NUS
National University
of Singapore

SCHOOL OF COMPUTING

RESEARCH REPORT 2013

Contents

1 Dean's Foreword

2 Highlights

3 Centres

4 Department of Computer Science
Head's Overview
- i Artificial Intelligence
- ii Computational Biology
- iii Database Management Systems
- iv Media
- v Programming Languages and Systems and Software Engineering
- vi Security
- vii Systems and Networking

5 Department of Information Systems
Head's Overview
- i IT-enabled Service Innovation
- ii Electronic and Social Commerce
- iii Healthcare Information Systems
- iv Business Intelligence

6 Innovation and Entrepreneurship

7 Appendices
- a) Faculty Members
- b) Graduate Students
- c) Research Projects

Blank Page

Dean's Foreword

It is my pleasure to introduce the fifth edition of the NUS School of Computing Research Report. In the past year, the School of Computing has been ranked among the top ten computing faculties in the world, due in large part to the quality and impact of our research, and the reputation of our world-class faculty. In addition, given the central role of computing in so many areas of human endeavor, we have experienced increasing opportunity to engage in high-impact, multi-disciplinary research with our colleagues in other faculties in NUS, and with leading research groups and institutions around the world.

Our Department of Computer Science has established research clusters in Artificial Intelligence, Computational Biology, Database Systems, Media, Programming Languages & Software Engineering, Security, and Systems & Networks, and we have several faculty members across these clusters making fundamental contributions to theoretical computer science. Our Department of Information Systems has long-term research strengths in Strategic Information Systems Management, Knowledge Management, and Electronic Commerce, and more recently it has established substantial research activity in Healthcare Informatics, Service Innovation, Social Computing and Business Analytics.

Our faculty publish their research in the top conferences and journals in their fields, and they are routinely recognized by their peers with best-paper awards, retrospective "test-of-time" paper awards, field medals, and recognition as Fellows in the leading professional societies in computing. Our research leadership is further reflected in the many important service roles our faculty are invited to fill, including editorships for leading journals, key organizational roles and membership on technical program committees for leading conferences, and service as officers of professional societies. We also continue to host some of the leading conferences in computing in Singapore and have inaugurated new conference series here as well.

Over the past three years, SoC has received research funding of approximately $66 million. Some of this funding has been awarded through competitive funding calls from A*STAR, DSO, MDA, MOE and NRF, while other funding has come from industrial sponsors, including CSR, Microsoft, IBM, Fuji Xerox and Huawei. We also have increased the number of large research centers we host, with the establishment of the CSIDM, NExT, COSMIC and SeSaMe centers funded by MDA and the CIIDAA project funded under NRF's Competitive Research Programme. Our research activity is underpinned by a large and diverse body of over 300 PhD students and 110 research staff. In addition, many of our more than 1600 undergraduate students contribute to our research, through our Undergraduate Research Opportunity Program (UROP) and through the Final Year Projects (FYPs) our faculty offer our students each year.

We also enjoy partnering with industry, not just in our core research activity but also in efforts to translate research outcomes to practice, particularly through our continued support for entrepreneurial activity among our faculty and students. Our incubation center, The Furnace, provides a supportive environment for developing proofs-of-concept and commercial prototypes from promising undergraduate FYPs, PhD projects, and faculty research. The companies inside The Furnace are supported through seed funding of start-ups, and The Furnace facilitates mentorship and networking with successful entrepreneurs, venture capitalists and other key players in the region's entrepreneurial ecosystem.

The following pages present many of the highlights of our research from the past three years, and you may find additional information at the SoC Web site (http://www.comp.nus.edu.sg/research/). I sincerely hope you find this report informative and stimulating, and I invite you to find ways of engaging with us as we continue to pursue research that we hope will have a lasting impact on science and society.

David S. ROSENBLUM
Professor and Dean
NUS School of Computing
Personal webpage: www.comp.nus.edu.sg/~david/

Research Highlights

We highlight here some of the awards, recognitions received by our Faculty and Students during the period 2011-2013.

Best Paper Awards

"Generating Representative Views of Landmarks via Scenic Theme Detection" by Yi-Liang Zhao, Yan-Tao Zheng, Xiangdong Zhou and Tat-Seng Chua received the Best Paper Award at the Multimedia Modeling Conference, 2011.

"Controllable Hand Deformation from Sparse Examples with Rich Details" by Haoda Huang, Ling Zhao, KangKang Yin, Yue Qi, Yizhou Yu and Xin Tong received the Best Paper Award at the ACM/Eurographics Symposium on Computer Animation, 2011.

"A Probabilistic Forest-to-String Model for Language Generation from Typed Lambda Calculus Expressions" by Lu Wei and Ng Hwee Tou received the Best Paper Award at the 2011 Conference on Empirical Methods in Natural Language Processing, 2011.

"Partial Social Network Disclosure and Crawlers" by Suhendry Effendy and Felix Halim (Yap Hock Chuan, Roland's student) won the Best Student Paper Award at the International Conference on Social Networks and Applications, 2011.

"Video Quality for Face Detection, Recognition and Tracking" by Pavel Korshunov and Wei Tsang Ooi received the TOMCCAP Nicolas D. Georganas Best Paper Award in ACM Transactions on Multimedia Computing, Communications and Applications, 2012.

"Recognition and Summarization of Chord Progressions and their Application to Music Information Retrieval" by Roger Zimmerman, Wang Ye, research fellow Yu Yi, and their co-author Vincent Oria won the Best Paper Award at the IEEE International Symposium on Multimedia, 2012.

"Graph Minor Approach for Application Mapping on CGRAs" by Tulika Mitra and her student Chen Liang won the Outstanding Paper Award at the International Conference on Field Programmable Technology, 2012.

"ABACUS: An Auction-Based Approach to Cloud Service Differentiation" by Zhenjie Zhang, Richard T. B. Ma, Jianbing Ding and Yin Yang won the Best Paper Award at the IEEE International Conference on Cloud Engineering, 2013.

Fellowships

Ooi Beng Chin was made a Fellow of the Association for Computing Machinery (ACM). The ACM Fellowship came in recognition of Beng Chin's contributions to spatio-temporal and distributed data management.

ACM, established in 1947, has been described as the world's first scientific and educational computing society. The ACM Fellows Program was set up in 1993 to recognize and honour outstanding ACM members for their contributions to computer science and information technology.

The Association for Information Systems (AIS) has named Professor Bernard Tan an AIS Fellow. The Award is presented in recognition of Prof Tan's outstanding contributions, both locally and globally, to the discipline of Information Systems in terms of research, teaching and service. Commenting on the award, AIS President Dov Te'eni said: "…awardees are esteemed for their high levels of professional and personal integrity. Professor Tan demonstrates these qualities and we congratulate him on receiving the honour." AIS was founded in 1994 with the purpose of serving as the premier global organisation for academics, students and professionals specialising in Information Systems. More about AIS may be found at: http://www.aisnet.org

Leong Tze Yun has been elected into the American College of Medical Informatics (ACMI) as an International Fellow. ACMI is a college of elected fellows who have made significant and sustained contributions to the field of biomedical informatics. The Fellowship is awarded through a formal election process when Voting Fellows of the College judge nominees. Each year only 10-20 new Fellows are elected. The College was founded in 1984 when five pioneers in informatics decided to establish an honorific society to recognise expertise in biomedical Informatics. Today, there are nearly 300 Fellows.

Ng Hwee Tou was elected into the Association for Computational Linguistics (ACL) as a Fellow in 2012. He was selected for his significant contributions to coreference resolution and semantic processing, and for the development of semantic corpora. The Association for Computational Linguistics is the international scientific and professional society for people working on problems involving natural language and computation. Fellowship is awarded through nomination by a fellow ACL member. Each year a small group of Fellows are elected. Founded in 2011, the ACL Fellowship Programme began with 17 founding members.

Tan Chew Lim has been elected as a fellow of the International Association for Pattern Recognition (IAPR). Since 1994, the IAPR Fellow Award has been biennially conferred on persons to acknowledge their distinguished contributions to the field of pattern recognition and to IAPR activities. This prestigious award is only presented to the top 0.25% of the IAPR membership. The award was presented to Chew Lim in recognition for his contributions to pattern recognition for document analysis.

Document analysis requires pattern recognition techniques to process and extract information from documents from different sources. Prof Tan's research in document analysis is in document image binarisation. This is usually performed in the pre-processing stage of any image processing application. It converts a gray-scale document image into a binary document image and accordingly facilitates the ensuing tasks such as document skew estimation and document layout analysis. As more and more text documents are scanned, fast and accurate document image binarisation is becoming increasingly important.

Awards

Ling Tok Wang won the 2011 Peter P. Chen Award for outstanding contributions to the field of conceptual modeling. Inaugurated in 2008, the Peter P Chen Award honours one person each year for his or her outstanding contributions to the field of conceptual modeling. Recipients are chosen for the contribution they make in their research, service, education and practice to the field of conceptual modelling as well as their international reputation.

Ooi Beng Chin and Kian Lee won the President's Science Award for 2011 for their outstanding contributions to database systems research. The President's Science Award is presented to research scientists and engineers in Singapore who have made outstanding contributions in basic research leading to the discovery of new knowledge or the pioneering development of scientific or engineering techniques and methods. The Award is part of the annual President's Science and Technology Awards, which are the highest honours presented to leading scientists and technologists in Singapore in recognition of their exceptional achievements in their chosen field of specializations.

Ooi Beng Chin received the 2012 IEEE Tsutomu Kanai Award from the IEEE Computer Society his pioneering research in distributed database management and peer-to-peer-based enterprise quality management. The Society's Tsutomu Kanai Award was established in 1997 to recognize major contributions to the state-of-the art distributed computing systems and their applications. It is named in honour of Hitachi Limited's long-serving president Dr Tsutomu Kanai.

Yu Haifeng received the NUS Young Researcher Award in 2011. He has made important scientific contributions in the general area of distributed computing/systems. In particular, his work on novel defense mechanisms against sybil attacks is among the first practical and effective solutions for the notoriously challenging problem of defending against such attacks. His research results have been widely cited by other scientists, taught in classes, and followed up by other research groups. He has also made contributions on other important topics such as error estimating coding and fault-tolerant communication complexity.

Rahul Jain received the NUS Young Researcher Award in 2012. Rahul Jain won the young researcher award in 2012. Rahul Jain has made significant contributions to our understanding of proofs and computation. His specialty is the interface between randomized, interactive, and quantum computing. One of his research focus is on the computational power of interactive protocols where a powerful prover tries to convince a verifier that a mathematical statement is true. This question is closely related to one of the seven Millennium Prize Problems put forward by the Clay Mathematics Institute in 2000. Rahul (and co-authors) fundamental paper on quantum verifiers won the "Best Paper Award" in 2010 at the ACM Symposium on Theory of Computing, one of the most prestigious conferences in theoretical computer science. The highly original techniques developed in this paper were later used by Rahul (and co-authors) to design fast parallel algorithms for an important class of problems.

Ooi Beng Chin received the NUS Outstanding Researcher Award in 2013. He is among the most highly regarded researchers in the database arena in the world, for his contributions to distributed data management and to the management of spatio-temporal and multimedia data. Beng Chin was one of the first to address issues concerning database support in peer-to-peer computing. He has made significant contributions in data distribution, search, and data management in peer-to-peer networks. He is especially noted for addressing the issue of providing enterprise-quality data sharing and processing in corporate networks.

Notable Achievements

"A Logical Foundation for Deductive Object-Oriented Databases" by Ling Tok Wang, Mengchi Liu, Gillian Dobbie received the Most Influential Paper of the Decade Award at the Database Systems for Advanced Applications Conference, DASFAA2011.

"Transport4You" Project by Yi Li, Hang Yang and Huanan Wu wins the Formal Methods Award in the Student Contest on Software Engineering at the International Conference on Software Engineering, 2011.

PhD student Wu Kegui, Lee Wee Sun and David Hsu won first place in the International Probabilistic Planning Competition at the International Conference on Automated Planning and Scheduling , 2011.

Sim Khe Chai won a Merit Award in the Open Category of the Tan Kah Kee Young Inventors' Award, TKKYIA 2012 for his work, "Perfecting Voice Input with a Magic Touch".

"Advancing Public Trust Relationships in Electronic Government: The Singapore E-Filing Journey" by Pan Shan Ling, (published in Information Systems Research) was awarded European Research Paper of the Year 2012 by CIOnet.

Abhik Roychoudhury received an appointment to the Association for Computing Machinery (ACM) Distinguished Speaker Program as of 13 May 2013. This appointment recognizes Abhik as a leading expert and researcher in his field and provides him a platform where he can share his research and ideas with professional organisations, peer institutions, governments, and industry bodies.

NUS-TSINGHUA CENTRE FOR EXTREME SEARCH

NExT is the NUS-Tsinghua Centre for Extreme Search, a leading centre jointly established by NUS and Tsinghua University, funded MDA (Media Development Authority) of Singapore. The Centre commenced in July 2010. The vision of the Centre is to organize, index and search for live, multi-source, multimedia, and multilingual social media data and events, currently not available on the traditional Web. Such user-generated social media data, also known as the user-generated contents or UGCs, is available in a myriad of sources, including live microblog sites like twitter; mobile sharing sites like 4Square and Instagram; information sharing sites like forums and blogs; image and video sharing sites like Flickr and YouTube; and the various community question-answering sites like Wiki-Answers and Yahoo!Answers; as well as their counterparts in China. The UGCs document the activities, interests and social interactions of people in the society.

Key research areas of NExT include:

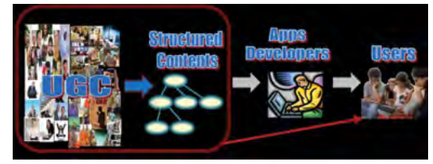
1) Reliable strategies for harvesting representative UGCs on any topic. This is particularly challenging when the "topic" is hot and there are various limitations that restrict the number of posts that can be crawled. To address this problem, we devise a strategy to crawl representative posts comprehensively by leveraging on: (a) multiple keywords; (b) key users; and (c) known accounts related to the "topic". We have formulated sub-optimal solutions to identify relevant and related keywords and key users, and we are integrating them towards intelligent and robust crawlers.

2) Indexing and retrieval of huge amount of media resources. We built a system to store the huge amount of UGC posts by adapting a distributed architecture that uses Hadoop, HBase and Elasticsearch index. For video and images, we developed our own media search engine to index the over 300 millions of images that we have crawled.

3) Framework for sharing of UGC resources and analytics. As it is meaningless to share the raw UGC data crawled, we are working with Web Science Trust to develop a framework for sharing the higher level analytics derived from these UGC posts. In the meantime, we are also working on several large scale test datasets with ground truth for the research community.

4) Organization of unstructured UGCs and users. It focuses on the construction of dynamic topic-specific knowledge structures by leveraging on a wide variety of UGC sources, ranging from structured information from Wikipedia or Blogs, semi-structured knowledge from cQA and forums, and the unstructured but live information sources from twitters. The resulting structures support browsing, question-answering, and knowledge inferencing.

(a)



(b)

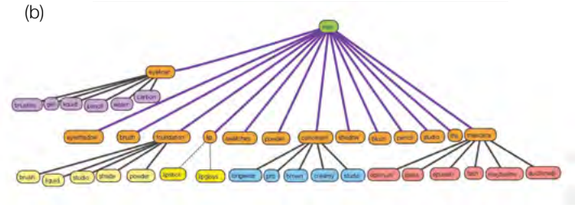


Figure 1. From unstructured UGC data to structured knowledge: (a) overall framework; and (b) the structures extracted from a collection of reviews on MAC Cosmetics.

5) Generation analytics related to location, people, topic and organization. The research includes the crawling of representative set of social media data related to an "entity", filtering out those irrelevant, detecting both the emerging and evolving events, and identifying hot events. Figure 2 shows some sample analytics that the system generates.

(a)

(b)

Figure 2: Analytics extracted on entities: (a) a live event in Singapore; and (b) semantics of venues

6) Other research includes: (a) live video event capturing and sharing; (b) differential news; (c) extreme database support; and (d) basic research on the analysis and retrieval of text, live discussion streams, images and videos.

Overall, the unique capability of NExT is its research on live, big, multi-source, multimedia and multilingual UGC data. Figure 3 presents the interface of NExT Live Observatory System, which offers access to five sub-systems: the “live crawler”, and analytics to “location”, “people”, “topics” and “organizations” (see http://www.nextcenter.org/).

Figure 3: Interface to NExT Live Observatory System

CHINA-SINGAPORE INSTITUTE OF DIGITAL MEDIA

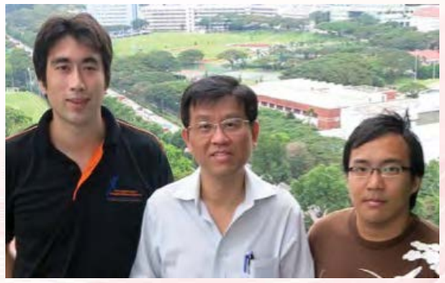
In June 2008, a research grant of 4 million dollars from the Singapore National Research Foundation (NRF), administered by the Interactive Digital Media Programme Office (IDMPO), was awarded to NUS to carry out research collaboration with the China-Singapore Institute of Digital Media (CSIDM). CSIDM is wholly owned by CASIA (Institute of Automation, Chinese Academy of Sciences), and it is the first R&D institute of the Chinese Academy of Sciences set up outside China. The co-directors of the CSIDM-NUS collaboration project are Professor Xu Bo of CASIA and Professor Ng Hwee Tou of the NUS Department of Computer Science. The collaboration project focuses on language mediation, aiming at breaking the language barrier to enable communication between speakers of different languages. It carries out research in natural language processing, speech processing, and multimedia processing.

First in Shared Task on Grammatical Error Correction

A team led by Professor Ng Hwee Tou and comprising his NGS PhD student Daniel Dahlmeier and Computer Science Department research assistant Eric Ng emerged as the best performing team in the HOO (Helping Our Own) 2012 shared task. The goal of the shared task was to come up with innovative algorithms that can automatically correct errors involving determiner and preposition errors made by English language learners.

Fourteen teams from academia and industry, including well-known institutions like Cambridge University and the Educational Testing Service, participated in the shared task. The teams were given a set of student essays with error annotations to develop their algorithms. Afterwards, the teams were given a blind test set of essays without error annotations and their grammatical error correction algorithms need to automatically correct all determiner or preposition errors in the test essays.

The system submitted by the NUS team achieved the highest score for determiner correction, preposition correction, and the overall correction task among all 85 systems submitted by the four-teen teams. The good result was sweetened by the fact that the NUS team only submitted a single system, and it gave the best scores. The success in the shared task shows that the innovative natural language processing technology developed at NUS can have practical impact on real-world problems.



Daniel Dahlmeier (left), Professor Ng Hwee Tou (centre) and Eric Ng

CENTRE OF SOCIAL MEDIA INNOVATIONS FOR COMMUNITIES

COSMIC (Centre of Social Media Innovations for Communities) aims to empower the next 10 million people in the community through social media innovations that improve the way they live, work, and play. Specifically, the people targeted are those involved in the informal economy and commonly labeled as “middle of the pyramid”. Numerous such people worldwide have access to mobile devices but are poorly supported by mobile services. The institutional members of COSMIC are National University of Singapore (NUS), Nanyang Technological University (NTU), and Indian Institute of Technology, Bombay (IITB). COSMIC is supported by the multi-agency Interactive Digital Media Programme Office (IDMPO) hosted by the Media Development Authority of Singapore.

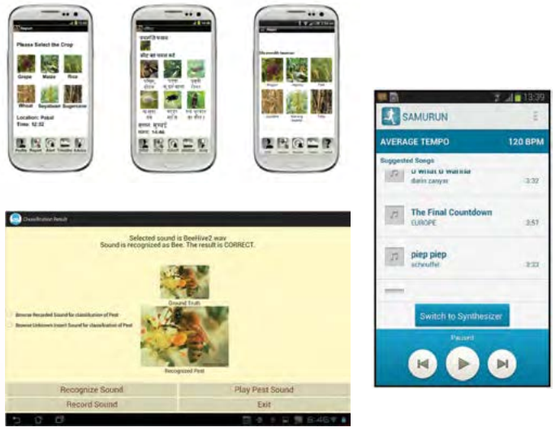


Figure 1. COSMIC mobile technologies for pest management, insect sound classification and interactive music synthesis.

COSMIC is structured into three project areas – the two application verticals in agriculture and healthcare, and a third infrastructure-focused area, which serves to explore and provide key technologies to the application verticals.

The goal of the agriculture project is to design, develop and deploy social media applications for agricultural process innovation that can improve productivity and increase earnings of farming communities. The ability to better manage agricultural process not only means lower costs and greater profits, but also improved food quality and generation capacity of human population. Major breakthroughs of the agriculture project are the unique applications of algorithms and techniques in image recognition and predictive analytics to identify, manage, and forecast pest problems based on information collected from users and automated sensors (e.g., pest observations, geographical locations, weather). The agriculture project leverages on crowd-sourced knowledge from farmers and agriculture experts by innovatively processing large amounts of image, audio and contextual data to create a network of valuable and futuristic insights. This allows stakeholders in the agriculture community to achieve greater efficiency in pest management, thereby reducing crop losses incurred through pest attacks. With pest infestation accounting for a 42% reduction ($244 billion) in yield annually, such a breakthrough provides extremely great value to government agencies, as well as companies and farms in the middle of the economic pyramid.

The goal for the technologies designed and developed under the infrastructure project is to be building blocks that will be integrated into other vertical applications, such as agriculture and healthcare. For example, algorithms on pest image recognition have been integrated into the mobile pest management system. Recent work has explored fundamental insect sound classification technology based on advanced machine learning algorithms. A mobile Android app allows a farmer to record insect sounds which are then classified and identified against an expert database on a server. In another sub-project, a mobile video live streaming component based on the GeoVid technology is been developed as an add-on for specialized social media platforms which are pursued as a joint project by the COSMIC partner organizations NUS, NTU and IITB.

In the healthcare project teams from both NUS and NTU are pursuing socially-mediated applications. For example, we have developed a real-time, interactive music synthesis application for running or jogging. A sedentary lifestyle is a major long-term health problem in many societies. A unique feature of the application is that it provides a flexible means for users to actively control music in real time while moving about. The rhythmic motion of the body is detected by the smartphone to set the tempo, and a computer-generated music ensemble performance is synthesized in real time to match—and react—to the body’s motion.

Department of **Computer Science**

HEAD'S OVERVIEW

The years 2010 – 2012 were exciting and fruitful in terms of research for the Department of Computer Science.

Several critical research areas in the department have been significantly strengthened with the arrival of a number of new faculty members. They are: David Rosenblum in software engineering, Prateek Saxena in web security, Richard T. B. Ma in internet economics and performance evaluation, Seth Gilbert in distributed systems, Bryan Low in multi-agent planning and coordination, Yin Kangkang in computer animation, and Stephanie Wehner in quantum security.

Several research initiatives in the department have led to the establishment of major research centres. In particular, the following centres were formed with significant funding from the Singapore Media Development Authority: The NUS-Tsinghua Extreme Search Centre (NExT) in 2010, the Centre of Social Media Innovations for Communities (COSMIC) in 2010, and the Sensor-enhanced Social Media Centre (SeSaMe) in 2012. In addition, the Felicitous Computing Institute was formed in 2012 with seed funding from the School of Computing and the Office of the Deputy President (Research & Technology) at NUS.

Many colleagues have been honoured for their contributions to the research community and their technical achievements. Professor Ng Hwee Tou was inducted a Fellow of the Association for Computational Linguistic. Professor Tan Chew Lim was inducted a Fellow of the International Association for Pattern Recognition. Associate Professor Leong Tze Yun was elected a Fellow of the American College of Medical Informatics. Professor Ooi Beng Chin was conferred the IEEE Computer Society Tsutomu Kanai Award. And many more colleagues won best paper awards, influential paper awards, and other awards at various conferences and journals.

Many interesting results were announced across a broad spectrum of computer science. Here are three examples: Professor Joxan Jaffar's work on interpolation has reached a major milestone with the completion of the TRACER system, and it gives a deep understanding of how to do interpolation on C programs of realistic size, and establishes many uses of interpolation for the first time. Associate Professor Stephanie Wehner's work that the degree of non-locality is determined by the uncertainty principle has been hailed as a breakthrough in quantum theory. Dr Seth Gilbert produced a mutual exclusion algorithm that is an exponential improvement over existing results, and has led to significant new interest by the research community in very fast (sub-logarithmic) protocols for multiprocessor coordination.

I hope the selected projects described in the following pages will convey to you an even greater sense of the depth, the breadth, and the vibrancy of our faculty members and our research clusters. Riding on these significant accomplishments, the department is working on inspiring leadership in computer science and evolving to deliver greater impact on the way we live, the way we work, and the way we have fun.

Wong Limsoon
KITHCT Chair Professor and Head
Department of Computer Science

ARTIFICIAL INTELLIGENCE

The term Artificial Intelligence (AI) was coined by John McCarthy in 1955 in a conference held in Dartmouth College, USA. Known as the Father of AI, McCarthy defines AI as the science and engineering of making intelligent machines. Today, AI is the key technology in many novel applications, ranging from intelligent robots, fraud detection in credit cards, speech understanding in telephone systems, face recognition in camera images, text understanding in documents, just to name a few. AI is important to us because it affects in many aspects of our lives such as in our daily use of smart home appliances like refrigerators and televisions and smart devices like hand phones and cameras, and in accessing intelligent web portals that understand users' preferences.

One of the recent remarkable advances in AI is the machine's ability to learn from a huge amount of data now so easily available on the internet to discover knowledge in a way that surpasses human learning abilities. In 2011, an Artificial Intelligent computer system named Watson developed by IBM beat two former champion winners in a television quiz show "Jeopardy!" winning a handsome prize of $1 million. Through AI machine learning techniques, Watson was able to answer questions posed in English by acquiring knowledge from 200 million pages of information gathered from the web. Bill Gates, the founder of Microsoft Corporation, alluding to the power of AI, once said this: "If you invent a breakthrough in artificial intelligence, so machines can learn, that is worth 10 Microsofts."

In the School of Computing, our AI group is actively involved in various aspects of AI research, particularly with respect to making machines learn from a host of different information sources to find interesting patterns, to gain insightful knowledge, to provide decision supports, to plan strategic tasks and to understand text in various media. Our AI research may be categorized into the following five main areas:

- Machine Learning
- Uncertainty Management
- Document and Scene Text Analysis
- Computational Natural Language Learning
- Planning and Decision Making

Machine Learning

Learning is one of the fundamental capabilities of intelligent systems. Machine learning, as it is often termed, has been studied since the dawn of AI. One of the earliest success stories in AI, Samuel's checkers player, was based on machine learning.

Machine learning methods currently give the best performance on many practical problems in computer vision, speech, natural language processing, and robot control. Along with the practical successes, much understanding of the fundamental issues of machine learning has also been gained.

