as a function of service chain networks and different loss distribution types. We show somewhat surprisingly (and discuss potential practical implications) that following a cyber-attack, the probability of a cyber-blackout and the increase in total service-related monetary losses across all organizations, due to the effect of (a) network interconnections, and (b) a wide range of loss distributions, are mostly very small, regardless of the network structure – the primary rationale behind the results being attributed to degrees of heterogeneity in wealth base among organizations, and Increasing Failure Rate (IFR) property of loss distributions.

UCAM-CL-TR-927

Robert N. M. Watson, Peter G. Neumann, Jonathan Woodruff, Michael Roe, Hesham Almatary, Jonathan Anderson, John Baldwin, David Chisnall, Brooks Davis, Nathaniel Wesley Filardo, Alexandre Joannou, Ben Laurie, A. Theodore Markettos, Simon W. Moore, Steven J. Murdoch, Kyndylan Nienhuis, Robert Norton, Alex Richardson, Peter Rugg, Peter Sewell, Stacey Son, Hongyan Xia:

Capability Hardware Enhanced RISC Instructions: CHERI Instruction-Set Architecture (Version 7)

June 2019, 496 pages, PDF

Abstract: This technical report describes CHERI ISAv7, the seventh version of the Capability Hardware Enhanced RISC Instructions (CHERI) Instruction-Set Architecture (ISA) being developed by SRI International and the University of Cambridge. This design captures nine years of research, development, experimentation, refinement, formal analysis, and validation through hardware and software implementation. CHERI ISAv7 is a substantial enhancement to prior ISA versions. We differentiate an architecture-neutral protection model vs. architecture-specific instantiations in 64-bit MIPS, 64-bit RISC-V, and x86-64. We have defined a new CHERI Concentrate compression model. CHERI-RISC-V is more substantially elaborated. A new compartment-ID register assists in resisting microarchitectural side-channel attacks. Experimental features include linear capabilities, capability coloring, temporal memory safety, and 64-bit capabilities for 32-bit architectures.

CHERI is a hybrid capability-system architecture that adds new capability-system primitives to commodity 64-bit RISC ISAs, enabling software to efficiently implement fine-grained memory protection and scalable software compartmentalization. Design goals include incremental adoptability within current ISAs and software stacks, low performance overhead for memory protection, significant performance improvements for software compartmentalization, formal grounding, and programmer-friendly underpinnings. We have focused on providing strong, non-probabilistic, efficient architectural foundations for the principles of least privilege and intentional use in the execution of software at multiple levels of abstraction, preventing and mitigating vulnerabilities.

The CHERI system architecture purposefully addresses known performance and robustness gaps in

commodity ISAs that hinder the adoption of more secure programming models centered around the principle of least privilege. To this end, CHERI blends traditional paged virtual memory with an in-address-space capability model that includes capability registers, capability instructions, and tagged memory. CHERI builds on the C-language fat-pointer literature: its capabilities can describe fine-grained regions of memory, and can be substituted for data or code pointers in generated code, protecting data and also improving control-flow robustness. Strong capability integrity and monotonicity properties allow the CHERI model to express a variety of protection properties, from enforcing valid C-language pointer provenance and bounds checking to implementing the isolation and controlled communication structures required for software compartmentalization.

CHERI's hybrid capability-system approach, inspired by the Capsicum security model, allows incremental adoption of capability-oriented design: software implementations that are more robust and resilient can be deployed where they are most needed, while leaving less critical software largely unmodified, but nevertheless suitably constrained to be incapable of having adverse effects. Potential deployment scenarios include low-level software Trusted Computing Bases (TCBs) such as separation kernels, hypervisors, and operating-system kernels, as well as userspace TCBs such as language runtimes and web browsers. We also see potential early-use scenarios around particularly high-risk software libraries (such as data compression, protocol parsing, and image processing), which are concentrations of both complex and historically vulnerability-prone code exposed to untrustworthy data sources, while leaving containing applications unchanged.

UCAM-CL-TR-928

Noa Zilberman, Łukasz Dudziak, Matthew Jadczak, Thomas Parks, Alessandro Rietmann, Vadim Safronov, Daniel Zuo:

Cut-through network switches: architecture, design and implementation

November 2018, 18 pages, PDF

Abstract: Cut-through switches are increasingly used within data-centres and high-performance networked systems. Despite their popularity, little is known of the architecture of cut-through switches used in today's networks. In this paper we introduce three different cut-through switch architectures, designed over the NetFPGA platform. Beyond exploring architectural and design considerations, we compare and contrast the different architectures, providing insights into real-world switch design. Last, we provide an evaluation of one successfully implemented cut-through switch design, providing constant switching latency regardless of packet size and cross-traffic, without compromising throughput.

UCAM-CL-TR-929

Khaled Baqer:

Resilient payment systems

November 2018, 115 pages, PDF
PhD thesis (St. Edmund's College, October 2018)

Abstract: There have been decades of attempts to evolve or revolutionise the traditional financial system, but not all such efforts have been transformative or even successful. From Chaum's proposals in the 1980s for private payment systems to micropayments, previous attempts failed to take off for a variety of reasons, including non-existing markets, or issues pertaining to usability, scalability and performance, resilience against failure, and complexity of protocols.

Towards creating more resilient payment systems, we investigated issues related to security engineering in general, and payment systems in particular. We identified that network coverage, central points of failure, and attacks may cripple system performance. The premise of our research is that offline capabilities are required to produce resilience in critical systems.

We focus on issues related to network problems and attacks, system resilience, and scalability by introducing the ability to process payments offline without relying on the availability of network coverage; a lack of network coverage renders some payment services unusable for their customers. Decentralising payment verification, and outsourcing some operations to users, alleviates the burden of contacting centralised systems to process every transaction. Our secondary goal is to minimise the cost of providing payment systems, so providers can cut transaction fees. Moreover, by decentralising payment verification that can be performed offline, we increase system resilience, and seamlessly maintain offline operations until a system is back online. We also use tamper-resistant hardware to tackle usability issues, by minimising cognitive overhead and helping users to correctly handle critical data, minimising the risks of data theft and tampering.

We apply our research towards extending financial inclusion efforts, since the issues discussed above must be solved to extend mobile payments to the poorest demographics. More research is needed to integrate online payments, offline payments, and delay-tolerant networking. This research extends and enhances not only payment systems, but other electronically-enabled services from pay-as-you-go solar panels to agricultural subsidies and payments from aid donors. We hope that this thesis is helpful for researchers, protocol designers, and policy makers interested in creating resilient payment systems by assisting them in financial inclusion efforts.

UCAM-CL-TR-930

Lucian Carata:

Provenance-based computing

December 2018, 132 pages, PDF
PhD thesis (Wolfson College, July 2016)

Abstract: Relying on computing systems that become increasingly complex is difficult: with many factors potentially affecting the result of a computation or its properties, understanding where problems appear and fixing them is a challenging proposition. Typically, the process of finding solutions is driven by trial and error or by experience-based insights.

In this dissertation, I examine the idea of using provenance metadata (the set of elements that have contributed to the existence of a piece of data, together with their relationships) instead. I show that considering provenance a primitive of computation enables the exploration of system behaviour, targeting both retrospective analysis (root cause analysis, performance tuning) and hypothetical scenarios (what-if questions). In this context, provenance can be used as part of feedback loops, with a double purpose: building software that is able to adapt for meeting certain quality and performance targets (semi-automated tuning) and enabling human operators to exert high-level runtime control with limited previous knowledge of a system's internal architecture.

My contributions towards this goal are threefold: providing low-level mechanisms for meaningful provenance collection considering OS-level resource multiplexing, proving that such provenance data can be used in inferences about application behaviour and generalising this to a set of primitives necessary for fine-grained provenance disclosure in a wider context.

To derive such primitives in a bottom-up manner, I first present Resourceful, a framework that enables capturing OS-level measurements in the context of application activities. It is the contextualisation that allows tying the measurements to provenance in a meaningful way, and I look at a number of use-cases in understanding application performance. This also provides a good setup for evaluating the impact and overheads of fine-grained provenance collection.

I then show that the collected data enables new ways of understanding performance variation by attributing it to specific components within a system. The resulting set of tools, Soroban, gives developers and operation engineers a principled way of examining the impact of various configuration, OS and virtualization parameters on application behaviour.

Finally, I consider how this supports the idea that provenance should be disclosed at application level and discuss why such disclosure is necessary for enabling the use of collected metadata efficiently and at a granularity which is meaningful in relation to application semantics.

UCAM-CL-TR-931

Jyothish Soman:

A Performance-efficient and practical processor error recovery framework

January 2019, 91 pages, PDF
PhD thesis (Wolfson College, July 2017)

Abstract: Continued reduction in the size of a transistor has affected the reliability of processors built using them. This is primarily due to factors such as inaccuracies while manufacturing, as well as non-ideal operating conditions, causing transistors to slow down consistently, eventually leading to permanent breakdown and erroneous operation of the processor. Permanent transistor breakdown, or faults, can occur at any point in time in the processor's lifetime. Errors are the discrepancies in the output of faulty circuits. This dissertation shows that the components containing faults can continue operating if the errors caused by them are within certain bounds. Further, the lifetime of a processor can be increased by adding supportive structures that start working once the processor develops these hard errors.

This dissertation has three major contributions, namely REPAIR, FaultSim and PreFix. REPAIR is a fault tolerant system with minimal changes to the processor design. It uses an external Instruction Re-execution Unit (IRU) to perform operations, which the faulty processor might have erroneously executed. Instructions that are found to use faulty hardware are then re-executed on the IRU. REPAIR shows that the performance overhead of such targeted re-execution is low for a limited number of faults.

FaultSim is a fast fault-simulator capable of simulating large circuits at the transistor level. It is developed in this dissertation to understand the effect of faults on different circuits. It performs digital logic based simulations, trading off analogue accuracy with speed, while still being able to support most fault models. A 32-bit addition takes under 15 micro-seconds, while simulating more than 1500 transistors. It can also be integrated into an architectural simulator, which added a performance overhead of 10 to 26 percent to a simulation. The results obtained show that single faults cause an error in an adder in less than 10 percent of the inputs.

PreFix brings together the fault models created using FaultSim and the design directions found using REPAIR. PreFix performs re-execution of instructions on a remote core, which pick up instructions to execute using a global instruction buffer. Error prediction and detection are used to reduce the number of re-executed instructions. PreFix has an area overhead of 3.5 percent in the setup used, and the performance overhead is within 5 percent of a fault-free case. This dissertation shows that faults in processors can be tolerated without explicitly switching off any component, and minimal redundancy is sufficient to achieve the same.

UCAM-CL-TR-932

Brooks Davis, Robert N. M. Watson, Alexander Richardson, Peter G. Neumann, Simon W. Moore, John Baldwin, David Chisnall, Jessica Clarke, Nathaniel Wesley Filardo, Khilan Gudka, Alexandre Joannou, Ben Laurie, A. Theodore Markettos, J. Edward Maste, Alfredo Mazzinghi, Edward Tomasz Napierala, Robert M. Norton, Michael Roe, Peter Sewell, Stacey Son, Jonathan Woodruff:

CheriABI: Enforcing valid pointer provenance and minimizing pointer privilege in the POSIX C run-time environment

April 2019, 40 pages, PDF

Abstract: The CHERI architecture allows pointers to be implemented as capabilities (rather than integer virtual addresses) in a manner that is compatible with, and strengthens, the semantics of the C language. In addition to the spatial protections offered by conventional fat pointers, CHERI capabilities offer strong integrity, enforced provenance validity, and access monotonicity.

The stronger guarantees of these architectural capabilities must be reconciled with the real-world behavior of operating systems, run-time environments, and applications. When the process model, user-kernel interactions, dynamic linking, and memory management are all considered, we observe that simple derivation of architectural capabilities is insufficient to describe appropriate access to memory. We bridge this conceptual gap with a notional abstract capability that describes the accesses that should be allowed at a given point in execution, whether in the kernel or userspace.

To investigate this notion at scale, we describe the first adaptation of a full C-language operating system (FreeBSD) with an enterprise database (PostgreSQL) for complete spatial and referential memory safety. We show that awareness of abstract capabilities, coupled with CHERI architectural capabilities, can provide more complete protection, strong compatibility, and acceptable performance overhead compared with the pre-CHERI baseline and software-only approaches. Our observations also have potentially significant implications for other mitigation techniques.

UCAM-CL-TR-933

Noa Zilberman:

An Evaluation of NDP performance

January 2019, 19 pages, PDF

Abstract: NDP is a novel data centre transport architecture that claims to achieve near-optimal completion times for short transfers and high flow throughput in a wide range of scenarios. This work presents a performance evaluation of NDP, both on the simulation and the hardware level. We show that NDP's implementation achieves lower switch throughput than simple TCP, and that the simulated performance is highly dependent on the selected parameters.

UCAM-CL-TR-934

Colin L. Rothwell:

Exploitation from malicious PCI Express peripherals

February 2019, 108 pages, PDF

Abstract: The thesis of this dissertation is that, despite widespread belief in the security community, systems are still vulnerable to attacks from malicious peripherals delivered over the PCI Express (PCIe) protocol. Malicious peripherals can be plugged directly into internal PCIe slots, or connected via an external Thunderbolt connection.

To prove this thesis, we designed and built a new PCIe attack platform. We discovered that a simple platform was insufficient to carry out complex attacks, so created the first PCIe attack platform that runs a full, conventional OS. To allows us to conduct attacks against higher-level OS functionality built on PCIe, we made the attack platform emulate in detail the behaviour of an Intel 82574L Network Interface Controller (NIC), by using a device model extracted from the QEMU emulator.

We discovered a number of vulnerabilities in the PCIe protocol itself, and with the way that the defence mechanisms it provides are used by modern OSs. The principal defence mechanism provided is the Input/Output Memory Management Unit (IOMMU). This remaps the address space used by peripherals in 4KiB chunks, and can prevent access to areas of address space that a peripheral should not be able to access. We found that, contrary to belief in the security community, the IOMMUs in modern systems were not designed to protect against attacks from malicious peripherals, but to allow virtual machines direct access to real hardware.

We discovered that use of the IOMMU is patchy even in modern operating systems. Windows effectively does not use the IOMMU at all; macOS opens windows that are shared by all devices; Linux and FreeBSD map windows into host memory separately for each device, but only if poorly documented boot flags are used. These OSs make no effort to ensure that only data that should be visible to the devices is in the mapped windows.

We created novel attacks that subverted control flow and read private data against systems running macOS,

Linux and FreeBSD with the highest level of relevant protection enabled. These represent the first use of the relevant exploits in each case.

In the final part of this thesis, we evaluate the suitability of a number of proposed general purpose and specific mitigations against DMA attacks, and make a number of recommendations about future directions in IOMMU software and hardware.

UCAM-CL-TR-935

Heidi Howard:

Distributed consensus revised

April 2019, 151 pages, PDF
PhD thesis (Pembroke College, September 2018)

Abstract: We depend upon distributed systems in every aspect of life. Distributed consensus, the ability to reach agreement in the face of failures and asynchrony, is a fundamental and powerful primitive for constructing reliable distributed systems from unreliable components.

For over two decades, the Paxos algorithm has been synonymous with distributed consensus. Paxos is widely deployed in production systems, yet it is poorly understood and it proves to be heavyweight, unscalable and unreliable in practice. As such, Paxos has been the subject of extensive research to better understand the algorithm, to optimise its performance and to mitigate its limitations.

