



Computing and Information Systems
2ND ANNUAL DOCTORAL COLLOQUIUM

Program
23rd July 2014

Welcome to the CIS 2014 Doctoral Colloquium!

Welcome to the 2nd Computing and Information Systems (CIS) Doctoral Colloquium (DC) at the University of Melbourne. CIS DC offers PhD students a tailored full-day forum where they are given the opportunity to present and discuss their research.

To our delight, we have experienced a 48% increase in the number of submissions this year. This extraordinary participation encouraged us to put our best foot forward in organising the event, so that all attendees would benefit highly from it.

This year, the CIS DC has crafted a stimulating environment for students to broaden their network and receive supportive and constructive feedback on their research plans. We have achieved this by advocating for extensive involvement from University industry partners along with internal and external senior researchers at the CIS department. The CIS DC 214 offers a diverse academic menu comprising of paper presentations, interactive posters, and flash talks, headlined by keynote speaker, and capped off by conferring awards to the best five paper and poster presentations.

CIS DIC 2014 was organized by a group of RHD students at the CIS department of Melbourne University. As chair of the Organising Committee, it is my privilege to thank the team for the countless hours of effort they have put in over the course of the past 5 months.

Also, a special thank you to Professor Justin Zobel and Rhonda Smithies for their continuous support.

Once again, welcome to CIS DC 2014!

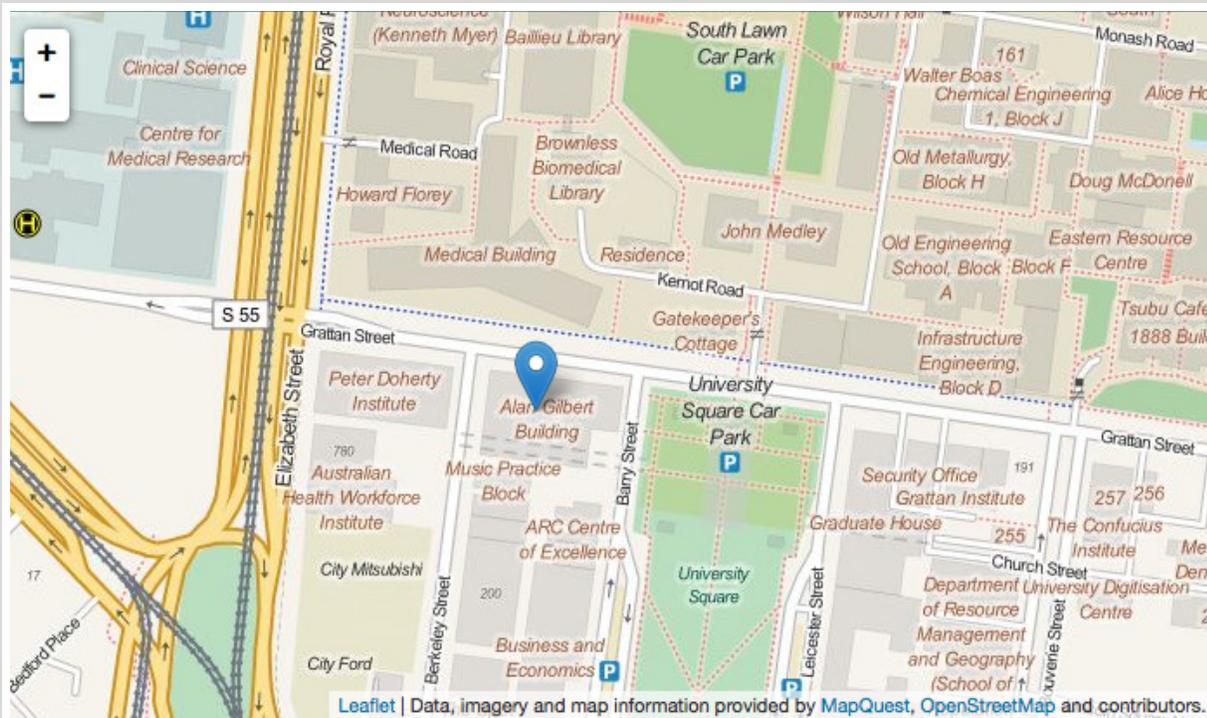
We all look forward to interacting with you today!