Our work spans both the theoretical and applied fronts of machine learning. One of the areas of focus is inductive inference, where we study learning in the limit. One of the projects involves learnability of automatic families (that is uniformly regular families of languages), where often the learners themselves are automatic, that is their behaviour can be modeled using finite state machines. Faculty members involved in this research topic have served on the program committees of important conferences such as Algorithmic Learning Theory (ALT), Computational Learning Theory (COLT), International Conference on Machine Learning (ICML) and Neural Information Processing Systems (NIPS). On a more applied side, we have developed algorithms for learning with F-score, a performance measure commonly used in information retrieval and extraction. Another area of intensive research in machine learning is reinforcement learning.

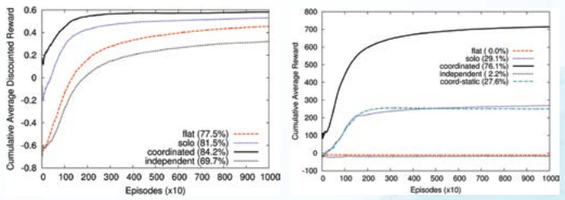
Reinforcement Learning

Online reinforcement learning (RL) in collaborative multi-agent domains is a challenging problem. The number of possible actions that can be considered at each time step is exponential to the number of agents. This curse of dimensionality poses serious obstacles for online learning as exploration requirements are huge. Further, the learning models for multiple agents can quickly become complex as the number of agents increase, and agents may have communication restrictions that vary dynamically with the state.

We design a centralized coordination guided reinforcement learning (CGRL) system that uses expert coordination knowledge to restrict the joint action space and to direct exploration towards more promising states. We propose a two-level reinforcement learning architecture where the top level learns to choose the coordination knowledge to constrain the bottom level's learning of joint actions. We also develop a distributed version of CGRL for domains where communication between agents changes over time. This necessitates that learned parameters are distributed among agents. To do so, we design localized learning for individual agents with coordination where possible. However, this increases the model complexity.

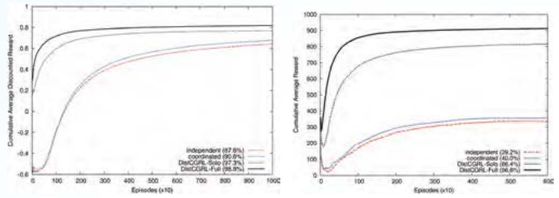
To deal with the issue of model complexity, we introduce a distributed form of relational temporal difference learning. This is achieved by an agent localized form of relational features and a message passing scheme. The solution allows agents to generalize learning respectively over its interactions with other agents and among groups of agents whenever a communication link is available.

We evaluate our proposed learners on a soccer game and a tactical real-time strategy game. Figure 1 shows that coordination constraints increase the learning rate and the two-level approach outperforms existing state-of-the-art methods. Figure 2 demonstrates that the distributed version is effective in the online setting where we wish to maximize rewards even while learning. Finally, Figure 3 shows that the relational generalization learning is able to greatly improve goal achievement with a significant reduction in the number of learned parameters.



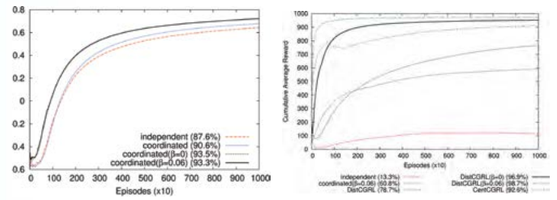
(a) Soccer (b) Realtime strategy game

Figure 1: Experiment results for centralized coordination guided reinforcement learner



(a) Soccer (b) Realtime strategy game

Figure 2: Experiment results for distributed coordination guided reinforcement learner



(a) Soccer (b) Realtime strategy game

Figure 3: Experiment results for distributed relational coordination guided learner

Representative Publications

Q. P. Lau, M. L. Lee and W. Hsu. Distributed Relational Temporal Difference Learning, AAMAS, 2013.

J. Case, S. Jain, T. D. Le, Y.S. Ong, P. Semukhin and F. Stephan. Automatic Learning of Subclasses of Pattern Languages. Information and Computation 218:17-35, 2012.

Q. P. Lau, M. L. Lee and W. Hsu. Coordination Guided Reinforcement Learning, AAMAS, 2012.

S. Jain and E. Kinber. Iterative Learning from Texts and Counterexamples Using Additional Information. Machine Learning 84(3):291-333, 2011.

Y. Wang, K. S. Won, D. Hsu and W. S. Lee. Monte Carlo Bayesian Reinforcement Learning, ICML, 2012.

N. Ye, K. M. A. Chai, W. S. Lee, and H. L. Chieu. Optimizing F-Measures: A Tale of Two Approaches, ICML, 2012.

Q. P. Lau, M. L. Lee and W. Hsu. Distributed Coordination Guidance in Multi-Agent Reinforcement Learning, ICTAI, 2011.

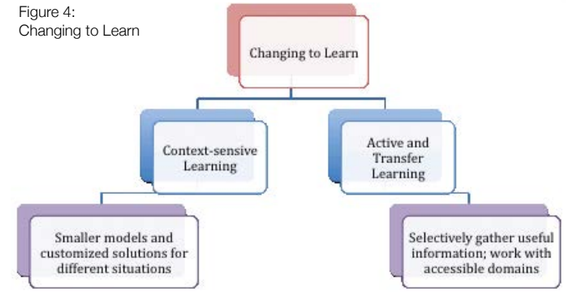
Uncertainty Management

Real-world problems are complex, uncertain, and changing. Problem solving and decision making in many critical areas including building sustainable environments, managing disaster recovery, designing personalized medicine, engineering devices to help the aged, are often complicated by incomplete information, varying assumptions, and limited resources. These problems require solutions that are adaptive, usable, and cost-effective. These solutions mandate computational methodologies that can effectively manage "change."

Most AI techniques that manage complexity and uncertainty assume stationary environments; the underlying assumptions, world dynamics, action alternatives, and potential outcomes are predetermined and unchanged over the course of the problem solving process. In many real-world situations, however, the environment is non-stationary and perceptions are often different from expectations. In other words, the world model as the problem solver understands it is often changing. It is also impractical to pre-enumerate all the possible scenarios or alternatives beforehand.

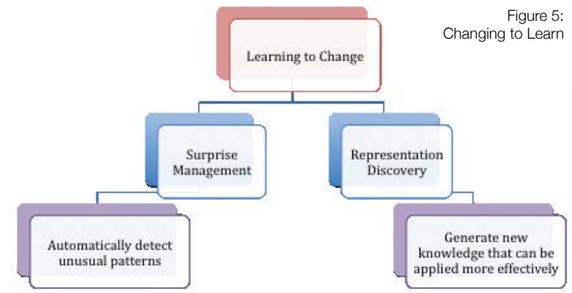
Our current research in Medical Computing Lab builds on our past and on-going work in dynamic decision making in complex and uncertain domains with limited resources. Together with our research collaborators, we are working on a family of representation, reasoning, and learning techniques and frameworks that can effectively reflect and better manage the incomplete and changing nature of real-world problems. The main theme of this work can be summarized as "change representation to learn effectively and learn to change representation when necessary", as depicted in Figures 4 and 5.

Figure 4:
Changing to Learn



In real world settings, it is often very difficult or even impossible to traverse the entire solution space for problem solving and decision making. The search space can be reduced by integrating control structure or domain knowledge, or by combining deterministic knowledge to guide learning under uncertainty. In face of incomplete and changing information, we can selectively and sequentially focus on the problem in a relevant context, or to transform the problem space to facilitate structural and parameter learning in model formulation.

Figure 5:
Changing to Learn



In real-world problems, e.g., in game playing, the preconceived notion of natural categories or opponent types may be misleading, or changing over time. It is therefore important for learning techniques to detect and effect such change. Anomaly detection, problem reformulation, e.g., from time and space to frequency domains, and approximate models are common approaches to effect representation change. We focus on an integrated approach that establishes an operational definition of "change" – When evidence mounts and the default assumptions are violated, a set of proper actions are taken to effect the change. For example, experience gained from first-principled decision making in a complex situation may induce a set of "fast and frugal" rules that can be applied more efficiently and effectively in routine decision making.

Application Areas

Intelligent Assistance

Our work on intelligent assistance is a good example of context sensitive learning in which context specific models are utilized based on different needs of the human the computer is trying to help. Recent development of ubiquitous sensing of our environment has created new kind of opportunities for building machines that make our daily lives easier. This holds a great potential for improved patient and eldercare, especially in countries like Singapore which, due to a long life expectancy and a low birth rate, feature a constrictive population pyramid.

A key challenge in this kind of assistive technology is to build a system that can quickly recognize the wishes and intentions of the person without need for the person to explicitly specify what the machine should do. This automatic detection of possibly changing human intentions requires elements of surprise management. Based on the theoretical framework of the partially observable Markov decision processes, and the idea of intention recognition as inverse planning we have devised a system which is capable of providing this kind of assistance. Solving the general decision theoretic aspect of the task is, however, not sufficient in practice since the provided assistance has to be done in timely manner which limits the amount of computation possible before the machine has to react to person's needs. Considerable amount of work has therefore been allocated to make the system fast enough to be practical. For research methodology and safety reasons, the system was developed in collaborative computer game domains which feature similar kind of complex environments, intention recognition criticality and real time performance requirements than the medical application might need. The goal is still to transfer this kind of technology to medical applications.

Representative Publication

T. Nguyen, Z. Li, T. Silander and T. Y. Leong. Online Feature Selection for Model-based Reinforcement Learning, ICML, 2013.

T. Nguyen, T. Silander and T. Y. Leong. Transferring Expectations in Model-based Reinforcement Learning, NIPS, 2012.

Transferable Decision Making in Complex Environments

While Markov decision processes (MDPs) form an elegant mathematical basis for sequential decision making required in clinical medical practice, many challenges remain to make them applicable in real situations. Formalizing a real decision problem as an MPD is a non-trivial task. Consequently, in machine learning community there have been attempts to learn aspects of the MDP from the data gathered during previous decision making episodes. However, in these attempts the state space of the situations in which the decision making occurs is usually assumed as given. However, psychological studies of expertise suggest that the very essence of the superior skill is often in the representation of the situations and actions. While traditional methods rely on human experts to form the representation of the state space, we have conducted research that aims at developing methods that also try to learn the representation of the state space from a very large set of relatively primitive features of states. The key challenge in this work is the amount of computation needed for evidence based learning of the MDPs that allow principled sequential decision making. To pass this computational bottleneck we have opted for gaining statistical power by actively learning simpler representations of the environment than those traditionally used in MDPs.

Another hurdle in deploying MDP-based decision support tools in practice is the brittleness of the mathematical construction of the MDP. The theoretical solution found for one MDP cannot be reused if the state space of the MDP changes even a little bit. In a changing real world this is likely to occur often. Our studies have therefore tried to tackle the issue of knowledge transfer and the detection of situations where new knowledge representation needs to be constructed since the old ones do not work anymore. This calls for principled surprise management. In series of studies we have been able to develop a framework that is both scalable to complex environments and that can transfer knowledge from situation to another. The system has been tested in artificial and real robotic domains, but it remains to be transferred to a medical context.

Representative Publications

T. Nguyen, Z. Li, T. Silander and T. Y. Leong. Online Feature Selection for Model-based Reinforcement Learning. International Conference on Machine Learning (ICML 2013), 16-21 June 2013, Atlanta, Georgia, USA.

T. Nguyen, T. Silander and T. Y. Leong. Transferring Expectations in Model-based Reinforcement Learning. Advances in Neural Information Processing Systems (NIPS 2012), 3-8 Dec 2012, Lake Tahoe, Nevada, USA

Medical Image Mining

Due to the development of imaging techniques, there are large amounts of medical imaging data and analysis available in hospitals. These images form a huge potential knowledge base, which at the moment is virtually unused due to the non-symbolic nature of the data which makes it difficult to retrieve and manipulate. Medical image mining research at the School of Computing stems for the group's original interests in data mining. In the context of medical images, image mining deals with the extraction of knowledge, image data relationships and other patterns not explicitly stored in the images. This information can be used for teaching, clinical diagnosis and prognosis.

In a series of studies we have investigated different representations of the medical images in order to annotate the images with symbolic descriptions that enable textual queries to databases. For example, in one of these studies we developed a classification method for categorizing hematoma type of traumatic brain injury (TBI). The proposed method made use of the Sparse Representation Classifier (SRC) and Gabor based features. By combining a set of weak classifiers, the proposed method can label the TBI images without segmentation and feature selection that usually require laborious manual processing conducted by a human expert. The unsupervised nature of this case based technique can significantly reduce the time and effort for system calibration.

In the brain computed tomography (CT) image project, a content-based CT image retrieval system – TBIDoc has been developed. The system helps to manage and retrieve traumatic brain injury (TBI) data. With this web-based system, doctors can query previous patient data by uploading brain CT scan images. Figure 6 shows the new interface of the TBIDoc system. Using TBIDoc, doctors can find relevant previous cases in a fast and convenient way.

It is also important to automatically quantify clinical features in brain CT images for TBI prognosis. We have proposed a few techniques on extracting features from the brain CT scan images. This extracted information will be useful for the building of an automated prognosis system, which aims to predict the possible outcome of TBI. Associations between the TBI images/extracted features and possible outcome can be linked by machine learning techniques.

As an important step in content-based image retrieval and computer-assisted diagnosis, the indexing of brain CT slices orders the slices and aligns each individual slice onto the corresponding brain axial height. We have proposed a fast indexing method using anatomical feature classification without registration. Our method is able to accurately and efficiently index brain CT slices with its corresponding label.

We have also developed a new method for automatically tracing and quantifying the midline shift in brain CT images using Midline landmarks. The midline is an imaginary line that separates the brain into left and right hemispheres. Visual inspection of midline shift allows the assessment of brain injury that caused the shift. The proposed method can help doctors in diagnosis and prognosis of TBI.

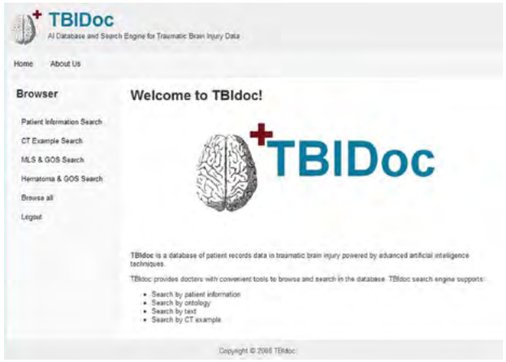


Figure 6: TBIDoc Interface

In addition, we proposed a labeling method to automatically classify the types of TBI. The proposed method makes use of the radiology report associated with the medical images. We first segment and reconstruct the 3D regions of interest (ROIs) from the medical images, and extract pathology and anatomy information from the associated report. We extract the volume, color, location, and shape features of the ROIs, and classify the types of ROIs using these features. The overall evaluation result is promising to doctors and medical professionals.

Representative Publications

T. A. Dinh, T. V. Silander, B Su, T Gong, B C Pang, C. C. T. Lim, C K Lee, C. L. Tan and T. Y. Leong . Unsupervised medical image classification by combining case-based classifiers, Medinfo, 2013.

T. A. Dinh, T. V. Silander, C. C. T. Lim and T. Y. Leong. An automated pathological class level annotation system for volumetric brain images, AMIA, 2012.

T. A. Dinh, T. V. Silander, C. C. T. Lim and T. Y. Leong. A generative model based approach to retrieving ischemic stroke images, AMIA, 2011.

T. Gong, S Li, J Wang, C L Tan, B C Pang, C C T Lim, C K Lee, Q Tian, Z Zhang, Automatic Labeling and Classification of Brain CT Images, ICIP, 2011.

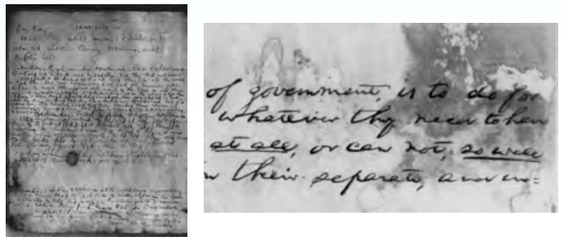
Document and Scene Text Analysis

Document analysis is the task of extracting information embedded in the document images captured using scanners or digital cameras. It is a great challenge to automatically retrieve and analyze document images. Document analysis research has a humble beginning 20 years ago when the first International Conference on Document Analysis and Recognition (ICDAR) was held in 1991 in Saint Malo, France. Today, document analysis research is a well-established research area with regular papers appearing in top journals such as IEEE Transactions on Pattern Analysis and Machine Intelligence, and Pattern Recognition. In the early years, document analysis research aimed at extracting and recognizing text in document images leading to the current matured technology of Optical Character Recognition (OCR). Another early research was in layout analysis of document images in order to enhance the correct detection of text regions in the documents. In recent years, document analysis research has moved into new areas including enhancement or correction of distorted or degraded document images, such as historical documents and text images captured by digital cameras and mobile devices. Another recent research interest is in the detecting and recognizing of the text in real scene images and videos. The document analysis group in the School of Computing is actively involved in these areas of research.

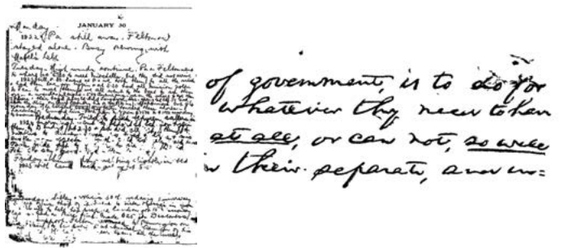
Document Image Binarization

Document image binarization is an important preprocessing technique, which aims to segregate the foreground text from the document background. A lot of binarization techniques have been proposed and work well on relatively clean documents. However, they are unable to perform well on degraded document images, includes smudges and smear, uneven illumination and intensity variation.

We have proposed a few techniques to binarize document images. But the thresholding of degraded document images is still an unsolved problem due to the high inter/intra-variation between the text stroke and the document background across different document images. We further developed a novel document image binarization technique that addresses these issues by using adaptive image contrast. The adaptive image contrast is a combination of the local image contrast and the local image gradient that is tolerant to text and background variation caused by different types of document degradations. In the proposed technique, an adaptive contrast map is first constructed for an input degraded document image. The contrast map is then binarized and combined with Canny's edge map to identify the text stroke edge pixels. The document text is further segmented by a local threshold that is estimated based on the intensities of detected text stroke edge pixels within a local window. The proposed method is simple, robust and involves minimum parameter tuning. Figure 7 shows some binarization results of our proposed method on some badly degraded document images.



(a) Degraded document image examples



(b) Binarization results of our proposed method

Figure 7: Binarization Examples

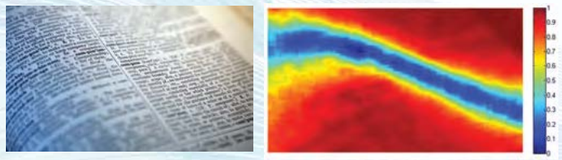
Besides developing the above new binarization techniques, we view document image binarization as a learning problem and proposed learning frameworks to improve the performance of existing binarization techniques. These proposed techniques first divide document pixels into three categories (foreground, background and uncertain) based on given document binarization methods. Then, learning techniques are applied to re-classify those uncertain pixels based on the foreground and background pixels.

Document Image Deblurring

Motion blur is a common distortion of document images, especially for those images taken with digital camera and mobile device. Blur often decreases the quality of document image and makes the text information within the document images unreachable by optical character recognition (OCR) or by a person.

To handle this problem, we first developed a simple and effective automatic image blurred region detection and classification technique. In the proposed technique, blurred image regions are first detected by examining singular value information for each image pixels. The blur types are then determined based on certain alpha channel constraint that requires neither image deblurring nor blur kernel estimation. The proposed technique can be used in many multimedia analysis applications such as image segmentation, depth estimation and information retrieval. Document image regions with different blur degrees can be detected and extracted using different thresholds, which is illustrated in Figure 8.
We then developed a blur correction technique that aims to correct motion blur within document images. Given a blurred document image, an alpha channel map is first constructed based on specific image

(a) (b)



(c) (d)

Figure 8: The document image in (a) contains defocus blur of different extents. Its corresponding singular value map is shown in (b), those regions with different blur degrees are highlighted in different color.

characteristics that are associated with text documents. Several blur parameters including blur direction and blur extent are then estimated from the constructed alpha channel map. Finally the blurred document image is restored by using a non-blind deconvolution technique based on the estimated blur parameters. Experiments on a number of document images with motion blur show that the proposed technique improves the document visual quality as well as the OCR performance significantly. Figure 9 shows two examples of deblurred document image results by our proposed method.

(a) The blurry images

(b) The deblurred results

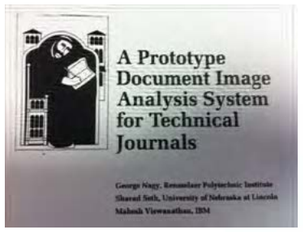


Figure 9: Examples of Document Image Deblurring

Scene Text Recognition

A large portion of images are taken in outdoor environments. In these cases, there is often text information, e.g., on road signs and bill boards. The ability for a machine to read such texts is of great importance because it facilitates a wide range of applications, both for individuals (e.g., text translation on mobile phones for tourists) and for large user communities (e.g., enhancing online maps with business names extracted automatically from images of major streets). Till today, this is still an unsolved problem due to difficult scenarios such as text with fancy fonts, text captured from a side angle, and unconstrained text content (e.g., a brand name does not have to be a word in the dictionary). Because of these challenges, applying OCR directly on scene images produces unsatisfactory recognition results.

A scene text recognition system typically consists of two main steps: detecting the locations of texts in a natural scene image, and recognizing the detected texts into alphanumeric strings. For the first step, text detection, we have developed two techniques. The first technique utilizes the fact that text exhibits a great deal of symmetry and regularity. Different from previous works which focus solely on the symmetry and regularity within the characters, we explore the same properties in the gaps between consecutive characters. This approach is useful for cases where the characters are hard to read (e.g., due to uneven illumination) but the gaps between them are still clearly visible. Experimentally, our technique is comparable to state-of-the-art on a public dataset for scene text detection. Some sample text detection results are shown in Figure 10.

Figure 10: Sample scene text detection results. The yellow bounding boxes indicate the text positions detected by our technique

For the second step, text recognition, most current research is still limited to recognizing texts that are horizontal and frontal to the image plane. However, in practice, scene texts can appear in any orientation, and with perspective distortion. Thus, we have developed a technique for recognizing perspective scene texts of arbitrary orientations. We also took into consideration the fact that it is labor-intensive and time-consuming to collect enough samples of perspective characters to train a classifier. For this reason, we used only frontal training data and extracted features that are robust to rotation and viewpoint change. This allows the extracted features to be used to recognize perspective characters. Experimentally, our technique shows significant improvement over the state-of-the-art on perspective scene texts. Figure 11 illustrates sample recognition results of our technique.

Figure 11: Sample recognition results of our technique on frontal, perspective and curved scene texts. The string below each image is the result returned by our technique for the given detected text

Video Text Recognition

Similar to natural scene images, videos often contain useful text information. A typical application of video text recognition is video indexing and retrieval, which becomes increasingly important with the rapid growth of online video databases.

Video text shares most of the challenges of scene text, with an additional problem: low resolution. This is due to the fact that videos are often compressed for fast transmission on the Internet.

The pipeline of a video text recognition system is similar to that of a scene text recognition system. However, the recognition step of the former needs to take into account the temporal information, because video text often appears on the screen for 2–3 seconds, in order to be readable by a human.
For the first step, text detection, we have developed a wide range of techniques for video text detection. One of our contributions to the community is to relax the assumption on text orientation. While most current research makes the assumption that text is horizontal, we have proposed techniques that are capable of handling multi-oriented texts as well as arbitrarily-oriented texts in video. Experimental results show that our techniques are able to detect video texts of a variety of orientation. Figure 12 shows two examples of video text successfully detected by one of our techniques.

Figure 12: Sample detection results of our technique for video text. In each pair of images, the one on the left shows the input video frame, and the one on the right shows the located text positions

For the second step, text recognition, we have developed several methods to reconstruct the complete shapes of video characters (which often have broken edges due to the problem of low resolution). Furthermore, to exploit the temporal information in video, we have proposed a technique to track linearly moving texts. The tracked text instances are then aligned and integrated into an enhanced and binarized text image. Experimental results show that using character reconstruction and temporal information improves the recognition accuracy significantly. Sample video text recognition results are shown in Figure 13.

Figure 13: Sample video text recognition result. The left hand side shows a few instances of the same word over time. The right hand side shows the integrated and binarized text image. By using temporal information, the final recognized string is correct, despite the low quality of the individual text instances.

In addition, a video may contain more than one language. Hence, we have developed a technique for video script identification. The texts of each language can then be sent to the respective OCR engine. This paves the way for research into multilingual video text recognition in the future, which is especially relevant in the Singapore context (with four languages: English, Chinese, Malay and Tamil).

Our binarization technique has achieved great performance in past years international competitions of document image bianrization, including the first prize in the Handwritten Document Image Binarization Contest (H-DIBCO 2010) which was held in conjunction with ICFHR 2010, the second position in performance in the Document Image Binarization Contest (DIBCO 2011) which was held under ICDAR 2011, and the third position in performance in H-DIBCO 2012 which was held in conjunction with ICFHR 2012. Our proposed method has been tested on several public dataset, and has achieved great performance and robustness.

Our proposed techniques also attract attentions from the industry. A

few start-up companies from Singapore and USA are keen in exploring the use of our document analysis techniques. They would like to make use of our document image binarization and scene text detection techniques for some applications.

Representative Publications

B. Su, S. Lu and C. L. Tan, A Robust Document Image Binarization for Degraded Document Images, IEEE Transactions on Image Processing 22(4):1408-1417, 2013.

P. Shivakumara, T. Q. Phan, S. Bhowmick, C. L. Tan and U. Pal, A Novel ring radius transform for video character reconstruction, Pattern Recognition 46(1):131-140, 2013.

P. Shivakumara, T. Q. Phan and C. L. Tan, Multi-oriented video scene text detection through Bayesian classification and boundary growing, IEEE Transactions on Circuits and Systems for Video Technology 22(8):1227-1235, 2012.

P. Shivakumara, T. Q. Phan and C. L. Tan, A Laplacian approach to multi-oriented text detection in video, IEEE Transactions on Pattern Analysis and Machine Intelligence 33(2):412-419, 2011.

T. Q. Phan, P. Shivakumara and C. L. Tan, Detecting Text in the Real World, ACM Multimedia, 2012.

B. Su, S. Lu and C. L. T an, Blurred Image Region Detection and Classification, ACM Multimedia, 2011.

Computational Natural Language Learning

Natural language processing (NLP) is an important AI subfield that studies and develops techniques to enable machines to understand human languages. Natural language expressions easily understood by humans pose great challenges to the machine in extracting the intended meaning from the text. Complex grammatical structures and idiomatic expression are barriers for the machine to overcome in extracting the correct information from text.

In recent year, enormous applications on the Internet have been developed which are affecting and shaping people's life. Among them, text based applications are particularly popular such as Search Engines, Blogs, Twitter, Facebook as well as other web social mining applications. Along with the development of these text-based web applications, natural language processing techniques have become increasingly more important, since they can help these web applications provide better user experience with more accurate results. This requirement strongly stimulates the research in Natural Language Processing (NLP) which has made quick progress in developing data driven applications. The NLP research domain also benefits from the availability of a large amount of natural language data and the improvement of the large scale statistical machine learning methods. The following are some of the machine learning research in the field of natural language processing.