In this thesis, we re-examine the foundations of how Paxos solves distributed consensus. Our hypothesis is that these limitations are not inherent to the problem of consensus but instead specific to the approach of Paxos. The surprising result of our analysis is a substantial weakening to the requirements of this widely studied algorithm. Building on this insight, we are able to prove an extensive generalisation over the Paxos algorithm.

Our revised understanding of distributed consensus enables us to construct a diverse family of algorithms for solving consensus; covering classical as well as novel algorithms to reach consensus where it was previously thought impossible. We will explore the wide reaching implications of this new understanding, ranging from pragmatic optimisations to production systems to fundamentally novel approaches to consensus, which achieve new tradeoffs in performance, scalability and reliability.

UCAM-CL-TR-936

Alexandre J. P. Joannou:

High-performance memory safety: optimizing the CHERI capability machine

May 2019, 132 pages, PDF
PhD thesis (Peterhouse College, September 2017)

Abstract: This work presents optimizations for modern capability machines and specifically for the CHERI architecture, a 64-bit MIPS instruction set extension for security, supporting fine-grained memory protection through hardware enforced capabilities.

The original CHERI model uses 256-bit capabilities to carry information required for various checks helping to enforce memory safety, leading to increased memory bandwidth requirements and cache pressure when using CHERI capabilities in place of conventional 64-bit pointers. In order to mitigate this cost, I present two new 128-bit CHERI capability formats, using different compression techniques, while preserving C-language compatibility lacking in previous pointer compression schemes. I explore the trade-offs introduced by these new formats over the 256-bit format. I produce an implementation in the L3 ISA modelling language, collaborate on the hardware implementation, and provide an evaluation of the mechanism.

Another cost related to CHERI capabilities is the memory traffic increase due to capability-validity tags: to provide unforgeable capabilities, CHERI uses a tagged memory that preserves validity tags for every 256-bit memory word in a shadow space inaccessible to software. The CHERI hardware implementation of this shadow space uses a capability-validity-tag table in memory and caches it at the end of the cache hierarchy. To efficiently implement such a shadow space and improve on CHERI's current approach, I use sparse data structures in a hierarchical tag-cache that filters unnecessary memory accesses. I present an in-depth study of this technique through a Python implementation of the hierarchical tag-cache, and also provide a hardware implementation and evaluation. I find that validity-tag traffic is reduced for all applications and scales with tag use. For legacy applications that do not use tags, there is near zero overhead.

Removing these costs through the use of the proposed optimizations makes the CHERI architecture more affordable and appealing for industrial adoption.

UCAM-CL-TR-937

Diana Andreea Popescu:

Latency-driven performance in data centres

June 2019, 190 pages, PDF
PhD thesis (Churchill College, December 2018)

Abstract: Data centre based cloud computing has revolutionised the way businesses use computing infrastructure. Instead of building their own data centres, companies rent computing resources and deploy their applications on cloud hardware. Providing customers with well-defined application performance guarantees is of paramount importance to ensure transparency and to build a lasting collaboration between users and cloud operators. A user's application performance is subject

to the constraints of the resources it has been allocated and to the impact of the network conditions in the data centre.

In this dissertation, I argue that application performance in data centres can be improved through cluster scheduling of applications informed by predictions of application performance for given network latency, and measurements of current network latency in data centres between hosts. Firstly, I show how to use the Precision Time Protocol (PTP), through an open-source software implementation PTPd, to measure network latency and packet loss in data centres. I propose PTPmesh, which uses PTPd, as a cloud network monitoring tool for tenants. Furthermore, I conduct a measurement study using PTPmesh in different cloud providers, finding that network latency variability in data centres is still common. Normal latency values in data centres are in the order of tens or hundreds of microseconds, while unexpected events, such as network congestion or packet loss, can lead to latency spikes in the order of milliseconds. Secondly, I show that network latency matters for certain distributed applications even in small amounts of tens or hundreds of microseconds, significantly reducing their performance. I propose a methodology to determine the impact of network latency on distributed applications performance by injecting artificial delay into the network of an experimental setup. Based on the experimental results, I build functions that predict the performance of an application for a given network latency. Given the network latency variability observed in data centres, applications' performance is determined by their placement within the data centre. Thirdly, I propose latency-driven, application performance-aware, cluster scheduling as a way to provide performance guarantees to applications. I introduce NoMora, a cluster scheduling architecture that leverages the predictions of application performance dependent upon network latency combined with dynamic network latency measurements taken between pairs of hosts in data centres to place applications. Moreover, I show that NoMora improves application performance by choosing better placements than other scheduling policies.

UCAM-CL-TR-938

Christopher Bryant:

Automatic annotation of error types for grammatical error correction

June 2019, 138 pages, PDF
PhD thesis (Churchill College, December 2018)

Abstract: Grammatical Error Correction (GEC) is the task of automatically detecting and correcting grammatical errors in text. Although previous work has focused on developing systems that target specific error types, the current state of the art uses machine translation to correct all error types simultaneously. A significant disadvantage of this approach is that machine translation does not produce annotated output and so error type information is lost. This means we can only evaluate a system in terms of overall performance and cannot carry out a more detailed analysis of different aspects of system performance.

In this thesis, I develop a system to automatically annotate parallel original and corrected sentence pairs with explicit edits and error types. In particular, I first extend the Damerau-Levenshtein alignment algorithm to make use of linguistic information when aligning parallel sentences, and supplement this alignment with a set of merging rules to handle multi-token edits. The output from this algorithm surpasses other edit extraction approaches in terms of approximating human edit annotations and is the current state of the art. Having extracted the edits, I next classify them according to a new rule-based error type framework that depends only on automatically obtained linguistic properties of the data, such as part-of-speech tags. This framework was inspired by existing frameworks, and human judges rated the appropriateness of the predicted error types as Good (85%) or Acceptable (10%) in a random sample of 200 edits. The whole system is called the ERRor ANnotation Toolkit (ERRANT) and is the first toolkit capable of automatically annotating parallel sentences with error types.

I demonstrate the value of ERRANT by applying it to the system output produced by the participants of the CoNLL-2014 shared task, and carry out a detailed error type analysis of system performance for the first time. I also develop a simple language model based approach to GEC, that does not require annotated training data, and show how it can be improved using ERRANT error types.

UCAM-CL-TR-939

Christine Guo Yu:

Effects of timing on users' perceived control when interacting with intelligent systems

August 2019, 284 pages, PDF
PhD thesis (Gonville & Caius Gollege, May 2018)

Abstract: This research relates to the usability of mixed-initiative interaction systems, in which actions can be initiated either through a choice by the user or through intelligent decisions taken by the system. The key issue addressed here is how to preserve the user's perceived control ('sense of agency') when the control of the interaction is being transferred between the system and the user in a back-and-forth manner.

Previous research in social psychology and cognitive neuroscience suggests timing is a factor that can influence perceived control in such back-and-forth interactions. This dissertation explores the hypothesis

that in mixed-initiative interaction, a predictable interaction rhythm can preserve the user's sense of control and enhance their experience during a task (e.g. higher confidence in task performance, stronger temporal alignment, lower perceived levels of stress and effort), whereas irregular interaction timing can have the opposite effect. Three controlled experiments compare alternative rhythmic strategies when users interact with simple visual stimuli, simple auditory stimuli, and a more realistic assisted text labelling task. The results of all three experiments support the hypothesis that a predictable interaction rhythm is beneficial in a range of interaction modalities and applications.

This research contributes to the field of human-computer interaction (HCI) in four ways. Firstly, it builds novel connections between existing theories in cognitive neuroscience, social psychology and HCI, highlighting how rhythmic temporal structures can be beneficial to the user's experience: particularly, their sense of control. Secondly, it establishes timing as a crucial design resource for mixed-initiative interaction, and provides empirical evidence of how the user's perceived control and other task experiences (such as reported levels of confidence, stress and effort) can be influenced by the manipulation of timing. Thirdly, it provides quantitative measures for the user's entrainment behaviours that are applicable to a wide range of interaction timescales. Lastly, it contextualises the design of timing in a realistic application scenario and offers insights to the design of general end-user automation and decision support tools.

UCAM-CL-TR-940

Kyndylan Nienhuis, Alexandre Joannou, Anthony Fox, Michael Roe, Thomas Bauereiss, Brian Campbell, Matthew Naylor, Robert M. Norton, Simon W. Moore, Peter G. Neumann, Ian Stark, Robert N. M. Watson, Peter Sewell:

Rigorous engineering for hardware security: formal modelling and proof in the CHERI design and implementation process

September 2019, 38 pages, PDF

Abstract: The root causes of many security vulnerabilities include a pernicious combination of two problems, often regarded as inescapable aspects of computing. First, the protection mechanisms provided by the mainstream processor architecture and C/C++ language abstractions, dating back to the 1970s and before, provide only coarse-grain virtual-memory-based protection. Second, mainstream system engineering relies almost exclusively on test-and-debug methods, with (at best) prose specifications. These methods have historically sufficed commercially for much of the computer industry, but they fail to prevent large numbers of exploitable bugs, and the security problems that this causes are becoming ever more acute.

In this paper we show how more rigorous engineering methods can be applied to the development of a new security-enhanced processor architecture, with its accompanying hardware implementation and software stack. We use formal models of the complete instruction-set architecture (ISA) at the heart of the design and engineering process, both in lightweight ways that support and improve normal engineering practice – as documentation, in emulators used as a test oracle for hardware and for running software, and for test generation – and for formal verification. We formalise key intended security properties of the design, and establish that these hold with mechanised proof. This is for the same complete ISA models (complete enough to boot operating systems), without idealisation.

We do this for CHERI, an architecture with hardware capabilities that supports fine-grained memory protection and scalable secure compartmentalisation, while offering a smooth adoption path for existing software. CHERI is a maturing research architecture, developed since 2010, with work now underway to explore its possible adoption in mass-market commercial processors. The rigorous engineering work described here has been an integral part of its development to date, enabling more rapid and confident experimentation, and boosting confidence in the design.

UCAM-CL-TR-941

Robert N. M. Watson, Simon W. Moore, Peter Sewell, Peter G. Neumann:

An Introduction to CHERI

September 2019, 43 pages, PDF

Abstract: CHERI (Capability Hardware Enhanced RISC Instructions) extends conventional processor Instruction-Set Architectures (ISAs) with architectural capabilities to enable fine-grained memory protection and highly scalable software compartmentalization. CHERI's hybrid capability-system approach allows architectural capabilities to be integrated cleanly with contemporary RISC architectures and microarchitectures, as well as with MMU-based C/C++-language software stacks.

CHERI's capabilities are unforgeable tokens of authority, which can be used to implement both explicit pointers (those declared in the language) and implied pointers (those used by the runtime and generated code) in C and C++. When used for C/C++ memory protection, CHERI directly mitigates a broad range of known vulnerability types and exploit techniques. Support for more scalable software compartmentalization

facilitates software mitigation techniques such as sandboxing, which also defend against future (currently unknown) vulnerability classes and exploit techniques.

We have developed, evaluated, and demonstrated this approach through hardware-software prototypes, including multiple CPU prototypes, and a full software stack. This stack includes an adapted version of the Clang/LLVM compiler suite with support for capability-based C/C++, and a full UNIX-style OS (CheriBSD, based on FreeBSD) implementing spatial, referential, and (currently for userspace) non-stack temporal memory safety. Formal modeling and verification allow us to make strong claims about the security properties of CHERI-enabled architectures.

This report is a high-level introduction to CHERI. The report describes our architectural approach, CHERI's key microarchitectural implications, our approach to formal modeling and proof, the CHERI software model, our software-stack prototypes, further reading, and potential areas of future research.

UCAM-CL-TR-942

Alexander Kuhnle:

Evaluating visually grounded language capabilities using microworlds

January 2020, 142 pages, PDF
PhD thesis (Queens' College, August 2019)

Abstract: Deep learning has had a transformative impact on computer vision and natural language processing. As a result, recent years have seen the introduction of more ambitious holistic understanding tasks, comprising a broad set of reasoning abilities. Datasets in this context typically act not just as application-focused benchmark, but also as basis to examine higher-level model capabilities. This thesis argues that emerging issues related to dataset quality, experimental practice and learned model behaviour are symptoms of the inappropriate use of benchmark datasets for capability-focused assessment. To address this deficiency, a new evaluation methodology is proposed here, which specifically targets in-depth investigation of model performance based on configurable data simulators. This focus on analysing system behaviour is complementary to the use of monolithic datasets as application-focused comparative benchmarks.

Visual question answering is an example of a modern holistic understanding task, unifying a range of abilities around visually grounded language understanding in a single problem statement. It has also been an early example for which some of the aforementioned issues were identified. To illustrate the new evaluation approach, this thesis introduces ShapeWorld, a diagnostic data generation framework. Its design is guided by the goal to provide a configurable and extensible testbed for the domain of visually grounded language understanding. Based on ShapeWorld data, the strengths and weaknesses of various state-of-the-art visual question answering models are analysed and compared in detail, with respect to their ability to correctly handle statements involving, for instance, spatial relations or numbers. Finally, three case studies illustrate the versatility of this approach and the ShapeWorld generation framework: an investigation of multi-task and curriculum learning, a replication of a psycholinguistic study for deep learning models, and an exploration of a new approach to assess generative tasks like image captioning.

UCAM-CL-TR-943

Mathew P. Grosvenor:

Latency-First datacenter network scheduling

January 2020, 310 pages, PDF
PhD thesis (April 2017)

Abstract: Every day we take for granted that, with the click of a mouse or a tap on a touchscreen, we can instantly access the Internet to globally exchange information, finances and physical goods. The computational machinery behind the Internet is found in datacenters scattered all over the globe. Each datacenter contains many tens of thousands of computers connected together through a datacenter network. Like the Internet, datacenter networks suffer from network interference. Network interference occurs when congestion caused by some applications, delays or interferes with traffic from other applications. Network interference makes it difficult to predict network latency: the time that any given packet will take to traverse the network. The lack of predictability makes it difficult to build fast, efficient, and responsive datacenter applications.

In this dissertation I address the problem of network interference in datacenter networks. I do so primarily by exploiting network scheduling techniques. Network scheduling techniques were previously developed to provide predictability in the Internet. However, they were complex to deploy and administer because few assumptions could be made about the network. Unlike the Internet, datacenter networks are administered by a single entity, and have well known physical properties, that rarely change.

The thesis of this dissertation is that it is both possible and practical to resolve network interference in datacenter networks by using network scheduling techniques. I show that it is possible to resolve network interference by deriving a simple, and general, network scheduler and traffic regulator. I take the novel step of basing my scheduler design on a simple, but realistic, model of a datacenter switch. By regulating the

flow of traffic into each switch, I show that it is possible to bound latency across the network. I develop the leaky token bucket regulator to perform this function. I show that my network scheduler design is practical by implementing a simple, and immediately deployable, system called QJUMP. QJUMP is thoroughly evaluated and demonstrated to resolve network interference between datacenter applications in both simulation and on a physical test-bed. I further show that QJUMP can be extended and improved upon in a variety of ways. I therefore conclude that it is both possible and practical to control network interference in datacenter networks using network scheduling techniques.

UCAM-CL-TR-944

Alexander Vetterl:

Honeypots in the age of universal attacks and the Internet of Things

February 2020, 115 pages, PDF
PhD thesis (Churchill College, November 2019)

Abstract: Today's Internet connects billions of physical devices. These devices are often immature and insecure, and share common vulnerabilities. The predominant form of attacks relies on recent advances in Internet-wide scanning and device discovery. The speed at which (vulnerable) devices can be discovered, and the device monoculture, mean that a single exploit, potentially trivial, can affect millions of devices across brands and continents.