Behnaz R. Yeganeh,
Committee Chair

Conference Location: Alan Gilbert Building



Map of Conference Venue



Conference Program

8:30	Registration and Arrival Tea/Coffee	Foyer, Level 1
9:00	Opening and Welcome Address Prof. Justin Zobel	Lecture Theatre 2
9:15	Keynote : Connecting theory & practice - a lesson from bio-informatics Dr. Tom Conway	Lecture Theatre 2
	Cloud Computing (Session 1A) Lecture Theatre 3	Information Systems (Session 1B) Lecture Theatre 2
10:00	A Performance Optimization Scheme for Mobile Cloud Bowen Zhou	The influence of personality traits on online information behaviour during political campaigns Miguel Wood, Simon Milton and Graeme Shanks
10:15	Resource Overbooking in IaaS Cloud Provider Chanh Nguyen	A conceptual model for analysing informal learning in online social networks for health professionals Xin Li and Kathleen Gray
10:30	Fault-Tolerant Workflow Scheduling Using Spot Instances on Clouds Deepak Poola, Kotagiri Ramamohanarao and Rajkumar Buyya	Task Complexity and Codification Marion Zalk, Rachele Bosua and Rajeev Sharma
10:45	Internet of Things: A Roadmap for Future Internet Direction Farzad Khodadadi	Online Medical Consultation: A review of current practices Ibrahim AL-Mahdi
		Life Science (Session 1C) Lecture Theatre 2
11:00	Energy Efficient Software Defined Clouds Jungmin Son	Performance evaluation of spectral unmixing method based on different ways of training data selection Liyan Liu, Andrey Kan and Christopher Leckie
11:15	Resource Provisioning and Scheduling Algorithm for Workflows on Clouds Maria A Rodriguez and Rajkumar Buyya	Retinal Cross Sectional Layer Segmentation using Optical Coherence Tomography Md Akter Hussain, Alauddin Bhuiyan and Kotagiri Ramamohanarao
11:30	Virtual Machine Customization and Task Mapping Model for Efficient Allocation of Cloud Data Center Resources S. F. Piraghaj, R. N. Calheiros and R. Buyya	An Effective Automated System for Grading Severity of Retinal Arteriovenous Nicking in Colour Retinal Images Pallab Kanti Roy, Uyen T. V. Nguyen, Alauddin Bhuiyan and Kotagiri Ramamohanarao
		Database, Machine Learning & Security (Session 1D) Lecture Theatre 2
11:45	Integrated Framework for Cloud-based Interoperability between Social Network Websites Safiollah Heidari and Rajkumar Buyya	Shooting the Breeze Involves neither Shooting nor a Breeze: Predicting the Compositionality of Multiword Expressions Bahar Salehi
12:00	SLA Based Resource Scheduling for Big Data Applications Yali Zhao	Semi-supervised Multilingual POS tagging Long Duong
12:15	Virtual Machine Allocation Policies against Co-resident Attacks in Cloud Computing Yi Han, Jeffrey Chan, Tansu Alpcan and Christopher Leckie	One-Way Games Andres Abeliuk, Gerardo Berbeglia and Pascal Van Hentenryck

12:30		Lunch Break	
Human Computer Interaction (Session 2A)		Database, Machine Learning & Security (Session 2B)	
Lecture Theatre 3		Lecture Theatre 2	
13:15	Developing an opportunistic online psychosocial screening tool for young people in general practice Marianne Webb, Greg Wadley and Sylvia Kauer, Lena Sanc		Improved k-centre Clustering with MapReduce Jessica McClintock
13:30	Help me keep my stuff safe Sarah Ellen Webber		Designing Organization-Aware Agents using OJazzIC Kathleen Keogh
13:45	Understanding the user experience of running with mixed reality stories Alexander Kan		Reliable power transmission networks Rodrigo R. Gumucio E., Carleton Coffrin and Pascal Van Hentenryck
14:00	Non-verbal Interactions in Domestic Video Conferencing Behnaz R. Yeganeh		Neighborhood Component Analysis for Feature Learning Zay Maung Maung Aye, Kotagiri Ramamohanarao and Benjamin Rubinstein
14:15	Won't you be my neighbour? Clustered borrowing in libraries as evidence for shelf browsing Dana McKay		Mining User Interactions and Activities using Mobile Data while Maintaining Privacy Anthony Quattrone
14:30	Supporting Tele-consultation through Tangible Interfaces Deepti Aggarwal		Privacy-Preserving Collaborative Anomaly Detection Sarah M. Erfani, Yee Wei Law, Shanika Karunasekera, Christopher A. Leckie and Marimuthu Palaniswami
14:45	Experiences with Mobile Mental Health Therapies Fernando Estrada		Automatically Recognizing Places of Interests/ Activities from Unreliable GPS Data and Smart Phone Sensors Tanusri Bhattacharya, Lars Kulik and James Bailey

Evening session

15:00	Afternoon Tea & Coffee	
	Human Computer Interaction (Session 3A) Lecture Theatre 3	Database, Machine Learning & Security (Session 3B) Lecture Theatre 2
15:30	Communication Technologies during Family Mealtime Hasan Shahid Ferdous	Principled Dictionary Pruning for Low-Memory Corpus Compression Jiancong Tong
15:45	Screen Ecologies and the Future of Domestic Gaming Marcus Carter, Bjorn Nansen and Martin Gibbs	Predicting at-risk students in Massive Open Online Courses Jiazhen He, James Bailey, Rui Zhang and Benjamin Rubinstein
	Software Engineering (Session 3C) Lecture Theatre 3	
16:00	Loop Untangling Kathryn Francis	Improved Feature Transformations for Classification using Density Estimation Yamuna Kankanige and James Bailey
16:15	Automated debugging using program spectra Neelofar, Lee Naish and Rao Kotagiri	FILTA : Better View Discovery from Collections of Clusterings via Filtering Yang Lei, Nguyen Xuan Vinh, Jeffrey Chan and James Bailey
16:30	Review of Interdependence in Coactive Design Ronal Singh	Bandwidth-Efficient Convergence of Gossip-based Recommendation Systems Irum Fahim Bukhari, Aaron Harwood and Shanika Karunasekera
16:45	Refreshments	
	Flash Talks Lecture Theatre 3	Poster Exhibition Foyer, Level 1
17:30	Closing and Award Ceremony	



What do you think about that?