Kernel Engineering for Natural Language Processing

In the field of NLP, the most familiar data representation is a sequence, since a document can be considered as a sequence of words. Many NLP applications rely on various fundamental text analysis techniques which have been developed for the sequential data. In addition, these techniques can be built on more complex data representations. For instance, syntactic parsing creates a tree structure over a sentence with each of the node in the tree denoting the grammatical function of the span of text it covers. Thus the sequence of words is enriched with a tree structure as the grammatical analysis. By means of these augmentations of the input data, the input data can be represented by more complex structures such as multiple sequence, trees and even graphs other than a sequence of plain works. Traditionally, statistical machine learning methods adopt a vector representation to express the features embedded in these complex structures. A preprocessing step is required to transform these complex structured data into a corresponding feature vector. The preprocessing often requires an expert to appropriately design and extract the features which is obviously non-trivial. In addition, this feature engineering process may lead to loss of information of the input data.

To better perform these structural data driven tasks, kernel methods are introduced which can directly take these structures as the input to kernel-based machine learning algorithms instead of explicitly representing the input data as a feature vector. However, existing tree kernels explore the structure features with respect to a single sub-tree representation. The structure of a large single sub-tree may be sparse in the data set, which prevents large structures from being effectively utilized. Sometimes, only certain parts of a large sub-tree are beneficial instead of the entire sub-tree. In this case, using the entire structure may introduce noisy information. To address the above efficiency, we investigate the phrase parse tree and attempt to design more sophisticated kernels to deeply explore the structure features embedded in the phrase trees other than the single sub-tree representation. Specifically, we propose tree sequence based kernels to explore the structure features in the phrase parse trees for various NLP applications. We show that our algorithm can benefit a variety of NLP applications such as Statistical Machine Translation, Question Classification and Entity Relation Extraction.

Event Co-reference Resolution

Coreference resolution, the task of resolving a given text expression to its referred expression in prior texts, is important for intelligent text processing and information extraction. Most previous work on coreference resolution aims at object coreferences in which both the two text expressions are mentions of the same concrete real world object (e.g. person, location and organization) as shown in the following example:

"I bought a house yesterday. It was in the sub-urban area".

Here the pronoun "it" refers to the object "house". In contrast, event corefrences are mentions that refer an abstract entity such as action, event and proposition. The following are three examples of event coreferences:

"I ran two miles. The run did me good."
"Fred wanted to go to the movies. But his mother wouldn't allow it."
"John's hitting Fred got everyone in trouble, for it led to a brawl"

In the above, the action "ran", the intention "to go to the movies", and the event "John's hitting Fred" are being referred by a later mention in the respective sentences. Event Coreference resolution is a very important task for event template extraction and other natural language processing tasks, but it is a much more challenging task than object coreference resolution.

In our research for event coreference resolution, we adopt a divide-and-conquer approach by separating the complicated event coreference resolution to three steps, namely, Event Mention Detection, Mention Pair Resolution, and Event Chain Formation. At event mention detection step, we propose a topic related approach to effective identify the potential event mentions with minimal noise included. At mention pair resolution step, we propose a collective decision process utilizing multiple resolution results and a new training strategy. At the last and most important step, event chain formation, we propose two graph partitioning methods (spectral graph partitioning and random walk partitioning) with deep linguistic constraints and preferences. Our proposed resolution system achieves the state-of-the-art performance.

Unsupervised Structure Induction

Many Natural Language Processing (NLP) tasks involve some kind of structure analysis, such as word alignment for machine translation, syntactic parsing for coreference resolution, semantic parsing for question answering, etc. Traditional supervised learning methods rely on manually labeled structures for training. The rising amount of available text data on the Internet can provide huge resources for these tasks and improve their performance. Unfortunately, manual annotations are often expensive and time-consuming for large amount of rich text. Therefore, it is of great value to automatically induce structures from un-annotated sentences for NLP research. In this research, we explore three unsupervised structure induction tasks: transliteration equivalence learning, constituency grammar induction, and dependency grammar induction

We first investigate transliteration equivalence learning where transliterated bilingual word pairs are given without internal syllable alignments. We study nonparametric Bayesian learning methods and propose a synchronous adaptor grammar for machine transliteration. Experiments demonstrate the proposed model outperforms the baseline EM-based model significantly on the transliteration task for four language pairs. Our proposed model provides a general framework to automatically learn syllable equivalents without heuristics or restrictions.

In linguistics, a constituent is a word or a group of words that represents some linguistic function as a single unit. There are many kinds of constituents according to their linguistic functions, such as sentence, noun phrase, verb phrase, and prepositional phrase. The hierarchical structure of constituents forms a constituency tree. From the constituency tree, we can extract context free transformation rules. To facilitate syntactic analysis, constituency tree banks have been created for some languages, such as the Peen English Treebank and the Penn Chinese Treebank. However, manually creating tree structures is expensive and time consuming. So far, annotated treebanks are only available for a few languages. Constituency grammar induction is therefore useful which allows us to learn constituency grammar from plain strings (words or part-of-speech tags). The induced grammar can be used to construct large treebanks, study language acquisition, improve machine translation, and so on. We propose a local feature-based model with gradient-based optimization for unsupervised constituency grammar induction. Various kinds of linguistic knowledge can be encoded into the model which shows improvements over baselines.

Constituency grammars perform well for languages with relatively strict word order (e.g. English). However, some free word order languages (e.g. Czech, Turkish) lack a finite verb phrase constituent, making constituency parsing difficult. In contrast, dependency grammars model the word-to-word dependency relations, which is more suitable for languages with free word order. Combinatory Categorial Grammar (CCG) is a linguistically expressive lexicalized grammar formalism which is able to capture long-range dependencies. We follow the state-of-the-art induction approach and propose a boundary model and Bayesian model, which improves the baseline on Penn treebank datasets.

Linking Entities to a Knowledge Base

The explosive growth in the amount of textual information brings a need for building a structured Knowledge Base (KB) to organize the knowledge scattered among these unstructured texts. On the other hand, the available KBs such as Wikipedia and Google Knowledge Graph which contain rich knowledge about the world's entities have been shown to form a valuable component for many NLP tasks. To populate or to utilize the KBs, we need to link the mentions of entities in text to their corresponding entries in the KB, which is called entity linking.

Most of state-of-the-art entity linking systems use annotated data to learn a classifier or ranker by supervised learning algorithms. Our research initially focuses on automatically labeling a large scale training corpus for the supervised learning algorithms, where we label the ambiguous mentions leveraging on their unambiguous synonyms. We also propose an instance selection strategy to select an informative, representative and diverse subset from the autogenerated data set.

Next, we introduce topic models to entity linking for measuring the context similarity between mention and KB entries. We propose a Wikipedia-LDA method to model the context as some hidden topics instead of only treating the context as literal terms. We investigate the effectiveness of five subsets from Wikipedia categories to represent the underlying topics. Besides, we propose a lazy learning model for entity linking, which can incorporate the query-specific information to the learning process by automatically labeling some data for the queried name. Then, instead of only using the labeled data set related with other names to train the linker, we propose to use the predictive structure shared by the two data sets which are related with queried name and other names respectively.

Finally, we perform entity linking task under a more challenging scenario, where we link the mentions in microblog to a KB in real time. We propose an unsupervised-supervised learning framework (USLF) which is based on three bipartite graphs to address the new challenges in microblog. Our USLF uses a Bayes method to model the three clues of disambiguation: the context information of query and entities, popularity knowledge of entities and clustering result on an additional tweet set. Besides, in our USLF, a tweet enrichment function is embedded based on the word similarity, which is calculated in the k principal component space of the word set with the help of auxiliary long text.

Our Entity Linking system achieved 2nd best performance among 21 teams in National Institute of Standards and Technology (NIST, US) 2011 entity linking task benchmarking.

From Semantic to Emotional Space in Sense Sentiment Analysis

Sentiment analysis/ Opinion mining aims to enable computers to derive sentiment from human language. Here, we aim to address sense sentiment similarity that is one of the fundamental tasks in sentiment analysis to infer the similarity between word pairs with respect to their senses and their underlying sentiments.

To achieve this aim, we analyze sentiments in Indirect Question Answer Pairs (IQAPs) to infer "yes" or "no" response, and show how

similarity between adjectives in question and answer pairs can be a main factor to address this inference task. Consider the following IQAPs as examples with the adjectives playing play pivotal roles in inferring different degrees of yes and no:

Row	IQAP	Answer
E1	Q: Do you think that's a great idea? A: I think it's acceptable	Weak-yes
E2	Q: Was she the best one on that old show? A: She was simply funny.	Strong-yes
E3	Q: He says he opposes amnesty, but…. Is he right? A: He is a bit funny.	Weak-no
E4	Q: ….Is that true? A: This is extraordinary and preposterous.	Strong-no

Traditional approach to finding similarity between words has been based on semantic similarity using methods such as Latent Semantic Analysis, Point-wise Mutual Information, and WordNet-based Similarity. In our research, we show that another type of similarity measure based on emotion rather than semantics is capable of capturing more accurately the underlying sentiment between words. Here, we propose an effective approach to compute sentiment similarity from a connection between semantic space and emotional space. Our model maps from senses of words to vectors of basic human emotions. We initially employ a fixed set of twelve basic human emotions to construct emotion vectors. However, our experiments show that there is little agreement on the number and types of basic emotions. This observation leads to our next model in which the number and types of basic emotions are not pre-defined. In other words, the emotions are deemed to be hidden. A probabilistic approach is thus proposed to infer sense sentiment similarity with respect to automatically learned hidden emotions. We learn hidden emotions from the interaction between reviews, rates, and words. We show that our emotion similarity based model significantly outperforms two popular semantic similarity measures.

Representative Publications

M. Mohtarami, M. Lan, H. Amiri and C. L. Tan. Probabilistic Sense Sentiment Similarity through Hidden Emotions, ACL, 2013.

M. Mohtarami, C. L. Tan. From Semantic to Emotional Space in Probabilistic Sense Sentiment Analysis, AAAI, 2013.

M. Mohtarami, H. Amiri, M. Lan, T. P. Tran, C. L. Tan. Sense Sentiment Similarity: An Analysis, AAAI, 2012.

Zhang W, J Su, C L Tan, Y Cao and C Y Lin. A Lazy Learning Model for Entity Linking Using Query-Specific Information, COLING, 2012.

M. Mohtarami, H. Amiri, M. Lan, C. L. Tan. Predicting the uncertainty of sentiment adjectives in indirect answers, CIKM, 2011.

J. Sun , M Zhang and C L Tan, Tree sequence kernel for natural language, AAAI, 2011.

Y Huang, M Zhang and C L Tan, Nonparametric Bayesian machine transliteration with synchronous adaptor grammars, ACL-HLT, 2011.

Planning and Decision Making

Decision-theoretic planning has recently received significant attention. Unlike classical AI planning, it allows sequential decision making problems to be modeled more realistically using the notion of uncertainty (e.g., in terms of uncertainties in sensing and actions). Furthermore, in contrast to classical AI planning that produces plans guaranteed to achieve certain goals, decision-theoretic planning aims to generate courses of actions (i.e., plans or policies) that achieve high expected utility.

There are three separate research efforts in decision-theoretic planning: (a) POMDP planning for robotic tasks, (b) information gathering agents for large-scale modeling and prediction of spatiotemporal traffic phenomena, and (c) planning under uncertainty for large-scale active multi-camera surveillance, which are discussed below:

POMDP Planning for Robotic Tasks

Partially observable Markov decision processes (POMDPs) provide a principled mathematical framework for planningvand decision making under uncertainty. POMDPs have broad applications in robotics, computer games, dialog systems, etc., but they are often avoided in practice, because solving POMDPs exactly is computationally intractable. Not long ago, the best algorithms could deal with POMDPs with a dozen states, which are woefully inadequate for any practical applications.

We are developing algorithmic tools for efficient approximate POMDP solutions. We are also developing theoretical foundations to identify and characterize sub-classes of POMDPs amenable to efficient algorithms. Our ultimate goal is to make POMDPs a practical technology for uncertainty planning in robotics and beyond. We have recently proposed POMDP algorithms that are able to handle continuous state spaces, continuous observations and Bayesian reinforcement learning problems. Our continuous state space POMDP solver, MCVI, has been released as a software package over the Internet.

Representative Publications

H. Bai, D. Hsu, M. J. Kochenderfer and W. S. Lee, Unmanned Aircraft Collision Avoidance using Continuous-State POMDPs, RSS, 2011.

Information Gathering Agents for Large-Scale Modeling and Prediction of Spatiotemporal Traffic Phenomena

Knowing and understanding the traffic conditions and phenomena over road networks has become increasingly important to the goal of achieving smooth-flowing, congestion-free traffic, especially in densely-populated urban cities. According to a 2011 urban mobility report, traffic congestion in the USA has caused 1.9 billion gallons of extra fuel, 4.8 billion hours of travel delay, and $101 billion of delay and fuel cost. Such huge resource wastage can potentially be mitigated if the spatiotemporally varying traffic phenomena (e.g., speeds and travel times along road segments) are predicted accurately enough in real time to detect and forecast the congestion hotspots; network-level (e.g., ramp metering, road pricing) and user-level (e.g., route re-planning) measures can then be taken to relieve this congestion, so as to improve the overall efficiency of road networks.

In practice, it is non-trivial to achieve real-time, accurate prediction of a spatiotemporally varying traffic phenomenon because the quantity of sensors that can be deployed to observe an entire road network is cost-constrained. Traditionally, static sensors such as loop detectors

are placed at designated locations in a road network to collect data for predicting the traffic phenomenon. However, they provide sparse coverage, incur high installation and maintenance costs, and cannot reposition by themselves in response to changes in the traffic phenomenon. GPS technology enables the collection of traffic data using passive mobile probes (e.g., taxis/cabs). Unlike static sensors, they can directly measure the travel times along road segments. But, they provide fairly sparse coverage due to low GPS sampling frequency (i.e., often imposed by taxi/cab companies) and no control over their routes, incur high initial implementation cost, pose privacy issues, and produce highly-varying speeds and travel times while traversing the same road segment due to inconsistent driving behaviors. In contrast, we propose the use of active mobile probes to overcome the limitations of static and passive mobile probes. In particular, they can be directed to explore any segments of a road network to gather traffic data at a desired GPS sampling rate while enforcing consistent driving behavior.

How then do the mobile sensing agents actively explore a road network to gather and assimilate the most informative observations for predicting the traffic phenomenon?

There are three key issues surrounding this problem:

(a) Models for predicting spatiotemporal traffic phenomena. The spatiotemporal correlation structure of a traffic phenomenon can be exploited to predict the traffic conditions of any unobserved road segment at any time using the observations taken along the sensors' paths. We assume the traffic phenomenon over an urban road network (i.e., comprising full range of road types like highways, arterials, slip roads, etc.) to be realized from a rich class of Bayesian non-parametric models called the Gaussian process (GP) that can formally characterize its spatiotemporal correlation structure and refine it with growing number of observations. More importantly, GP can provide formal measures of predictive uncertainty (e.g., based on variance or entropy criterion) for directing the sensors to explore highly uncertain areas of the road network. This work proposes a relational GP whose correlation structure exploits the geodesic distance between segments based on the topology of a directed road network with vertices denoting road segments and edges indicating adjacent segments weighted by dissimilarity of their features, hence tightly integrating the features and relational information.

(b) Data fusion. The observations are gathered distributedly by each sensor along its path in the road network and have to be assimilated in order to predict the traffic phenomenon. Since a large number of observations are to be collected, a centralized approach to GP prediction cannot be performed in real time due to its cubic time complexity. To resolve this, we propose a decentralized data fusion approach to efficient and scalable approximate GP prediction.

(c) Active sensing. The sensors have to coordinate to actively gather the most informative observations for minimizing the uncertainty of modeling and predicting the traffic phenomenon. Existing centralized and decentralized active sensing algorithms scale poorly with large observations and sensors. We propose a partially decentralized active sensing algorithm that overcomes these issues of scalability.

This work presents a novel Decentralized Data Fusion and Active Sensing (D2FAS) algorithm for sampling spatiotemporally varying environmental phenomena with mobile sensing agents. Note that the decentralized data fusion component of D2FAS can also be used for static and passive mobile sensors. Its predictive performance is theoretically guaranteed to be equivalent to that of a sophisticated centralized sparse approximation for the GP model. This result implies that the computation of such a sparse approximate GP model can be parallelized and distributed among the mobile sensors (in a Google-like MapReduce paradigm), thereby achieving efficient and scalable prediction. We also theoretically guarantee its active sensing performance that improves under various practical environmental conditions. The practical applicability of D2FAS is not restricted to traffic monitoring; it can be used in other environmental sensing applications such as monitoring of ocean and freshwater phenomena, pollution, or contamination. Empirical evaluation on real-world urban road network data shows that our D2FAS algorithm is significantly more time-efficient and scalable than state-of-the-art centralized algorithms while achieving comparable predictive performance.

Representative Publication

J. Chen, K. H. Low, C. K.-Y. Tan, A. Oran, P. Jaillet, J. M. Dolan, and G. S. Sukhatme. Decentralized Data Fusion and Active Sensing with Mobile Sensors for Modeling and Predicting Spatiotemporal Traffic Phenomena, UAI, 2012.

Planning under Uncertainty for Large-Scale Active Multi-camera Surveillance

The problem of surveillance has grown to be a critical concern in many urban cities worldwide following a recent series of security threats like Mumbai terrorist attacks and London bomb blasts. Central to the problem of surveillance is that of monitoring, tracking, and observing multiple mobile targets of interest distributed over a large-scale obstacle-ridden environment (e.g., airport terminals, railway and subway stations, bus depots, shopping malls, school campuses, military bases). It is often necessary to acquire high-resolution videos/images of these targets for supporting real-world surveillance applications like activity/intention tracking and recognition, biometric analysis like target identification and face recognition, surveillance video mining, forensic video analysis/retrieval, among others.

Traditional surveillance systems consist of a large network of fixed/static CCTV (Closed Circuit Television) cameras that are placed to constantly focus at selected important locations in the buildings like entrance/exit, lobby, etc. Unfortunately, the maximum resolution of these cameras is limited to 720 x 480 pixels. So, they cannot capture high-resolution images/videos of the targets, especially when the targets are far away from the cameras. As a result, they perform poorly in acquiring the close-up views of the targets and their activities. HDTV/Megapixel cameras have recently been introduced to overcome this resolution issue. Similar to CCTV cameras, these fixed/static HDTV/megapixel cameras are placed to constantly focus at specific locations in the environment. A relatively large network of such cameras has to be installed in order to observe the targets in any region of the environment at high resolution, which is impractical in terms of equipment, installation, and maintenance costs.

The use of active PTZ (Pan/Tilt/Zoom) cameras is becoming an increasingly popular alternative to that of fixed/static cameras for surveillance because the active cameras are endowed with pan-tilt-zoom capabilities that can be exploited to focus on and observe the targets at high image/video resolution. Hence, fewer active cameras need to be deployed to be able to capture high-resolution images/videos of the targets in any region of the environment. In order to achieve effective real-time surveillance, an efficient automated mechanism is required to autonomously coordinate and control these cameras' actions.

The objective of this work is thus to address the following central surveillance problem: "How can a network of active cameras be coordinated and controlled to maximize the number of targets observed with a guaranteed image resolution?"

This work presents a novel principled decision-theoretic planning under

uncertainty approach to coordinating and controlling a large-scale network of active cameras for performing high-quality surveillance of large crowds of moving targets. In particular, our approach addresses the following practical issues affecting the surveillance problem:

(a) Multiple sources of uncertainty. A typical surveillance environment is fraught with multiple sources of uncertainty such as noisy cameras' observations, stochastic targets' motion, and unknown targets' locations, etc. These uncertainties make it difficult for the active cameras to know where to observe in order to keep the targets within their fields of view (fov). Consequently, they may lose track of the observed targets. To resolve this, our approach models a belief over the targets' states (i.e., locations, directions, and velocities) and updates the belief in a Bayesian paradigm based on probabilistic models of the targets' motion and the active cameras' observations;

(b) Camera-target ratio. In crowded environments, the number of targets to be observed is usually much greater than the number of available cameras. When the number of targets increases, a surveillance system, if poorly designed, tends to incur exponentially increasing computation time, which degrades the real-time performance of the entire surveillance system;

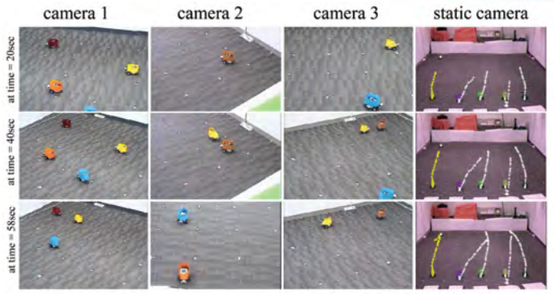
(c) Trade-off between maximizing the expected number of observed targets and the image resolution of observing them. Increasing the resolution of observing some targets through panning, tilting, or zooming may result in the loss of other targets being tracked. To address this trade-off, the cameras' actions are coordinated to simultaneously improve the belief over the targets' states and maximize the expected number of targets observed with a guaranteed pre-defined resolution;

(d) Scalability. By exploiting the inherent structure of the surveillance problem, our approach can scale linearly in the number of targets to be observed;

(e) Real-time requirement. The cameras' actions are computed in real time;

(f) Occlusions. Many real-world surveillance environments contain obstacles that occlude the fov of some or perhaps even all of the cameras, thus preventing the cameras from persistently tracking their observed targets. The regions where the targets cannot be observed by any of the cameras due to obstacles are said to be occluded. When the targets reside in these occluded regions or are not within the fov of any camera, the surveillance system loses track of them, thus degrading the surveillance performance. Such environments are called partially observable in the sense that the exact locations of the targets may not be observed directly by the cameras at all times.

As demonstrated empirically through simulations, our approach can achieve high-quality surveillance of a large number of targets in real time and its surveillance performance degrades gracefully with an increasing number of targets. We have also conducted real camera experiments to show the practicality of our decision-theoretic approach to coordinate and control 3 AXIS 214 PTZ cameras to monitor up to 6 Lego robots (i.e., targets) in a surveillance environment with the size of 11 × 9 grid cells. The size of each grid cell is 0.5 m2. Each active camera has 3 states. The states of the cameras are determined such that all the cells of the environment can be observed at high resolution by at least one camera. We have a low-resolution static camera that can track these robots based on OpenCV Camshift tracker. The static camera is calibrated to obtain the approximate locations of the robots at every time step. The direction and velocity of the robots are determined based on their previous and current estimated locations. We have tested our implementation up to 6 robots while keeping one of the robots static. It can be observed in Figure 14 that cameras 2 and 3 coordinate to observe the brown static robot as follows: Camera 2 pans to another state (see bottom two rows of Figure 14) only when camera 3 takes over the observation of the static target (see top two rows of Figure 14). This static target can be replaced by a portion of the surveillance environment like the entrance/exit or reception where we need to pay more attention. The experimental setup and the detailed results are shown in our demo video at http://www.comp.nus.edu.sg/~lowkh/camera.html



Representative Publications

P. Natarajan, T. N. Hoang, K. H. Low, and M. Kankanhalli. Decision-Theoretic Coordination and Control for Active Multi-Camera Surveillance in Uncertain, Partially Observable Environments, ICDSC, 2012.

P. Natarajan, T. N. Hoang, K. H. Low, and M. Kankanhalli. Decision-Theoretic Approach to Maximizing Observation of Multiple Targets in Multi-Camera Surveillance, AAMAS, 2012.

The faculty members in AI research are:

- Michael Brown
- David Hsu
- Wynne Hsu
- Sanjay Jain
- Lee Mong Li
- Lee Wee Sun
- Leong Tze Yun
- Low Kian Hsiang, Bryan
- Frank Christian Stephan
- Tan Chew Lim
- Tan Keng Yan, Colin
- Yap Hock Chuan, Roland

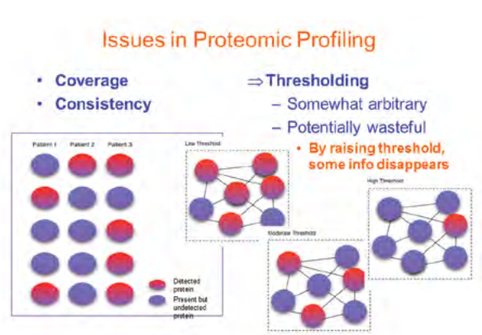
COMPUTATIONAL BIOLOGY

Present-day biomedical researchers are confronted by vast amounts of data from genome sequencing; proteomics; microscopy; high-throughput analytical techniques for DNA, RNA, and proteins; and a host of other new experimental technologies. Coupled with advances in computing power, this flow of information enables scientists to computationally model and analyze biological systems in novel ways.

Accordingly, the Computational Biology Lab in SOC works towards fundamental advances in knowledge discovery, database management, combinatorial algorithms, and modeling and simulation, as well as in the applications of these technologies to problems in biology and medicine. Some of our research projects and activities are described below.

More Sophisticated Proteomic Profile Analysis

Mass spectrometry (MS)-based proteomics is a powerful tool for profiling systems-wide protein expression changes. It can be applied for various purposes, e.g., biomarker discovery in diseases and study of drug responses. However, MS-based proteomics tend to have consistency (poor reproducibility and inter-sample agreement) and coverage (inability to detect the entire proteome) issues. We deal with these two challenges by proposing approaches that analyze proteomic profiles in the context of biological networks.



We highlight here one of our recent results:

- We develop a novel method, PSP (Proteomics Signature Profiling), which utilizes all detected proteins. In this method, for each patient, a hit rate is calculated for each protein complex based on its reported proteins in the patient, in the absence of fold change threshold. Although consistency of individual proteins between patients is low, we found the reported proteins tend to hit protein complexes in a meaningful and informative manner. By extracting this information in the form of a Proteomics Signature Profile, we confirm that this information is conserved and can be used for (1) clustering of patient samples, (2) identification of significant clusters based on real biological complexes, and (3) overcoming consistency and coverage issues prevalent in proteomics data sets.

Representative Publications

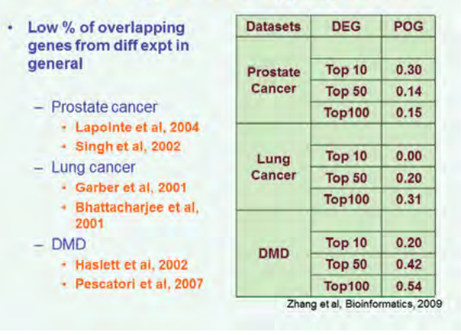
Wilson Goh, Yie Hou Lee, Maxey Chung, Limsoon Wong. How advancement in biological network analysis methods empowers proteomics. Proteomics, 12(4-5):550-563, 2012.