In an attempt to detect and profile the growing threat of autonomous and Internet-scale attacks against the Internet of Things, we revisit honeypots, resources that appear to be legitimate systems. We show that this endeavour was previously limited by a fundamentally flawed generation of honeypots and associated misconceptions.

We show with two one-year-long studies that the display of warning messages has no deterrent effect in an attacked computer system. Previous research assumed that they would measure individual behaviour, but we find that the number of human attackers is orders of magnitude lower than previously assumed.

Turning to the current generation of low- and medium-interaction honeypots, we demonstrate that their architecture is fatally flawed. The use of off-the-shelf libraries to provide the transport layer means that the protocols are implemented subtly differently from the systems being impersonated. We developed a generic technique which can find any such honeypot at Internet scale with just one packet for an established TCP connection.

We then applied our technique and conducted several Internet-wide scans over a one-year period. By logging in to two SSH honeypots and sending specific commands, we not only revealed their configuration and patch status, but also found that many of them were not up to date. As we were the first to knowingly authenticate to honeypots, we provide a detailed legal analysis and an extended ethical justification for our research to show why we did not infringe computer-misuse laws.

Lastly, we present honware, a honeypot framework for rapid implementation and deployment of high-interaction honeypots. Honware automatically processes a standard firmware image and can emulate a wide range of devices without any access to the manufacturers' hardware. We believe that honware is a major contribution towards re-balancing the economics of attackers and defenders by reducing the period in which attackers can exploit vulnerabilities at Internet scale in a world of ubiquitous networked 'things'.

UCAM-CL-TR-945

Maxwell Jay Conway:

Machine learning methods for detecting structure in metabolic flow networks

March 2020, 132 pages, PDF
PhD thesis (Selwyn College, August 2018)

Abstract: Metabolic flow networks are large scale, mechanistic biological models with good predictive power. However, even when they provide good predictions, interpreting the meaning of their structure can be very difficult, especially for large networks which model entire organisms. This is an underaddressed problem in general, and the analytic techniques that exist currently are difficult to combine with experimental data. The central hypothesis of this thesis is that statistical analysis of large datasets of simulated metabolic fluxes is an effective way to gain insight into the structure of metabolic networks. These datasets can be either simulated or experimental, allowing insight on real world data while retaining the large sample sizes only easily possible via simulation. This work demonstrates that this approach can yield results in detecting structure in both a population of solutions and in the network itself.

This work begins with a taxonomy of sampling methods over metabolic networks, before introducing three case studies, of different sampling strategies. Two of these case studies represent, to my knowledge, the largest datasets of their kind, at around half a million points each. This required the creation of custom software to achieve this in a reasonable time frame, and is necessary due to the high dimensionality of the sample space.

Next, a number of techniques are described which operate on smaller datasets. These techniques, focused on pairwise comparison, show what can be achieved with these smaller datasets, and how in these cases, visualisation techniques are applicable which do not have simple analogues with larger datasets.

In the next chapter, Similarity Network Fusion is used for the first time to cluster organisms across several levels of biological organisation, resulting in the detection of discrete, quantised biological states in the underlying datasets. This quantisation effect was maintained across both real biological data and Monte-Carlo simulated data, with related underlying biological correlates, implying that this behaviour stems from the network structure itself, rather than from the genetic or regulatory mechanisms that would normally be assumed.

Finally, Hierarchical Block Matrices are used as a model of multi-level network structure, by clustering reactions using a variety of distance metrics: first standard network distance measures, then by Local Network Learning, a novel approach of measuring connection strength via the gain in predictive power of each node on its neighbourhood. The clusters uncovered using this approach are validated against pre-existing subsystem labels and found to outperform alternative techniques.

Overall this thesis represents a significant new approach to metabolic network structure detection, as both a theoretical framework and as technological tools, which can readily be expanded to cover other classes of multilayer network, an under explored datatype across a wide variety of contexts. In addition to the new techniques for metabolic network structure detection introduced, this research has proved fruitful both in its use in applied biological research and in terms of the software developed, which is experiencing substantial usage.

UCAM-CL-TR-946

Michael Schaarschmidt:

End-to-end deep reinforcement learning in computer systems

April 2020, 166 pages, PDF
PhD thesis (Sidney Sussex College, September 2019)

Abstract: The growing complexity of data processing systems has long led systems designers to imagine systems (e.g. databases, schedulers) which can self-configure and adapt based on environmental cues. In this context, reinforcement learning (RL) methods have since their inception appealed to systems developers. They promise to acquire complex decision policies from raw feedback signals. Despite their conceptual popularity, RL methods are scarcely found in real-world data processing systems. Recently, RL has seen explosive growth in interest due to high profile successes when utilising large neural networks (deep reinforcement learning). Newly emerging machine learning frameworks and powerful hardware accelerators have given rise to a plethora of new potential applications.

In this dissertation, I first argue that in order to design and execute deep RL algorithms efficiently, novel software abstractions are required which can accommodate the distinct computational patterns of communication-intensive and fast-evolving algorithms. I propose an architecture which decouples logical algorithm construction from local and distributed execution semantics. I further present RLgraph, my proof-of-concept implementation of this architecture. In RLgraph, algorithm developers can explore novel designs by constructing a high-level data flow graph through combination of logical components. This dataflow graph is independent of specific backend frameworks or notions of execution, and is only later mapped to execution semantics via a staged build process. RLgraph enables high-performing algorithm implementations while maintaining flexibility for rapid prototyping.

Second, I investigate reasons for the scarcity of RL applications in systems themselves. I argue that progress in applied RL is hindered by a lack of tools for task model design which bridge the gap between systems and algorithms, and also by missing shared standards for evaluation of model capabilities. I introduce Wield, a first-of-its-kind tool for incremental model design in applied RL. Wield provides a small set of primitives which decouple systems interfaces and deployment-specific configuration from representation. Core to Wield is a novel instructive experiment protocol called progressive randomisation which helps practitioners to incrementally evaluate different dimensions of non-determinism. I demonstrate how Wield and progressive randomisation can be used to reproduce and assess prior work, and to guide implementation of novel RL applications.

UCAM-CL-TR-947

Robert N. M. Watson,
Alexander Richardson, Brooks Davis,
John Baldwin, David Chisnall, Jessica Clarke,
Nathaniel Filardo, Simon W. Moore,
Edward Napierala, Peter Sewell,
Peter G. Neumann:

CHERI C/C++ Programming Guide

June 2020, 33 pages, PDF

Abstract: This document is a brief introduction to the CHERI C/C++ programming languages. We explain the principles underlying these language variants, and their grounding in CHERI's multiple architectural instantiations: CHERI-MIPS, CHERI-RISC-V, and Arm's Morello. We describe the most commonly encountered differences between these dialects and C/C++ on conventional architectures, and where existing software may require minor changes. We document new compiler warnings and errors that may be experienced compiling code with the CHERI Clang/LLVM compiler, and suggest how they may be addressed through typically minor source-code changes. We explain how

modest language extensions allow selected software, such as memory allocators, to further refine permissions and bounds on pointers. This guidance is based on our experience adapting the FreeBSD operating-system userspace, and applications such as PostgreSQL and WebKit, to run in a CHERI C/C++ capability-based programming environment. We conclude by recommending further reading.

UCAM-CL-TR-948

Dylan McDermott:

Reasoning about effectful programs and evaluation order

June 2020, 150 pages, PDF
PhD thesis (Homerton College, October 2019)

Abstract: Program transformations have various applications, such as in compiler optimizations. These transformations are often effect-dependent: replacing one program with another relies on some restriction on the side-effects of subprograms. For example, we cannot eliminate a dead computation that raises an exception, or a duplicated computation that prints to the screen. Effect-dependent program transformations can be described formally using effect systems, which annotate types with information about the side-effects of expressions.

In this thesis, we extend previous work on effect systems and correctness of effect-dependent transformations in two related directions.

First, we consider evaluation order. Effect systems for call-by-value languages are well-known, but are not sound for other evaluation orders. We describe sound and precise effect systems for various evaluation orders, including call-by-name. We also describe an effect system for Levy's call-by-push-value, and show that this subsumes those for call-by-value and call-by-name. This naturally leads us to consider effect-dependent transformations that replace one evaluation order with another. We show how to use the call-by-push-value effect system to prove the correctness of transformations that replace call-by-value with call-by-name, using an argument based on logical relations. Finally, we extend call-by-push-value to additionally capture call-by-need. We use our extension to show a classic example of a relationship between evaluation orders: if the side-effects are restricted to (at most) nontermination, then call-by-name is equivalent to call-by-need.

The second direction we consider is non-invertible transformations. A program transformation is non-invertible if only one direction is correct. Such transformations arise, for example, when considering undefined behaviour, nondeterminism, or concurrency. We present a general framework for verifying noninvertible effect-dependent transformations, based on our effect system for call-by-push-value. The framework includes a non-symmetric notion of correctness for effect-dependent transformations, and a denotational semantics based on order-enriched category theory that can be used to prove correctness.

UCAM-CL-TR-949

Alexander Richardson:

Complete spatial safety for C and C++ using CHERI capabilities

June 2020, 189 pages, PDF
PhD thesis (Emmanuel College, October 2019)

Abstract: Lack of memory safety in commonly used systems-level languages such as C and C++ results in a constant stream of new exploitable software vulnerabilities and exploit techniques. Many exploit mitigations have been proposed and deployed over the years, yet none address the root issue: lack of memory safety. Most C and C++ implementations assume a memory model based on a linear array of bytes rather than an object-centric view. Whilst more efficient on contemporary CPU architectures, linear addresses cannot encode the target object, thus permitting memory errors such as spatial safety violations (ignoring the bounds of an object). One promising mechanism to provide memory safety is CHERI (Capability Hardware Enhanced RISC Instructions), which extends existing processor architectures with capabilities that provide hardware-enforced checks for all accesses and can be used to prevent spatial memory violations. This dissertation prototypes and evaluates a pure-capability programming model (using CHERI capabilities for all pointers) to provide complete spatial memory protection for traditionally unsafe languages.

As the first step towards memory safety, all language-visible pointers can be implemented as capabilities. I analyse the programmer-visible impact of this change and refine the pure-capability programming model to provide strong source-level compatibility with existing code. Second, to provide robust spatial safety, language-invisible pointers (mostly arising from program linkage) such as those used for functions calls and global variable accesses must also be protected. In doing so, I highlight trade-offs between performance and privilege minimization for implicit and programmer-visible pointers. Finally, I present CheriSH, a novel and highly compatible technique that protects against buffer overflows between fields of the same object, hereby ensuring that the CHERI spatial memory protection is complete.

I find that the byte-granular spatial safety provided by CHERI pure-capability code is not only stronger than most other approaches, but also incurs almost negligible performance overheads in common cases (0.1% geometric mean) and a worst-case overhead of only 23.3% compared to the insecure MIPS baseline. Moreover, I show that the pure-capability programming model provides near-complete source-level compatibility with existing programs. I evaluate this based

on porting large widely used open-source applications such as PostgreSQL and WebKit with only minimal changes: fewer than 0.1% of source lines.

I conclude that pure-capability CHERI C/C++ is an eminently viable programming environment offering strong memory protection, good source-level compatibility and low performance overheads.

UCAM-CL-TR-950

Hugo Paquet:

Probabilistic concurrent game semantics

August 2020, 156 pages, PDF
PhD thesis (Homerton College, September 2019)

Abstract: This thesis presents a variety of models for probabilistic programming languages in the framework of concurrent games.

Our starting point is the model of concurrent games with symmetry of Castellan, Clairambault and Winskel. We show that they form a symmetric monoidal closed bicategory, and that this can be turned into a cartesian closed bicategory using a linear exponential pseudo-comonad inspired by linear logic.

Then, we enrich this with probability, relying heavily on Winskel's model of probabilistic concurrent strategies. We see that the bicategorical structure is not perturbed by the addition of probability. We apply this model to two probabilistic languages: a probabilistic untyped λ-calculus, and Probabilistic PCF. For the former, we relate the semantics to the probabilistic Nakajima trees of Leventis, thus obtaining a characterisation of observational equivalence for programs in terms of strategies. For the latter, we show a definability result in the spirit of the game semantics tradition. This solves an open problem, as it is notoriously difficult to model Probabilistic PCF with sequential game semantics.

Finally, we introduce a model for measurable game semantics, in which games and strategies come equipped with measure-theoretic structure allowing for an accurate description of computation with continuous data types. The objective of this model is to support computation with arbitrary probability measures on the reals. In the last part of this thesis we see how this can be done by equipping strategies with parametrised families of probability measures (also known as stochastic kernels), and we construct a bicategory of measurable concurrent games and probabilistic measurable strategies.

UCAM-CL-TR-951

Robert N. M. Watson, Peter G. Neumann, Jonathan Woodruff, Michael Roe, Hesham Almatary, Jonathan Anderson, John Baldwin, Graeme Barnes, David Chisnall, Jessica Clarke, Brooks Davis, Lee Eisen, Nathaniel Wesley Filardo, Richard Grisenthwaite, Alexandre Joannou, Ben Laurie, A. Theodore Markettos, Simon W. Moore, Steven J. Murdoch, Kyndylan Nienhuis, Robert Norton, Alexander Richardson, Peter Rugg, Peter Sewell, Stacey Son, Hongyan Xia:

Capability Hardware Enhanced RISC Instructions: CHERI Instruction-Set Architecture (Version 8)

October 2020, 590 pages, PDF

Abstract: This technical report describes CHERI ISAv8, the eighth version of the CHERI architecture being developed by SRI International and the University of Cambridge. This design captures ten years of research, development, experimentation, refinement, formal analysis, and validation through hardware and software implementation.

CHERI introduces an architecture-neutral capability-based protection model, which has been instantiated in various commodity base architectures to give CHERI-MIPS, CHERI-RISC-V, Arm's prototype Morello architecture, and (sketched) CHERI-x86-64. It enables software to efficiently implement fine-grained memory protection and scalable software compartmentalization, by providing strong, non-probabilistic, efficient mechanisms to support the principles of least privilege and intentional use in the execution of software at multiple levels of abstraction, preventing and mitigating vulnerabilities. Design goals include incremental adoptability from current ISAs and software stacks, low performance overhead for memory protection, significant performance improvements for software compartmentalization, formal grounding, and programmer-friendly underpinnings.

CHERI blends traditional paged virtual memory with an in-address-space capability model that includes capability values in registers, capability instructions, and tagged memory to enforce capability integrity. This hybrid approach, inspired by the Capsicum security model, addresses the performance and robustness issues that arise when trying to express more secure programming models, minimising privilege, above conventional architectures that provide only MMU-based protection. CHERI builds on the C-language fat-pointer literature: its capabilities can describe fine-grained regions of memory, and can be substituted for data or code pointers in generated code, protecting data and improving control-flow robustness. Strong capability integrity and monotonicity properties allow CHERI to express a variety of protection idioms, from enforcing valid C-language pointer provenance and bounds checking to implementing the isolation and controlled

communication structures required for software compartmentalization.

CHERI's hybrid approach allows incremental adoption of capability-oriented design: critical components can be ported and recompiled to use capabilities throughout, providing fine-grain memory protection, or be largely unmodified but encapsulated in ways that permit only controlled interaction. Potential early deployment scenarios include low-level software Trusted Computing Bases (TCBs) such as separation kernels, hypervisors, and operating-system kernels, userspace TCBs such as language runtimes and web browsers, and particularly high-risk software libraries such as data compression, protocol parsing, and image processing (which are concentrations of both complex and historically vulnerability-prone code exposed to untrustworthy data sources).