Tweet us [@cis_dc](#) and share your thoughts on the talk / presentation / poster you are listening to!

The hashtag to use is [#cisdc2014](#)

Keynote



Dr. Tom Conway

Connecting theory & practice - a lesson from bio-informatics

A divide often exists between the theoreticians and the “practical” researchers in a Computer Science department, with the former stereotypically seeing their work as real computer science, and the work of their colleagues as mere programming. The converse view is that the work of the theoreticians is irrelevant compared to the useful work of their colleagues. This talk will show by example how the work of the theoreticians is vital to the production of useful methodologies for real problems. The concrete example is the problem of de novo assembly – a core problem in bioinformatics. Along the way we will discuss a bit of biology, a bit of information theory, a bit about computer architectures, and correctness.

Bio Sketch

Dr. Tom Conway undertook his doctoral research in the area of Logic Programming Language implementation, specifically looking at models for automatic parallelization of programs, semantics of threads and profiling. After completing his PhD, he worked for Multimedia Database Systems – a spin-off from RMIT University developing advanced text search techniques. After leaving MDS, he then took up a bioinformatics position at NICTA, delving in to advanced data structures and algorithms and applying them to bioinformatic analysis problems. Tom is currently leading the Genomics team at IBM Research Australia.

Posters

Analysis of Road Traffic Using Contrast Mining and GPS Trajectories

Xiaoting Wang, Christopher Leckie, Tharshan Vaithianathan

Answering Complex Questions in Neuroimaging Informatics

Aref Eshghishargh, Simon Milton, Andrew Lonie, Gary Egan

Audience Experience in Domestic Videogaming

John Downs

Compression and Visualization of Large Scale Graphs

Lida Rashidi, Christopher Leckie, Sutharshan Rajasegarar

Data Quality and Data Cleansing in Bio-informatic Databases

Qingyu Chen, Justin Zobel, Karin Verspoor

Energy and Carbon-Efficient Resource Management for Geo-Distributed Cloud Data Centers

Atefeh Khosravi and Rajkumar Buyya

Fragment Based Planning using Column Generation

Toby Davies, Adrian Pearce, Peter Stuckey, Harald Søndergaard

Information Security Strategy in Organisations

Craig Horne

Intelligence Driven Information Security Risk Management

Jeb Webb, Atif Ahmad, Sean Maynard, Graeme Shanks

Lazy and Eager Approaches for the Set Cover Problem

Lim Ching Lih, Alistair Moffat, Tony Wirth

Memory Efficient Local Outlier Detection in Data Streams

Mahsa Salehi, Christopher Leckie, Tharshan Vaithianathan

Preserving Location Privacy in Location-based Services - Obfuscation through Locality Preserving Hashing

Maryam Fanaeepour, Lars Kulik, Egemen Tanin

Privacy Aware Dynamic Ride Sharing

Preeti Goel, Lars Kulik, Kotagiri Ramamohanarao

Read Sequence Alignment Compression

Rodrigo Canovas, Alistair Moffat, Andrew Turpin

Secure Repair in Large Scale Distributed Storage Systems

Lakshmi J Mohan, Udaya Parampalli, Aaron Harwood

Spatio-Temporal Trajectory Simplification for Inferring Travel Paths

Hengfeng Li, Lars Kulik, Kotagiri Ramamohanarao

Transfer learning of a temporal bone performance model via anatomical feature registration

Yun Zhou, Ioanna Ioannou, Sudanthi Wijewickrema, James Bailey, Patorn Piroomchai, Stephen O'Leary, Gregor Kennedy

Paper Abstracts

Cloud Computing

A Performance Optimization Scheme for Mobile Cloud

Bowen Zhou

Computation outsourcing is one of the Cloud-based mobile augmentation approaches that offload the resource intensive computation to the cloud to enhance mobile devices computing capabilities. In this paper, we propose a performance optimization partitioning scheme for mobile cloud computation outsourcing.

Resource Overbooking in IaaS Cloud Provider

Chanh Nguyen

Infrastructure-as-a-service Cloud provider offers different purchasing models along with various service guarantees respectively to give customer the flexibility to optimize their costs. As a consequence, the diversification of purchasing options introduces a far more trivial problem for providers in which they need to decide how many available resource should be segmented into each pricing model so that the expected revenue is maximized. In this work, We address problem of maximized revenue through developing the optimal algorithms in which leveraging information from resource demand prediction to make efficiently decision on resource segmentation respecting to available resource on data center and the SLA promised to the customers so that the final objective is maximized revenue.

Fault-Tolerant Workflow Scheduling Using Spot Instances on Clouds

Deepak Poola, Kotagiri Ramamohanarao and Rajkumar Buyya

A scheduling algorithm that schedules tasks on Cloud resources using two different pricing models (spot and on-demand instances) to reduce the cost of execution whilst meeting the workflow deadline is proposed. This algorithm is fault tolerant against the premature termination of spot instances(SI) and also robust against performance variations of Cloud resources. Experimental results demonstrate that our heuristic reduces up to 70% execution cost as against using only on-demand instances.