Wilson Goh, Yie Hou Lee, Ramdzan Zubaidah, Jingjing Jin, Difeng Dong, Qingsong Lin, Maxey Chung, Limsoon Wong. A Network-based pipeline for analyzing MS data---An application towards liver cancer. Journal of Proteome Research, 10(5):2261-2272, 2011.

Wilson Goh, Yie Hou Lee, Zubaidah Ramdzan, Marek Sergot, Maxey Chung, Limsoon Wong. Proteomics Signature Profiling (PSP): A novel contextualization approach for cancer proteomics. Journal of Proteome Research, 11(3):1571-1581, 2012.

More Sophisticated Analysis of Gene Expression Profiles

Many approaches have been proposed for the inference of differentially expressed genes that are useful for understanding diseases and treatment response. However, the statistical significance of the selected genes and the reproducibility of the resulting diagnosis system have a high degree of uncertainty. In particular, many of these methods produce gene lists that do not significantly overlap when they are applied to different data sets of the same phenotypes. Furthermore, the transition from the selected genes to the understanding of the sequences of causative molecular events is unclear.



Datasets	DEG	POG
Prostate Cancer	Top 10	0.30
	Top 50	0.14
	Top100	0.15
Lung Cancer	Top 10	0.00
	Top 50	0.20
	Top100	0.31
DMD	Top 10	0.20
	Top 50	0.42
	Top100	0.54

In order to qualitatively improve the statistical power of microarray analysis methods and the reliability of the results, additional dimensions present in the problem have to be brought into consideration. In particular, each disease generally has an underlying cause. So there should be a unifying biological theme---a biological pathway or a subnetwork of protein interactions---for genes that are truly associated with the disease. We thus envision an advanced integrated framework, and are developing a system based on it, to provide biologically inspired solutions. It comprises: (1) automated analysis and extraction of information from biomedical texts; (2) targeted construction of known pathways; and (3) direct hypothesis generation based on logical reasoning on, and tests for, consistencies of observed data against known pathways and models.

We highlight here one of our recent results:

- Contemporary methods of microarray analysis have a tendency to produce different results from different datasets of the same disease. We solve this reproducibility problem by introducing a technique (SNet). SNet provides both quantitative and descriptive analysis of microarray datasets by identifying specific subnetworks of pathways that are significant. We tested SNet on independent

datasets of several diseases, including childhood ALL, DMD and lung cancer. In each case, SNet consistently produced almost the same list of significant nontrivial subnetworks from two independent sets of microarray data.

Representative Publications

Donny Soh, Difeng Dong, Yike Guo, Limsoon Wong. Finding consistent disease subnetworks across microarray datasets. BMC Genomics, 12(Suppl. 13):S15, 2011.

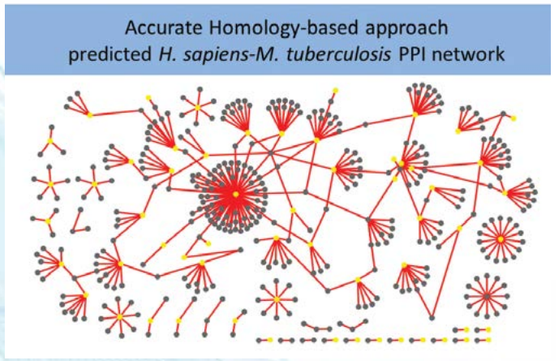
Chuan Hock Koh, Sucheendra K. Palaniappan, P. S. Thiagarajan, Limsoon Wong. Improved statistical model checking methods for pathways analysis. BMC Bioinformatics, 13(Suppl 17):S15, 2012.

Chuan Hock Koh, Limsoon Wong. Embracing noise to improve cross-batch prediction accuracy. BMC Systems Biology, 6(Suppl 2):S3, 2012.

Reliable Protein Interactomes for Infectious Diseases

There is a critical need to address the emergence of drug resistant varieties of pathogens for several infectious diseases. Approaches to counter drug resistance have so far achieved limited success. It has been proposed that this lack of success is due to a lack of understanding of how resistance emerges in bacterial upon drug treatment and that a systems-level analysis of the proteins and interactions involved is essential to gaining insights into routes required for drug resistance.

The challenges in the systems-level analysis of proteins and interactions in pathogens of infectious diseases for identifying drug resistance pathways include: (1) unreliability of bacterial genome sequence annotations; (2) unreliability of bacterial interactome maps; (3) paucity of bacterial interactome maps; (4) prediction of novel protein interactions in bacteria and between bacteria and host; and (5) identification candidate mutations and pathways to drug resistance in a bacterium.



Our recent results include:

- The effective use of public databases (e.g., KEGG, WikiPathways, and BioCyc) is hindered by issues such as incompatible data formats, inconsistent molecular representations, inconsistent molecular relationship representations, inconsistent referrals to pathway names, and incomprehensive data from different databases. We overcome these issues through extraction, normalization and integration of pathway data from several major public databases in a “full unification” approach. This results in a database, IntPath (Integrated Pathway gene relationship database for model organisms and important pathogens), which contains much richer pathway-gene and pathway-gene pair relationships and much larger number of non-redundant genes and gene pairs than any of the single-source databases.

- The large amount of inconsistency in gene structure annotations of bacterial strains impedes effective comparative genomic analysis of bacterial strains in promising applications such as gaining insights into bacterial drug resistance. We develop CAMBer as an approach to support comparative analysis of multiple bacterial strains. CAMBer produces what we called multigene families. Each multigene family reveals genes that are in one-to-one correspondence in the bacterial strains, thereby permitting their annotations to be integrated. As a result, more accurate and more comprehensive annotations of the bacterial strains can be produced.

Representative Publications

Hufeng Zhou, Jingjing Jin, Haojun Zhang, Bo Yi, Michal Wozniak, Limsoon Wong. IntPath---an integrated pathway gene relationship database for model organisms and important pathogens. BMC Systems Biology, 6(Suppl 2):S2, 2012.

Michal Wozniak, Limsoon Wong, Jerzy Tiuryn. CAMBer: An appproach to support Comparative Analysis of Multiple Bacterial strains. BMC Genomics, 12(Suppl. 2):S6, 2011.

Michal Wozniak, Jerzy Tiuryn, Limsoon Wong. An approach to identifying drug resistance associated mutations in bacterial strains. BMC Genomics, 13(Suppl 7):S23, 2012.

Protein Complex Prediction

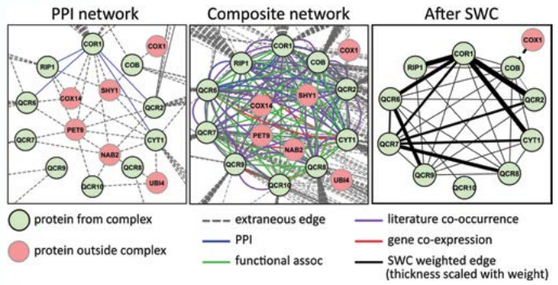
Progress in high-throughput experimental techniques in the past decade has resulted in a rapid accumulation of protein-protein interaction (PPI) data. However, interaction data obtained by the popular high-throughput assays such as two-hybrid experiments may contain as much as 50% false positives and false negatives. Also, PPI networks resulting from these assays are still essentially an in vitro scaffold. Further progress in computational analyses techniques and experimental methods is needed to reliably deduce in vivo protein interactions, to distinguish between permanent and transient interactions, to distinguish between direct protein binding from membership in the same protein complex and to distinguish protein complexes from functional modules.

We aim to advance computational techniques for: (1) assessing the reliability of PPIs detected by high-throughput methods; (2) deriving protein complexes from PPI networks; and (3) exploring approaches for distinguishing the various forms of interactions mentioned above.

Our recent results include:

- Existing algorithms for discovering protein complexes from PPI networks generally do not take into consideration the fact that not all the interactions in a PPI network take place at the same time. As a result, predicted complexes often contain many spuriously included proteins. We tackle this problem by using localization GO terms and hubs to decompose a PPI network before complex discovery, which achieves considerable improvement.

- Existing algorithms for discovering protein complexes from PPI networks are also hindered by the high rate of spurious and missing interactions in high-throughput PPI data. As a result, predicted complexes often do not match true complexes well, and many true complexes go undetected. We address these challenges by integrating PPI data with other heterogeneous data sources to construct a composite protein network, and using a supervised maximum-likelihood approach, SWC, to weight each edge based on its posterior probability of belonging to a complex. Our method recalls more known complexes, achieves higher precision at all recall levels, and generates novel complexes of greater functional similarity than earlier methods.



Representative Publications

Guimei Liu, Chern Han Yong, Hon Nian Chua, Limsoon Wong. Decomposing PPI networks for complex discovery. Proteome Science, 9(Suppl. 1):S15, 2011.

Min Wu, Xiaoli Li, Chee-Keong Kwoh, See-Kiong Ng, Limsoon Wong. Discovery of protein complexes with core-attachment structures from TAP data. Journal of Computational Biology, 19(9):1027-1042, 2012.

Chern Han Yong, Guimei Liu, Hon Nian Chua, Limsoon Wong. Supervised maximum-likelihood weighting of composite protein networks for complex prediction. BMC Systems Biology, 6(Suppl 2):S13, 2012.

Data Mining Algorithms for Pharmacogenomics

Human genome harbors millions of common single nucleotide polymorphisms (SNPs) and other types of genetic variations. These genetic variations play an important role in understanding the correlation between genetic variations and human diseases and the body’s responses to prescribed drugs. The discovery of such genetic factors has come to be known as pharmacogenomics. In this project, we explore several pharmacogenomic-related applications of database and data mining technologies.

Our recent results include:

- Genotyping all SNPs is very expensive. Fortunately, adjacent SNPs are often not independent. It is thus desirable to select a subset of SNPs that are sufficient to infer all the other SNPs. These selected SNPs are called tag SNPs. We develop the FastTagger algorithm to select tag SNPs based on multi-marker correlations. FastTagger is many times faster, consume much less memory, and also reduce the number of selected tag SNPs, than existing tag SNP selection algorithms. At the same time, we also develop techniques to use tagging rules (discovered in the process of tag SNP selection) to impute untyped SNPs at significantly higher accuracy and sensitivity than existing methods.

- Recent studies suggested that a combination of multiple single nucleotide polymorphisms (SNPs) could have more significant associations with a specific phenotype. However, to discover these epistatic interactions of SNPs is a computationally challenging task. We develop an efficient Cloud-based Epistasis cOmputing (eCEO) model for large-scale epistatic interaction in genome-wide association study (GWAS). Given a large number of combinations of SNPs, our eCEO model is able to distribute them to balance the load across the processing nodes. Moreover, our eCEO model can efficiently process each combination of SNPs to determine the significance of its association with the phenotype. In addition, we have also deployed our eCEO model on the Amazon Elastic Compute Cloud. Our study further confirms its efficiency and ease of use in a public cloud.

Representative Publications

Guimei Liu, Yue Wang, Limsoon Wong. FastTagger: An efficient algorithm for genome-wide tag SNP selection. BMC Bioinformatics, 11:66, 2010.

Yue Wang, Guimei Liu, Mengling Feng, Limsoon Wong. An empirical comparison of several recent epistatic interaction detection methods. Bioinformatics, 27(21):2936--2943, 2011.

Zhengkui Wang, Yue Wang, Kian-Lee Tan, Limsoon Wong, Divyakant Agrawal. eCEO: An efficient Cloud Epistasis cOmputing model in genome-wide association study. Bioinformatics, 27(8):1045--1051, 2011.

Computational Systems Biology

Mathematical modeling is being increasingly recognized as a useful tool in the biomedical sciences. They can uncover new phenomena to explore, identify key factors of a system, link different level of details, enable a formalization of intuitive understanding, screen unpromising hypotheses, predict variable inaccessible to measurement, and expand the range of questions that can meaningfully be asked. Our overall goal in this project is to develop powerful scalable computational techniques to model and analyze biochemical networks.

A major theme is to use probabilistic graphical models to approximate the dynamics and to deploy probabilistic formal verification methods to perform key analysis tasks such as parameter estimation and sensitivity analysis. Currently we are extending our methods to handle multi-mode signaling pathways. We are also building powerful GPU based implementations of our algorithms to achieve scale.

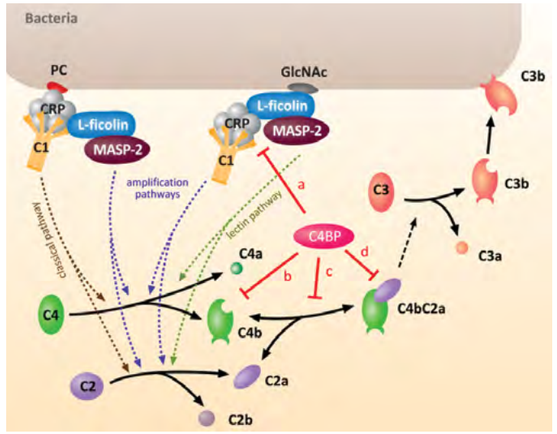
Our recent results include:

- Dynamic Bayesian Networks (DBNs) can serve as succinct probabilistic dynamic models of biochemical networks. To analyze these models, one must compute the probability distribution over system states at a given time point. Doing this exactly is infeasible for large models. We present an approximate algorithm called the Hybrid Factored Frontier (HFF) algorithm. At each time slice, in addition to maintaining probability distributions over local states, HFF explicitly maintains the probabilities of a number of global states called spikes. By increasing the number of spikes one can reduce errors, while the additional computational effort required is only quadratic in the number of spikes.

- Biopathways are often modeled as systems of ordinary differential equations (ODEs). Such systems usually have many unknown parameters. Since the data available for calibration have limited precision, an approximate representation of the ODEs dynamics should suffice. One must, however, be able to efficiently construct such approximations for large models and perform model calibration and subsequent analysis. We present a graphical processing unit (GPU) based scheme by which a system of ODEs is approximated as a dynamic Bayesian network (DBN). We then

construct a model checking procedure for DBNs based on a simple probabilistic linear time temporal logic. The GPU implementation shows significant gains in performance and scalability whereas the model checking framework turns out to be convenient and efficient for specifying and verifying interesting pathways properties.

- The complement system is key to innate immunity and its activation is required for the clearance of bacteria and apoptotic cells. To quantitatively understand the modulatory mechanisms of the complement system, we built a computational model involving the enhancement and suppression mechanisms that regulate complement activity. Our model consists of a large system of Ordinary Differential Equations (ODEs) accompanied by a dynamic Bayesian network as a probabilistic approximation of the ODE dynamics. Our combined computational and experimental study elucidates the regulatory mechanisms of the complement system and demonstrates how the bio-pathway machinery maintains the balance between activation and inhibition.



Representative Publications

Sucheendra Kumar Palaniappan, S. Akshay, Bing Liu, B. Genest and P.S. Thiagarajan. A hybrid factored frontier algorithm for dynamic Bayesian networks with biopathways application. ACM Transactions on Computational Biology and Bioinformatics, 9(5):1352-1365, 2012.

Bing Liu, Andrei Hagiescu, Sucheendra K. Palaniappan, Bipasa Chattopadhyay, Zheng Cui, Weng-Fai Wong and P.S. Thiagarajan. Approximate probabilistic analysis of biopathways dynamics. Bioinformatics, 28(11):1508-1516, 2012.

Bing Liu, Jing Zhang, Pei Yi Tan, David Hsu, Anna M. Blom, Benjamin Leong, Sunil Sethi, Bow Ho, Jeak Leak Ding and P.S. Thiagarajan. A combined computational and experimental study of the regulatory mechanisms of the complement system. PLoS Computational Biology, 7(1): e1001059, 2011.

Genome Assembly and Genome-Wide Alignment

Modern high-throughput sequencing (or next-generation sequencing) technologies parallelize the sequencing process, producing thousands or millions of short sequences at once. Genome assembly and alignment from these sequences lead to a number of computationally challenging problems. We have worked on some of these problems, including:

- Scaffolding, the problem of ordering and orienting contigs, typically using paired-end reads, is a crucial step in the assembly of high-quality draft genomes. We explore the feasibility of an exact solution for scaffolding and present a first tractable solution for this problem (Opera). We also describe a graph contraction procedure that allows the solution to scale to large scaffolding problems and demonstrate this by scaffolding several large real and synthetic datasets.

- Second-generation sequencing (SGS) generates millions of reads that need to be aligned to a reference genome allowing errors. Although current aligners can efficiently map reads allowing a small number of mismatches, they are not well suited for handling a large number of mismatches. We introduce Basic Alignment tool for Mismatches (BatMis), an efficient exact method to align short reads to a reference allowing k mismatches. BatMis performs better than other aligners in solving the k-mismatch problem. Furthermore, it can compete favourably even when compared with the heuristic modes of other aligners.

- DNA methylation plays a crucial role in higher organisms. Coupling bisulfite treatment with next-generation sequencing enables the interrogation of 5-methylcytosine sites in the genome. However, bisulfite conversion introduces mismatches between the reads and the reference genome, which makes mapping of Illumina and SOLiD reads slow and inaccurate. BatMeth is an algorithm that integrates novel mismatch counting, list filtering, mismatch stage filtering and fast mapping onto two indexes to improve unique mapping rate, speed and precision. Experimental results show that BatMeth is faster and more accurate than existing tools.

Representative Publications

W. K. Sung, H. Zheng, S. Li, R. Chen, X. Liu, Y. Li, N. P. Lee, W. H. Lee, P. N. Ariyaratne, C. Tennakoon, F. H. Mulawadi, K. F. Wong, A. M. Liu, R. T. Poon, S. T. Fan, K. L. Chan, Z. Gong, Y. Hu, Z. Lin, G. Wang, Q. Zhang, T. D. Barber, W. C. Chou, A. Aggarwal, K. Hao, W. Zhou, C. Zhang, J. Hardwick, C. Buser, J. Xu, Z. Kan, H. Dai, M. Mao, G. Reinhard, J. Wang, J. M. Luk. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. Nature Genetics, 44(7):765-769, 2012.

J. Q. Lim, C. Tennakoon, G. Li, E. Wong, Y. Ruan, C. L. Wei, W. K. Sung. BatMeth: Improved mapper for bisulfite sequencing reads on DNA methylation. Genome Biology, 13(10):R82, 2012.

C. Tennakoon, R. W. Purbojati, W. K. Sung. BatMis: A fast algorithm for k-mismatch mapping. Bioinformatics, 28(16):2122-2128, 2012.

Pramila Nuwantha Ariyaratne, Wing-Kin Sung. PE-Assembler: de novo assembler using short paired-end reads. Bioinformatics, 27(2): 167-174, 2011.

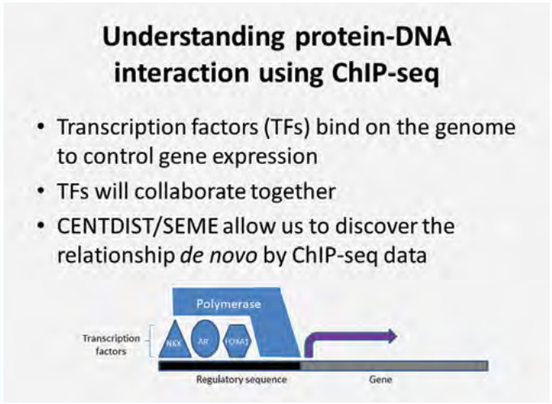
S. Gao, W. K. Sung, N. Nagarajan. Opera: Reconstructing Optimal Genomic Scaffolds with High-Throughput Paired-End Sequences. Journal of Computational Biology, 18(11):1681-1691, 2011.

ChIP-Seq and Motif Finding

Interactions between macromolecules play many essential roles---e.g., metabolic reactions and signal transduction---and occur in many combinations, such as protein-protein, protein-DNA, and protein-RNA. Protein interactions with DNA and RNA are the primary mechanisms for controlling gene expressions. What is needed is a recognition code that maps from the protein sequence to a pattern that describes the family of DNA binding sites---the functional elements. Identification of functional elements in the human genome is fundamental to our understanding of cell functions---how these codes orchestrate the complex network of gene transcription, the transcriptome, and interactions in distinct locations. We aim to: (1) develop methods for accurate identification of transcription factor binding sites and other regulatory sites; and (2) develop methods for inferring the interactions of transcription factors and other functional elements.
Our recent results include the following:

- Transcription factors (TFs) do not function alone but work together with other TFs (called co-TFs) in a combinatorial fashion to precisely control the transcription of target genes. We have developed a novel web-based co-motif scanning program called CENTDIST. CENTDIST automatically determines the best set of parameters and ranks co-TF motifs based on their distribution around ChIP-seq peaks. We applied CENTDIST on an Androgen Receptor (AR) ChIP-seq data set from a prostate cancer cell line and correctly predicted all known co-TFs (eight TFs) of AR in the top 20 hits as well as discovering AP4 as a novel co-TF of AR (which was missed by existing methods).

- Although de novo motifs can be discovered through mining over-represented sequence patterns, this approach misses some real motifs and generates many false positives. We develop a de novo motif discovery algorithm called SEME (sampling with expectation maximization for motif elicitation), which uses pure probabilistic mixture model to model the motif's binding features and uses expectation maximization (EM) algorithms to simultaneously learn the sequence motif, position, and sequence rank preferences without asking for any prior knowledge from the user. SEME is applied to a difficult problem of finding the co-regulated TF (coTF) motifs in 15 ChIP-Seq libraries. It identifies significantly more correct coTF motifs and, at the same time, predicted coTF motifs that better match to known motifs.



Representative Publications

ZhiZhuo Zhang, Cheng Wei Chang, Hugo Willy, Edwin Cheung, Wing-Kin Sung. Simultaneously learning DNA motif along with its position and sequence rank preferences through EM algorithm. Proceedings of 16th Annual International Conference on Research in Computational Molecular Biology (RECOMB), pages 355-370, 2012.

Z. Zhang, C. W. Chang, W. L. Goh, W. K. Sung, E. Cheung. CENTDIST: Discovery of co-associated factors by motif distribution. Nucleic Acids Research, 39(Suppl 2):W391-W399, 2011.

Biological Sequence Mining

The increasing infectious disease outbreaks has led to a need for new research to better understand the disease's origins, epidemiological features and pathogenicity caused by fast-mutating, fast-spreading viruses. Mutations in influenza virus isolates have been found to be responsible for outbreaks. Existing mutation analysis methods use sequence alignment and sequence comparison to identify single point mutations. Single-point mutations are small mutations that alter only one nucleotide at a time. They are responsible for the so-called "antigenic drift", in which the virus gradually accumulates more and more such mutations, eventually causing them to become new strains. In contrast, "antigenic shift" is caused by a large and sudden mutation that involved the changing of many nucleotides, often leading to major outbreaks. Vaccines developed in advance of a pandemic to become less effective over time. This is because the more a virus has mutated, the more likely that the population's immune systems would not recognize it.

A recent trend is to develop region-specific vaccines which require the spatial and temporal dynamics of the viral mutations. In other words, we need to know how highly-mutated regions in virus sequences at one geographical location change when the virus moves to another location over time. We introduce the notion of mutation chains to capture mutation patterns with geographical spread over time. We develop algorithms to discover not only single-point mutations (1-mutation), but also larger genetic changes that involve k (>1) consecutive nucleotides (k-mutations) in the viral sequences as they spread over time and space. We applied the algorithms on the influenza A virus dataset and successfully identified interesting mutation patterns that were found to correspond to previously known pandemic influenzas.

We observe that with the frequent mutation of influenza viruses, new mutations may emerge while old mutations disappear over time. Further, some mutations may be dominant in one sub-population but not in the other. Discovering such mutations can help to customize vaccines to increase the effectiveness for targeted group of people. We develop an algorithm called DMMiner to discover discriminative mutation chains. Figure 1 shows the discriminative mutation chains found in the real world Influenza A virus protein datasets before and after 2007. We observe that different mutations dominate over the years, indicating that DMMiner is able to find emerging and outdated mutations which is an active area of research in vaccinology. Figure 2 gives the discriminative mutation chains discovered over Asian and non-Asian countries. In particular, the discriminative mutation <H5N1: NA : 116, T➔K > has been independently validated by another study.

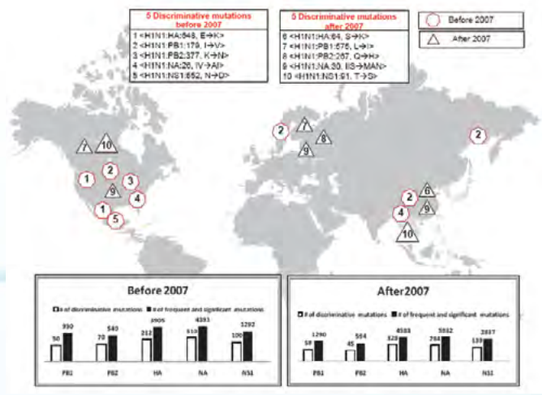


Figure 1: Discriminative mutation chains before and after 2007.

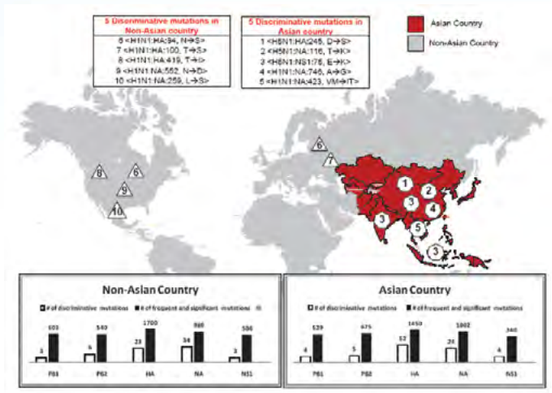


Figure 2: Region-specific mutations (Asian vs. non-Asian countries).

Representative Publications

Dhaval Patel, Wynne Hsu, Mong Li Lee. Discriminative mutation chains in virus sequences. IEEE ICTAI, 2011.

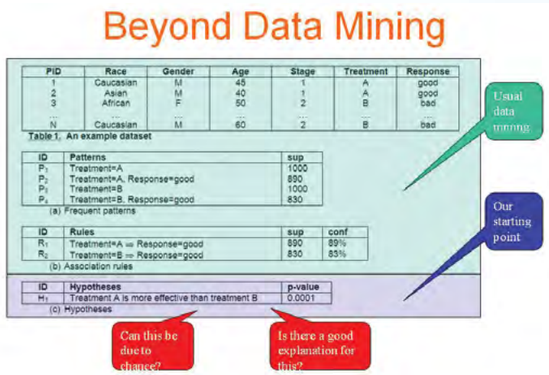
Exploratory Hypothesis Testing and Analysis

More and more data have been accumulated in digital format in various applications. These data provide rich sources for making new discoveries. Data mining has become an important tool to transform data into knowledge. Finding useful and actionable knowledge is the main objective of diagnostic data mining. Most existing works tackle the problem by discovering patterns and rules and then studying their interestingness. We use a different paradigm which represents the discovered knowledge in the form of hypotheses. A hypothesis involves a comparison of two or more samples. Compared with patterns and rules, hypotheses provide the context in which a piece of information is interesting, thus hypotheses are more intuitive and informative than patterns and rules. More importantly, users can take actions more easily based on what a hypothesis indicates. We further analyse the discovered significant hypotheses and identify the reasons behind them so that users not only get to know what is happening but also have some rough ideas on when or why it is happening. This new data mining paradigm has the potential to make diagnostic data mining as successful as predictive data mining in real-life applications.