CHERI ISAv8 is a substantial enhancement to prior ISA versions. Capability compression is now part of the abstract model. Both 32-bit and 64-bit architectural address sizes are supported. Various previously experimental features, such as sentry capabilities and CHERI-RISC-V, are now considered mature. We have defined a number of new temporal memory-safety acceleration features including MMU assistance for a load-side-barrier revocation model. We have added a chapter on practical CHERI microarchitecture. CHERI ISAv8 is synchronized with Arm Morello.

UCAM-CL-TR-952

David Adeboye:

Exploring the effect of spatial faithfulness on group decision-making

October 2020, 64 pages, PDF
MPhil thesis (Jesus College, 2020)

Abstract: Remote working is becoming increasingly popular and many large organisations are asking their employees to work from home. However, several studies have shown that groups who make decisions over videoconferencing take longer to complete tasks, are less effective and are less satisfied with the result. The ability for a communication medium to convey information, cues or symbols to its users has been theorised to influence team/communication performance. Videoconferencing fails to communicate these non-verbal behaviours, which provide complementary information to speech. For example, the inability to use gaze to help indicate the next speaker means that conversations over videoconferencing typically contain more explicit handovers such as names.

This thesis presents Spatial, a new spatially faithful videoconferencing application that captures the aspects of face-to-face conversations that are not available on standard systems. Unlike previous work, which requires specialised equipment or setups, Spatial focuses on work-from-home environments. Spatial aims to replicate the spatial characteristics of face-to-face conversations, using commodity hardware. It builds environments that ensure that both visual and auditory communication can be transmitted directionally and as wholly as possible. Using Spatial they can calibrate their working environments to ensure that their experience is free from perspective distortions.

We show that under Spatial, groups replicate conversation characteristics of face-to-face interactions. They completed a cooperative decision-making task in a shorter amount of time, took less time for turns and interrupted each other less.

UCAM-CL-TR-953

Robert N. M. Watson, Jonathan Woodruff, Alexandre Joannou, Simon W. Moore, Peter Sewell, Arm Limited:

DSbD CHERI and Morello Capability Essential IP (Version 1)

December 2020, 25 pages, PDF

Abstract: The CHERI protection model extends contemporary Instruction Set Architectures (ISAs) with support for architectural capabilities. The UKRI Digital Security by Design (DSbD) programme is supporting the creation of Arm's prototype Morello processor, System-on-Chip (SoC), and board. Morello experimentally incorporates the CHERI protection model, developed at the University of Cambridge and SRI International, into the ARMv8-A architecture. This document declares a set of capability essential IP – ideas essential to the creation of a contemporary CHERI capability system in architecture and microarchitecture. Arm and Cambridge agree that they have made this IP available for use without restriction. This document also identifies a set of CHERI background documents that may be of value as prior art.

UCAM-CL-TR-954

Myoung Jin Nam:

Inline and sideline approaches for low-cost memory safety in C

February 2021, 124 pages, PDF
PhD thesis (Selwyn College, November 2020)

Abstract: System languages such as C or C++ are widely used for their high performance, however the allowance of arbitrary pointer arithmetic and typecast introduces a risk of memory corruptions. These memory errors cause unexpected termination of programs, or even

worse, attackers can exploit them to alter the behavior of programs or leak crucial data.

Despite advances in memory safety solutions, high and unpredictable overhead remains a major challenge. Accepting that it is extremely difficult to achieve complete memory safety with the performance level suitable for production deployment, researchers attempt to strike a balance between performance, detection coverage, interoperability, precision, and detection timing. Some properties are much more desirable, e.g. the interoperability with pre-compiled libraries. Comparatively less critical properties are sacrificed for performance, for example, tolerating longer detection delay or narrowing down detection coverage by performing approximate or probabilistic checking or detecting only certain errors. Modern solutions compete for performance.

The performance matrix of memory safety solutions has two major assessment criteria – run-time and memory overheads. Researchers trade-off and balance performance metrics depending on its purpose or placement. Many of them tolerate the increase in memory use for better speed, since memory safety enforcement is more desirable for troubleshooting or testing during development, where a memory resource is not the main issue. Run-time overhead, considered more critical, is impacted by cache misses, dynamic instructions, DRAM row activations, branch predictions and other factors.

This research proposes, implements, and evaluates MIU: Memory Integrity Utilities containing three solutions – MemPatrol, FRAMER and spaceMiu. MIU suggests new techniques for practical deployment of memory safety by exploiting free resources with the following focuses: (1) achieving memory safety with overhead $< 1\%$ by using concurrency and trading off prompt detection and coverage; but yet providing eventual detection by a monitor isolation design of an in-register monitor process and the use of AES instructions (2) complete memory safety with near-zero false negatives focusing on eliminating overhead, that hardware support cannot resolve, by using a new tagged-pointer representation utilising the top unused bits of a pointer.

UCAM-CL-TR-955

Hongyan Xia:

Capability memory protection for embedded systems

February 2021, 145 pages, PDF
PhD thesis (Hughes Hall, May 2019)

Abstract: This dissertation explores the use of capability security hardware and software in real-time and latency-sensitive embedded systems, to address existing memory safety and task isolation problems as well as providing new means to design a secure and scalable real-time system. In addition, this dissertation looks into how practical and high-performance temporal memory safety can be achieved under a capability architecture.

State-of-the-art memory protection schemes for embedded systems typically present limited and inflexible solutions to memory protection and isolation, and fail to scale as embedded devices become more capable and ubiquitous. I investigate whether a capability architecture is able to provide new angles to address memory safety issues in an embedded scenario. Previous CHERI capability research focuses on 64-bit architectures in UNIX operating systems, which does not translate to typical 32-bit embedded processors with low-latency and real-time requirements. I propose and implement the CHERI CC-64 encoding and the CHERI-64 coprocessor to construct a feasible capability-enabled 32-bit CPU. In addition, I implement a real-time kernel for embedded systems atop CHERI-64. On this hardware and software platform, I focus on exploring scalable task isolation and fine-grained memory protection enabled by capabilities in a single flat physical address space, which are otherwise difficult or impossible to achieve via state-of-the-art approaches. Later, I present the evaluation of the hardware implementation and the software run-time overhead and real-time performance.

Even with capability support, CHERI-64 as well as other CHERI processors still expose major attack surfaces through temporal vulnerabilities like use-after-free. A naïve approach that sweeps memory to invalidate stale capabilities is inefficient and incurs significant cycle overhead and DRAM traffic. To make sweeping revocation feasible, I introduce new architectural mechanisms and micro-architectural optimisations to substantially reduce the cost of memory sweeping and capability revocation. Another factor of the cost is the frequency of memory sweeping. I explore tradeoffs of memory allocator designs that use quarantine buffers and shadow space tags to prevent frequent unnecessary sweeping. The evaluation shows that the optimisations and new allocator designs reduce the cost of capability sweeping revocation by orders of magnitude, making it already practical for most applications to adopt temporal safety under CHERI.

UCAM-CL-TR-956

Jianxin Zhao:

Optimisation of a modern numerical library: a bottom-up approach

April 2021, 96 pages, PDF
PhD thesis (Corpus Christi College, September 2019)

Abstract: Numerical libraries lie in the heart of modern applications, from machine learning, scientific computation and to Internet of Things (IoT). It has dominated many aspects of our daily lives. Numerical library used to lie in the low level of applications, and only need to focus on provide fast calculation. However, with social awareness of privacy and personal data arising,

computation is gradually moved to devices in heterogeneous environment. Recently development of edge devices such as Edge TPU also promotes a trend of decentralised computation. Given this trend, a new understanding of the full stack of computation is required to optimise computation at various levels.

In this thesis, based on my experience participating in the development of a numerical library, I present a bottom-up approach that centres on numerical library to describe the optimisation of computation at various levels. I present the low-level design of numerical operations and show the related impact on performance optimisation. I create new algorithms for key operations, and build an automatic tuning module to further improve performance. At the graph level, which consists of multiple operations, I present the idea of using graph as common Intermediate Representation to enable interoperability on other computation frameworks and devices. To demonstrate this approach, I build the TFgraph system that provides a symbolic representation layer to exchange the computation between Owl and other frameworks. At a higher level, the computation graph can be seen as a unit, and at this level I identify the problems of computation composition and deployment. I build the Zoo system to address these two problems. It provides a small Domain-specific Language to enable composition of advanced data analytics services. Benefiting from OCaml's powerful type system, the Zoo provides type checking for the composition. Besides, the Zoo DSL supports fine-grained version control in composing. It also involves deploying composed services to multiple backends. Finally, the top level involves collaboration of multiple deployed computations. At this level, I focus on the barrier control methods, propose two quantitative metrics to evaluate existing barrier control methods, and bring new insights into their design. I have also built a simulation platform and real-world experiments to perform thorough evaluation of the PSP compared to existing barrier methods.

UCAM-CL-TR-957

Daniel Hugenroth, Martin Kleppmann, Alastair R. Beresford:

Rollercoaster: an efficient group-multicast scheme for mix networks

27 pages, PDF
Originally published in Proceedings of the 30th USENIX Security Symposium

Abstract: Mix network designs such as Loopix provide strong metadata anonymity guarantees that are crucial across many applications. However, because they limit the rate at which messages can be sent by each user, they incur high delays when sending many messages to multiple recipients – for instance, in decentralised collaborative apps.

In this paper we present an efficient multicast scheme named Rollercoaster that reduces the time for delivering a message to all members of a group of size m from O(m) to O(log m). Rollercoaster can be deployed without modifications to the underlying mix network, allowing it to benefit from the anonymity set provided by existing users. We further develop an extension that achieves the same asymptotic guarantees in the presence of unreliable group members.

While the scheme is applicable to many mix network designs, we evaluate it for the Loopix network, which is the most advanced and practical design to date. For this evaluation we developed a network simulator that allows fast, reproducible, and inspectable runs while eliminating external influences.

UCAM-CL-TR-958

Francisco Vargas:

Machine-learning approaches for the empirical Schrödinger bridge problem

June 2021, 114 pages, PDF
MPhil thesis (Girton College, 24 June 2020)

Abstract: The Schrödinger bridge problem is concerned with finding the most likely stochastic evolution between two probability distributions given a prior/reference stochastic evolution. This problem was posed by Schrödinger (1931, 1932) and solved to a large extent. Problems of this kind, whilst not popular in the machine learning community, have direct applications such as domain adaptation, hypothesis testing, semantic similarity, and others.

Thus, the focus of this thesis is to carry out a preliminary study on computational approaches for estimating the Schrödinger bridge between two distributions, when these distributions are available (or can be available) through samples, as most problems in machine learning are.

Due to the mathematical nature of the problem, this manuscript is also concerned with restating and re-deriving theorems and results that seem to be considered communal knowledge within the mathematical community or hidden in type-written textbooks behind paywalls. Part of the aim of this thesis is to make the mathematical machinery behind these approaches more accessible to a broader audience.

UCAM-CL-TR-959

Thomas Bauereiss, Brian Campbell, Thomas Sewell, Alasdair Armstrong, Lawrence Esswood, Ian Stark, Graeme Barnes, Robert N. M. Watson, Peter Sewell:

Verified security for the Morello capability-enhanced prototype Arm architecture

September 2021, 24 pages, PDF

Abstract: Memory safety bugs continue to be a major source of security vulnerabilities in our critical infrastructure. The CHERI project has proposed extending conventional architectures with hardware-supported capabilities to enable fine-grained memory protection and scalable compartmentalisation, allowing historically memory-unsafe C and C++ to be adapted to deterministically mitigate large classes of vulnerabilities, while requiring only minor changes to existing system software sources. Arm is currently designing and building Morello, a CHERI-enabled prototype architecture, processor, SoC, and board, extending the high-performance Neoverse N1, to enable industrial evaluation of CHERI and pave the way for potential mass-market adoption. However, for such a major new security-oriented architecture feature, it is important to establish high confidence that it does provide the protections it intends to, and that cannot be done with conventional engineering techniques.

In this paper we put the Morello architecture on a solid mathematical footing from the outset. We define the fundamental security property that Morello aims to provide, reachable capability monotonicity, and prove that the architecture definition satisfies it. This proof is mechanised in Isabelle/HOL, and applies to a translation of the official Arm Morello specification into Isabelle. The main challenge is handling the complexity and scale of a production architecture: 62,000 lines of specification, translated to 210,000 lines of Isabelle. We do so by factoring the proof via a narrow abstraction capturing the essential properties of instruction execution in an arbitrary CHERI ISA, expressed above a monadic intra-instruction semantics. We also develop a model-based test generator, which generates instruction-sequence tests that give good specification coverage, used in early testing of the Morello implementation and in Morello QEMU development. We also use Arm's internal test suite to validate our internal model.

This gives us machine-checked mathematical proofs of whole-ISA security properties of a full-scale industry architecture, at design-time. To the best of our knowledge, this is the first demonstration that that is feasible, and it significantly increases confidence in Morello.

UCAM-CL-TR-960

Dionysis Manousakas:

Data summarizations for scalable, robust and privacy-aware learning in high dimensions

September 2021, 130 pages, PDF
PhD thesis (Darwin College, October 2020)

Abstract: The advent of large-scale datasets has offered unprecedented amounts of information for building statistically powerful machines, but, at the same time, also introduced a remarkable computational challenge: how can we efficiently process massive data? This thesis presents a suite of data reduction methods that make learning algorithms scale on large datasets, via extracting a succinct model-specific representation that summarizes the full data collection—a coreset. Our frameworks support by design datasets of arbitrary dimensionality, and can be used for general purpose Bayesian inference under real-world constraints, including privacy preservation and robustness to outliers, encompassing diverse uncertainty-aware data analysis tasks, such as density estimation, classification and regression.

We motivate the necessity for novel data reduction techniques in the first place by developing a reidentification attack on coarsened representations of private behavioural data. Analysing longitudinal records of human mobility, we detect privacy-revealing structural patterns, that remain preserved in reduced graph representations of individuals' information with manageable size. These unique patterns enable mounting linkage attacks via structural similarity computations on longitudinal mobility traces, revealing an overlooked, yet existing, privacy threat.

We then propose a scalable variational inference scheme for approximating posteriors on large datasets via learnable weighted pseudodata, termed pseudocoresets. We show that the use of pseudodata enables overcoming the constraints on minimum summary size for given approximation quality, that are imposed on all existing Bayesian coreset constructions due to data dimensionality. Moreover, it allows us to develop a scheme for pseudocoresets-based summarization that satisfies the standard framework of differential privacy by construction; in this way, we can release reduced size privacy-preserving representations for sensitive datasets that are amenable to arbitrary post-processing.

Subsequently, we consider summarizations for large-scale Bayesian inference in scenarios when observed datapoints depart from the statistical assumptions of our model. Using robust divergences, we develop a method for constructing coresets resilient to model misspecification. Crucially, this method is able to automatically discard outliers from the generated data summaries. Thus we deliver robustified scalable representations for inference, that are suitable for applications involving contaminated and unreliable data sources.

We demonstrate the performance of proposed summarization techniques on multiple parametric statistical models, and diverse simulated and real-world datasets, from music genre features to hospital readmission records, considering a wide range of data dimensionalities.