Internet of Things: A Roadmap for Future Internet Direction

Farzad Khodadadi

Internet is continuously growing in multiple dimensions and as it gets bigger, the number of devices using it as a means for communication has been magnified. Currently, Cloud Computing is referred to as a de-facto standard for dealing with Big Data that requires elastic computing and on-demand escalation and such potential can be leveraged for the sake of IoT applications, if it is defined and used properly. To feel the gap in this area, I have modified Aneka, which is a platform for Cloud Computing as a service, to address IoT needs and provide features such as load balancing, resource provisioning, and scheduling to IoT applications, besides serving other cloud-based applications.

Energy Efficient Software Defined Clouds

Jungmin Son

In this paper, we discuss energy efficient data center for clouds through software defined clouds.

Resource Provisioning and Scheduling Algorithm for Workflows on Clouds

Maria A Rodriguez and Rajkumar Buyya

Cloud computing is the latest distributed computing paradigm and it offers tremendous opportunities to solve large-scale scientific problems. This paper proposes a resource provisioning and scheduling strategy for scientific workflows on Infrastructure as a Service (IaaS) Clouds. We present an algorithm based on Particle Swarm Optimization (PSO), which aims to minimize the overall workflow execution cost while meeting deadline constraints. Our heuristic is evaluated using various well-known scientific workflows of different sizes and the results show that it performs better than the current state-of-the-art algorithms.

Virtual Machine Customization and Task Mapping Model for Efficient Allocation of Cloud Data Center Resources

S. F. Piraghaj, R. N. Calheiros and R. Buyya

Although there has been significant research efforts in decreasing the energy consumed by data centers, most approaches lack the analysis of real cloud backend traces and their variance in the proposed solutions. To counter this issue, this paper proposes a new energy-aware methodology for mapping groups of tasks to customized virtual machine (VM) types based on the predicted task resource usage obtained from the analysis of the historical data. The objectives of this approach are efficient resource allocation and energy consumption improvement via the provisioning of a smaller number of efficiently utilized virtual machines.

Integrated Framework for Cloud-based Interoperability between Social Network Websites

Safiollah Heidari and Rajkumar Buyya

In this study, we try to find a framework for interoperability problem in the next generation of cloud-based social networks that are not only under control of a particular provider, but also more than one provider.

SLA Based Resource Scheduling for Big Data Applications

Yali Zhao

Big data refer to various and voluminous data sets which are continuously generated at high velocity that are difficult to process using traditional data processing tools. Big data applications are developed to analyze and find potential value and “secret” from big data to enable government, enterprises and academic make wise decisions, i.e., product trends prediction, disaster prediction and management, business strategy making.

Big Data applications require large amount of Cloud resources, i.e., CPU, RAM, Storage, etc., to process big data. How to efficiently schedule resources from multiple Clouds that can on the one hand satisfy specified Quality of Service (QoS) requirements of applications and on the other hand reduce resource consumption and cost is an urgent research problem raised for Big Data computing. To solve this, we design a Service Level Agreement (SLA) Based Resource Scheduling Platform for Big data applications.

Virtual Machine Allocation Policies against Co-resident Attacks in Cloud Computing

Yi Han, Jeffrey Chan, Tansu Alpcan and Christopher Leckie

While the services-based model of cloud computing makes more and more IT resources available to a wider range of customers, the massive amount of data in cloud platforms is also becoming a target for malicious users. Previous studies show that attackers can co-locate their virtual machines (VMs) with target VMs on the same server, and obtain sensitive information from the victims using side channels. This paper investigates VM allocation policies and practical countermeasures against this kind of attack by developing a set of security metrics and a quantitative model.

A security analysis of three commonly used VM allocation policies reveals that the server’s configuration, oversubscription and background traffic have a large impact on the ability to prevent an attacker from co-locating with the targets. If the servers are properly configured, and oversubscription is enabled, the best policy is to allocate new VMs to the server with the most VMs.

Based on these results, a new strategy is introduced that effectively decreases the probability of achieving co-residence. The proposed solution only requires minor changes to current allocation policies, and can be easily integrated into existing cloud platforms to mitigate the threat of co-resident attacks.

The influence of personality traits on online information behaviour during political campaigns

Miguel Wood, Simon Milton and Graeme Shanks

The adoption of social analytics heralds the arrival of a new paradigm in political campaign practice, driven by both behavioural science and social networks. Yet these domains have historically been in tension with opposite perspectives on political behaviour. The increasing focus on online social networks offers an opportunity to synthesise complementary research traditions on political behaviour to better understand the relationship between individual personality and information behaviour. We propose a model based on core dispositional personality traits, political participation as information consumption and information diffusion theory to explain online information behaviour during election events. Information diffusion requires users to perform several actions in combination - engage, rate content, comment, share and interpret information received. The research model can be used in empirical research to explore how personality type guides interaction with and diffusion of political information across online social networks. The learned behaviors' and direct responses of participants may help political campaigns in the design of future behavioural targeting strategies that activate social network effects to achieve campaign goals.