- Help users understand their data and gain new insights: The system enables users to use computational methods to examine large amounts of data automatically. It summarizes data in a comparative way, which makes it easier for users to identify interesting information and the context in which the information is interesting.

- Find actionable knowledge: The system automatically identifies interesting phenomena from data as well as possible reasons behind the phenomena, which either suggests possible actions that users can take or provide clues that users can follow for further investigation. For example, given a product engineering dataset, the proposed system can answer the following questions: which product has higher failure rate than other products? Under which situation, the product is more likely to fail?

- Make it to real-life applications: Most existing diagnostic data mining systems are not easy to use, and few of them have made it to real-life applications. Our system provides a new diagnostic data mining paradigm that is easy to use and represents knowledge in a more intuitive and informative way. This new data mining paradigm has the potential to make diagnostic data mining as successful as predictive data mining in real-life applications.



Representative Publications

Guimei Liu, Haojun Zhang, Limsoon Wong. Finding minimum representative pattern sets. ACM SIGKDD, 2012.

Guimei Liu, Mengling Feng, Yue Wang, Limsoon Wong, See-Kiong Ng, Tzia Liang Mah, Edmund Jon Deoon Lee. Towards exploratory hypothesis testing and analysis. IEEE ICDE, 2011.

Guimei Liu, Haojun Zhang, Limsoon Wong. Controlling false positives in association rule mining. Proceedings of the VLDB Endowment, 5(2):145--156, 2011.

The faculty members involved in computational biology research are:

- David Hsu
- Wynne Hsu
- Lee Mong LI
- Leong Hon Wai
- Sung Wing-Kin, Ken
- Tan Kian Lee
- P. S. Thiagarajan
- Tung Kum Hoe, Anthony
- Wong Limsoon

For more information, please visit
http://www.comp.nus.edu.sg/~cbl.

DATABASE

Relational Query Processing and Optimization

In many database applications, there is a need to evaluate complex queries that involve multiple instances of a relation. For example, we may be interested to find out the list of employees who earn more than their managers. At the same time, it is not uncommon to have a workload of "similar" queries that access the same table. For example, organizations typically generate multiple reports from a single table but the reports may be sorted based on different attributes (e.g., one by product listing, another by stock-on-hand). As an example of a more complex query, consider the case of finding the rankings of each employee's salary within his department as well as the whole company. This query requires multiple analytic window functions on a single table. We have developed efficient algorithms that process and optimize a variety of such queries. For self-join, we have designed a novel scheme that scans the table once. We have developed sort-aware query processing schemes that optimize the ordering and processing of multiple sort orders over a table. New methods for processing multiple analytic functions over a table have also been developed. Finally, we are the first to investigate the processing of multiple online aggregation queries – in such queries, approximate answers are returned to users instantaneously; these answers are continuously being refined as processing continues.

Representative Publications

Optimization of Analytic Window Functions Y. Cao, C.Y. Chan, J. Li, K.L. Tan. VLDB, 2012.

On Optimizing Relational Self-Joins Y. Cao, Y. Zhou, C.Y. Chan, K.L. Tan. EDBT, 2012.

Sort-Sharing-Aware Query Processing Y. Cao, C.Y. Chan, K.L. Tan, R. Bramandia, The VLDB Journal 21 (3): 411-436, 2012.

MapReduce-based Data Processing Systems

One of our goails is to allow users of MapReduce-based systems to keep the programming model of the MapReduce framework, and yet to empower them with data management functionalities at an acceptable performance. We achieved this in two directions. First, we sought to identify key design factors of MapReduce (Hadoop) that affect its performance. We conducted a comprehensive and in-depth study of Hadoop, and found that, by carefully tuning these factors, we can achieve much better performance.

Second, we have developed query processing engine under the MapReduce framework. At the operator level, we have developed join algorithms. In particular, our proposed MapReduce-based similarity (kNN) join exploits Voronoi diagram to minimize the number of objects to be sent to the reducer node to minimize computation and communication overheads. We also designed several schemes for processing multi-join queries efficiently - while the Map-Join-Reduce mechanism, introduces a join operator to combine multiple datasets, the multi-join scheme in AQUA exploits replication to expand the plan space. We have also developed an automatic query analyzer that accepts an SQL query, optimizes it and translates it into a set of MapReduce jobs. Finally, to support data warehousing, we have leveraged on column store, and proposed Concurrent Join to support multi-way join over the partitioned data. In all these works, we target to reduce the number of MapReduce jobs to minimize the initialization overheads.

Representative Publications

Wei Lu, Yanyan Shen, Su Chen, Beng Chin Ooi. Efficient Processing of k Nearest Neighbor Joins using MapReduce. VLDB, 2012.

S. Wu, F. Li, S. Mehrotra, B. C. Ooi. Query Optimization for Massively Parallel Data Processing. ACM SOCC, 2011.

Y. Lin, D. Agrawal, C. Chen, B. C. Ooi, S. Wu. Llama: Leveraging Columnar Storage for Scalable Join Processing in the MapReduce. ACM SIGMOD, 2011.

epiC: elastic power-aware data intensive Cloud

To overcome the limitations of MapReduce-based systems to cope with a variety of workloads, we decided to focus on OLAP nd OLTP workloads. The epiC project, a joint system project between researchers from NUS and Zhejiang University, is initiated to investigate new cloud platforms that can support both OLAP and OLTP workloads. This is challenging as the two types of workload have different performance requirements - while OLAP workloads comprise data intensive analytical jobs that are long running, OLTP workloads consist of online transactions that demand short response time.

To deal with these challenges, we developed a novel elastic storage system (ES2) and elastic execution engine (E3). ES2 employs vertical partitioning to group columns that are to be accessed together, and horizontal partitioning to further split these column groups across a cluster of nodes. A number of novel cloud-based indexing structures supporting structured data (e.g. B+-tree and bitmap indexes) and multidimensional data (e.g., R-tree index) have been developed.

E3 adopts a "middle" approach between MapReduce and Dryad in that E3 has a simpler communication model than Dryad yet it can support multi-stages job better than MapReduce. E3 avoids reprocessing intermediate results by adopting a stage-based evaluation strategy and collocating data and user-defined (map or reduce) functions into independent processing units for parallel execution. Furthermore, E3 supports block-level indexes, and built-in functions for specifying and optimizing data processing flows. We have also examined how transactions can be supported. This led to the design of ecStore, a system that provides automated data partitioning and replication, load balancing, efficient range query processing and transactional access. ecStore exploits multi-version optimistic concurrency control and provides adaptive read consistency on replicated data.

Representative Publications

Gang Chen, Ke Chen, Dawei Jiang, Beng Chin Ooi, Lei Shi, Hoang Tam Vo, Sai Wu: E3. an Elastic Execution Engine for Scalable Data Processing. JIP 20(1): 65-76, 2012.

Yu Cao, Chun Chen, Fei Guo, Dawei Jiang, Yuting Lin, Beng Chin Ooi, Hoang Tam Vo, Sai Wu, Quanqing Xu. ES2: A Cloud Data Storage System for Supporting both OLTP and OLAP, ICDE, 2011.

Peer-to-Peer-based Cloud Data Management

One direction that we are pursuing is the integration of cloud computing, database and peer-to-peer (P2P) technologies. Exploiting a P2P architecture on a cluster of nodes offers several advantages over the MapReduce framework: (a) It offers more robust query processing mechanisms as nodes can now communicate with one another; (b) It removes the single point-of-failure in the master/slave architecture of MapReduce; (c) It facilitates elastic design as peers can be readily added and removed in a P2P architecture.

We have developed BestPeer++, a cloud-enabled evolution of BestPeer. BestPeer++ is enhanced with distributed access control, multiple types of indexes, and pay-as-you-go query processing for delivering elastic data sharing services in the cloud. BestPeer++ instances are organized as a structured P2P overlay network. We have used BATON, developed at NUS, as it can support range queries efficiently. The data are indexed by the table name, column name and data range for efficient retrieval.

The Katana framework is a novel peer-to-peer (P2P) based generalized data processing framework. It can be deployed on many of the currently known structured P2P overlays. The framework provides a programming model in which processing logic may be implicitly distributed with universality and expressiveness, much like the MapReduce framework. The programming model can be distinguished into a data model and a processing model. We adopt a key/value data model with possible duplicated keys to represent the data elements. However, the data model is conceptually a graph-based model, i.e., data elements can be organized into a graph structure. This facilitates the mapping from the data elements to the Cayley graphs which in turn can be mapped to the structured P2P overlays. Like MapReduce, the processing model hides the parallelism mechanism from the users. Instead, Katana provides two MapReduce-like functions: kata and ana. However, unlike MapReduce, the kata and ana functions are independent from one another and are not required to be executed one after another.

Representative Publications

Katana: Generalized Data Processing on Peer-to-Peer Overlays. W.X. Goh, K.L. Tan, IC2E, 2013.

BestPeer++: A Peer-to-Peer based Large-scale Data Processing Platform. G. Chen, T. Hu, D. Jiang, P. Lu, K.L. Tan, H. T. Vo, S. Wu, IEEE Transactions on Knowledge and Data Engineering (Special Issue on ICDE 2012 Best Papers), IEEE CS, 2012.

BestPeer++: A Peer-to-Peer based Large-scale Data Processing Platform G. Chen, T. Hu, D. Jiang, P. Lu, K.L. Tan, H. T. Vo, S. Wu, ICDE, 2012.

Query Reverse Engineering

While database system research has made tremendous advances on functionality and performance related issues over the years, research on improving database usability has not attracted as much attention as it deserves. To help users with constructing queries and understanding query results, we have developed an approach, termed Query by Output (QBO), to reverse engineer queries given an input pair of database and query output. Given a database D and a result table T = Q(D), which is the output of some query Q on D, the goal of QBO is to construct candidate queries Q', referred to as instance-equivalent queries, such that the output of query Q' on database D is equal to Q(D).

We have applied QBO to improve database usability in two contexts. In the first scenario [1], QBO is used to help users better understand their query results by augmenting the result of a query Q (w.r.t. a database) with instance-equivalent queries that describe alternative characterizations of their query results. As an example, suppose that a university physician issues a query to his clinic's database to find students who have been infected with a skin rash over the past week. Besides returning the query result, if the database system had also computed and returned an equivalence-instance query that revealed the additional information that all the students in the query result either had recently returned from an overseas trip to region X or are staying in the same dormitory as those students, then the physician could have been alerted about a potential skin rash outbreak in those dormitories. Thus, it is useful for a database system to augment a query's result with alternative characterizations of the query's result to provide additional insightful information about the result.

In the second scenario, QBO is used to generate explanations for unexpected query results that have missing expected result tuples [2]. As an example, suppose that a manager issues a query to compute the annual sales figures for each of her regional sales agents and she is surprised to find that Alice's sales performance is lower than that of Bob's, which is inconsistent with her impression of their results. The manager could issue a follow-up "why-not" question to clarify why Alice's sales figure is not higher than that of Bob's. Using QBO, the database system could respond to this why-not question with an explanation in the form of an alternative query (e.g., compute total sales for each sales agent excluding the period when Alice was on sick leave) which would have returned an output result that is consistent with the manager's why-not question. Thus, providing a capability to explain why-not questions would be very useful to help users understand their query results.

We are currently implementing a query acquisition tool based on QBO that enables users to construct queries from examples of database and query result pairs.

Representative Publications

Q.T. Tran, C.Y. Chan, G. Wang, Evaluation of Set-based Queries with Aggregation Constraints, ACM CIKM, 2011.

Q.T. Tran, C.Y. Chan, How to ConQueR Why-not Questions, ACM SIGMOD, 2010.

Integrated Mining and Visualization of Complex Data

The drive to find gold nuggets in data has resulted in the explosion of discovery algorithms in the past decade. Many of these discovery algorithms focus on specific data type. However, with the advances of technology, many applications now involve records with attributes of diverse data types, ranging from categorical, to numerical, to time series, to trajectories. Figure 1 shows a sample of electronic medical record of a patient. The record consists of categorical data, numerical data, interval data and time series data. In addition, the trajectories of the patient can be captured with GPS-enabled mobile devices and through the use of check-in apps and location-based services.

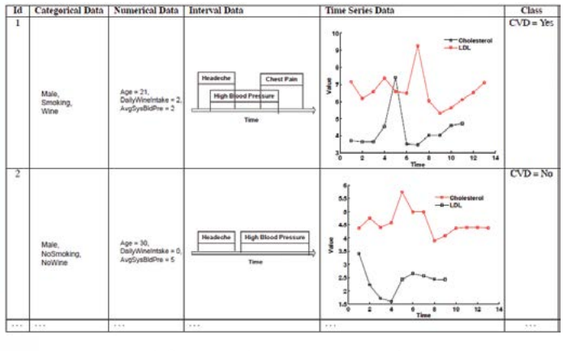


Figure 1: Sample electronic medical record of a patient

Knowing the relationships among all the different types of data can aid in the understanding of a patient's health condition. For example, suppose we have a frequent itemset {Male, Smoker} and an interval-based temporal pattern {Headache Overlap HighBloodPressure}. Separately, these patterns are not interesting since we have many male smokers in the population, and many people suffer from headache with elevated blood pressure. However, if these two patterns occur together, it may raise an alarm as studies have shown that a male smoker who experiences headache with elevated blood pressure has a high risk of having cardiovascular disease.

Handling datasets with such variety is a challenge as the complexity of the problem can quickly grow out of hand. We have developed a framework to perform the integrated mining of big data with diverse data types. The framework consists of algorithms for mining patterns from interval-based events, lag patterns involving motifs in time series data, spatial interaction patterns. We have also designed algorithms to discover duration-aware region rules and path rules from trajectories. With this, we are able to capture the associations among different complex data types and demonstrate how these patterns can be used to improve the classification accuracy in various real world datasets.

We have also developed a tool to generate and highlight interesting patterns discovered from the different data types. This tool will also allow the visualization of event incidences, clusters and heat maps. Figure 2 shows a screenshot of our system. Ongoing research aims to develop an interactive system for the visualization and analysis of trajectories.

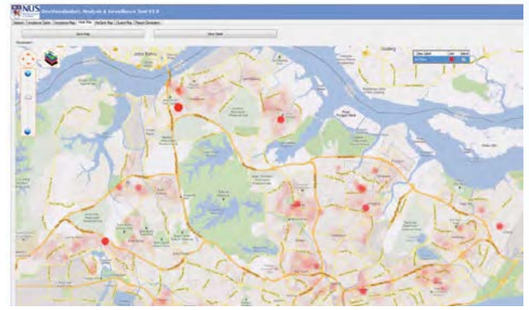


Figure 2: Screenshot of Geovisualization System

Representative Publications

Dhaval Patel, Wynne Hsu, Mong Li Lee. Integrating Frequent Pattern Mining from Multiple Data Domains for Classification, ICDE, 2012.

Dhaval Patel, Chang Sheng, Wynne Hsu, Mong Li Lee. Incorporating Duration Information for Trajectory Classification, ICDE, 2012.

Answering MaxBRkNN Queries

The MaxBRNN problem finds a region such that setting up a new service site within this region would guarantee the maximum number of customers by proximity. This problem assumes that each customer only uses the service provided by his/her nearest service site. However, in reality, a customer tends to go to his/her k nearest service sites. This leads to the MaxBRkNN problem which finds an optimal region such that setting up a service site in this region guarantees the maximum number of customers who would consider the site as one of their k nearest service locations. We generalize the MaxBRkNN problem to reflect the real world scenario where customers may have different preferences for different service sites, and at the same time, service sites may have preferred targeted customers [1]. We design an efficient solution called MaxFirst to solve this generalized MaxBRkNN problem. The algorithm works by partitioning the space into quadrants and searches only in those quadrants that potentially contain an optimal region. During the partitioning process, we compute the upper and lower bounds for each quadrant and use these bounds to prune the unpromising quadrants. Experiment results show that MaxFirst is two to three orders of magnitude faster than the state-of-the-art algorithm.

Representative Publication

Zenan Zhou, Wei Wu, Xiaohui Li, Mong Li Lee, Wynne Hsu. MaxFirst for MaxBRkNN, ICDE, 2011.

Personalized Search and Recommendation using Tags

The amount of information on the Web has increased significantly with online social networks (OSNs) such as Delicious, Facebook and Flicker. Users of such web sites can use tags to annotate various resources and share them with their friends. With this new way of sharing and managing resources in the social systems, the challenge is to incorporate this additional information to retrieve relevant resources for the querying users. Existing user-centric friendship-based solutions where the strength of a friendship is determined by the degree of overlapping tags or resources among users may not retrieve relevant results as they do not consider whether these friends are qualified to answer the specific query. We introduce the notion of authoritative users who have frequently tag resources that are relevant to the issued query [1]. Determining these authoritative users for each query in a large social system requires an efficient and scalable algorithm. We design an incremental algorithm to build a social network graph to identify the authoritative users for each query. Experiment results on the real world Delicious annotation dataset demonstrate the effectiveness of our query-based personalized approach.

Another way to cope with the information overload is to utilize tags to retrieve relevant resources to the users automatically via recommender systems. Till now, most of these systems consider only ternary relationships such as (user, item, rating) or (user, item, tag) to generate their recommendations. We show that there is a need to consider the quaternary relationship among users, tags, ratings, and items together as this can reveal hidden semantics that cannot be obtained otherwise. For example, two users may use the same tag for an item but their ratings differ greatly, giving extra insights into their preferences. In order to capture this quaternary relationship, we propose a model based on 4-order tensor and apply the higher order singular value decomposition to reveal the latent semantic associations among users, items, tags and rating [2]. Experiment results on the real world MovieLens dataset shows that the proposed approach outperforms state-of-the-art methods for item rating prediction as well as user, item and tag recommendations. We also develop a system with online relevance feedback to increase user acceptance of recommendations. The system utilizes tags to provide intuitive explanations of why the items are recommended. Figure 3 shows a screenshot of the system.

Figure 3: Quaternary analysis for movie recommendation with relevance feedback

Representative Publications

Gang Zhao, Mong Li Lee, Wynne Hsu, Jiawei Zhang. Query-based Personalized Search in Tag Social Systems, Workshop on Social Web Search and Mining, 2011.

Wei Chen, Wynne Hsu, Mong Li Lee. Quaternary Semantic Analysis: Providing Recommendations based on Explicit and Implicit Feedbacks, ACM SIGIR, 2011.

Management of Location-based Data: Analytic and Real-Time Services

The wide adoption of GPS-equipped mobile phones has resulted in new data services – location analytic services and location-based services. In location analytics services, trajectories of movement data are analyzed to discover the behaviors of objects tracked. For example, scientists may be interested to study the migratory behavior of animals. As another example, taxi companies may be interested to find frequent routes that were drawn upon the experiences and local knowledge of the drivers. In our research, we have developed algorithms to mine and discover a variety of interesting patterns: the convoy patterns identify a group of objects that move together, the frequent route patterns retrieves routes that are frequently traversed, and the meeting patterns uncover regions where moving objects meet.

For real-time services, we have developed TsingNUS, a comprehensive location-based service system. Unlike traditional location-aware applications, TsingNUS facilitates location-aware keyword search. TsingNUS (a) offers a location-aware search-as-you-type feature that enables answers to be continuously returned and refined as the query is typed letter-by-letter; (b) supports spatial-keyword queries for both AND and OR semantics; (c) incorporates continuous spatial-keyword search that automatically refines the answers as user moves around. The algorithms and data structures we developed resulted in superior real-time performance even for very large datasets (with over millions of point-of-interests), making TsingNUS a practical system for supporting large number of end-users.

Representative Publications

Scalable Top-K Spatial Keyword Search D. Zhang, K.L. Tan, A.K.H. Tung, EDBT, 2013.

Location-Aware Instant Search R. Zhong, J. Fan, G. Li, K.L. Tan, L. Zhou, CIKM, 2012.

Efficient Safe-region Construction for Moving top-K Spatial Keyword Queries. W. Huang, G. Li, K.L. Tan, J. Feng, CIKM, 2012.

Finding Closed MEMOs H.H. Aung, K.L. Tan, SSDBM, 2011.

Collaborative Visual Analytics

In this research, we investigate how people can collaboratively achieve certain tasks by sharing their data and analytics result through the social network. Our "Internet Observatory: Visual Analytics for the Masses (www.trendspedia.com)" project is catered to the observation of the contents on the Internet. This includes contents that are generated on conventional websites and social media.

To bring proper context to the contents which are streaming in from the Internet, we aim to index these dynamic contents via Wikipedia, a well-established online encyclopaedia which have entries for large number of entities and concepts. Consider an example is the "Senkaku Island Dispute". The UI offers a choice of languages. The visual display also shows a set of social websites that users can choose to login. By doing so, users will be allowed to pull in information from these social websites and try to link/compare them to other information and opinions that are available on the Internet. Implicitly, by doing so, the user is also adding his/her private data into a public pool for general analysis. The dynamic information that are related to "Senkaku Island Dispute" including URLs, images, tag summarization, community view and geographical view will also be displayed.
Another project entitled "ReadPeer: Sharing Your Thoughts on What You Are Reading (www.readpeer.com)" aims to promote reading as a larger scale social activities by integrating ebooks and social networks to encourage more people to read and discuss about the materials they read. We are developing the ReadPeer system that allows users to make annotations on ebooks, research articles or any documents in PDF format. These annotations can be linked to various multimedia contents like blogs, videos, images, web links etc. and shared to friends in a social network. Comments and discussions can also be done in a forum-like manner for each annotation. Such a system not only help to enhance ebooks with rich multimedia information but can also bring together various groups of people with common interest and allow them to collaborate with each other.

As can be seen, both our effort in collaborative visual analytics involve reorganizing social media messages around a center of focus like Wikipedia articles or ebooks instead of putting these messages in a plain newsfeed. This allow users of common interest to come together to share their insights and analysis. Central to this is the design of visual interface that allow users to communicate and understand the perspective of each other. In the process, these interactions generated databases that capture a lot of interesting semantic through linkage of social media messages into a rich information network.

Representative Publications

Chen Liu, Sai Wu, Shouxu Jiang, Anthony K. H. Tung. Cross Domain Search by Exploiting Wikipedia, ICDE, 2012.

Feng Zhao and Anthony K. H. Tung. Large Scale Cohesive Subgraphs Discovery for Social Network Visual Analysis, VLDB, 2012.

XML Data Management

XML is emerging as a de facto standard for information exchange over the Internet. Business and enterprises generate and exchange XML data more often. Querying information in the XML data is an important research area in XML data management. We study XML node labeling schemes for querying dynamic XML database efficiently. We also study two different approaches to query XML data: twig pattern query and keyword query.

Node Labeling Schemes for Dynamic XML Data. The method of assigning labels to the nodes of an XML data is called a node labeling scheme. Based on the labels only, both ordered and un-ordered queries can be processed without accessing the original XML data file. The core issue for XML query is to efficiently determine the following four basic relationships: ancestor-descendant (A-D), parent-child (P-C), sibling and ordering relationships. One more important point for the labeling scheme is to process updates when nodes are inserted into or deleted from the XML data. All the existing node labeling schemes, i.e. range-based and prefix-based labeling schemes have high update cost.

We first propose a novel Compact Dynamic Binary String (CDBS) encoding to encode the labels of different labeling schemes and based on CDBS encoding, updates can be efficiently processed. However, the fixed size length field of CDBS will encounter the overflow problem. We then improve CDBS to Compact Dynamic Quaternary String (CDQS) encoding.

We later introduce vector order which is the foundation of the dynamic labeling schemes we propose. Compared with previous solutions that are based on natural order or lexicographical order, vector order is a simple and yet most effective solution to process updates in XML database. We show that vector order can be gracefully applied to both range-based and prefix-based labeling schemes with little overhead introduced. Moreover, vector order-based labeling schemes are not only efficient to process, but also resilient to skewed insertions.

Twig pattern Query Processing. An XML query describes a tree-shaped search pattern, which is often referred to as a twig pattern, with additional query conditions (if any) described as predicates on the three nodes. XML queries thus are called tree queries or twig pattern queries. Answering a twig pattern query is a core operation in XML query processing which is essentially to find all matching instances from an XML database that match the twig pattern query.

Twig pattern matching is a critical operation for XML query processing, and the holistic computing approach has shown superior performance over other methods. Prior works demonstrate that holistic twig pattern matching algorithm is an efficient technique to answer twig pattern queries with parent-child (P-C) and ancestor-descendant (A-D) relationships, as it can effectively control the size of intermediate results during query processing. We establish a theoretical framework about "matching cross" which demonstrates the intrinsic reason in the proof of optimality on holistic algorithms.

Practical XML queries often require support for more general twig patterns, such as the ones that allow arbitrary occurrences of an arbitrary number of logical connectives (AND, OR, and NOT); such types of twigs are referred to as B-twigs (i.e., Boolean-Twigs) or AND/OR/NOT-twigs. We investigate novel mechanisms for efficient B-twig pattern matching.

XML Keyword Query Processing. Inspired by the great success of information retrieval (IR) style keyword search on the web, keyword search on XML has emerged recently. The difference between text database and XML database results in three new challenges: 1) Identify the user search intention, i.e., identify the XML node types that user wants to search for and search via. 2) Resolve keyword ambiguity problems: a keyword can appear as both a tag name and a text value of some node; a keyword can appear as the text values of different XML node types and carry different meanings; a keyword can appear as the tag name of different XML node types with different meanings. 3) As the search results are subtrees of the XML document, new scoring function is needed to estimate its relevance to a given query. However, existing methods cannot resolve these challenges, thus return low result quality in term of query relevance. We propose an IR-style approach which basically utilizes the statistics of underlying XML data to address these challenges. We first propose specific guidelines that a search engine should meet in both search intention identification and relevance oriented ranking for search results. Then, based on these guidelines, we design novel formulae to identify the search for nodes and search via nodes of a query, and present a novel XML TF*IDF ranking strategy to rank the individual matches of all possible search intentions. To complement our result ranking framework, we also take the popularity into consideration for the results that have comparable relevance scores.

Representative Publications

Dunren Che, Tok Wang Ling, Wen-Chi Hou. Holistic Boolean-Twig Pattern Matching for Efficient XML Query Processing. IEEE Trans. Knowl. Data Eng. 24(11): 2008-2024, 2012.

Junfeng Zhou, Zhifeng Bao, Wei Wang, Tok Wang Ling, Ziyang Chen, Xudong Lin, Jingfeng Guo: Fast SLCA and ELCA Computation for XML Keyword Queries Based on Set Intersection, ICDE, 2012.

Liang Xu, Tok Wang Ling, Huayu Yu: Labeling Dynamic XML Documents: An Order-Centric Approach. IEEE Trans. Knowl. Data Eng. 24(1): 100-113, 2012.

Jiaheng Lu, Tok Wang Ling, Zhifeng Bao, Chen Wang: Extended XML Tree Pattern Matching: Theories and Algorithms. IEEE Trans. Knowl. Data Eng. 23(3): 402-416, 2011.