UCAM-CL-TR-961

Lawrence G. Esswood:

CheriOS: designing an untrusted single-address-space capability operating system utilising capability hardware and a minimal hypervisor

September 2021, 195 pages, PDF
PhD thesis (Churchill College, July 2020)

Abstract: This thesis presents the design, implementation, and evaluation of a novel capability operating system: CheriOS. The guiding motivation behind CheriOS is to provide strong security guarantees to programmers, even allowing them to continue to program in fast, but typically unsafe, languages such as C. Furthermore, it does this in the presence of an extremely strong adversarial model: in CheriOS, every compartment – and even the operating system itself – is considered actively malicious. Building on top of the architecturally enforced capabilities offered by the CHERI microprocessor, I show that only a few more capability types and enforcement checks are required to provide a strong compartmentalisation model that can facilitate mutual distrust. I implement these new primitives in software, in a new abstraction layer I dub the nanokernel. Among the new OS primitives I introduce are one for integrity and confidentiality called a Reservation (which allows allocating private memory without trusting the allocator), as well as another that can provide attestation about the state of the system, a Foundation (which provides a key to sign and protect capabilities based on a signature of the starting state of a program). I show that, using these new facilities, it is possible to design an operating system without having to trust the implementation is correct.

CheriOS is fundamentally fail-safe; there are no assumptions about the behaviour of the system, apart from the CHERI processor and the nanokernel, to be broken. Using CHERI and the new nanokernel primitives, programmers can expect full isolation at scopes ranging from a whole program to a single function, and not just with respect to other programs but the system itself. Programs compiled for and run on CheriOS offer full memory safety, both spatial and temporal, enforced control flow integrity between compartments and protection against common vulnerabilities such as buffer overflows, code injection and Return-Oriented-Programming attacks. I achieve this by designing a new CHERI-based ABI (Application Binary Interface) which includes a novel stack structure that offers temporal safety. I evaluate how practical the new designs are by prototyping them and offering a detailed performance evaluation. I also contrast with existing offerings from both industry and academia.

CHERI capabilities can be used to restrict access to system resources, such as memory, with the required dynamic checks being performed by hardware in parallel with normal operation. Using the accelerating features of CHERI, I show that many of the security guarantees that CheriOS offers can come at little to no cost. I present a novel and secure IO/IPC layer that allows secure marshalling of multiple data streams through mutually distrusting compartments, with fine-grained authenticated access control for end-points, and without either copying or encryption. For example, CheriOS can restrict its TCP stack from having access to packet contents, or restrict an open socket to ensure data sent on it to arrives at an endpoint signed as a TLS implementation. Even with added security requirements, CheriOS can perform well on real workloads. I showcase this by running a state-of-the-art webserver, NGINX, atop both CheriOS and FreeBSD and show improvements in performance ranging from 3x to 6x when running on a small-scale low-power FPGA implementation of CHERI-MIPS.

UCAM-CL-TR-962

Harri Bell-Thomas:

Trusted reference monitors for Linux using Intel SGX enclaves

October 2021, 62 pages, PDF
MEng dissertation (Jesus College, June 2020)

Abstract: Information Flow Control (IFC) is a powerful tool for protecting data in a computer system, enforcing not only who may access it, but also how it may be used throughout its lifespan. Intel's Software Guard Extension (SGX) affords complementary protection, providing a general-purpose Trusted Execution Environment for applications and their data. To date, no work has been conducted considering the overlap between the two, and how they may mutually reinforce each other.

This dissertation presents Citadel, a modular, SGX-backed reference monitor to securely and verifiably implement IFC methods in the Linux kernel. Its prototype externalises policy decisions from its enforcement security module, providing a userspace promise-of-access model with asynchronous fulfilment. By aliasing system calls, the system transparently integrates with unmodified applications, and amortises the performance cost of integration by inferring processes' underlying security contexts.

Observed results are promising, demonstrating a worst-case median performance overhead of 25%. In addition, the Nginx webserver is demonstrated running under Citadel; high bandwidth transfers exhibit near parity with the native Linux kernel's performance. This work illustrates the potential viability of a symbiotic enclave-kernel relationship for security implementations, something that may, in the long run, benefit both.

UCAM-CL-TR-963

Michael G. Dodson:

Capability-based access control for cyber physical systems

October 2021, 127 pages, PDF
PhD thesis (Queens' College, July 2021)

Abstract: Cyber Physical Systems (CPS) couple digital systems with the physical environment, creating technical, usability, and economic security challenges beyond those of information systems. Their distributed and hierarchical nature, real-time and safety-critical requirements, and limited resources create new vulnerability classes and severely constrain the security solution space. This dissertation explores these challenges, focusing on Industrial Control Systems (ICS), but demonstrating broader applicability to the whole domain.

We begin by systematising the usability and economic challenges to secure ICS. We fingerprint and track more than 10,000 Internet-connected devices over four years and show the population is growing, continuously-connected, and unpatched. We then explore adversarial interest in this vulnerable population. We track 150,000 botnet hosts, sift 70 million underground forum posts, and perform the largest ICS honeypot study to date to demonstrate that the cybercrime community has little competence or interest in the domain. We show that the heterogeneity, cost, and expertise required for large-scale attacks on ICS are economic deterrents when targets in the IoT domain are available.

The ICS landscape is changing, however, and we demonstrate the imminent convergence with the IoT domain as inexpensive hardware, commodity operating systems, and wireless connectivity become standard. Industry's security solution is boundary defence, pushing privilege to firewalls and anomaly detectors; however, this propagates rather than minimises privilege and leaves the hierarchy vulnerable to a single boundary compromise.

In contrast, we propose, implement, and evaluate a security architecture based on distributed capabilities. Specifically, we show that object capabilities, representing physical resources, can be constructed, delegated, and used anywhere in a distributed CPS by composing hardware-enforced architectural capabilities and cryptographic network tokens. Our architecture provides defence-in-depth, minimising privilege at every level of the CPS hierarchy, and both supports and adds integrity protection to legacy CPS protocols. We implement distributed capabilities in robotics and ICS demonstrators, and we show that our architecture adds negligible overhead to realistic integrations and can be implemented without significant modification to existing source code.

UCAM-CL-TR-964

Hayk Saribekyan:

Information dissemination via random walks

November 2021, 134 pages, PDF
PhD thesis (St John's College, July 2021)

Abstract: Information dissemination is a fundamental task in distributed computing: How to deliver a piece of information from a node of a network to some or all other nodes? In the face of large and still growing modern networks, it is imperative that dissemination algorithms are decentralised and can operate under unreliable conditions. In the past decades, randomised rumour spreading algorithms have addressed these challenges. In these algorithms, a message is initially placed at a source node of a network, and, at regular intervals, each node contacts a randomly selected neighbour. A message may be transmitted in one or both directions during each of these communications, depending on the exact protocol. The main measure of performance for these algorithms is their broadcast time, which is the time until a message originating from a source node is disseminated to all nodes of the network. Apart from being extremely simple and robust to failures, randomised rumour spreading achieves theoretically optimal broadcast time in many common network topologies.

In this thesis, we propose an agent-based information dissemination algorithm, called Visit-Exchange. In our protocol, a number of agents perform independent random walks in the network. An agent becomes informed when it visits a node that has a message, and later informs all future nodes it visits. Visit-Exchange shares many of the properties of randomised rumour spreading, namely, it is very simple and uses the same amount of communication in a unit of time. Moreover, the protocol can be used as a simple model of non-recoverable epidemic processes.

We investigate the broadcast time of Visit-Exchange on a variety of network topologies, and compare it to traditional rumour spreading. On dense regular networks we show that the two types of protocols are equivalent, which means that in this setting the vast literature on randomised rumour spreading applies in our model as well. Since many networks of interest, including real-world ones, are very sparse, we also study agent-based broadcast for sparse networks. Our results include almost optimal or optimal bounds for sparse regular graphs, expanders, random regular graphs, balanced trees and grids. We establish that depending on the network topology, Visit-Exchange may be either slower or faster than traditional rumour spreading. In particular, in graphs consisting of hubs that are not well connected, broadcast using agents can be significantly faster. Our conclusion is that a combined broadcasting protocol that simultaneously uses both traditional

rumour spreading and agent-based dissemination can be fast on a larger range of topologies than each of its components separately.

UCAM-CL-TR-965

Daniel M. Fisher, Jon A. Crowcroft:

Improving commercial LiFi network feasibility through rotation invariance, motion prediction, and bandwidth aggregation at the physical layer

November 2021, 79 pages, PDF
MPhil thesis (Downing College, May 2020)

Abstract: In recent years, the number of devices per household has exponentially increased. Additionally, Internet traffic surged during the 2020–2021 calendar years due to the COVID-19 pandemic. Both of these occurrences have served as reminders that the spectrum available for radio frequency (RF) based networking is running thin, as internet service providers had to find creative ways to serve this growth of data demand. Introduced in 2011 by Dr. Harald Haas, Light Fidelity, or LiFi, offers a suitable supplement to ameliorate this challenge as well as offer faster data rates to consumers. However, due to the line-of-sight nature of data transmission required by LiFi, path blockage to the photodetector and natural light obstruction have proven to be significant challenges in commercially implementing the computer network. Hybrid networks have been proposed, serving as a middle ground where the networking load is balanced between LiFi and WiFi depending on which is available. However, many challenges exist with this concept, in particular with optimizing the logic so it is worthwhile.

The goal of this thesis is to further develop this field and push LiFi closer to being commercially practical for the next generation of computer networks. Three lines of effort are pursued to do so. First, the path blockage problem is mitigated through a novel rotationally invariant photodetector configuration that allows for significantly better sensor visibility than previous solutions. Additionally, horizontal handovers from one LiFi hub to another, as a user moves through a space, is improved via leveraging probabilistic machine learning (ML) and other motion tracking algorithms to more accurately predict human movement and identify "hot spots." Lastly, a novel bandwidth aggregation scheme at the physical layer is proposed and preliminarily evaluated through a simulation to significantly improve data rates when both LiFi and WiFi are available.

UCAM-CL-TR-966

Helen Oliver:

Obstacles to wearable computing

December 2021, 318 pages, PDF
PhD thesis (Murray Edwards College, June 2020)
Funded and supported by The Alan Turing Institute Doctoral Scheme

Abstract: In the year 2021, wearable technology could look beautiful and feel magical, but instead is exemplified by a plain wristband that looks suspiciously like a prison monitor.

How can we make wearable technology that respects our privacy, enhances our daily lives, integrates with our other connected devices without leashing us to a smartphone, and visually expresses who we are?

This study uses a novel method of participatory design fiction (PDFi) to understand potential users of everyday wearable technology through storytelling. I recruited participants from the general public and gave them a five-point prompt to create a design fiction (DF), which inspired the user-centred design of an everyday connected wearable device. The participants each received a technology probe to wear in the wild for a year. They then updated their DFs as a way to reflect on the implications of the technology. For the purposes of privacy, augmenting device functionality through interoperability, and integration into an Internet of Things (IoT) ecosystem, I used the Hub-of-All-Things personal data store to provide the software infrastructure.

By listening to their stories, we can elicit design concepts directly from the users, to help us create wearable IoT devices that put the wearer at the centre of the design process, and are satisfying both functionally and emotionally.

UCAM-CL-TR-967

Primož Fabiani:

Gaussian Pixie Autoencoder: Introducing Functional Distributional Semantics to continuous latent spaces

January 2022, 50 pages, PDF
MPhil thesis (Hughes Hall, June 2021)

Abstract: Functional Distributional Semantics (FDS) is a recent lexical semantics framework that represents word meaning as a function from the latent space of entities to a probability for each word. This thesis examines previous FDS models, highlighting the advantages and drawbacks. A new Gaussian Pixie Autoencoder model is proposed to introduce FDS to continuous latent modelling. The proposed model improves on the predecessors in terms of simplicity and efficiency setting a new baseline for continuous FDS models. The

thesis shows dropout is necessary for context learning with this model type. Evaluated on contextual similarity the proposed model outperforms the discrete autoencoder and BERT baseline on one task with satisfactory performance on the other.

UCAM-CL-TR-968

James Thorne:

Evidence-based verification and correction of textual claims

February 2022, 231 pages, PDF
PhD thesis (Peterhouse College, September 2021)

Abstract: This thesis considers the task of fact-checking: predicting the veracity of claims made in written or spoken language using evidence. However, in previous task formulations, modelling assumptions ignore the requirement for systems to retrieve the necessary evidence. To better model how human fact-checkers operate, who first find evidence before labelling a claim's veracity, the methodology proposed in this thesis requires automated systems to retrieve evidence from a corpus to justify the veracity predictions made when modelling this task. The primary contribution of this thesis is the development and release of FEVER, a large-scale collection of human-written claims annotated with evidence from Wikipedia. Analysis of systems trained on this data highlights challenges in resolving ambiguity and context, as well as being resilient to imperfect evidence retrieval. To understand the limitations of models trained on datasets such as FEVER, contemporary fact verification systems are further evaluated using adversarial attacks – instances constructed specifically to identify weaknesses and blind spots. However, as automated means for generating adversarial instances induce their own errors, this thesis proposes considering instances' correctness, allowing fairer comparison. The thesis subsequently considers how biases captured in these models can be mitigated with finetuning regularised with elastic weight consolidation. Finally, the thesis presents a new extension to the verification task: factual error correction. Rather than predicting the claim's veracity, systems must also generate a correction for the claim so that it is better supported by evidence, acting as another means to communicate the claim's veracity to an end-user. In contrast to previous work on explainable fact-checking, the method proposed in this chapter does not require additional data for supervision.

UCAM-CL-TR-969

Martin Kleppmann:

Assessing the understandability of a distributed algorithm by tweeting buggy pseudocode

May 2022, 15 pages, PDF

Abstract: Designing algorithms for distributed systems has a reputation of being a difficult and error-prone task, but this difficulty is rarely measured or quantified in any way. This report tells the story of one informal experiment, in which users on Twitter were invited to identify the bug in an incorrect CRDT algorithm. Over the following 11 hours, at least 16 people (many of whom are professional software engineers) made attempts to find the bug, but most were unsuccessful. The two people who did identify the bug were both PhD students specialising in CRDTs. This result may serve as evidence of the difficulty of designing correct CRDT algorithms.

UCAM-CL-TR-970

Anita L. Verő:

Transparent analysis of multi-modal embeddings

May 2022, 202 pages, PDF
PhD thesis (King's College, November 2021)

Abstract: Vector Space Models of Distributional Semantics — or Embeddings — serve as useful statistical models of word meanings, which can be applied as proxies to learn about human concepts. One of their main benefits is that not only textual, but a wide range of data types can be mapped to a space, where they are comparable or can be fused together.

Multi-modal semantics aims to enhance Embeddings with perceptual input, based on the assumption that the representation of meaning in humans is grounded in sensory experience. Most multi-modal research focuses on downstream tasks, involving direct visual input, such as Visual Question Answering. Fewer papers have exploited visual information for meaning representations when the evaluation tasks involve no direct visual input, such as semantic similarity. When such research has been undertaken, the results on the impact of visual information have been often inconsistent, due to the lack of comparison and the ambiguity of intrinsic evaluation.

Does visual data bolster performance on non-visual tasks? If it does, is this only because we add more data or does it convey complementary quality information compared to a higher quantity of text? Can we achieve comparable performance using small-data if it comes from the right data distribution? Is the modality, the size or the distributional properties of the data that matters? Evaluating on downstream or similarity-type tasks is a good start to compare models and data sources. However, if we want to resolve the ambiguity of intrinsic evaluations and the spurious correlations of downstream results, creating more transparent and human interpretable models is necessary.