A conceptual model for analysing informal learning in online social networks for health professionals

Xin Li and Kathleen Gray

Online social networking (OSN) provides a new way for health professionals to communicate, collaborate and share ideas with each other for informal learning on a massive scale. It has important implications for ongoing efforts to support Continuing Professional Development (CPD) in the health professions. However, the challenge of analysing the data generated in OSNs makes it difficult to understand whether and how they are useful for CPD. This paper presents a conceptual model for using mixed methods to study data from OSNs to examine the efficacy of OSN in supporting informal learning of health professionals. It is expected that using this model with the dataset generated in OSNs for informal learning will produce new and important insights into how well this innovation in CPD is serving professionals and the healthcare system.

Task Complexity and Codification

Marion Zalk, Rachelle Bosua and Rajeev Sharma

The importance of knowledge and knowledge creation within an organisation is widely recognised, however the process of creating and managing knowledge is not well understood [7]. This research will address a key element in the knowledge creation process, specifically knowledge sharing behaviour within organisations. This act of knowledge sharing is a key precursor for the actual usage of knowledge [2]. Knowledge transfer has two components a source and a recipient [6]. This research is concerned with the source, specifically with his/her perceived task complexity of codification. The issue here is twofold as we have two task complexities - the task complexity of the primary work task and then the secondary task of codifying the primary task. A quantitative research method is employed in order to consider the relationship between the complexities of the different tasks.

Online Medical Consultation: A review of current practices

Ibrahim AL-Mahdi

Online Medical Consultation (OMC) has been offered by many providers since beginning of this century. There is a growing trend towards using the internet as a source of health information as well as for medical consultation. This study performs an empirical analysis on a group of existing OMC sites to explore their major themes, modalities, geographical coverage, and claimed value. These features have been illustrated for a better understanding of the promise on which these services operate. The study found out that OMC is a growing phenomenon featuring several interaction modalities, serving various medical consultation purposes, and accessible to consumers throughout the world. The contribution of this work is to synthesise a list of features and benefits of OMC services and to understand this phenomenon within the wider context of the rise of Internet services.

Performance evaluation of spectral unmixing method based on different ways of training data selection

Liyan Liu, Andrey Kan and Christopher Leckie

Fluorescence multi-channel imaging suffers from inherent problem of spectral bleed-through. Spectral unmixing method has been widely adopted to solve that problem. However, almost all the methods of evaluating the performance of spectral unmixing are based on the reference objects which are unfortunately not available in most applications. Moreover, there is little discussion about how to select training data as an initialization of spectral unmixing. So our purpose is to address the challenge of missing reference objects by proposing a new performance evaluation approach.

Retinal Cross Sectional Layer Segmentation using Optical Coherence Tomography

Md Akter Hussain, Alauddin Bhuiyan and Kotagiri Ramamohanarao

In this paper, we demonstrate an automated method for the layer segmentation of the retinal OCT image for reliable and repeatable analysis for predicting diseases like glaucoma, AMD.

An Effective Automated System for Grading Severity of Retinal Arteriovenous Nicking in Colour Retinal Images

Pallab Kanti Roy, Uyen T. V. Nguyen, Alauddin Bhuiyan and Kotagiri Ramamohanarao

In this paper, an effective method is proposed for the analysis of retinal venular widths to automatically classify the severity level of Retinal Arteriovenous nicking.

Shooting the Breeze Involves neither Shooting nor a Breeze: Predicting the Compositionality of Multiword Expressions

Bahar Salehi

In this paper, we propose a simple, language independent and highly effective method for measuring the compositionality of multiword expressions (MWEs). The proposed method uses the translations of the MWE and its components.

Semi-supervised Multilingual POS tagging

Long Duong

In this paper we address the problem of multilingual part-of-speech tagging for resource-poor languages. We use parallel data to transfer part-of-speech information from resource-rich to resource-poor languages. Additionally, we use a small amount of annotated data to learn to “correct” errors from projected approach such as tagset mismatch between languages, achieving state-of-the-art performance (91.3%) across 8 languages. Our approach is based on a modest data requirements, and uses minimum divergence technique. For situations where no universal tagset mapping is available, we propose an alternate method, resulting in state-of-the-art 87.2% accuracy on the resource-poor language Malagasy.

One-Way Games

Andres Abeliuk, Gerardo Berbeglia and Pascal Van Hentenryck

This paper takes a first step into finding ways of achieving a better social welfare outcome in applications such as large-scale power restoration.

Improved k-centre Clustering with MapReduce

Jessica McClintock

We present an intuitive MapReduce scheme for the k-centre clustering problem that requires few iterations and gives a 4-approximation guarantee under certain conditions. Experimentally, we show that solutions obtained by this algorithm are comparable to a sequential 2-approximation algorithm.

Designing Organization-Aware Agents using OJazzIC

Kathleen Keogh

In this paper, we highlight design considerations when using the OJazzIC architecture in multi-agent system design. We list a number of steps to be considered in the specification of agent requirements. We have used these in the design and creation of an agent based incident response simulation system.

Reliable power transmission networks

Rodrigo R. Gumucio E., Carleton Coffrin and Pascal Van Hentenryck

In the context of the Transmission Network Expansion Planning (TNEP) problem, a recent study shows that there exists a gap, between high cost expansion plans produced by AC-heuristics and potentially infeasible expansion plans produced by DC-approximations, that can be bridged by using the recently proposed LPAC power flow approximation. The work in progress for this paper aims to incorporate the $n - 1$ reliability criterion to the mentioned expansion planning study by proposing a genetic algorithm which may scale to large networks.