Big Data

One challenge of Big Data problems is how to scale algorithms and processing as the size of the data increases. The nature of big data problems vary depending on the form of the data and processing needed. For sequence and one dimensional data such as real-time streams, as the size of the data grows, it is convenient to work with a smaller approximation of the data.

Optimal algorithms are known which give smaller summaries while minimizing the approximation error, however, they do not scale well as the data sizes increase. We have developed a novel state using local search ideas from AI which gives very good summaries, often optimal, while only being linear time, thus, giving up optimality guarantees of the slower algorithms.

As data increases, traditional database indexing techniques may no longer be applicable. The cost of indexing the data fully may be too high. Adaptive indexing techniques try to avoid the upfront cost of indexing with the main representative being Cracking used in main-memory column stores. Cracking only indexes as it needs to based on the query workload. We show that unfortunately that existing Cracking algorithms are not robust and may fail to give the desired indexing adaptation, degnerating to linear scan.

We present an improvement which proposes the new idea of stochastic adaptation. Stochastic cracking gives the desired cracking behavior under a wide variety of workloads while being significantly faster than cracking on some workloads.

The Internet gives rise to many important graphs which are extremely large. The challenge is how to process such large graphs as conventional algorithms are unsuited to do so give the limitations of existing machines.

We demonstrate that it is feasible to process very large graphs on the order of hundreds of millions of nodes using the MapReduce framework. More specifically, we have focused on how to compute the maximum flow on very large small world graphs such as social networks or the web.

Representative Publications

Felix Halim, Stratos Idreos, Panagiotis Karras, Roland H.C. Yap. Stochastic Database Cracking: Towards Robust Adaptive Indexing in Main-Memory Column-Stores, VLDB, 2012.

Felix Halim, Rolamd H.C. Yap, Yongzheng Wu. A MapReduce-Based Maximum-Flow Algorithm for Large Small-World Network Graphs, ICDCS, 2011.

The faculty members in database management research are:

- Stephane Bressan
- Chan Chee Yong
- Wynne Hsu
- Lee Mong Li
- Ling Tok Wang
- Ooi Beng Chin
- Tan Kian Lee
- Tung Kum Hoe, Anthony
- Yap Hock Chuan, Roland

MEDIA

Media research focuses on the analysis, generation, manipulation, understanding, transmission and storage of digital media including traditional forms such as text, computer, video, audio, graphics, as well as newer forms from social media and sensor data. The media group is composed of researchers focusing on both basic and applied research targeting areas including multimedia systems, computer vision, medical imaging, geometry processing, computer graphics, human computer interaction, animation, natural language processing, social media analytics, speech recognition and sound and music computing. Members of the group are internationally active and participate in chairing major international conferences, serving on editorial boards and technical programme committees. Media group members also participate in a variety of national level technical committees and activities.

Media Search

Lab for Media Search (LMS) focuses on research on large-scale real-life problems arising from social media. In particular, it focuses on the analysis of live, big, multi-source and multi-faceted data arising from user-generated contents (UGCs). Such data is available in a myriad of sources, including live sharing sites like twitter, mobile sharing sites like 4Square, Instagram, forums and blogs, traditional image and video sharing sites like Flickr and YouTube, and the various community question-answering sites like Wiki-Answers & Yahoo! Answers etc., as well as their Chinese counterparts. To effectively analyze these contents, we need to tackle not only the multimedia, multilingual and multimodal aspects of data sources, but also the social aspects of these contents, such as user relations, communities, and key users with respect to any topics. In addition, we need to deal with the issues of cross lingual and cross domain data types.

Key research focuses under this lab includes: (a) reliable strategies for harvesting UGCs; (b) indexing and retrieving of huge media resources arising from these media; (c) fusion of UGCs to generate analytics related to locations, people, topics and organizations; and (d) architecture for accessing and sharing of UGCs and the analytics. The research emphasizes on the handling of large and live data streams. Here we outline four of the research topics carry out in the lab.

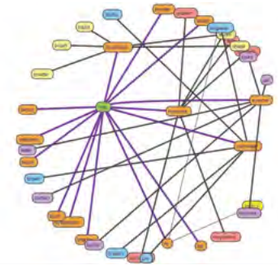
The Generation of Structured Knowledge from Heterogonous UGCs

Given the sheer volume of UGCs on any given topic, it is extremely difficult and time-consuming for users to grasp and follow the evolution of knowledge even in their interested domains. To better enhance the aggregation and corroboration of insights and knowledge of the crowd, this research explores techniques to automatically analyse and organize the large amount of UGCs on any specific topic so as to encourage macro-level and micro-level information access and knowledge creation. The generation of knowledge structures on topics exploits a wide range of knowledge available on the Web, including, structured knowledge from Wikipedia or Blogs, semi-structured knowledge from cQA and forums, and the unstructured information sources from twitters. Figure 1 shows an example of structures generated on the topic of MAC cosmetics. The resulting structures facilitate user browsing, querying and question-answering, and most importantly, they are dynamic and can be re-generated from the latest UGC sources. Our other research includes the summarization, evolution and predictive analytics of topics.

Mining the Sense of Organizations, People and Other Entities from Social Media

Social media contains the important and timely indicators on the spontaneous and often genuine views of users on any entity such as the organization. This research aims to discover users' views about an entity in terms of emerging and evolving topics/events, and offer them as alerts and general online feedbacks. One key challenge of this research involves the elicitation of representative data from different sources including fixed and dynamic keywords, known users, and automatically identified key-users of the target entity. To achieve this, we first need to develop techniques to learn and track the evolution of keywords, key topics and users of the target entity online through time. For active users, we identify those who regularly tweets about the entity, initiates major discussions, and has many followers within the entity. For evolving key topics, we propose a sparse coding algorithm that can quickly identify the emerging topics, keep track of the evolving topics, and purge the trivial ones as time passes. We further develop techniques to identify hot topics before they become viral. In addition to organizations, the same approach can be applied to mine analytics of people, locations and other entities. Our other research includes techniques and formal models to harvest representative data from UGC sources, and real-time algorithms for data filtering, emerging event detection, and summarization.

(a)



(b)

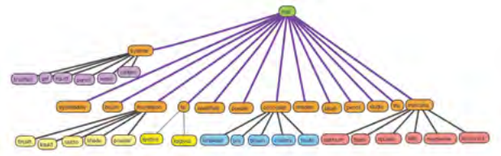


Figure 1: Knowledge structures extracted from UGCs on the topic of MAC Cosmetic presented as: (a) graph; and (b) tree structures.

Large-scale Analysis, Indexing and Retrieval of Social Image

There is an increasing trend that people are sharing more images and videos on live social media sites such as Twitter and Weibo. Our preliminary investigation on Sina Weibo indicates that about 30% of these microblogs have accompanying images, but only close to 30% of these images have meaningful text annotations. In order to fully comprehend the social signals propagated in microblog sites, we need to analyze both text and image contents. The first challenge is therefore to develop a visual search engine that can handle millions of fast changing social media images. Specifically, we built a hash-based image search engine that utilizes spatial pyramid visual feature with extended Spectral Hashing to index the images. Figure 2 presents the flowchart of the system. Current research explores other image representation and hashing approaches, and develops guidelines for efficient large-scale image indexing and retrieval.

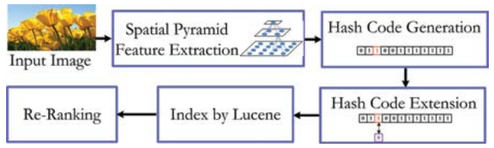


Figure 2: The flowchart for the large scale image indexing system.

Building on top of the visual search engine, we investigate the propagation of social signals in microblog sites through brand/event tracking task. In order to overcome the limitation of text-based approaches, we explore not only text and image contents, but also the social and location context of social signals. Figure 3 presents a framework for brand tracking in micrologs that comprises 4 stages. First, given a query brand in text such as “Puma”, the text-based search returns a preliminary set of relevant tweets. Second, we extract the “seeds” by performing visual logo detection on the preliminary results to find relevant tweets with also high visual confidence. Third, an extended search is conducted by utilizing the social and visual content of “seeds” to find other relevant tweets without containing relevant text information. The final step re-ranks all candidate tweets by jointly modeling the text-visual contents of all tweets using a hypergraph model. Current research extends the brand tracking to other more events over a variety of UGC sites.

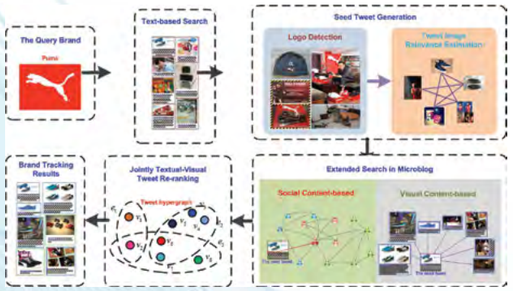


Figure 3: The framework for brand search in microblog sites.

Bridging the “Intention Gap” in Multimedia Search

The explosive proliferation of UGCs in the form of text, images, and videos has resulted in an urgent need for “user-centric” multimedia applications. Pivotal to these applications is the understanding of users’ search intents. However, there exists an “Intention Gap” between a user’s search intent and the query, leading to poor search results. The research here aims to tackle the users’ “Intension Gap” along two directions. One is to exploit user feedbacks to infer user search intent, and the other is to sense user’s intent from UGCs. Figure 4 shows the snapshots of some of our proposed techniques, including (a) Visual Query Suggestion, a new query suggestion scheme that provides users with both keyword and image suggestions; (b) Related Sample Feedback, that introduces a new category of samples termed “related samples“, which are related to part of the query and may help to find more relevant samples in a relevance feedback framework; and (c) Attribute Feedback, a new interactive framework that goes beyond conventional relevance feedback by allowing the users to provide feedbacks on semantic attributes, which act as the bridge connecting user’s search intent and low-level visual features.

Figure 4: Snapshots of our proposed techniques towards bridging the “Intention Gap”.

In parallel with the above research, we are working on: (a) large-scale ontology-based learning of semantics from multimedia content; and (b) location analytics to extract multimedia semantics of venues and users travel patterns. Given the huge amount of UGC’s that we and many research groups have gathered, one key challenge is how we can share and leverage on each other’s databases to promote large-scale , cross-domain and cross region social media research and applications. Towards this objective of building global social media observatories, we will tackle: (a) architecture for sharing of dataset and analytics; and (b) development of large-scale social media test datasets to facilitate research, such as the development of NUS–WIDE dataset to support image search.

Representative Publications

Yan Chen, Hadi Amiri, Zhoujun Li and Tat-Seng Chua: Emerging Topic Detection for Organizations from Microblogs. ACM SIGIR, 2013.

Hadi Amiri, Tat-Seng Chua: Mining Sentiment Terminology through Time. CIKM 2012.

Tat-Seng Chua, Huan-Bo Luan, Maosong Sun, Shiqiang Yang: NExT: NUS-Tsinghua Center for Extreme Search of User-Generated Content. IEEE MultiMedia 19(3): 81-87 (2012)

Jianxing Yu, Zheng-Jun Zha, Tat-Seng Chua: Answering Opinion Questions on Products by Exploiting Hierarchical Organization of Consumer Reviews. EMNLP-CoNLL, 2012.

Hanwang Zhang, Zheng-Jun Zha, Jingwen Bian, Yue Gao, Huan-Bo Luan, Tat-Seng Chua: Attribute feedback. ACM MM, 2012.

Multimedia Analysis and Synthesis

Analysis: A diversity of media types such as text, audio, video and novel sensory forms is proliferating in a variety of applications. This is true of television, Internet, home photos & videos, and emerging areas such as cyber-physical systems, social media and social networks. This calls for media analysis, which is the precursor of other forms of processing such as archiving, querying, retrieval and transcoding. For example, the ability to analyse and fuse information from different sources and in multiple media types in order to detect and recognize the semantic events occurring has become critically important in many media handling tasks. Handling live data (be it symbolic text feed or signal sensor feed) is particularly challenging in this environment. Security and privacy issues also influence the architecture of these systems. There are two broad themes of our research work: content-based multimedia information processing and multimedia information security. Both are systems research areas which have fundamental conceptual issues arising out of real-world problems. So their flavour is a blend of both basic and applied research. Basic research is related to ideas about media semantics, visual saliency, autonomous systems and information fusion. The application areas considered are surveillance, computational advertisement and social cyber-physical systems. There are three basic thrusts:

Multi-Sensor Computing: Embedded systems and advanced sensing are producing sensors that can do self-processing and communication with other sensors (e.g. smart cameras). Usage of multiple heterogeneous sensors is becoming more practical for multimedia applications. Sensors in the current systems do not actively share information with other sensors for making any decisions; instead they are controlled by the human operator or by a centralized server. The decision making capability of these sensors for tracking, best-view computation, etc. are not explored much. In order to automate such sensor systems, there is a need for making the sensors to collaborate with each other. To make group decisions among sensors, there should be a framework or protocol for the sensors to talk to other sensors. Our research develops the theoretical background for the sensors to involve themselves in group actions and group decision for multimedia applications. Many open problems in multimedia and computer vision systems can be solved robustly by making the sensors to collaborate with each other. For example, events are a fundamental entity in many applications e.g., the “left luggage” event is important in surveillance systems. An event is captured for the purpose of further analysis including detection, storage, reasoning, mining, exploration and action. In the past, most of the event processing research has been conducted based on the data obtained from single and static sensor. Our research utilizes additional data (via multiple sensors) and assumes control over sensors (via active sensors). We would like to build advanced sensor systems that are intelligent such that they continuously learn from the environment and make their decisions autonomously. We have developed such systems for surveillance using MDP and POMDP controllers for handling multiple active cameras.

Semantics and Saliency: Our group considers media and human perception together, in a more holistic way to look at problems

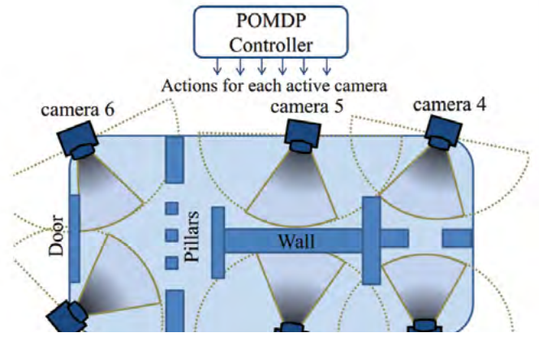


Figure 1: An Active Multimedia Surveillance Architecture.

relating to image and video understanding than to try and understand visual content alone in isolation. We are trying to correlate human understanding of scenes to the underlying semantics, affect and aesthetics. We have obtained a better understanding of visual perception and attention as people interact with digital images and video. One problem was on finding how low level global and local information in images influence category discrimination and aesthetic value in images. We also have investigated how semantic and affective cues relating to objects and their interactions influence scene semantics in static and dynamic scenes. Eye-tracking is used as proxy for human visual attention. We have also built better techniques for affect representation and modeling. Salience is the distinct subjective perceptual quality which makes some items in the world stand out from others and immediately grab our attention. Visual salience, where the underlying signal is image or video, has immense potential in active processing and analysis of visual multimedia, where a human-like approach of understanding a scene by looking at certain important parts of it and the relation between them is considered. Designing a reliable automatic and unsupervised visual salience computation system is a non-trivial task. We have developed a comprehensive bio-mimicking image salience computation system, which faithfully imitates many known important functions of retina, lateral geniculate nucleus, primary cortex and superior colliculus. Non-linear responses and interaction among cells in retina and lateral geniculate nucleus, divisive signal suppression and normalization for salience computation in superior colliculus, retino-cortical interaction for texture and orientation discrimination, and intra-cortical interaction for better signal abstraction are some phenomena considered unlike other many existing simple bio-mimicking image salience computation systems.

Multimedia Security: The group’s distinctive contributions in the area of multimedia security have been in exploiting the interplay of signal processing and cryptography. For scalability of businesses, multiparty multilevel digital rights management (DRM) architecture, where some multimedia content is delivered by an owner to a consumer through several levels of distributors has been suggested as an alternative to the traditional two-party (buyer-seller) DRM architecture. A combination of cryptographic and watermarking techniques are usually used for secure content delivery and protecting the rights of seller and buyer in the two party DRM architecture. In a multiparty multilevel DRM architecture the cryptographic and watermarking mechanism need to ensure the secure delivery of the content as well as the security concerns of the owner, multiple levels of the distributors and the consumer. However, existing solutions are either impractical or degrade the content to an unacceptable degree. We have come up with a novel CRT-based joint watermarking scheme which takes care of the above security issues, for delivering multimedia content through multiparty multilevel DRM architecture. We have recently starting focusing on the issue of privacy. We have successfully argued that privacy in the multimedia setting need not be a binary question (full

privacy or no privacy). We have come up with a formal notion of partial privacy for multimedia data such as video. We have now extended the model for both single camera surveillance setting and multi-camera setting. We have also developed the anonymous surveillance architecture which preserves privacy.

Synthesis: A well-produced video makes a strong impression on the viewer. However, amateur home video makers are unaware of the principles of cinema grammar. The videos they shoot are meant to convey some specific intent, although often, their inexperience and lack of editing skills, or the limitations of the means of video capture they have at their disposal belie their intentions. We have developed an automated post-processing method for home produced videos based on frame "interestingness". The input single video clip is treated as a long take, and film editing operations for sequence shot are performed. The proposed system automatically adjusts the distribution of interestingness, both spatially and temporally, in the video clip. We use the idea of video retargeting to introduce fake camera work and manipulate spatial interestingness. Meanwhile, video re-projection introduces motion rhythm and modifies the temporal distribution of interestingness. In another example of multimedia synthesis, we have developed a method to automatically create adaptive video presentations of home videos using affective analysis. The method adaptively creates presentations based on three properties: emotional tone, local main character and global main character. A novel sparsity-based affective labeling method is used to identify the emotional content of the videos. The local and global main characters are determined by applying face recognition in each shot. Three kinds of presentations are created for family, acquaintance and outsider. Experimental evaluations show that the proposed approach is very effective.

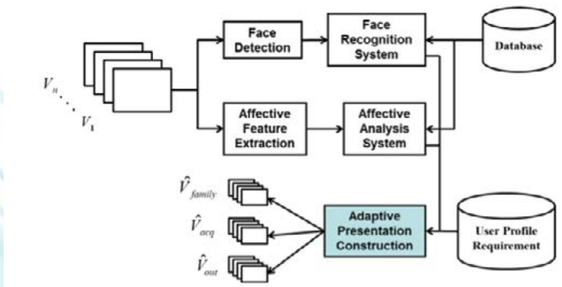


Figure 2: Multimedia Synthesis – Adaptive Presentation of Home Videos.

Representative Publications

Multimedia Fusion with Mean-Covariance Analysis, X.Y. Wang and M.S. Kankanhalli, IEEE Transactions on Multimedia, 2013.

Interactive Video Advertising: A Multimodal Affective Approach, K. Yadati, H. Katti and M.S. Kankanhalli, MMM 2013.

Aggregate Licenses Validation for Digital Rights Violation Detection, A. Sachan, S. Emmanuel and M.S. Kankanhalli, ACM Transactions on Multimedia Computing, Communications and Applications, 2012.

Concept Based Near Duplicate Video Clip Detection for Novelty Re-ranking of Web Video Search Results, C. Bhatt, P. K. Atrey and M.S. Kankanhalli, Multimedia Systems Journal, 2012.

Adaptive Workload Equalization in Multi-camera Surveillance Systems, M. Saini, X. Wang, P. K. Atrey, and M S. Kankanhalli, IEEE Transactions on Multimedia, 2012.

Decision-Theoretic Approach to Maximizing Observation of Multiple Targets in Multi-Camera Surveillance, P. Natarajan, T.N. Hoang, K.H. Low and M.S. Kankanhalli, AAMAS, 2012.

Robust Watermarking of Compressed and Encrypted JPEG2000 Images, A.V. Subramanyam, S. Emmanuel and M.S. Kankanhalli, IEEE Transactions on Multimedia, 2012.

W3-Privacy: Understanding What, When, and Where Inference Channels in Multi-camera Surveillance Video, M. Saini, P. K. Atrey, S. Mehrotra and M S. Kankanhalli. Multimedia Tools and Applications, 2012.

Depth Matters: Influence of Depth Cues on Visual Saliency, C.Y. Lang, T. Nguyen, H. Katti, K. Yadati, S.C. Yan and M.S. Kankanhalli, European Conference on Computer Vision (ECCV2012), Florence, 2012.

Computer Vision

Our research in computer vision spans many areas, including: face modelling and recognition, human motion analysis, medical image analysis, and computational photography.

Computational Photography

Unlike film cameras, digital cameras perform a wide range of processing to convert the camera's raw sensor response to a visually pleasing output image. This on-camera processing is often termed photo-finishing. On most modern digital cameras, there are a number of settings that a user can manipulate to control aspects of how this photo-finishing is applied: the primary two settings being white-balance and picture style (e.g. landscape, portrait, vivid, etc).

Each camera manufacturer has its own proprietary photo-finishing algorithms. In fact, it is often a camera's photo-finishing algorithms, and not the image sensor (i.e. CCD/CMOS), that distinguishes its quality and performance.

Figure 1 shows a schematic diagram for a mathematical model we have developed to describe the imaging pipeline of a digital camera. This model is the results of a thorough study involving more than 30 cameras and over 10,000 images taken with various camera settings and under various lighting conditions. Our model offers a powerful mechanism to mathematically describe how a camera processes its initial sensor response (called RAW) to the final sRGB output.

Our approach breaks down the photo-finishing pipeline into two stages. The first stage considers the white balance setting on the camera together with the necessary transform to convert the camera's RAW-RGB values to the corresponding values in the canonical CIE color-space.

The next stage involves two steps, gamut mapping and tone mapping, which are associated with the picture style of the camera. Tone mapping and gamut mapping are employed to transform the colorimetry of the source image to one that produces a visually pleasing image on the actual reproduction medium, i.e. sRGB. Tone mapping is modeled by what is referred to as the camera's response function, f, and aims to compress the dynamic range of the luminance (i.e. brightness) recorded from the imaged scene. The gamut mapping (h) acts on the color itself and brings the colors that are outside the sRGB gamut to within the gamut. As previously mentioned, this gamut-mapping varies based on the different picture styles. In Figure 1, the gamut mapping is represented as the magnitude of displacement applied to each pixel within the gamut, e.g. blue means low change, while red means higher change. These mathematical transformations can be used to model the photo-finishing steps for any camera.

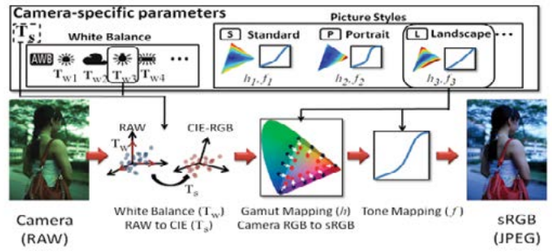


Figure 1: This figure illustrates our photo-finishing model for digital cameras. The first stage is related to the white balance setting on the camera, while gamut mapping and tone mapping are related to the picture style.

Representative Publications

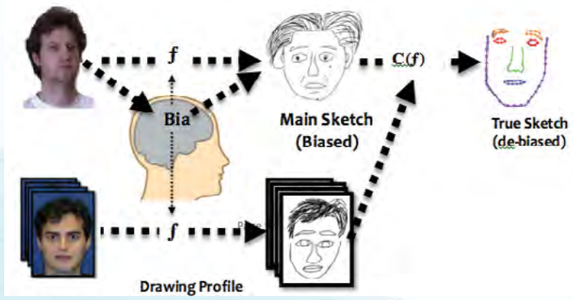
Kim S. J., Lin H.T., Lu Z, Susstrunk S, Lin S., Brown M. S. "A New In-Camera Imaging Model for Color Computer Vision and its Application", IEEE T-PAMI, 34(12), 2012.

Lin H.T, Lu Z., Kim S. J., Brown, M. S. "Nonuniform Lattice Regression for Modeling the Camera Imaging Pipeline", ECCV, 2012.

Lin H.T., Kim S. J., Susstrunk S, Brown M. S. (2011) "Revisiting Radiometric Calibration for Color Computer Vision", IEEE ICCV, 2011.

Improving Eyewitness Face/Sketch Recognition

Eyewitness testimonies have played an important role in police investigations for many decades. Pencil sketches of the suspect's face, obtained from eyewitnesses, are routinely used to help police and the public in identifying and convicting criminals. However, psychological studies are increasingly showing that such sketches are unreliable, biased, and often lead to wrong arrests.



To tackle this problem, we are combining psychological insights into human memory with modern computer vision techniques. Specifically, we attempt to compensate for the eyewitness' mental bias by eliciting his/her drawing profile based on photographs of known faces. After "de-biasing" the initial face sketch, we then detect deviations from a mental norm (called the Exception Report Model in psychology) to extract relevant facial features. These are then matched against a database of known criminals to shortlist candidates for further police follow-up. Experiments confirm the superiority of our approach, and we are now preparing to trial with the Singapore Police Force.

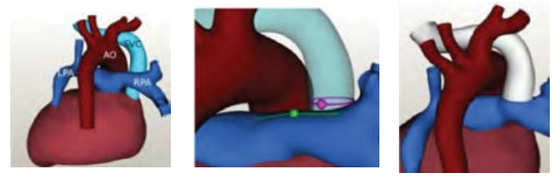
Representative Publications

Kiani Galoogahi, H and T Sim, "Face Photo Retrieval by Sketch Example". ACM Multimedia, 2012.

Hossein, N, T Sim and E. M. Martinez, "Do You See What I See? A More Realistic Eyewitness Sketch recognition". IEEE Biometrics Compendium, 2011.

Medical Imaging

Medical imaging research is often multi-disciplinary, involving computer vision, computer graphics, medicine, surgery, and anatomy. For example, in simulating cardiovascular surgeries (Figure1), a new hollow-tube model of blood vessel is developed, which integrates computer graphics and mechanical engineering to achieve physically correct and yet efficient modeling. In skull model analysis for craniofacial surgery (Figure 2), anatomical planes and landmarks on a skull model are automatically identified based on their definitions in anatomy and forensics. Computational geometry methods are also used to compute surface curvatures on the skull model.



Reconstructed TGA heart model

Cut paths on SVC and RPA

Predicted surgical result

Figure 1. Simulation of cardiovascular surgery.

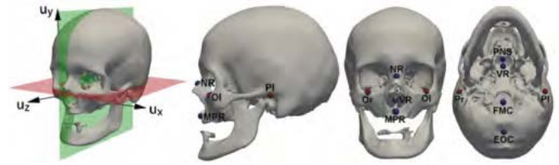


Figure 2. Anatomical planes and landmarks on skull model.

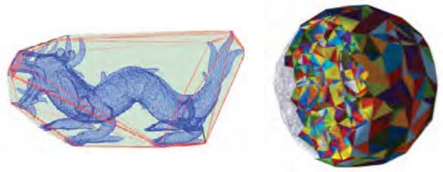
Representative Publications

Y. Cheng, W. K. Leow, T. C. Lim. Automatic Identification of Frankfurt Plane and Mid-Sagittal Plane of Skull. IEEE Workshop on Applications of Computer Vision, 2012.