This thesis proposes diverse studies to scrutinize the inner "cognitive models" of Embeddings, trained on various data sources and modalities. Our contribution

is threefold. Firstly, we present comprehensive analyses of how various visual and linguistic models behave in semantic similarity and brain imaging evaluation tasks. We analyse the effect of various image sources on the performance of semantic models, as well as the impact of the quantity of images in visual and multi-modal models. Secondly, we introduce a new type of modality: a visually structured, text based semantic representation, lying in-between visual and linguistic modalities. We show that this type of embedding can serve as an efficient modality when combined with low resource text data. Thirdly, we propose and present proof-of-concept studies of a transparent, interpretable semantic space analysis framework.

UCAM-CL-TR-971

Aaron Stockdill:

Automating representation change across domains for reasoning

June 2022, 268 pages, PDF
PhD thesis (Selwyn College, September 2021)

Abstract: Representing a problem well can make it trivial to solve; represent it poorly, and it becomes impossible. But what makes a representation suitable for a problem, and how can we automatically choose the most suitable from a set of alternatives? Choosing an appropriate representation is a difficult, long-standing problem in artificial intelligence; we want to support people in making an appropriate representation selection based on the problem they are solving, their own cognitive strengths, and the representational systems available. A large part of the challenge in choosing alternative representations stems from not knowing what is 'the same': which parts in the problem statement correspond to parts of an analogous statement in a different representation. If instead this choice was automated, users could better understand the problem, and work towards a solution when given a more appropriate representation.

This dissertation contributes a novel approach for the identification of alternative representations of problems through the idea of correspondences. This is a key step towards being able to select representations that are well-suited to enabling problem solutions. Exploiting correspondences, we demonstrate how to compute the informational suitability of alternative representational systems; the practical utility of this is shown with a software implementation. The generality of this theory and implementation is demonstrated by applying both to a domain that is distinct from the one it was developed in. We evaluate our theory and implementation with an empirical study, where we present experts with a similar challenge of evaluating representational system suitability, and comparing their responses with that of our implementation.

The work described in this dissertation creates possibilities for software tools that react to the problem and user: intelligent tutoring systems with multiple ways of explaining concepts to students; or interactive theorem provers that create analogies to help the human prover in finding key insights. The resulting tools centre on the representational needs of the human, not the computer.

UCAM-CL-TR-972

Xuan Guo, Daniel Bates, Robert Mullins, Alex Bradbury:

Muntjac multicore RV64 processor: introduction and microarchitectural guide

June 2022, 27 pages, PDF

Abstract: Muntjac is an open-source collection of components which can be used to build a multicore, Linux-capable system-on-chip. This includes a 64-bit RISC-V core, a cache subsystem, and TileLink interconnect supporting cache-coherent multicore configurations and I/O. Each component is easy to understand, verify, and extend, with most being configurable enough to be useful across a wide range of applications. In its current state, Muntjac achieves 2.17 DMIPS/MHz and 3.01 CoreMark/MHz, and can achieve 80+ MHz (with FPU enabled) when targeting a Xilinx Kintex 7 FPGA. This document provides an overview of Muntjac, the standards it follows and an explanation of design decisions and implementation details.

UCAM-CL-TR-973

Tiago M. L. Azevedo:

Data-driven representations in brain science: modelling approaches in gene expression and neuroimaging domains

July 2022, 136 pages, PDF
PhD thesis (Churchill College, 13 February 2022)

Abstract: The assumptions made before modelling real-world data greatly affect performance tasks in machine learning. It is then paramount to find a good data representation in order to successfully develop machine learning models. When no considerable prior assumption exists on the data, values are directly represented in a "flatten", 1-Dimensional vector space. However, it is possible to go one step further and perceive more complex relational patterns: for example, a Graph-Dimensional space is used to illustrate the more structured way to represent data and their relational inductive bias.

This thesis is focused on these two computational data dimensions across two scales of human

biology: the micro scale of molecular biology using gene expression data, and the macro scale of neuroscience using neuroimaging data. Different modelling approaches will be explored to understand how one can model and represent high-dimensional brain data across the specific needs in the applied fields of these two scales. Specifically, for Graph-Dimensional data two approaches will be developed. Firstly, specific and shared genetic profiles that can be generalisable to external datasets will be extracted by applying multilayer co-expression networks across 49 human tissues. Then, a novel deep learning model will be introduced to leverage the entirety of resting-state fMRI data (i.e., spatial and temporal dynamics), as opposed to previous approaches in the literature that simplify and condense this type of data, while illustrating its robustness in an external multimodal dataset and explainability capacities. For 1-Dimensional data, an interpretable model will be developed for understanding cognitive factors using multimodal brain data.

Overall, the research adopted in this thesis explores explainable data-driven representations and modelling approaches across the multidisciplinary scientific fields of machine learning, molecular biology, and neuroscience. It also helps highlight the contributions of these fields when modelling the brain and its intra- and inter-dynamics across the human body.

UCAM-CL-TR-974

Indigo J. D. Orton:

Dynamic analysis for concurrency optimisation

August 2022, 166 pages, PDF
PhD thesis (Hughes Hall, October 2021)

Abstract: Modern software engineering is, broadly, a continuous activity – many pieces of industrial software are constantly developed, they are never "finished". This process of continuous improvement necessitates small, incremental changes to ensure stability and maintainability of the software and its codebase. This includes incremental changes to improve performance. In this thesis I focus on improvements to the efficiency of concurrency usage, at a source-code level, within a piece of software. These improvements are challenging to indentify and implement as cuncurrency and performance behaviour is only exhibited at runtime, thus requiring dynamic analysis, whilst making incremental changes requires static source-code patches. The challenge is compounded as a codebase evolves, as the efficiency of various concurrency uses may change – an instance of concurrency that was previously beneficial may become inefficient due to the evolution of the software. I present an automatic-program-analysis methodology to indentify potential performance improvements, estimate their quantitative effect, and generate static source-code patches to implement them. Using a proof-of-concept implementation, I present evaluations that demonstarate the methodology's efficacy.

Multicore processors are the default for modern computers and leveraging this concurrency is a key aspect of modern software. Effective use of concurrency can significantly improve software performance, though the inverse is also true – ineffective use can impair software performance. However, as the saying goes "concurrency is hard"; it is fundamentally difficult to statically reason about, let alone optimise. Indeed, many of its properties, especially those related to performance, are only exhibited at runtime.

In this thesis I explore the use of dynamic analysis for concurrency optimisation. I argue this field is under-explored, yet represents a substantial opportunity for improving software performance. A key challenge within this field, and one that extends beyond concurrency, is the generation of static changes (e.g. source-code changes) from dynamic analysis. The gap between static and dynamic domains is well studied in terms of using static analysis to improve dynamic analysis efficiency and using dynamic analysis to confirm static analysis hypothesises (e.g. race-condition detection), however, I argue the gap is understudied when transitioning from the dynamic to the static domain.

UCAM-CL-TR-975

Brett Gutstein:

Memory safety with CHERI capabilities: security analysis, language interpreters, and heap temporal safety

November 2022, 119 pages, PDF
PhD thesis (Trinity College, July 2022)

Abstract: CHERI (Capability Hardware Enhanced RISC Instructions) is a promising research processor-architecture protection model that facilitates memory safety and fine-grained compartmentalization for software. The architecture has reached a mature state and been integrated into Arm's industrial-scale Morello system-on-chip, a large corpus of software has been adapted to support CHERI, and prior work has demonstrated that replacing integer pointers with CHERI capabilities can make C and C++ programs spatially safe. In this dissertation, I identify gaps that limit the ability of current mitigations based on CHERI to deliver real-world vulnerability protection, and I work towards addressing them.

I develop the memory-operations framework (MOF) for reasoning about memory-safety mitigations and the types of attacks they prevent. I apply the MOF to analyze CheriABI, the most sophisticated memory-safety mitigation built atop CHERI. I also evaluate CheriABI's effectiveness in mitigating a set of real-world attacks that targeted devices running

Apple's iOS. Based on this evaluation, I identify two key areas in CHERI-supported memory safety that require improved protections.

One of these areas involves support for contemporary programming language interpreters, which have not previously been adapted to CHERI. Using Apple's JavaScriptCore as a case study, I evaluate the feasibility, source-code compatibility, and security properties of adapting an interpreter that supports just-in-time compilation to CHERI. I determine that such an adaptation is feasible, practical, and can achieve parity with more typical applications in terms of memory protection.

The other area is providing temporal safety for userspace heaps, which CheriABI does not currently support. I introduce novel algorithms and software components that constitute a fully elaborated system for CHERI-based userspace heap temporal safety. I implement the system, which includes the Cornucopia kernel subsystem for sweeping capability revocation and a generic userspace library that encapsulates changes required for memory allocators, in CheriBSD for Morello. Relative to the CHERIvoke algorithm for heap temporal safety, which has previously been published but not implemented on CHERI hardware, the novel algorithms reduce application runtimes by up to 23.5% and pause times by up to 11,000x, potentially making temporal safety with CHERI feasible for large, real-world workloads.

UCAM-CL-TR-976

Hesham Almatary:

CHERI compartmentalisation for embedded systems

November 2022, 142 pages, PDF
PhD thesis (St John's College, February 2022)

Abstract: Embedded system designers are facing an inexorable pressure to add more features and leverage connectivity. This creates potential attack vectors in areas that were not subject to security concerns before. Individuals' privacy could be violated, cars and planes could crash, credit-card details could be stolen, and medical devices could critically malfunction, affecting vital life-concerning tasks or leaking sensitive patients' details. Software compartmentalisation has the potential to manage the attack vector better, defend against unknown future software vulnerabilities, and limit the consequences of potential successful attacks to the compromised component without affecting the rest of the system. Unfortunately, the current state-of-the-art security technologies for embedded systems (e.g., MPUs) are not well-designed for implementing fine-grained software compartmentalisation while meeting embedded systems requirements. They suffer from inherent design issues that limit scalability, compatibility, security, and performance.

This dissertation proposes CompartOS as a new lightweight hardware-software compartmentalisation model building on CHERI (a hardware capability architecture) to secure mainstream and complex embedded software systems. CompartOS is an automatic, linkage-based compartmentalisation model that isolates mutually distrusting linkage modules (e.g., third-party libraries) executing in a single-address-space and single-privilege-ring environment. Further, CompartOS enables the management of faults within software components by introducing support for partial recovery, thus improving availability while maintaining compatibility by requiring minimal development efforts—a critical requirement for many embedded systems.

We have implemented multiple prototypes of compartmentalisation models, including MPU-based protection and CompartOS, in FreeRTOS and compared them in performance, compatibility, security, and availability. Microbenchmarks show that CompartOS' protection-domain crossing is 95% faster than MPU based IPC. We applied the CompartOS model, with low effort, to complex, mainstream systems, including TCP servers, Amazon's OTA updates, and a safety-critical automotive demo. CompartOS not only catches 10 out of 13 FreeRTOS-TCP published vulnerabilities that MPU-based protection (e.g., uVisor) cannot catch but can also recover from them, maintaining the availability of safety critical systems. Further, our TCP throughput evaluations show that our CompartOS prototype is 52% faster than the most relevant and advanced MPU-based compartmentalisation model (e.g., ACES), with a 15% overhead compared to an unprotected system. This comes at an FPGA's LUTs overhead of 10.4% to support CHERI for an unprotected baseline RISC-V processor, compared to 7.6% to support MPU, while CHERI only incurs 1.3% of the registers area overhead compared to 2% for MPU.

UCAM-CL-TR-977

Akshay Jindal:

Motion quality models for real-time adaptive rendering

January 2023, 132 pages, PDF
PhD thesis (St Edmund's College, October 2022)

Abstract: The demand for compute power and transmission bandwidth is growing rapidly as the display technologies progress towards higher spatial resolutions and frame rates, more bits per pixel (HDR), and multiple views required for 3D displays. Advancement in real-time rendering has also made shading incredibly complex. However, GPUs are still limited in processing capabilities and often have to work at a fraction of their available bandwidth due to hardware constraints.

In this dissertation, I build upon the observation that the human visual system has a limited capability to perceive images of high spatial and temporal frequency, and hence it is unnecessary to strive to meet these computational demands. I propose to model the

spatio-temporal limitations of the visual system, specifically the perception of image artefacts under motion, and exploit them to improve the quality of rendering.

I present four main contributions: First, I demonstrate the potential of existing motion quality models in improving rendering quality under restricted bandwidths. This validation is done using an eye tracker through psychophysical experiments involving complex motion on a G-Sync display. Second, I note that the current models of motion quality ignore the effect of displayed content and cannot take advantage of recent shading technologies such as variable-rate shading which allows for more flexible control of local shading resolution. To this end, I develop a new content-dependent model of motion quality and calibrate it through psychophysical experiments on a wide range of content, display configurations, and velocities. Third, I propose a new rendering algorithm that utilises such models to calculate the optimal refresh rate and local shading resolution given the allowed bandwidth. Finally, I present a novel high dynamic range multi-focal stereo display that will serve as an experimental apparatus for next-generation of perceptual experiments by enabling us to study the interplay of these factors in achieving perceptual realism.

UCAM-CL-TR-978

Alexander G. Fraser:

A Next Generation Internet Architecture

February 2023, 118 pages, PDF
Edited by Elisabeth Fraser and foreword by Anil Madhavapeddy and David J. Scott.

Abstract: This report is the unmodified work-in-progress monograph written by Sandy Fraser from 2003 onwards to capture his vision for a global network architecture that would scale to levels of quality and resilience beyond that of anything existing at the time. There are multiple areas of the document that are marked as "to do" or are otherwise incomplete. We believe it to be a valuable historical record of the original research that occurred first at AT&T Labs Research and subsequently at Fraser Research, and have preserved it in its entirety in this technical report.

UCAM-CL-TR-979

Xuan Guo:

Efficient virtual cache coherency for multicore systems and accelerators

February 2023, 202 pages, PDF
PhD thesis (Peterhouse College, September 2022)

Abstract: There is a paradigm shift from general-purpose cores to specialised hardware, which has vastly different programming models. It will be helpful if the existing programming model can be kept and new hardware can co-exist and cooperate with existing userspace software. A virtual cache coherence protocol can be helpful for such task, allowing individual components to perform virtual address accesses without having to include their own hardware for address translation and memory protection. This thesis presents such a protocol, together with tooling and hardware infrastructure that are developed in the process of creating it.

This thesis makes three contributions. The first contribution is in the area of processor simulation techniques. A high-performance simulator is presented for exploring just the behaviour of translation lookaside buffers (TLBs). This is then extended to provide fast cycle-level simulator. The simulator employs an innovative technique to combine binary translation with cycle-level simulation and therefore significantly speedup the simulation process compared to traditional interpretation-based simulators. The simulator built, R2VM, can achieve ˜30 million instructions per second (MIPS) in cycle-level simulation in lockstep execution mode, more than 100x the performance of gem5 in a similar mode of operation. For non-cycle-level fast-forward execution, R2VM can achieve >400 MIPS per thread. This is significantly faster than gem5's 3 MIPS fast-forward execution, and even better than emulators that exploit dynamic binary translation (DBT), such as QEMU.

The second contribution is a collection of open-source processor and system-on-chip (SoC) components, called Muntjac. Muntjac contains implementation of a RISC-V (RV64GC) core with machine and supervisor privilege levels, as well as cache subsystems and interconnect components that utilise TileLink. An untethered Linux-capable example SoC is implemented with these components. Muntjac core can achieve a Dhrystone score of 2.17 DMIPS/MHz and CoreMark score of 3.01 CoreMark/MHz. Muntjac is designed to be modular, verifiable and extendable, and be a good starting point for education, research and industrial applications.