Neighborhood Component Analysis for Feature Learning

Zay Maung Maung Aye, Kotagiri Ramamohanarao and Benjamin Rubinstein

In this paper, we describe how Neighborhood Component Analysis can be modified as the feature learning step for learning from high dimensional and class imbalanced data.

Mining User Interactions and Activities using Mobile Data while Maintaining Privacy

Anthony Quattrone

Mobile privacy has recently become a well-researched topic in pervasive computing. However more research is needed to determine if personal information can be inferred from raw mobile data. Examples include if users are meeting other users, calls a user is making and receiving, trajectories of individuals accessing location-based services, social media services used and if mobile sensors can be used to identify a person. We demonstrate sophisticated algorithms that reveal a user's interactions and activities based on data stored on a mobile device. Mobile users typically keep their devices with them at all times. Near-field technologies (Bluetooth, WiFi) can be used as sensors to determine if user devices are in proximity to each other. Through the collection of very granular mobile data, we demonstrate that it can be inferred when users are meeting one another.

Privacy-Preserving Collaborative Anomaly Detection

Sarah M. Erfani, Yee Wei Law, Shanika Karunasekera, Christopher A. Leckie and Marimuthu Palaniswami

In collaborative anomaly detection, multiple data sources submit their data to an on-line service, in order to detect anomalies with respect to the wider population. A major challenge is how to achieve reasonable detection accuracy without disclosing the actual values of the participants' data. We propose a lightweight and scalable privacy-preserving collaborative anomaly detection scheme called RMP, which is a combination of nonlinear and participant-specific linear perturbation. A privacy analysis is given for Bayesian Estimation and ICA attacks. Experimental results on various datasets using an autoencoder show that RMP yields comparable results to non-privacy preserving anomaly detection.

Automatically Recognizing Places of Interests/Activities from Unreliable GPS Data and Smart Phone Sensors

Tanusri Bhattacharya, Lars Kulik and James Bailey

Modern smartphones are equipped with many sensors, such as GPS, Wi-Fi, Bluetooth, ambient temperature, pressure, humidity sensors, etc. This has created opportunity to the researchers in pervasive community to mine useful semantic context about a user's behaviour. One important user's context is to mine the significant places visited by a user. GPS provides broad coverage for positioning worldwide. It provides excellent accuracy at outdoor environments and does not incur any infrastructural cost for indoor positioning. Significant places are often indoor locations and sometimes buildings in urban areas where GPS is unreliable. In this research, we develop a novel algorithm POI-ID to identify a user's Place of interests (POIs) at building level accuracy from highly inaccurate GPS data. We further extend our research to explore the ambient sensors embedded in a smart phone to recognize a user's places or activities in real time.

Principled Dictionary Pruning for Low-Memory Corpus Compression

Jiancong Tong

Compression of collections, such as text databases, can both reduce space consumption and increase retrieval efficiency, through better caching and better exploitation of the memory hierarchy. A promising technique is relative Lempel-Ziv coding, in which a sample of material from the collection serves as a static dictionary; in previous work, this method demonstrated good compression ratios, while allowing extremely fast random access to individual items. However, there is a trade-off between dictionary size and compression ratio, motivating the search for a compact, yet similarly effective, dictionary. In previous work it was observed that, since the dictionary is generated by sampling, some of it (selected substrings) may be discarded with little loss in compression. Unfortunately, simple dictionary pruning approaches are ineffective. We generate measures for identification of low-value substrings in the dictionary, and show on a variety of sizes of text collection that halving the dictionary size leads to only marginal loss in compression ratio. This is a dramatic improvement on previous approaches.

Predicting at-risk students in Massive Open Online Courses

Jiazhen He, James Bailey, Rui Zhang and Benjamin Rubinstein

Massive Open Online Courses (MOOCs) have received wide public attention for their potential to scale higher education with multiple platforms such as Coursera, edX and Udacity. However, a big problem MOOCs face is the high rates of attrition. We explore to predict at-risk students early and accurately, and provide possible interventions for them.

Improved Feature Transformations for Classification using Density Estimation

Yamuna Kankanige and James Bailey

Density based logistic regression (DLR) is a classification technique that has been introduced recently. It can be explained as a one-to-one non-linear transformation of the original feature space based on density estimations, which is particularly suitable for learning a logistic regression model. Performance gains, good interpretability and time efficiency make DLR an attractive technique. However, there are some open questions regarding the use of DLR. In our work, we tackle these limitations and propose several new extensions which are evaluated using several synthetic and publicly available datasets.

FILTA : Better View Discovery from Collections of Clusterings via Filtering

Yang Lei, Nguyen Xuan Vinh, Jeffrey Chan and James Bailey

Traditional clustering methods have focused on finding a single 'best' clustering solutions. However, in real applications, there may exist more than one reasonable set of clusters (a clustering) to explain the data. Also, users are often uncertain about what is a good clustering for their purpose. Meta clustering has been successfully used to discover multiple reasonable and different clusterings (views) in a dataset by navigating and refining a large collection of base clusterings. However, the effectiveness of meta-clustering is highly dependent on the distribution of the base clusterings and open challenges exist with regard to its stability and noise tolerance. In this paper we propose a simple and effective filtering algorithm (FILTA) that can be flexibly used in conjunction with any meta-clustering method and get better views by removing base clusterings having poor quality or high redundancy. We evaluate FILTA on both synthetic and real world datasets, and see how its use can enhance view discovery for complex scenarios.