Computer Graphics and Geometry Processing

The computational capability of the GPU (Graphics Processing Unit) has surpassed that of the CPU in recent years. Consequently GPU has been used extensively to solve large scale problems such as physical and biological simulation. Our research further exploits the enormous power of the GPU to solve fundamental problems in computational geometry such as distance transform, Voronoi diagram, convex hull (CH), and Delaunay triangulation (DT). These geometric structures are useful in, for example, scientific visualization, computer games, robotics and astronomy. The goal of the research is to design and develop a suite of efficient computational geometry algorithms for GPU in 2D and higher dimensions. All our computer programs developed are made available for public download in our project websites.

A raster image & its constrained DT of the edge map



3D CH of the dragon model

Cut-away of a 3D DT

Representative Publications

Gao M C, T T Cao, T S Tan and Z Y Huang. “Flip-Flop: Convex Hull Construction via Star-Shaped Polyhedron in 3D”. ACM SIGGRAPH, 2013.

Qi M, T T Cao and T S Tan. “Computing 2D Constrained Delaunay Triangulation Using the GPU”. IEEE Transactions on Visualization and Computer Graphics, 19 (5): 736-748, 2013.

Character Animation

Physics-based character animation can guarantee the physical realism of the synthesized motions. We have continued our effort on developing physics-based dynamic controllers for fundamental motion tasks such as locomotion. In addition, we have also started investigating highly dynamic motion skills, as shown in Figure 1, including running, jumping, speed vaulting, and drop-rolling. Our system demonstrates the first complete control system that starts from reference motion capture data all the way to controllable dynamic avatars that are able to accomplish parkour-style fast terrain crossing.

Figure 1. A simulated character runs, vaults, jumps, and drop-rolls across a parkour terrain.

Performance-based avatar control is an essential part in human-centric Augmented Reality applications. We have developed a virtual try-on system using a single Microsoft Kinect sensor and a High-Definition Camera. This system provides innovative and interactive 3D retail solution, and was officially launched on April 12 at Raffles City mall, one of Singapore’s largest shopping centers with approximately 3 million visitors passing through every month.

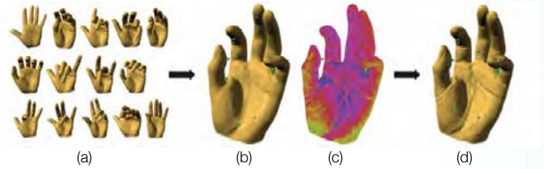


Figure 2: Given sparse training examples (a) and a new configuration of the control points, our deformation models can generate smooth large-scale deformations (b) and high resolution displacements (c), which are then combined to produce deformed models with rich details (d).

A data-driven approach can achieve high quality animations. However, capturing performance from human subjects can be expensive and time consuming. We have investigated deformation models that can produce high quality results from sparse example data. Figure 2 shows a hand animation result. Figure 15 shows a hand animation result from as few as fourteen training hand models.

Representative Publications

Libin Liu, KangKang Yin, Michiel van de Panne, Baining Guo. Terrain Runner: Control, Parameterization, Composition, and Planning for Highly Dynamic Motions. ACM SIGGRAPH Asia, 2012.

Stevie Giovanni, Yeun Chul Choi, Jay Huang, Eng Tat Khoo, and KangKang Yin. Virtual Try-on using a Kinect with an HD Camera. Motionin Games, 2012.

Haoda Huang, KangKang Yin, Ling Zhao, Yue Qi, Yizhou Yu, Xin Tong. Detail-Preserving Controllable Deformation from Sparse Examples. IEEE Transactions on Visualization and Computer 18 (8): 1215-1227, 2012.

Stevie Giovanni and KangKang Yin. LocoTest: Deploying and Evaluating Physics-based Locomotion on Multiple Simulation Platforms. Motion in Games, 2011.

Haoda Huang, Ling Zhao, KangKang Yin, Yue Qi, Yizhou Yu, Xin Tong. Controllable Hand Deformation from Sparse Examples with Rich Details. ACM/Eurographics Symposium on Computer Animation, 2011.

Human Computer Interaction

Human Computer Interaction (HCI) is the study of interaction between people (users) and computer. Our lab focuses on making contributions towards solving real world problems. In keeping with the diverse research field of HCI, our lab has been successful in making contributions to various research questions.

Works of art have been an outlet for human expressions since times immemorial. As the times change, the methods for creating and perceiving art and creativity are also changing. With this view, we have created system and tools for artists to aid them in their work. The project SandCanvas brings the art of sand animation to a multi-touch tabletop. Vignette is an interactive system that facilitates texture creation in pen-and-ink illustrations.

Mobile devices are now ubiquitous, and an integral part of our daily life. We are performing more tasks than ever on the move, and a dedicated visual attention from the user might not always be available. Eyes-free interaction becomes an interesting proposition in such a scenario. Vibration Notification is a step towards using tactile feedback for eyes free interaction. earPod is an eyes free menu selection technique based on audio feedback.

Almost everyone with a computer uses social media to stay connected with friends and family. Farmer’s Tale is a facebook game that exploits the hours spent by people on social media to promote volunteerism in the society. We have also studied the web page revisitation habits of users during tabbed browsing.

We are interested in other research areas as well. Robots in my contact list is a step towards developing communication between humans and robots. We have also contributed to the field of information visualization, wherein we compared plotting techniques for visualizing multidimensional data.

Representative Publications

Designing an Effective Vibration-based Notification Interface for Mobile Phones, Bahador Saket, Chrisnawan Prasojo, Yongfeng Huang, Shengdong Zhao, CSCW, 2013.

Vignette: Interactive Texture Design and Manipulation with Freeform Gestures for Pen-and-Ink Illustration, Rubaiat Habib Kazi, Takeo Igarashi, Shengdong Zhao, Richard Davis, ACM CHI, 2012.

Tracing Tuples Across Dimensions: A Comparison of Scatterplots and Parallel Coordinate Plots, Xiaole Kuang, Haimo Zhang, Shengdong Zhao, Michael J. McGuffin, EuroVis, 2012.

Farmer’s Tale: A Facebook Game to Promote Volunteerism, Don Sim Jianqiang, Xiaojuan Ma, Shengdong Zhao, Jing Ting Khoo, Swee Ling Bay, Zhenhui Jiang, ACM CHI, 2011.

Measuring Web Page Revisitation in Tabbed Browsing, Haimo Zhang, Shengdong Zhao , ACM CHI 2011, 1831-1834

Computational and Interactive Techniques for Artistic Creations

Our goal is to use new and existing computational techniques and user-interface designs to automate or assist the design and creation of artworks. Some of the past and present research projects include

- SandCanvas. A digital interactive medium for visually realistic sand painting and animation on a large touch table. Computer assistance has enabled interesting extensions to the traditional artform.

- Automatic Photo Recomposition. Our method automatically enhances photographs by warping the photographs to optimize the visual dominance of the main subjects. Our new image warping technique is much better at preserving the spatial semantics in the photographs after recomposition.

Figure 1. Automatic photo recomposition.

Figure 2. Automatic sliceform paper pop-up design.

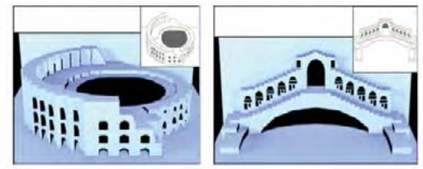


Figure 3. Automatic origamic architecture paper pop-up design.

- Automatic Paper Pop-up Design. We have developed mathematical models and computational methods for two common types of paper pop-ups—paper sliceforms and origamic architectures. Our systems take in ordinary 3D models and produce paper pop-up designs that visually depict the input models.

Representative Publications

Lai-Kuan Wong and Kok-Lim Low. Tearable Image Warping for Extreme Image Retargeting. CGI, 2012.

Lai-Kuan Wong and Kok-Lim Low. Enhancing Visual Dominance by Semantics-Preserving Image Recomposition. ACM Multimedia, 2012.

Rubaiat Habib Kazi, Kien-Chuan Chua, Shengdong Zhao, Richard Davis, Kok-Lim Low. SandCanvas: A Multi-Touch Art Medium Inspired by Sand Animation. ACM CHI, 2011.

Sound and Music Computing

From Napster to YouTube and iTunes, music has always been a major driving force of Internet technologies. A huge amount of music content is now accessible to the public. Organizing and categorizing this content to support an effective recommendation and retrieval has become a significant challenge. We have been developing new technologies to address this challenge in the field of healthcare. We seek to harness the synergy of sound and music computing (SMC), mobile computing, and cloud computing technologies to promote healthy lifestyles and to facilitate disease prevention, diagnosis, and treatment in both developed countries and resource-poor developing countries. This line of research is interdisciplinary in nature (See Figure 1). In addition to being part of Institute for Felicitous Computing (IFC), Centre of Social Media Innovations for Communities (COSMIC), we are in close collaboration with Singapore General Hospital (SGH), National University Hospital (NUH), as well as Beth Israel Deaconess Medical Center (BIDMC) / Harvard Medical School (HMS). With Music Information Retrieval (MIR) as one of our chief research interests, we built a mobile and cloud computing research infrastructure to support the development of large-scale music recommendation systems and retrieval systems. In collaboration with SGH and BIDMC/HMS, we are pioneering high-performance and low-cost gait sensors to facilitate gait training programs using rhythmic auditory stimulation (RAS). In collaboration with SGH and Canossian School for Hearing Impaired, we have created mobile gaming systems MOGAT for auditory training/ habilitation of children with cochlear implants and MOGCLASS for music edutainment of school-aged children. Combining SMC technologies and user-centered design, we strive to develop real-life applications to improve the quality of life for millions of people around the world.

Figure 1. Interdisciplinary research between computing, music and healthcare

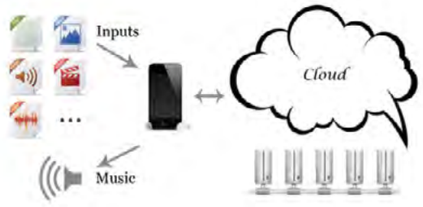


Figure 2. Fusing SMC with mobile and cloud computing

Figure 3. Mobile Games with Auditory Training (MOGAT) System

Music Information Retrieval

Music information retrieval has become a critical but challenging research topic, especially for the real-time search and recommendation of songs according to various user needs. In our recent work, we have designed novel algorithms to satisfy the emerging demands related to music application scenarios. Two of them are as follows. First, many music movie soundtracks, with the same or similar melody but sung and recorded by different people, are uploaded to YouTube every year. Our goal is to help a user to find audio tracks similar to his favorite melody using an audio example, or music companies can recommend new music albums to users with similar melodies according to their listening habits. Our framework for a scalable content-based music information retrieval system is shown in Figure 1 (left).

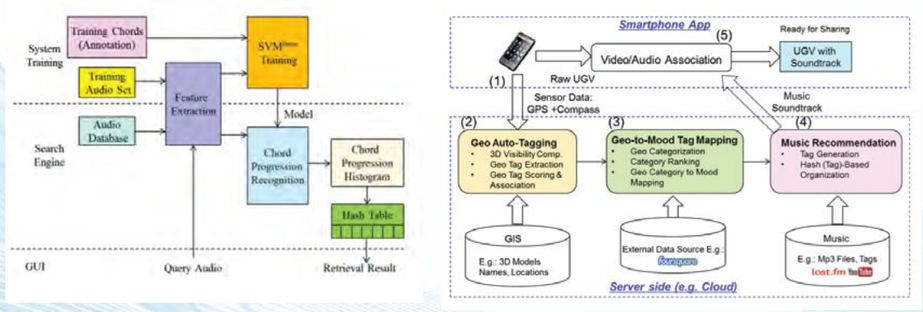


Figure 1: Framework for music information retrieval (left) and automatic soundtrack recommendation system (right).

The architecture consists of four main parts: (1) chord model training, (2) chord progression (CP) recognition, (3) chord progression histogram (CPH) computation, and (4) hash table organization. For training we use the SVMhmm model, which considers both the spectral structure in each feature and CP embedded in adjacent features. With the trained model, CP recognition is performed for all

The second application considers that many user generated videos (UGV) are recorded without much interesting audio and therefore lack appeal when sharing. This is especially true for outdoor, scenic videos (e.g., captured by tourists on vacation) which may be accompanied mostly environmental sounds. We designed a system that enhances UGVs by automatically augmenting them with a music soundtrack that is inspired by the geo-location of where the video was taken. For example a busy city scene that is bustling with people should convey a different atmosphere than a majestic view of a mountain range. Our system, provides a number of key contributions, such as the fusion of video and sensor information, textual geo-tags that are associated with the relevant sub-segments of a video and then translated in to emotionally relevant mood tags. The mood-tags provide the input into a sophisticated music retrieval engine that returns corresponding audio clips that form the soundtrack to be associated with the UGVs.

Representative Publications

Yi Yu, Roger Zimmermann, Ye Wang, and Vincent Oria. Scalable Content-Based Music Retrieval Using Chord Progression Histogram and Tree-Structure LSH. IEEE Transactions on Multimedia, 2013.

Xinxi Wang, David Rosenblum, Ye Wang, “Context-Aware Mobile Music Recommendation for Daily Activities,” ACM Multimedia, 2012

Zhonghua Li and Ye Wang, “A Domain-Specific Music Search Engine for Gait Training,” ACM Multimedia, 2012.

Shenggao Zhu, Hugh Anderson, and Ye Wang, “A Real-Time On-Chip algorithm for IMU-based Gait Measurement,” IEEE PCM, 2012

Yinsheng Zhou, Khe Chai Sim, Patsy Tan, Ye Wang, “MOGAT: Mobile Games with Auditory Training for Children with Cochlear Implants”, ACM Multimedia, 2012.

Wang Feng Ng, Yinsheng Zhou, Patsy Tan, and Ye Wang, “Using the MOGCLASS in Group Music Therapy with Individuals with Muscular Dystrophy: A Pilot Study” Music and Medicine, 4(4) 199-204, 2012.

Yi Yu, Roger Zimmermann, Ye Wang, and Vincent Oria. Recognition and Summarization of Chord Progressions and Their Application to Music Information Retrieval. IEEE Int. Symposium on Multimedia, 2012.

Yi Yu, Zhijie Shen and Roger Zimmermann. Automatic Music Soundtrack Generation for Outdoor Videos from Contextual Sensor Information. ACM Multimedia, 2012.

Interactive Multimedia Systems

We can now pervasively access, view, and interact with large volume of media data on a variety of end devices, enabling new applications such as telemedicine, remote education, virtual meetings, and entertainment. Our research aims to improve the user experience and systems performance of such applications.

Zoomable Video

Mismatch between the display sizes and the resolution of cameras has driven us to rethink how user should interact with video. We have developed techniques and systems that allow users to zoom into either a stored or a live video to view a selected region of interest (RoI) in greater details and to pan around the video to move the RoI.

To improve efficiency, our system transmits and decodes only the bits necessary to decode the pixels within the current RoI. We proposed two approaches to achieve this goal: (i) We analyze the dependencies across different video frames and different video macroblocks, and transmit or decode only the bits that would contribute to the decoding the pixels within the given RoI; (ii) We reencode the videos to reduce the dependencies in the video bitstreams. Both techniques offer a different set of tuning parameters and trade offs, which we have studied extensively.

To improve usability, we conducted a study to learn about the user behavior when they zoom and pan in a video. Our study revealed that: (i) users tend to zoom into the same regions, and (ii) there are many consecutive pans due either to tracking a moving target, or repositioning of the RoI after an imprecise zoom. To reduce the number of pans, we have developed a new UI, called COZI, that combines content analysis and frequently zoomed-in regions to automatically recommend potentially interesting regions for users to zoom into. Further, the UI automatically pans to track a moving object selected by users, significantly reducing the number of user interactions needed. Our algorithm to infer interesting regions can also be used to retarget a video to a small display, by automatically zooming into the most interesting region.

Representative Publications

F. Liu and W. T. Ooi, “Zoomable Video Playback on Mobile Devices by Selective Decoding,” Pacific-Rim Conference on Multimedia, 2012.

Carlier, G. Ravindra, V. Charvillat, and W. T. Ooi, “Combining Contentbased Analysis and Crowdsourcing to Improve User Interaction with Zoomable Video,” ACM Multimedia, 2011.

Carlier, G. Ravindra, and W. T. Ooi, “Towards Characterizing Users’ Interaction with Zoomable Video,” Int. Workshop on Social, Adaptive, and Personalized Multimedia Interaction and Access, 2011.

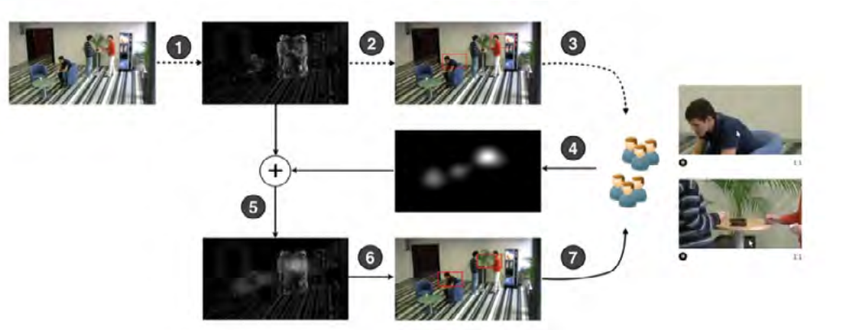


Figure 1: How COZI recommends potential ROIs to users. 1: Compute saliency. 2: Recommend regions. 3: Display to users. 4: Collect user inputs. 5: Fused saliency with user inputs. 6: Recommend new regions. 7: Display to new users.

Peer-Assisted Systems

The need for scalability has lead to the rise of peer-assisted media streaming systems. Supporting user interactivity in such systems, however, leads to tighter performance requirements and lower predictability. Our group addresses these issues from several angles. We have designed a new content discovery middleware called APRICOD that learns from past access patterns to correlate between media object accesses, allowing a peer to pre-connect to others with the data the peer will likely need in the future, thereby reducing the latency during interaction. We investigate how to allocate bandwidth and schedule the transmissions among peers for on-demand and prefetch requests, so that more content requests can be received within the interaction deadline. Finally, we mathematically modeled a peer-assisted VoD system that supports user interaction, to study the impact of interactions on the system.

A peer may also assist by offloading computations from a server. We considered such scenario in the PeAR project, in the context of a networked virtual environment, where multiple users are accessing the same 3D scene simultaneously. Instead of performing server-based rendering for thin clients, part of the scene that was rendered in another client is reused for a thin client. Our preliminary study shows the promise of such technique. We are currently exploring efficient methods to store and index previously rendered data in a distributed manner among peers to support fast content discovery and bandwidth efficient data exchange among peers.

Representative Publications

Z. W. Zhao, S. Samarth, and W. T. Ooi, “Modeling the Effect of User Interactions on Mesh-based P2P VOD Streaming Systems,” ACM Transactions of Multimedia Computing, Communications and Applications 9 (2), 2013.

Minhui Zhu, S. Mondet, G. Morin, W. T. Ooi, and W. Cheng, “Towards Peer-Assisted Rendering in Networked Virtual Environments,” ACM Multimedia, 2011.

K. Liang, R. Zimmermann, and W. T. Ooi, “Peer-Assisted Texture Streaming in Metaverses,” ACM Multimedia, 2011.

Networked Graphics

Availability of 3D scanning devices and improvement in scanning techniques have led to an increasing availability of high resolution, voluminous, 3D data. Sharing this data across the network is important for applications such as tele-medicine, digital museum, and e-commerce.

Our group proposes a new paradigm for sharing high-resolution 3D meshes called 3D preview streaming. Our technique straddles in between interactive 3D mesh streaming and non-interactive video streaming -- the 3D meshes are previewed by users along a pre-defined camera path and vertices visible along the path is streamed. This technique overcomes many deficiencies in video-based preview of 3D objects -- it is flexible enough to adapt to bandwidth fluctuations without significant degrading of perceptual quality, and it allows users to modify the rendering parameters dynamically while previewing a 3D object.

Another application we tackled is remote rendering and storage of medical images and 3D data. Medical institutions now have the option of storing medical data in cloud-based remote storage. The cloud servers may serve as a rendering server, rendering 3D volumes and transmitting the resulting images or videos to remote thin clients. In such systems, storing sensitive medical data on third party servers leads to privacy and security concerns. We are investigating into techniques that allow the medical data to be hidden from the servers they are stored on, yet allow the servers to operate on the data.

Figure 1: Previewing a 3D model as a mesh leads to better quality when bandwidth is low (Right) compared to viewing the model as a video (Left).

Representative Publications

S. Zhao, W. T. Ooi, A. Carlier, G. Morin, and V. Charvillat, “3D Mesh Preview Streaming,” ACM Multimedia Systems Conference, 2013.

M. Mohanty, P. K. Atrey, and W. T. Ooi, “Secure Cloud-based Medical Data Visualization,” ACM Multimedia, 2012.

Sensor-Rich Video Management

The rapid adoption and deployment of ubiquitous mobile devices with video cameras, such as smartphones and tablets, has led to the collection of voluminous amounts of media data. However, the managing of large video archives remains a very challenging task. In the GeoVid project we focus on exploitation of fully automatic, contextual annotations of videos from built-in sensors in mobile devices such as compass, GPS and accelerometers. In our work we have investigated numerous aspects of sensor-rich videos, for example we have implemented a search portal (http://geovid.org) which allows various, very efficient video retrievals based on spatio-temporal properties (e.g., directional searches, surround queries). We have further explored solutions for real-world issues such as sensor meta-data correction (raw sensor data is often very noisy and spurious).

In addition to a management portal (see Figure 1) the project team has created acquisition apps for iOS and Android which are available in the Apple iTunes and Google Play stores. The GeoVid project is a collaboration with the Integrated Media Systems Center (IMSC) at the University of Southern California, and the system has been used in a number of experiments, notably at the NATO Summit in Chicago on 21/22 May 2012 and with PBS NewsHour during the Presidential Inauguration of Mr Barack Obama’s second term in Washington DC, on 21 January 2013

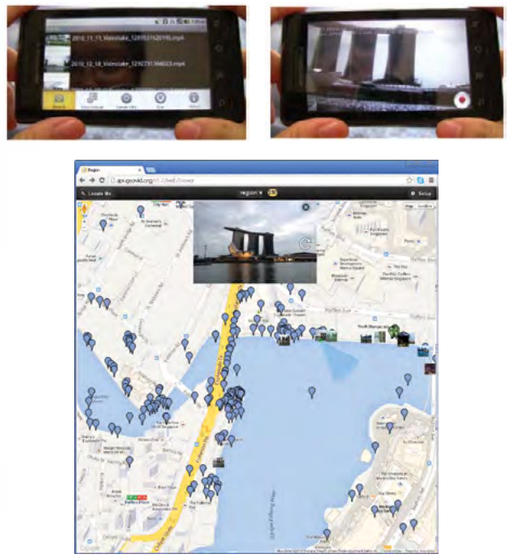


Figure 1. The GeoVid Android app is available from the Google Play store. The GeoVid portal supports a number of innovative spatio-temporal queries, such as directional and surround queries.

Representative Publications

Ying Zhang and Roger Zimmermann. Dynamic Multi-Video Summarization of Sensor-rich Videos in Geo-Space, MMM, 2013.

He Ma and Roger Zimmermann. HUGVid: Handling, Indexing and Querying of Uncertain Geo-tagged Videos, ACM SIGSPATIAL, 2012.

Guanfeng Wang, Haiyang Ma, Beomjoo Seo and Roger Zimmermann. Sensor-Assisted Camera Motion Analysis and Motion Estimation Improvement for H.264/AVC Video Encoding, ACM SIGMM Workshop on Network and Operating Systems Support for Digital Audio and Video, 2012.

Ying Zhang, Guanfeng Wang, Beomjoo Seo, and Roger Zimmermann. Multi-Video Summary and Skim Generation of Sensor-rich Videos in Geo-Space, ACM Multimedia Systems Conference, 2012.

Weiwei Cui, Beomjoo Seo, and Roger Zimmermann. Efficient Video Uploading from Mobile Devices in Support of HTTP Streaming, ACM Multimedia Systems Conference, 2012.

Natural Language Processing

Semantic Processing

We continued our research on word sense disambiguation (WSD), the task of determining the correct meaning or sense of a word in context. We released our open-source word sense disambiguation system, IMS, which performs WSD for English texts. We found that word sense disambiguation improves the performance of information retrieval (IR). We also embarked on research related to semantic similarity or semantic equivalence, which determines whether two natural language sentences are semantically similar or equivalent. This work was carried out in the context of automatic machine translation (MT) evaluation and paraphrase evaluation. We came up with a family of novel automatic evaluation metrics for machine translation, called TESLA-*. These metrics employ linear programming-based matching, which enables optimal matching of weighted n-grams. We participated in WMT shared task evaluation on MT evaluation metrics in 2010 and 2011. Our TESLA-M metric achieved the highest average system level correlation with human judgments when translating from English, and TESLA-F when translating into English. Two additional improved variants of TESLA were used in automatic machine translation evaluation of output languages with ambiguous word boundaries (such as Chinese), and in summarization evaluation. We have also released the open-source code of the TESLA family of metrics for public use. In addition, we have developed PEM, the first fully automatic metric to evaluate the quality of paraphrases and paraphrase generation systems.

Representative Publications

Zhong, Zhi, & Ng, Hwee Tou. Word Sense Disambiguation Improves Information Retrieval, ACL, 2012.

Liu, Chang, & Ng, Hwee Tou. Character-Level Machine Translation Evaluation for Languages with Ambiguous Word Boundaries, ACL, 2012.

Dahlmeier, Daniel, & Liu, Chang, & Ng, Hwee Tou. Translation Evaluation and Tunable Metric, WMT, 2011.

Li, Maoxi, & Zong, Chengqing, & Ng, Hwee Tou. Automatic Evaluation of Chinese Translation Output: Word-Level or Character-Level? ACL, 2011.

Natural Language Generation

We came up with the first probabilistic model for generating natural language sentences from typed lambda calculus expressions. The contributions of this work include a novel forest-to-string generation algorithm and a novel grammar induction algorithm. The new model outperforms previous proposed models when evaluated on common benchmark data sets. This work is significant in that it addresses the fundamental task of natural language generation from semantic representation, linking logic-based typed lambda calculus expressions to natural language sentences. A paper describing this work won the Best Paper Award at the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP 2011).

Representative Publications

Lu, Wei, & Ng, Hwee Tou. A Probabilistic Forest-to-String Model for Language Generation from Typed Lambda Calculus Expressions, EMNLP, 2011.

Discourse Processing

We have successfully built the first end-to-end discourse parser that outputs PDTB-style discourse relations given any input English text. Our discourse parser is automatically trained on the discourse relations annotated in PDTB (Penn Discourse Treebank). We have released our open-source discourse parser. The discourse relations output automatically by our parser have been used in two ways. First, we came up with a novel coherence metric that measures text coherence based on the discourse relations automatically identified by our discourse parser. Second, our coherence metric is combined with a variant of the TESLA metric to form a novel summarization evaluation metric CREMER, which gives highly competitive performance when evaluated on the TAC (Text Analysis Conference) summarization evaluation task.