The final contribution is a virtual cache coherence protocol that permits the use of virtually-indexed virtually-tagged (VIVT) L1 caches. The protocol is designed to allow commonly used read-only synonyms to reside in caches while still maintaining correctness in hardware when writable synonyms occur. The protocol is designed and described in detail, and implemented and evaluated on a field programmable gate array (FPGA) using Muntjac. Caches that communicate using the protocol are implemented, and support has been added to a Linux kernel port. Systems with the protocol have lower resource utilisation and higher maximum frequency compared to the physically coherent counterpart as the TLB is removed from the L1 and the critical path of memory access, while still be-

ing comparable in terms of performance per MHz. The flexibility and advantages of the protocol are demonstrated by the creation and integration of easy-to-use accelerators that can be accessed from the general-purpose cores with a low latency.

UCAM-CL-TR-980

Marno van der Maas:

Protecting enclaves from side-channel attacks through physical isolation

March 2023, 120 pages, PDF
PhD thesis (Clare Hall, September 2021)

Abstract: The digital world is taking an increasingly crucial role in our lives. Digital systems control our calendars, how we gain access to our devices and even the vehicles we use for transportation. It is therefore no surprise that security solutions like trusted execution environments have been introduced in many systems ranging from small embedded networking devices to large server racks. One of the main challenges of this ever growing functionality is keeping the trusted computing base small and manageable. Enclave systems are a way to do exactly that: they allow applications to run on the same system as a rich OS while ensuring the confidentiality and integrity of enclave data.

In this thesis I explore the difficulty in protecting enclaves from side-channel attacks in the face of privileged software. I propose a threat model, a methodology to analyze side channels and a new enclave system that adheres to this threat model. Due to the complexities of modern superscalar processors, I conclude that it is undesirable to run enclaves on the same cores as untrusted software due to the performance degradation this would have on regular applications. My new enclave system uses a heterogeneous multi-core processor to physically isolate enclaves on secure cores while regular applications run on fast cores. I show that this system works with a conventional OS by implementing a Linux driver that facilitates management of enclaves and communication between untrusted applications and enclaves. The enclave subsystem only requires a small trusted computing base: a trusted management shim to interface the Linux driver with the enclave hardware. I evaluate hardware implementation approaches in simulation and on a field-programmable gate array. The evaluation shows that this system is reasonable in communication overhead, memory footprint, runtime and hardware area. Thus, physical isolation is a feasible way to protect enclaves from side-channel attacks in modern enclave systems.

UCAM-CL-TR-981

Kayvan Memarian:

The Cerberus C semantics

May 2023, 290 pages, PDF
PhD thesis (Wolfson College, October 2022)

Abstract: The C programming language, has since its introduction fifty years ago, become central to our computing infrastructure. It would therefore be desirable to have a precise semantics, that in particular could serve as a reference for implementers of compiler, analysis tools, etc. The ISO standard that notionally defines C suffers from two issues. First, as an inevitable result of being written in prose, it is imprecise. Second, it does not really attempt to precisely define the memory model. These shortcomings leave C's many obscure corners open to differing interpretations, and this is especially apparent when it comes to the memory model. While system programmers often rely on a very concrete view of pointers (even more concrete than what the ISO standard actually offers), compiler implementers take a more abstract view. Some optimisations, in particular ones based on alias analysis, reason about how pointer values are constructed during the program execution instead of only considering their representation, and perform transformations that would not be sound with respect to a concrete view of memory.

In this thesis, we present Cerberus, an executable model for a substantial fragment of C11. The dynamics of C is expressed as a compositional translation to a purpose-built language called Core. With this semantics by elaboration, we make the subtleties of C's expressions and statements explicit in the form of syntax in the Core representation. For these aspects of the semantics of C, the existing ISO standard has remained in agreement with de facto practice, and our model follows it. The elaboration allows for a model of the dynamics that is relatable to the ISO prose, and that is tractable despite the complexity of C.

For the memory model, as the de facto standards do not exist as coherent specifications that we could formalise, we opted at the start of this work for an empirical study of the design space for a realistic memory model. We surveyed the mainstream practice in C system programming and the assumptions made by compiler implementers. From this study and through engagement with WG14, the working group authoring the ISO standard, we have designed a family of memory models where pointer values have a provenance. At the time of writing one of these models is being published in collaboration with some members of WG14 as a ISO technical specification to accompany the standard.

We have dedicated significant effort in the executability of the model, both in term of performance and the scope of our frontend, which allows Cerberus to be used on medium scale off-the-self C programs with only limited amount of modification.

With this work we show that by suitably tailoring the target language, a semantics by elaboration produces a tractable definition of a large fragment of C.

UCAM-CL-TR-982

Robert N. M. Watson, Graeme Barnes, Jessica Clarke, Richard Grisenthwaite, Peter Sewell, Simon W. Moore, Jonathan Woodruff:

Arm Morello Programme: Architectural security goals and known limitations

July 2023, 8 pages, PDF

Abstract: Arm's Morello prototype incorporates a first-generation CHERI-enabled Armv8-A CPU prototype. We have developed Morello to enable CHERI-based research by a growing community of researchers seeking access to potentially transformative architectural security improvement. This includes supporting experimentation, evaluation, and demonstration across microarchitecture and software. Morello is an exciting opportunity to work with – and improve – CHERI, and we seek your help and collaboration in preparing CHERI for mainstream use.

The purpose of this document is to lay out the specific architectural security objectives of the Arm Morello prototype, as well as areas that fell out of scope for the project. We invite not only your feedback, but also your collaboration, in helping us to create a future class of CHERI-extended processors that dramatically enhance software security.

UCAM-CL-TR-983

Maria Bada, Alice Hutchings, Yanna Papadodimitraki, Richard Clayton:

An evaluation of police interventions for cybercrime prevention

July 2023, 80 pages, PDF

Abstract: This study explores the effects of the cybercrime intervention workshops organised by the National Crime Agency between March 2018 and January 2020 in the UK. The interventions are designed to deter those suspected of or involved in cybercrime from engaging in (further) offending. We use a retrospective, cross-sectional, anonymous survey to collect data on the interventions and criminal justice; we compare the intervention group (workshops) to a control group which includes police visits and cease and desist letters. Based on this, we explore harmful online behaviours and perceptions relating to criminal justice. In the technical report, we present our research questions and hypotheses, followed by our methodological approach and analysis.

UCAM-CL-TR-984

Peter David Rugg:

Efficient spatial and temporal safety for microcontrollers and application-class processors

July 2023, 189 pages, PDF
PhD thesis (Churchill College, December 2022)

Abstract: This thesis discusses the implementation of Capability Hardware Enhanced RISC Instructions (CHERI) secure capabilities for RISC-V microarchitectures. This includes implementations for three different scales of core, including microcontrollers and the first open application of CHERI to a superscalar processor. Tradeoffs in developing the architecture and performant microarchitecture are investigated. The processors are then used as a platform to conduct research in reducing the overheads for achieving temporal safety with CHERI.

CHERI offers a contemporary cross-architecture description of capabilities. The initial design was previously carried out in a single MIPS processor. Based on its success in this context, this thesis investigates the microarchitectural implications across a wider range of processors. To improve adoption, this work is performed on the more contemporary RISC-V architecture. The thesis also explores the microarchitectural implications of architectural decisions arising from the adaptation of CHERI to this new context.

The first implementations are to the Piccolo and Flute microcontrollers. They present new tradeoffs, for example being the first CHERI implementations supporting a merged register file and capability mode bit. The area and frequency implications are evaluated on FPGA, and the performance and power overheads are investigated across a range of benchmarks. To validate correctness, the processors are integrated into a new TestRIG infrastructure.

This thesis also develops the first open instantiation of CHERI for a superscalar out-of-order application-class core: RiscyOO. This presents new questions due to the very different design of the more sophisticated microarchitecture, and highlights more architectural tradeoffs. Again, the processor is evaluated on FPGA, investigating area, frequency, power, and performance. This allows the first analysis of how the overheads scale differently across different sizes of core.

Finally, the augmented processors are used as a platform to refine the use of CHERI for temporal safety. Significant improvements are made to the architecture-neutral model used for revocation sweeps. In addition, processor-specific acceleration of revocation is performed, including new approaches for caching capability tags.

UCAM-CL-TR-985

Sharan S. Agrawal:

Scalable agent-based models for optimized policy design: applications to the economics of biodiversity and carbon

August 2023, 84 pages, PDF
MPhil thesis (Darwin College, June 2023)

Abstract: As the world faces twinned crises of climate change and biodiversity loss, the need for integrated policy approaches addressing both is paramount. To help address this, a new agent-based model (ABM), the VDSK-B, was developed. Using Dasgupta's review of the economics of biodiversity, it builds on the Dystopian Schumpeter meets Keynes (DSK) climate economics model to link together the climate, economy and biosphere. To our knowledge, this is the first ABM proposed that integrates all 3 key elements.

Existing ABM frameworks struggled with global policy design needs due to their inability to scale to planetary-sized models, and optimize model parameters at the large scales needed for policy design. A new ABM framework called SalVO was built using a formalism for ABMs that expressed agent updates as recursive applications of pure agent functions. This formalism differs from existing computational ABM models but is shown to be expressive enough to emulate a Turing complete language. SalVO is built on a JAX backend and designed to be scalable, vectorized, and optimizable. Employing hardware acceleration, tests showed it was more performant and more able to scale on a single machine than any existing ABM framework, such as FLAME (GPU).

Techniques for using backpropagation to create optimized policies differentiable, deterministic ABMs were further extended and implemented in SalVO. A novel protocol, GP-ABM, using William's REINFORCE algorithm, was developed to optimize parameters in non-differentiable, stochastic ABMs. Both approaches are shown to be able to optimize ABMs for thousands of parameters, with backpropagation learning a highly non-trivial policy to move the centroid of a flock to a target location. This represents an innovation over current state-of-the-art techniques, such as Simulated Minimum Distance, which do not scale past fifty at most.

Finally, the VDSK-B model was implemented in SalVO, showing its capability of expressing highly complex ABMs. SalVO proved to be highly scalable, running a 5x bigger version of VDSK-B using just 4% of the time taken by the current open-source implementation, significantly strengthening its position as a preferred tool for large-scale ABM studies. While further work remains to be done on VDSK-B's calibration and correctness, SalVO's marriage of speed, scale and optimization has the potential to reshape how we approach, design, and apply agent-based models.

UCAM-CL-TR-986

Robert N. M. Watson, Jessica Clarke, Peter Sewell, Jonathan Woodruff, Simon W. Moore, Graeme Barnes, Richard Grisenthwaite, Kathryn Stacer, Silviu Baranga, Alexander Richardson:

Early performance results from the prototype Morello microarchitecture

September 2023, 19 pages, PDF

Abstract: Arm's Morello is a first-generation, CHERI-enabled prototype CPU based on Arm's Neoverse N1, as found in the N1SDP evaluation board. CHERI is an architectural feature that promises to dramatically improve software security through fine-grained memory protection and scalable compartmentalization. Supported by UKRI, Morello is a research platform to evaluate CHERI at an industrial scale through composition with a rich, contemporary, high-performance microarchitecture and full software stack at a scale unobtainable via ISA emulators or hardware simulators. This report provides a first look at software performance on Morello working with both the baseline microarchitecture and modified designs on FPGA, exploring the performance of the SPECint benchmark suite using multiple code generation models. This technical report is a snapshot of a living document tracking ongoing performance investigation and experimentation.

UCAM-CL-TR-987

Robert N. M. Watson, Peter G. Neumann, Jonathan Woodruff, Michael Roe, Hesham Almatary, Jonathan Anderson, John Baldwin, Graeme Barnes, David Chisnall, Jessica Clarke, Brooks Davis, Lee Eisen, Nathaniel Wesley Filardo, Franz A. Fuchs, Richard Grisenthwaite, Alexandre Joannou, Ben Laurie, A. Theodore Markettos, Simon W. Moore, Steven J. Murdoch, Kyndylan Nienhuis, Robert Norton, Alexander Richardson, Peter Rugg, Peter Sewell, Stacey Son, Hongyan Xia:

Capability Hardware Enhanced RISC Instructions: CHERI Instruction-Set Architecture (Version 9)

September 2023, 523 pages, PDF

Abstract: This technical report describes CHERI ISAv9, the ninth version of the CHERI architecture being developed by SRI International and the University of Cambridge. This design captures thirteen years of research, development, experimentation, refinement, formal analysis, and validation through hardware and software implementation.

CHERI introduces an architecture-neutral capability-based protection model, which has been instantiated in various commodity base architectures to give CHERI-RISC-V, Arm's prototype Morello architecture, and (sketched) CHERI-x86-64. It enables software to efficiently implement fine-grained memory protection and scalable software compartmentalization, by providing strong, deterministic, efficient mechanisms to support the principles of least privilege and intentional use in the execution of software at multiple levels of abstraction, preventing and mitigating vulnerabilities. Design goals include incremental adoptability from current ISAs and software stacks, low performance overhead for memory protection, significant performance improvements for software compartmentalization, formal grounding, and programmer-friendly underpinnings.

CHERI blends traditional paged virtual memory with an in-address-space capability model that includes capability values in registers, capability instructions, and tagged memory to enforce capability integrity. This hybrid approach addresses the performance and robustness issues that arise when trying to express more secure, privilege minimising programming models, above conventional architectures that provide only MMU-based protection. CHERI builds on the C-language fat-pointer literature: its capabilities can describe fine-grained regions of memory, and can be substituted for data or code pointers in generated code, protecting data and improving control-flow robustness. Strong capability integrity and monotonicity properties allow CHERI to express a variety of protection idioms, from enforcing valid C-language pointer provenance and bounds checking to implementing the isolation and controlled communication structures required for software compartmentalization.

CHERI's hybrid approach allows incremental adoption of capability-oriented design: critical components can be ported and recompiled to use capabilities throughout, providing finegrain memory protection, or be largely unmodified but encapsulated in ways that permit only controlled interaction. Potential early deployment scenarios include low-level software Trusted Computing Bases (TCBs) such as separation kernels, hypervisors, and operating-system kernels, userspace TCBs such as language runtimes and web browsers, and particularly high-risk software libraries such as data compression, protocol parsing, and image processing (which are concentrations of both complex and historically vulnerability-prone code exposed to untrustworthy data sources).

CHERI ISAv9 is a substantial enhancement to prior ISA versions. CHERI-RISC-V has replaced CHERI-MIPS as the primary reference platform, and CHERI-MIPS has been removed from the specification. CHERI architectures now always use merged register files where existing general-purpose registers are extended to support capabilities. CHERI architectures have adopted two design decisions from Arm Morello: 1) CHERI architectures now clear tags rather than raising exceptions if an instruction attempts a non-monotonic modification of a capability; and 2) DDC and PCC no longer relocate legacy memory accesses by default. CHERI-RISC-V has received numerous updates to serve as a better baseline for an upstream standard proposal including a more mature definition of compressed instructions in capability mode. CHERI-x86-64 now includes details of extensions to existing x86 instructions and proposed new instructions in a separate ISA reference chapter along with various other updates.

UCAM-CL-TR-988

Vadim Zaliva, Kayvan Memarian, Ricardo Almeida, Jessica Clarke, Brooks Davis, Alex Richardson, David Chisnall, Brian Campbell, Ian Stark, Robert N. M. Watson, Peter Sewell:

CHERI C semantics as an extension of the ISO C17 standard

October 2023, 11 pages, PDF

Abstract: This document provides a specification for CHERI C, adhering to the style, conventions, and terminology found in the ISO C17 standard. Alongside the ISO/IEC 9899:2018 standard text and the "A Provenance-aware Memory Object Model for C" draft specification, it offers a comprehensive specification of the CHERI C language.