Bandwidth-Efficient Convergence of Gossip-based Recommendation Systems

Irum Fahim Bukhari, Aaron Harwood and Shanika Karunasekera

In this work, we analyzed the effect of similarity distributions on the convergence of the protocol and found the trade-off between convergence time and bandwidth to decide optimal settings that lead to fast convergence and affordable bandwidth from recommendation systems perspective.

Human-Computer Interaction

Developing an opportunistic online psychosocial screening tool for young people in general practice

Marianne Webb, Greg Wadley and Sylvia Kauer, Lena Sanci

Online self-administered multidimensional screening tools are a quick and effective way for General Practitioners (GPs) to identify areas of health concern for young people. However, there is currently little research into how they can be designed in way that places young people as more active participants in their consultation. There is also a paucity of research that explores the design needs of GPs and how to successfully integrate these tools into existing office technology systems. The aim of this research is to use user-centered design to develop and evaluate an online multidimensional assessment tool for young people and GPs in the general practice setting.

Help me keep my stuff safe

Sarah Ellen Webber

This paper reports on research into the design of an online repository for the personal possessions of young people in care, and resulting design heuristics for supporting collaboration around digital content in highly dynamic social contexts.

Understanding the user experience of running with mixed reality stories.

Alexander Kan

The notion of running with mixed reality stories (MRS) has emerged in recent years. Although different aspects of these MRS have already received interest from researchers, to date still little is known about how MRS work as a whole in the context of running experience. This projects aims to provide a holistic understanding of MRS and their effects on the experience of running.

Non-verbal Interactions in Domestic Video Conferencing

Behnaz R. Yeganeh

This research aims to understand how the design of video conferencing technologies should be improved to support non-verbal activities in domestic video conferencing. Recent research shows that people are increasingly using video conferencing tools to go 'beyond talking heads' to engage in a range of activities with their remote family members and friends. These activities include show and tell, watching video programs, cooking and eating, etc. The findings of initial studies emphasise the importance of these non-verbal interactions, however they demonstrate the inability of current video conferencing tools to fully support them.

Won't you be my neighbour? Clustered borrowing in libraries as evidence for shelf browsing

Dana McKay

In this paper, I describe a study designed to evaluate the impact of browsing on borrowing patterns in physical libraries, with a view to understanding the importance of browsing in online systems.

Supporting Tele-consultation through Tangible Interfaces

Deepti Aggarwal

This research explores the use of tangible interfaces to support and enrich tele-consultation between a patient and health professional.

Experiences with Mobile Mental Health Therapies

Fernando Estrada

Mental health disorders have been identified by the World Health Organisation (WHO) as one of the leading causes of disability worldwide. The WHO has raised concerns about how those suffering with mental health disorders (users) may be assisted. As a resource tool, mental health therapies (MHTs) using mobile phones (mobiles) have become increasingly popular to reach users. However, despite their rapidly increasing numbers, high levels of attrition and/or unacceptability have been reported. In this regard, the aim of this research is firstly to understand the user experience of mobile MHTs. Secondly, develop a theoretical evaluation and testing model of mobile MHTs, and thirdly, to present practical recommendations for the design of mobile MHTs based on engagement and user experience.

Communication Technologies during Family Mealtime

Hasan Shahid Ferdous

The aim of this research is to understand the use of communication technologies during family mealtime, whether they are used as information source, entertainment, or enabling social connection, among many other uses. We aim to understand the comprehensive relationship between technology and people, between the people themselves, and also the interaction among multiple technologies or devices in the context of family mealtime activities.

Screen Ecologies and the Future of Domestic Gaming

Marcus Carter, Bjorn Nansen and Martin Gibbs

In contrast to conceptualizations of gaming as immersive, fully-engaging experiences, contemporary domestic play occurs across a variety of fluctuating frames of activity and registers of engagement. In this talk, I will show how this is attributable to what I call the domestic screen ecology; the emerging coexistence and simultaneous engagement with activities on multiple monitors, televisions, laptop computers and smart phones. I argue this has considerable consequence for the future design of engaging interactive experiences and challenges the fallacy of immersion purported by new VR platforms like the Oculus Rift.

Software Engineering

Loop Untangling

Kathryn Francis

This paper concerns the translation of program code into constraints (mathematical equations defining the effects of the code), in order to facilitate automated reasoning about program behaviour. In particular a new technique is presented for the translation of bounded for and while loops.

Automated debugging using program spectra

Neelofar, Lee Naish and Rao Kotagiri

This paper presents an automated approach for software fault localization using program spectra obtained by dynamic execution of a software system. We present few metrics obtained by genetic programming to rank program statements or blocks according to how likely they are to be buggy. The metrics are tested on 92 faults from four unix utilities and the results shows that they outperforms other metrics, such as Jaccard, Trantula and Ochiai in single bug program.

Review of Interdependence in Coactive Design

Ronal Singh

Interdependence is defined as the set of complementary relationships between two or more parties. This paper reviews the conditions that lead to interdependence between agents and highlights that there is a need to further investigate it.