UCAM-CL-TR-989

Dimitrios Los:

Balanced allocations under incomplete information: New settings and techniques

November 2023, 230 pages, PDF
PhD thesis (St John's College, April 2023)

Abstract: In the balanced allocations framework, there are m balls to be allocated into n bins with the aim of minimising the maximum load of any of the bins, or equivalently minimising the "gap", i.e., the difference between the maximum load and the average load. In this dissertation, we focus on the "heavily-loaded case" where $m \gg n$, which tends to be more challenging to analyse.

In a decentralised setting, the simplest process is One-Choice, which allocates each ball to a bin sampled uniformly at random. It is well-known that w.h.p. Gap(m) = Θ(sqrt($m/n \cdot \log n$)) for any $m \gg n$. A great improvement over this is the Two-Choice process [ABKU99, KLM96], which allocates each ball to the least loaded of two bins sampled uniformly at random. Berenbrink, Czumaj, Steger, and Vöcking (2006) showed that w.h.p. Gap(m) = $\log_2 \log n + \Theta(1)$ for any $m \geq n$. This improvement is known as the "power of two choices". It has found several applications in hashing, load balancing and routing; and its importance was recently recognised in the 2020 ACM Theory and Practice Award.

In this dissertation, we introduce a set of techniques based on "potential functions". These enable us to analyse (both in terms of gap and load distribution) a wide range of processes and settings in the heavily-loaded case and to establish interesting insights in the balanced allocations framework:

• We analyse variants of the Two-Choice process which trade sample efficiency, completeness of information and gap guarantees. For the (1+β)-process which mixes One-Choice and Two-Choice with probability β in (0, 1], we prove tight bounds for small and large β, extending the results of Peres, Talwar and Wieder (2015). Another sample efficient family is that of Two-Thinning processes, which allocate to the two sampled bins in an online manner. For Two-Thinning processes that use as a decision function thresholds relative to the average load or thresholds in the rank domain, we establish tight bounds and also resolve a conjecture by Feldheim and Gurel-Gurevich (2021). We also quantify trade-offs for two-sample processes between the number of queries and the gap bound, establishing a "power of two queries" phenomenon.

• We analyse the Two-Choice process with random, adversarial and delay noise, proving tight bounds for various settings. In the adversarial setting, the adversary can decide in which of the two sampled bins the ball is allocated to, only when the two loads differ by at most g. The analysis of this setting implies bounds for settings with random noise and delay.

For the setting where load information is updated periodically every b steps, for $b = n$ we tighten the bound of [BCEFN12] to Θ($\log n$ / $\log \log n$) and prove that Two-Choice is optimal in this setting for any in [$n \cdot \exp(-\log^c n)$, $n \log n$] for any constant $c > 0$. For b in [$n \log n$, n^3], we show that Two-Choice achieves w.h.p. a $\Theta(b/n)$ gap, while surprisingly the (1+β)-process with appropriately chosen β achieves w.h.p. a Θ(sqrt($b/n \cdot \log n$)) gap, which is optimal over a large family of processes. This proves that in the presence of outdated information, less aggressive strategies can outperform the greedy processes (such as Two-Choice), which has been empirically observed in the queuing setting [D00, M00] for centralised processes since 2000, but to the best of our knowledge has not been formally proven.

• Next we analyse Two-Choice in the graphical setting, where bins are vertices of a graph and each ball is allocated to the lesser loaded of the vertices adjacent to a randomly sampled edge. We extend the results of Kenthapadi and Panigrahy (2006) proving that for dense expanders in the heavily-loaded case the gap is w.h.p. O(log log n). In the presence of weights, we make progress towards [Open Problem 1, PTW15] by proving that for graphs with conductance ϕ, the gap is w.h.p. O(log n / ϕ).

• Further, we introduce and analyse processes which can allocate more than one balls to a sampled bin. We prove that these processes achieve w.h.p. an O(log n) gap (which also applies for any d-regular graph), while still being more sample-efficient than One-Choice ("power of filling").

• For the Memory process that can store bins in a cache, we generalise the O(log log n) gap bound by Mitzenmacher, Prabhakar and Shah (2002) to the heavily-loaded case and prove a matching lower bound. Further, in the presence of heterogeneous sampling distributions, we establish a striking difference between Two-Choice (or even d-Choice with d = O(1)) and Memory, showing that for the later the gap is bounded, while for the former it is known to diverge [W07] ("power of memory").

UCAM-CL-TR-990

Euan Ong:

Probing the foundations of neural algorithmic reasoning

December 2023, 75 pages, PDF
BA dissertation (Trinity College, May 2023)

Abstract: While the field of neural algorithmic reasoning (NAR) — training neural networks to imitate algorithms and using them as algorithmic inductive biases in real-world problems — has risen in popularity, there has been no investigation confirming that its fundamental claims hold in general. Indeed, we argue that such an investigation has so far been infeasible, due to the lack of a general extensible library creating a very high barrier to entry for reproductions and systematic studies.

As such, we develop an extensible laboratory for NAR, by introducing a novel framework for multi-domain, type-driven, declarative ML, and using its components to derive flexible NAR pipelines from first principles through the paradigm of representations-as-types. We use this laboratory to perform systematic analyses, reproductions and comparisons of prior work in NAR, matching (and often beating) state-of-the-art performance across various domains by identifying and alleviating bottlenecks across popular NAR frameworks and architectures.

We then conduct a systematic investigation into the fundamental claims of NAR, in the context of a new synthetic dataset inspired by recent work in neural algorithmics. Through a series of statistically-robust ablation tests, while we confirm the established result

that algorithmic modules beat non-algorithmic baselines, we find evidence to refute one of the central claims of NAR, showing that neural algorithmic processors (NAPs) do not overcome the 'scalar bottleneck' of differentiable algorithmic black-boxes (sDABs).

Based on our observations, we develop a new hypothesis to replace this claim: that sDABs instead suffer from an 'ensembling bottleneck' of not being able to execute multiple instances of the same algorithm in parallel, which is alleviated not by NAPs, but by simply using an unfrozen, structurally-aligned neural network. And, through exploring the effects of parallelising sDABs, we not only find strong evidence in support of this hypothesis, but also achieve a long-standing goal of neural algorithmics: developing a way to deterministically distill an algorithm into a robust, high-dimensional processor network that preserves both the efficiency and correctness guarantees of sDABs while avoiding their performance bottleneck.

UCAM-CL-TR-991

Jacky W. E. Kung:

Porting a mix network client to mobile

December 2023, 57 pages, PDF
BA dissertation (St Edmund's College, May 2023)

Abstract: This project set out to investigate the feasibility of mix network clients on the mobile ecosystem. It considers the Android operating system, and Nym, a production-grade mix network infrastructure based on the abstract Loopix architecture first presented in 2017. The goal of the project was to produce a minimal working prototype, and present an evaluation of the trade-offs necessary for an efficient implementation in the Android ecosystem. Nym's client codebase written in Rust has been successfully ported over to Android after adjusting parts of the code and constructing the compilation toolchain. An exploration of the performance effects of compilation parameters and mixnet parameters is presented. Two extension tasks were completed: a semi-automated compilation pipeline, and further evaluation using measurements taken using the custom hardware provided by my supervisor. The repository also contains, as a side-product, a Rust library that provides a friendly interface between code that runs across the Rust and Kotlin languages.

UCAM-CL-TR-992

Allison Randal:

Transient execution vulnerabilities in the security context of server hardware

December 2023, 145 pages, PDF
PhD thesis (Robinson College, July 2023)

Abstract: Many mitigations have been proposed and implemented for many variants of the transient execution vulnerabilities, and while the Meltdown-type exception-based transient execution vulnerabilities have proven to be tractable, Spectre-type vulnerabilities and other speculation-based transient execution vulnerabilities have been far more resistant to countermeasures. For smaller-scale embedded systems or security-focused hardware such as a cryptographic system or a root-of-trust (RoT), eliminating speculation is widely accepted as a reasonable approach to improving security. But, for larger-scale and general-purpose hardware, eliminating speculation is often dismissed as inconceivable, though the claim that speculation is required for adequate performance is rarely supported by concrete performance results. The performance results we do have from several independent strands of research over the past few decades have shown that speculation features on large-scale server hardware do not offer the same performance advantages as on smaller-scale hardware, so eliminating speculation on large-scale server hardware does not harm performance as much as we might suspect. And selective speculation techniques have shown that speculation-based transient execution vulnerabilities can be mitigated by a partial elimination of speculation, so we can preserve some of the performance of speculation while subduing the security risk. In order to demonstrate that eliminating speculation is a feasible approach to mitigating the transient execution vulnerabilities on large-scale server hardware, this work considers three alternative approaches that partially or completely eliminate speculative execution: heterogeneous multicore systems combining speculative and non-speculative cores; entirely non-speculative microarchitectures; and selective speculation microarchitectures.

UCAM-CL-TR-993

Paul M. Scherer:

Distributional and relational inductive biases for graph representation learning in biomedicine

April 2024, 164 pages, PDF
PhD thesis (Gonville and Caius College, May 2023)

Abstract: The immense complexity in which DNAs, RNAs, proteins and other biomolecules interact amongst themselves, with one another, and the environment to bring about life processes motivates the mass collection of biomolecular data and data-driven modelling to gain insights into physiological phenomena. Recent predictive modelling efforts have focused on deep representation learning methods which offer a flexible modelling paradigm to handling high dimensional data at scale and incorporating inductive biases. The emerging field of representation learning on

graph structured data opens opportunities to leverage the abundance of structured biomedical knowledge and data to improve model performance.

Grand international initiatives have been coordinated to organise and structure our growing knowledge about the interactions and putative functions of biomolecular entities using graphs and networks. This dissertation considers how we may use the inductive biases within recent graph representation learning methods to leverage these structures and incorporate biologically relevant relational priors into machine learning methods for biomedicine. We present contributions in two parts with the aim to foster research in this multidisciplinary domain and present novel methods that achieve strong performance through the use of distributional and relational inductive biases operating on graph-structured biomedical knowledge and data.

The first part is concerned with consolidating and expanding the current ecosystem of practical frameworks dedicated to graph representation learning. Our first contribution presents Geo2DR, the first practical framework and software library for constructing methods capable of learning distributed representations of graphs. Our second contribution, Pytorch Geometric Temporal, is the first open source representation learning library for dynamic graphs, expanding the scope of research software on graph neural networks that were previously limited to static graphs.

The second part presents three methods wherein each contribution tackles an active biomedical research problem using relational structures that exist within different aspects of the data. First, we present a methodology for learning distributed representations of molecular graphs in the context of drug pair scoring. Next, we present a method for leveraging structured knowledge on the variables of gene expression profiles to automatically construct sparse neural models for cancer subtyping. Finally, we present a state-of-the-art cell deconvolution model for spatial transcriptomics data using the positional relationships between observations in the dataset.

UCAM-CL-TR-994

Nicholas Boucher:

Deception and defense from machine learning to supply chains

May 2024, 161 pages, PDF
PhD thesis (Clare College, December 2023)

Abstract: Broad classes of modern cyberattacks are dependent upon their ability to deceive human victims. Given the ubiquity of text across modern computational systems, we present and analyze a set of techniques that attack the encoding of text to produce deceptive inputs to critical systems. By targeting a core building block of modern systems, we can adversarially manipulate dependent applications ranging from natural language processing pipelines to search engines to code compilers. Left undefended, these vulnerabilities enable many ill effects including uncurtailed online hate speech, disinformation campaigns, and software supply chain attacks.

We begin by generating adversarial examples for text-based machine learning systems. Due to the discrete nature of text, adversarial examples for text pipelines have traditionally involved conspicuous perturbations compared to the subtle changes of the more continuous visual and auditory domains. Instead, we propose imperceptible perturbations: techniques that manipulate text encodings without affecting the text in its rendered form. We use these techniques to craft the first set of adversarial examples for text-based machine learning systems that are human-indistinguishable from their unperturbed form, and demonstrate their efficacy against systems ranging from machine translation to toxic content detection. We also describe a set of defenses against these techniques.

Next, we propose a new attack setting which we call adversarial search. In this setting, an adversary seeks to manipulate the results of search engines to surface certain results only and consistently when a hidden trigger is detected. We accomplish this by applying the encoding techniques of imperceptible perturbations to both indexed content and queries in major search engines. We demonstrate that imperceptibly encoded triggers can be used to manipulate the results of current commercial search engines, and then describe a social engineering attack exploiting this vulnerability that can be used to power disinformation campaigns. Again, we describe a set of defenses against these techniques.

We then look to compilers and propose a different set of text perturbations which can be used to craft deceptive source code. We exploit the bidirectional nature of modern text standards to embed directionality control characters into comments and string literals. These control characters allow attackers to shuffle the sequence of tokens rendered in source code, and in doing so to implement programs that appear to do one thing when rendered to human code reviewers, but to do something different from the perspective of the compiler. We dub this technique the Trojan Source attack, and demonstrate the vulnerability of C, C++, C#, JavaScript, Java, Rust, Go, Python, SQL, Bash, Assembly, and Solidity. We also explore the applicability of this attack technique to launching supply chain attacks, and propose defenses that can be used to mitigate this risk. We also describe and analyze a 99-day coordinated disclosure that yielded patches to dozens of market-leading compilers, code editors, and code repositories.

Finally, we propose a novel method of identifying software supply chain attacks that works not only for Trojan Source attacks, but for most forms of supply chain attacks. We describe an extension to compilers dubbed the Automated Bill of Materials, or ABOM, which embeds dependency metadata into compiled binaries. Specifically, hashes of each source code file consumed by a compiler are embedded into its emitted binary, and these hashes are included recursively into all

downstream dependencies. They are stored in a highly space and time efficient probabilistic data structure that requires an expected value of just 2.1 bytes to represent each unique dependency source code file. With ABOMs, it becomes possible to detect all naturally occurring and most adversarially induced vulnerabilities used for supply chain attacks in downstream software by querying binaries for the presence of poisoned dependencies without the need to locate tangible indicators of compromise.

In this thesis, we therefore demonstrate how weaknesses in a core building block of modern systems – text encodings – can cause failures in a wide range of domains including machine learning, search engines, and source code. We propose defenses against each variant of our attack, including a new tool to identify most generic software supply chain attacks. We believe that these techniques will be useful in securing software ecosystems against the next generation of attacks.

UCAM-CL-TR-995

Henry Batchelor:

Fragment-template power-analysis attacks against microcontroller implementations of the 32-bit stream cipher ChaCha

July 2024, 59 pages, PDF
MEng dissertation (Selwyn College, May 2024)

Abstract: ChaCha is a widely adopted stream cipher, used for both random number generation and encryption. I propose a factor graph of ChaCha to improve the success rate of side-channel attacks that provide leakages throughout the entire execution of the algorithm. I also assess (fragment) template attacks against several implementations of ChaCha to demonstrate that the factor graph is helpful when working with actual side-channel attacks.

These attacks could fully recover the correct key from an 8-bit implementation. In contrast, a 32-bit implementation, with most of the state held in registers, was significantly more challenging to attack. An adversary with access to 10 power traces and an incremented counter could achieve a success rate of 14.6%. For a 32-bit implementation, with lots of SRAM activity, an attacker could successfully recover the key in 2.6% of cases from a single trace.