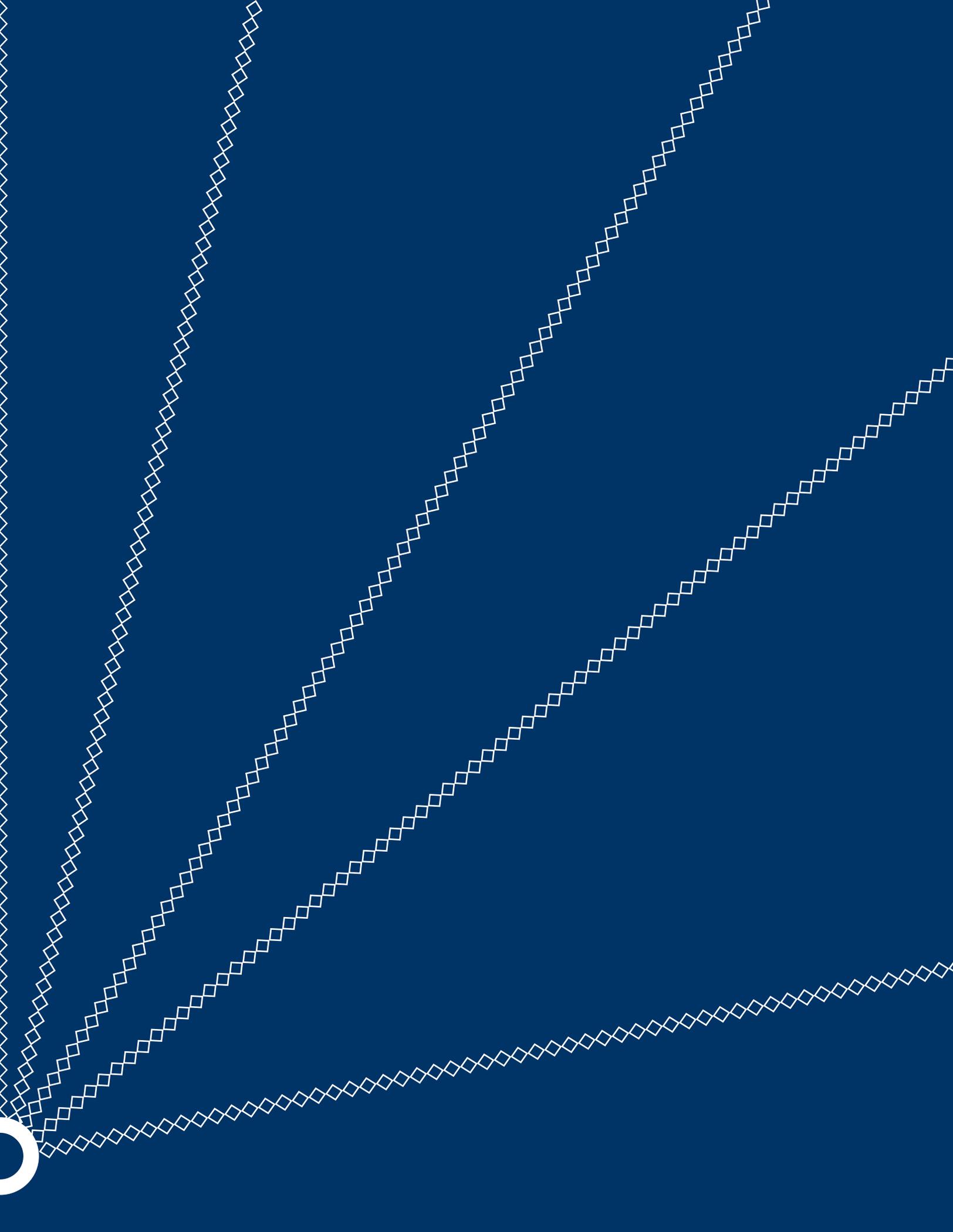
Download the program and
proceedings from our website!

<http://www.cis.unimelb.edu.au/colloquium/>



cis-dc@unimelb.edu.au





Thank You!

What a great time of planning it has been!

The 2014 CIS DC would not have been possible without the support and participation of students and staff at the Computer and Information Systems Department.

We would especially like to acknowledge Justin Zobel, the head of the CIS department and Rhonda Smithies for helping the CIS DC committee plan and co-ordinate the event.

Our dear supervisors who pushed us to submit to the colloquium and have supported us in our studies thus far.

Our sponsors, for making the CIS DC possible.

Our keynote speaker Dr. Conway, industry partners and colleagues from other institutions.

The panelists and judges at each session, who volunteered their time to make this event a success.

A heartfelt thank you to all students and staff who participated in the CIS DC.

Finally, we would like to acknowledge the 2014 planning committee for volunteering their time, skills and cheer, and also for the 2013 committee for their support.

The 2014 CIS DC Planning Committee!



Behnaz R.
Yeganeh



Fernando
Estrada



Lida
Rashidi



Maryam
Fanaeepour



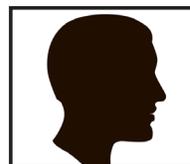
Yali
Zhao



Kagonya
Awori



Yi
Han



Alexander
Kan



Oliver
Adams



Ida
Someh

A special thanks to our sponsors!