Representative Publications

Lin, Ziheng, & Liu, Chang, & Ng, Hwee Tou, & Kan, Min-Yen. Combining Coherence Models and Machine Translation Evaluation Metrics for Summarization Evaluation, ACL, 2012.

Lin, Ziheng, & Ng, Hwee Tou, & Kan, Min-Yen. Automatically Evaluating Text Coherence Using Discourse Relations. ACL, 2011.

Grammatical Error Correction

We have embarked on research on grammatical error correction. We came up with a novel beam-search decoder that can correct multiple, interacting grammatical errors. This work is unique in that it is the first that attacks grammatical error detection and correction of multiple interacting errors, whereas the current state of the art in the field focuses on using individual classifiers to correct single error types. We also came up with a novel approach for grammatical error correction that exploits both learner and non-learner corpora, based on alternating structure optimization. Another invention uses information about the writer's first language (native language) to infer which words or phrases are confusable for the writer.

We participated in the HOO (Helping Our Own) 2012 shared task, on grammatical error detection and correction of determiners and prepositions. Our team emerged as the best team among 14 participating teams, and the best system among 85 systems submitted to the shared task. Our participating system makes use of large-scale web corpus statistics, which improves the robustness of our system. We also came up with a better evaluation metric for grammatical error correction that improves upon the metric used in the HOO shared task. We have filed three US patents for our work on grammatical error correction.

Representative Publications

Dahlmeier, Daniel, & Ng, Hwee Tou. A Beam-Search Decoder for Grammatical Error Correction, EMNLP-CoNLL, 2012.

Dahlmeier, Daniel, & Ng, Hwee Tou, & Ng, Eric Jun Feng. NUS at the HOO 2012 Shared Task. Workshop on the Innovative Use of NLP for Building Educational Applications, 2012.

Dahlmeier, Daniel, & Ng, Hwee Tou. Better Evaluation for Grammatical Error Correction, NAACL, 2012

Dahlmeier, Daniel, & Ng, Hwee Tou. Correcting Semantic Collocation Errors with L1-induced Paraphrases, EMNLP, 2011.

Dahlmeier, Daniel, & Ng, Hwee Tou. Grammatical Error Correction with Alternating Structure Optimization, ACL, 2011.

Machine Translation

Our research on statistical machine translation (SMT) focuses on translation of resource poor languages such as Malay and Indonesian. We came up with a novel language-independent approach for improving machine translation for resource-poor languages by exploiting their similarity to resource-rich ones, taking advantage of vocabulary overlap and similarities between the languages in spelling, word order, and syntax.

Most existing SMT approaches treat a word as the basic unit of translation. However, this leads to data sparseness when translating morphologically more complex languages such as Malay and Indonesian. We proposed a novel approach for translating Malay into English, by treating morphologically related words as paraphrases. Our proposed approach emphasizes derivational morphology, which has not been actively addressed in the past.

Leveraging our research on automatic MT evaluation, we have found that tuning an SMT system using TESLA-M or TESLA-F outperforms tuning using BLEU, a widely used MT evaluation metric. This enables us to build better SMT systems.

Representative Publications

Wang, Pidong, & Nakov, Preslav, & Ng, Hwee Tou. Source Language Adaptation for Resource-Poor Machine Translation, EMNLP-CoNLL, 2012.

Nakov, Preslav, & Ng, Hwee Tou. Improving Statistical Machine Translation for a Resource-Poor Language Using Related Resource-Rich Languages. Journal of Artificial Intelligence Research, 44: 179 - 222, 2012.

Nakov, Preslav, & Ng, Hwee Tou. Translating from Morphologically Complex Languages: A Paraphrase-Based Approach, ACL, 2011.

Liu, Chang, & Dahlmeier, Daniel, & Ng, Hwee Tou (2011). Better Evaluation Metrics Lead to Better Machine Translation, EMNLP, 2011.

Social Media Analytics and Digital Libraries

Digital library and social media research are diverse fields, embodying the research, development and deployment of digital libraries and social network. In the area of social media preservation, our groups have created an open access repository for both English and Chinese Weibo social media content (microblogs, check-ins, tweets, and photos), for both enabling their preservation, as well as creating a rich repository for in-depth integrative analysis. Our work on social media analytics help people predict the popularity of their posts and optimize how to present their posts to their readership.

We have also furthered the state of the art research and specific domain specific digital libraries. In the medical vertical, we have created state-of-the-art joint text classification algorithms that parse, understand, and extract key patient demographic metadata from MedLine abstracts, at both the sentence- and word-levels.

In a separate line of work, we examine the problem of recommending research papers to authors themselves. We have found that enriching the citation network with text extracted from paper citing over the work, as well as papers referenced by the work, improve this recommendation accuracy significantly.

Representative Publications

Jin Zhao, Praveen Bysani and Min-Yen Kan. Exploiting Classification Correlations for the Extraction of Evidence-based Practice Information, AMIA Annual Symposium, 2012.

Anqi Cui, Liner Yang, Dejun Hou, Min-Yen Kan, Yiqun Liu, Min Zhang and Shaoping Ma. PrEV: Preservation Explorer and Vault for Web 2.0 User-Generated Content, TPDL 2012.

Aobo Wang, Tao Chen and Min-Yen Kan. Re-tweeting from a Linguistic Perspective, NAACL-HLT, 2012, Workshop on Language in Social Media.

Speech Recognition

Research in the area of automatic speech recognition (ASR) aims at improving the accuracy and efficiency of converting speech into text. ASR is an essential component for many voice-enabled applications, notably the voice search and intelligent personal assistant capabilities on the mobile devices. Despite its long history of research and development, ASR is still quite vulnerable when operating in an adverse acoustic environment. One of the major challenges of noise-robust speech recognition is how to reliably compensate the acoustic models trained on clean speech data so that they can perform well on noisy speech. Dr. Sim and his research team have developed a novel trajectory-based model compensation framework that can serve as the foundation for existing compensation methods by offering a better model for the dynamics (temporal correlations) in the acoustic signals. Trajectory-based compensation can achieve about 10% relative word error rate reduction compared to existing state-of-the-art techniques.

Nevertheless, there is only so much one can improve the performance of ASR in noisy environment by relying on acoustic signals alone. Multimodal input methods are commonly used to improve the robustness by introducing redundancies with multiple input streams. Along this direction, Dr. Sim has also developed Haptic Voice Recognition (HVR), an innovative multimodal input framework specifically designed for modern portable devices that combines both voice and touch inputs. The goal of HVR is to enable reliable and efficient offline speech recognition, where touch inputs are augmented to standard ASR system to reduce the recognition computation and improve the recognition accuracy. A probabilistic framework has been developed to tightly integrate multiple touch-based and voice-based inference models. Empirical studies show that by using the initial letters of each word as redundant cues, HVR offers more than 50% relative improvement to recognition accuracy and more than twice the speedup in recognition time compared to conventional ASR.

Representative Publications

K.C. Sim, S. Zhao, K. Yu and H. Liao, "ICMI'12 Grand Challenge Haptic Voice Recognition", ICMI, 2012.

G. Wang, B. Li, S. Liu, X. Wang, X. Wang and K. C. Sim, "Improving Mandarin Predictive Text Input By Augmenting Pinyin Initials with Speech and Tonal Information", ICMI, 2012, Grand Challenge - Haptic Voice Recognition Workshop.

K.C. Sim, "Probabilistic Integration Of Partial Lexical Information For Noise Robust Haptic Voice Recognition", ACL, 2012,

S. Liu and K.C. Sim, "Implicit Trajectory Modelling Using Temporally Varying Weight Regression For Automatic Speech Recognition", ICASSP, 2012.

K.C. Sim and M. Luong, "A Trajectory based Parallel Model Combination With A Unified Static and Dynamic Parameter Compensation for Noisy Speech Recognition", ASRU, 2011.

The faculty members of the Media Group are:

- Michael Brown
- Chua Tat-Seng
- Kan Min-Yen
- Mohan Kankanhalli
- Leow Wee Kheng
- Low Kok Lim
- Ng Hwee Tou
- Ooi Wei Tsang
- Sim Khe Chai
- Sim Mong Cheng, Terence
- Tan Tiow Seng
- Wang Ye
- Yin Kang Kang
- Zhao Shengdong
- Roger Zimmermann

PROGRAMMING LANGUAGES AND SYSTEMS

Programming Languages and Systems is a central and fundamental topic in computer science which concerns all aspects that have to do with programming. The research in the programming language group includes work on particular programming languages, theoretical aspects of programming languages such as types and semantics, program analysis and verification of software, paradigms of programming, and aspects arising from concurrent, parallel and distributed programming. There is also overlap between the work in programming languages with the software engineering and formal methods and also security areas. The research can be classified along the following dimensions:

- Constraint Programming: constraint logic programming (CLP), algorithms for Constraint Satisfaction Problems (CSP), applications to real-life combinatorial problems, constraint solving frameworks and systems, etc.
- Dependable Software: detection of software vulnerabilities, systematic construction of dependable software
- Formal Methods: verification, model checking and theorem proving
- Functional Programming: extension of type systems, enhancing functional programming languages, etc.
- Program Analysis and Transformations: abstract interpretation, reasoning about programs and partial evaluation.
- Programming paradigms: Principles and Practice of Aspect Oriented Programming
- Theoretical Foundations: models of concurrency, semantics and logics of programming languages, etc.
- Understanding software behavior: specification mining, software visualization

Constraint Programming

Constraint Programming (CP) is a programming paradigm where constraints, i.e. relations over variables, are used as basic entities for problem solving. It differs from imperative programming in that the constraints express declarative relationships and thus do not specify how the constraints or the problem should be solved. Instead that is the work of the constraint programming system or language as well as the constraint solver. Constraint programming has strong links with logic programming because of its declarative nature and much of the work in constraints has arisen from Constraint Logic Programming. Work in CP has strong links with the artificial intelligence community and CP (and CSP) is a major area in AI. The main areas of research in CP at the School of Computing are in new forms of constraints, consistency algorithms, local search and applications of CP such as scheduling, and CLP for program reasoning.

Arbitrarily defined finite domain constraints, also known as non-binary CSPs or table constraints, pose a challenge for solvers because even local consistency becomes intractable in the worst case. We have devised a new global constraint called mddc which uses multi-valued decision diagrams (MDD) allowing it to incorporate implicit structure as well as compression. This gives both a compact representation as well as an efficient representation.

In earlier work, we demonstrated the usefulness of mddc constraints which can give very good performance when the MDD is a compact representation for the constraint. The regular constraint is a relation whose values are defined by a regular expression. A regular expression can be represented in many forms such as the regular expression, a non-deterministic finite automata (NFA) and a deterministic finite automata. In the case of the regular constraint some possible representations are as a MDD, NFA and DFA. We study the space and time tradeoffs of these representations and show that although the MDD representation is fast, as the constraint becomes larger, the even more compact representations become desirable. We also extend the approach to work with grammar constraints, i.e. constraints defined by grammars.

A related approach to mddc are the Simple Tabular Reduction (STR) which use the idea that as search progresses and variables are instantiated, the table defining the constraint also becomes smaller. Both the mddc and STR approaches perform less work as variables become instantiated. Together with mddc, the STR2 GAC (Generalised Arc Consistency) algorithm are among the best for achieving GAC. We have improved the STR2 algorithm with a new algorithm, STR3, which has better optimality properties when used during search.

Representative Publications

Chavalit Likitvivatanavong and Roland H.C. Yap: Many-to-many interchangeable sets of values in CSPs, SAC: ACM Symp. On Applied Computing, 2013.

Kenil C. Kenil C.K. Cheng, Wei Xia and Roland H.C. Yap: Space-Time Tradeoffs for the Regular Constraint, CP, 2012.

Christophe Lecoutre, Chavalit Likitvivatanavong and Roland H.C. Yap: A Path-Optimal GAC Algorithm for Table Constraints, ECAI, 2012.

Symbolic Execution

Formal reasoning of programs for discovering and proving properties of programs is crucially needed and yet their development and large-scale deployment is in its infancy. We are developing a highly flexible framework for general program reasoning that attempts to bridge this gap. Our work starts by modeling programs in the framework of Constraint Logic Programming (CLP), which is highly expressive for behavioral modeling of systems, including those of programs and specifications. In particular, the support for constraints allows for a flexible modeling of complex behavior. By executing the model CLP program, the system runs can be straightforwardly simulated.

Our goal is to automatically prove certain properties of programs, as well as perform program analysis. The work is founded on the use of symbolic execution enabled with the concept of interpolation, for the purpose of pruning during verification, and the concept of dynamic programming re-use, for the purpose of discovering the best solutions. Present research focusses both on sequential programming languages C and LLVM programs, and present analyses are focused on resource bounds in embedded programs, as well as complex properties about mutable data structures. We also work on concurrent programs, with a focus on interleaving reduction based on symmetries and partial-order reductions. In addition to verification and analysis, our core technology is being used to perform program transformations, e.g. slicing, for the purpose of creating more new programs that can be more efficiently processed by a third-party application. Testing is a diving example of for our program transformations.

A system TRACER is under development. The input is a C program, and some target properties. The system then produces a special control flow graph from which we can extract correctness properties. User intervention is in the form of local abstractions which serve to accelerate the time it takes to cover the symbolic execution search space. While TRACER serves as support for research into new algorithms, it will ultimately be a stand-along tool for program understanding.

Representative Publications

Joxan Jaffar, Vijayaraghavan Murali, Jorge Navas and Andrew Santosa: TRACER: A Symbolic Execution Tool for Verification, CAV, 2012.

Joxan Jaffar, Vijayaraghavan Murali, Jorge Navas and Andrew Santosa: Path Sensitive Backward Slicing, SAS, 2012.

Chu Duc Hiep and Joxan Jaffar: A Complete Method for Symmetry Reduction in Safety Verification, CAV, 2012.

Joxan Jaffar, Jorge Navas and Andrew Santosa: Unbounded Symbolic Execution for Program Verification, RV, 2011.

Chu Duc Hiep and Joxan Jaffar: Symbolic Simulation on Complicated Loops for WCET Path Analysis, EMSOFT, 2011.

Separation Logic

There are a number of lines of research on separation logic. One is to work towards the automated verification of software to ensure safety and dependability. It aims to address one of the grand challenge issues in computer science, to make verified software a reality. Two tools have been developed: SLEEK is an entailment prover for separation logic formula, while HIP is a verifier for an imperative language that supports specification of data structures with strong invariant properties.

Another line of research is on semantic modeling focused on models for separation logic, particularly separation logic with higher-order features. This included work on separation algebras and share accounting, as well as methods to approximate an important class of recursive domain equations. Separation logic has also been used to reason about programs that manipulate data structures with intrinsic sharing, e.g. graphs and DAGs.

Representative Publications

Aquinas Hobor and Jules Villard: The Ramifications of Sharing in Data Structures. POPL, 2013.

Wei-Ngan Chin, Cristina David, Huu Hai Nguyen and Shengchao Qin: Automated verification of shape, size and bag properties via user-defined predicates in separation logic. Science of Computer Programing, 77(9), 1006-1036, 2012.

Cristian Gherghina, Cristina David, Shengchao Qin and Wei-Ngan Chin: Structured Specifications for Better Verification of Heap-Manipulating Programs. Formal Methods, 2011.

Wei-Ngan Chin, Cristian Gherghina, Razvan Voicu, Quang Loc Le, Florin, Craciun and Shengchao Qin: A Specialization Calculus for Pruning Disjunctive Predicates to Support Verification, CAV, 2011.

Cristina David and Wei-Ngan Chin: Immutable specifications for more concise and precise verification. OOPSLA, 2011.

Specification Mining and Software Understanding

It's best if all programs and software projects are developed with clear, precise and documented specifications. However, due to hard deadlines and 'short-time-to-market' requirement, software products often come with poor, incomplete and even without any documented specifications. This situation is aggravated by a phenomenon termed as software evolution. As software evolves the documented specification is often not updated. This might render the original documented specification of little use after several cycles of program evolution.

In our research, we investigate various means to discover software specifications from its associated program execution traces, as well as applying discovered specifications to help improve programmers' productivity. In the frontier of specification discovery, we devise a novel mining technique to generate a "difference specification" from a pair of execution traces obtained from two versions of a program. This difference specification concisely identifies the differences across program versions at the level of message sequence graphs.

In the frontier of improving programmers' productivity, we collaborate with the University of Copenhagen and Singapore Management University to develop an efficient tool for inferring transformation specifications from a few examples of original and updated code. We ensure that the inferred transformations are context sensitive, easing the task of understanding and applying collateral evolutions. We also develop a runtime anomaly monitoring system, NORT, that verifies on-the-fly whether running applications comply to their expected normal behavior. The model of normal behavior is created based on various indicators, including a rich set of discriminators, which are patterns of system calls. Experiments run on malware samples have shown that our approach is able to effectively detect a broad range of attacks with very low overheads. We are also looking at the problem of understanding the behavior and interactions of complex real-world software. A different approach to mining is to use visualization.

Representative Publications

Sandeep Kumar, Siau-Cheng Khoo, Abhik Roychoudhury and David Lo: Inferring class level specifications for distributed systems, ICSE, 2012.

Jesper Andersen, Anh Cuong Nguyen, David Lo, Julia L. Lawall and Siau-Cheng Khoo: Semantic patch inference, ASE, 2012.

Narcisa Andreea Milea, Siau-Cheng Khoo, David Lo and Cristian Pop: NORT: Runtime Anomaly-Based Monitoring of Malicious Behavior for Windows, RV, 2011

The faculty members in programming languages research are:

- Norman Hugh Anderson
- Chin Wei Ngan
- Martin Henz
- Aquinas Hobor
- Joxan Jaffar
- Khoo Siau Cheng
- Abhik Roychoudhury
- Yap Hock Chuan, Roland

SOFTWARE ENGINEERING

Software engineering is concerned with the problem of producing high quality software. It deals with issues encompassing design aspects to the implementation and subsequent maintenance of the software. It includes the use of formal methods as well as the pragmatics of actual software development. The research in the practical aspects of software engineering group ranges from using software reuse techniques for designing high-variability software components to formal methods. In formal methods, the research looks at the problems of specification and verification of software. There is synergy with the faculty and research in the programming languages area in program verification, debugging and system specification.

The research can be classified along the following dimensions:

- integrated modeling techniques
- mining software systems
- model checking
- model-driven engineering
- program and software composition
- scalability in software systems
- software engineering education
- software engineering methodologies
- software reuse
- software specification
- ubiquitous computing

Clone Detection and Analysis

Software clones are similar code fragments that meet certain threshold of similarity metrics. We introduced the concept of structural clones that represent design-level similarity patterns in programs.
For example, classes, source files, directories with source files, components of any kind, patterns of collaborating components or even subsystems can be recognized as structural clones if they meet pre-defined threshold of similarity metrics. We developed Clone Miner that can detect certain types of structural clones, and Clone Analyzer to help a human expert to analyze structural clones reported by Clone Miner. Knowledge of structural clones is useful in re-engineering of legacy software systems into Software Product lines for reuse, and in software maintenance.

Representative Publications

Basit, H., Usman, A. and Jarzabek, S.: Things Structural Clones Tell that Simple Clones Don't, ICSM, 2012.

Xing, Z., Xue, Y. and Jarzabek, S.: CloneDifferentiator: Analyzing Software Clones by Differentiation, ASE, 2011.

Basit, H., Ali, U. and Jarzabek, S.: Viewing Simple Clones from a Structural Clones' Perspective, IWSC, 2011

Formal Verification

One of the most difficult aspects of developing software is to ensure that programs are free from bugs. Formal verification seeks to prove the correctness of the software. Usually, it is a particular aspect or model of the software. Verification also addresses the issue whether the software (or model of the software) conforms to its specification.
The PAT (Process Analysis Toolkit) verification system supports composing, simulating and reasoning of concurrent, real-time systems and other possible domains. It comes with user friendly interfaces, featured model editor and animated simulator. Most importantly, PAT implements various model checking techniques catering for different properties such as deadlock-freeness, divergence-freeness, reachability, LTL properties with fairness assumptions, refinement checking and probabilistic model checking. To achieve good performance, advanced optimization techniques are implemented in PAT, e.g. partial order reduction, symmetry reduction, process counter abstraction, parallel model checking. So far, PAT has 2500+ registered users from 600+ organizations in 62 countries and regions.

The USMMC, a UML State Machine Model Checker, supports editing, simulation and automatic verification of UML state machines. It follows a formal operational semantics covering all features of the latest version (2.4.1) of the UML state machines specification including synchronous and asynchronous communications between state machines.

Representative Publications

Jun Sun, Yang Liu, Jin Song Dong, Yan Liu, Ling Shi, Etienne Andre: Modeling and Verifying Hierarchical Real-time Systems using Stateful Timed CSP, TOSEM: ACM Trans. on Software Engineering and Methodology, 22(1), 2013.

Tian Huat Tan, Etienne Andre, Jun Sun, Yang Liu, Jin Song Dong and Manman Chen: Dynamic Synthesis of Local Time Requirement for Service Composition. ICSE, 2013.

Etienne Andre, Yang Liu, Jun Sun, Jin Song Dong and Shang-Wei Lin: Parameter Synthesis for Hierarchical Concurrent Real-Time Systems. CAV, 2013.

Lin Gui, Jun Sun, Yang Liu, Yuanjie Si, Jin Song Dong and Xinyu Wang: Combining Model Checking and Testing with an Application to Reliability Prediction and Distribution. ISSTA, 2013.

Shuang Liu, Yang Liu, Étienne André, Christine Choppy, Jun Sun, Bimlesh Wadhwa and Jin Song Dong: A Formal Semantics for Complete UML State Machines with Communications. IFM, 2013.

Scalability of Software Systems

Scalability is an important quality that pervades many aspects of computing, but it is a quality that is poorly defined. Recent research has attempted to develop a sound basis for systematic characterization and analysis of scalability. In this work, scalability is regarded as a cross-cutting quality of software systems characterized by the operational impact that scaling aspects of the execution environment and system design have on certain measured software qualities, as these aspects are varied over expected operational ranges. If the system can accommodate this variation in a way that is acceptable to appropriate stakeholders, then it is a scalable system. The research produced a framework for precise characterization and analysis of software scalability, which views scalability analysis as a form of experimental design, and it produced systematic techniques for scalability analysis and elicitation of scalability requirements. Case studies with a real financial fraud detection system demonstrated that systematic scalability analysis could have revealed scalability limitations in the first generation of the system's design.

The most recent research addresses the problem of systematically determining which elements of the execution environment are most relevant to system requirements. The research produced an extension to the popular KAOS method for goal-oriented requirements engineering, in which a suite of tactics is used to systematically identify scalability obstacles in a goal-oriented software requirements specification and to resolve the identified obstacles with refined goals. Case studies with the same financial fraud detection system demonstrated the increased precision and understanding that can be achieved when applying the method for elicitation of scalability requirements.

Representative Publication

Leticia Duboc, Emmanuel Letier and David S. Rosenblum: Systematic Elaboration of Scalability Requirements through Goal-Obstacle Analysis. IEEE Trans. Software Eng, 39(1), 2013.

Software Engineering Education

While software engineering fundamentals have been established, how to effectively teach them in the frame of university courses remains a challenge. In particular: how to teach design fundamentals in a way that is relevant to a plethora of changing software technologies and emerging new application domains? Today programs are built on top of functionalities provided by software platforms. Most often, developers extend existing systems rather than develop from scratch. Programming with application program interfaces (API) that allow newly written code to call middleware or existing application software has become a norm in software industries. In 2000, we developed a methodology to teach API concepts and skills in a one-term software engineering project course (CS3215). Since then the course evolved to full-year course, employing a wide range of teaching methods that we apply in small groups of students.

Another focus is on refining Software Engineering pedagogy tools and techniques for team-based SE courses. For example, we have started an open source project called TEAMMATES (http://teammatesonline.info) with the ambition to make it one of the biggest student-built projects in the world. TEAMMATES is used as a training environment for students to work in a non-trivial real-world software with a large user bases. In addition, TEAMMATES serves as an online peer evaluation/feedback tool for educators everywhere. As of April 2013, TEAMMATES user base spans over 30 universities and its implementation contains contributions from over 50 students.

Representative Publications

Stanislaw Jarzabek: Teaching Advanced Software Design in Team-Based Project Course, CSEET, 2013.

Damith Chatura Rajapakse: Some Observations from Releasing Student Projects to the Public, CSEET, 2011.

G. Goh, X. Lai and Damith Chatura Rajapakse: TEAMMATES: A Cloud-Based Peer Evaluation Tool for Student Team Projects, CSEET, 2011.

Software Reuse

A family of similar software products, managed from a common set of reusable components is called software Product Line (SPL). All products in SPL are similar to each other, but also differ in some customer-specific features. Similarities among products create reuse opportunities. But to reuse, the variability related to customer-specific features must be managed. A common SPL reuse practice is to design a component architecture to be shared by all the products, and then to apply variation mechanisms to parameterize components to make them reusable across many products we wish to build. Commonly used variation mechanisms (such as preprocessors, configuration files or design patterns) are not sufficient - they are not compatible with one another, difficult to use, do not scale well.

We have developed a uniform variation mechanism for reuse called XVCL (XML-based Variant Configuration Language, http:xvcl.comp.nus.edu.sg). XVCL' mechanisms can do what other ad hoc variation mechanism do and more. For example, in XVCL we can represent any group of similar program structures in a generic, adaptable form, suitable for reuse. We have built reuse solutions with XVCL in lab studies and industrial projects, developed tools supporting XVCL reuse method, integrated XVCL into Model-Driven Development to facilitate better use of software models in the reuse context. XVCL has been applied in industrial projects with promising results.

Representative Publications

Xue, Y., Xing, Z. and Jarzabek, S.: Feature Location in a Collection of Product Variants, WCRE, 2012.

Jarzabek, S. and Trung, H. D.: Flexible Generators for Software Reuse and Evolution, ICSE, 2011.

Jarzabek, S, Pettersson and U., Zhang, H.: University-Industry Collaboration Journey towards Product Lines, ICSR, 2011.

Zhu, J., Peng, X., Jarzabek, S., Xing, Z., Xue, Y. and Zhao, W.: Improving Product Line Architecture Design and Customization by Raising the Level of Variability Modeling, ICSR, 2011

Xue, Y, Jarzabek, S., Ye, P., Peng, X. and Zhao, W.: Scalability of Variability Management: An Example of Industrial Practice and Some Improvements, SEKE, 2011.

Ubiquitous Computing

The Felicitous Computing Institute began operation in the School of Computing in January 2012. The Institute is conducting a broad research program on next-generation ubiquitous computing systems from the perspectives of design, interaction, dependability and evaluation. One of the research projects carried out in the Institute is a context-aware mobile music recommendation system, which is described in the MEDIA section of this report. The Institute also is building on previous work in software engineering problems for ubiquitous computing.

A fundamental software engineering problem lies in testing and verification of the adaptation behavior of ubiquitous computing systems as such systems respond to changes in execution context. Adaptation typically is expressed in some rule-based form, with each rule expressing some context condition to be sensed plus an associated adaptive change to be made to the state of the application. While the rules themselves are conceptually simple, the adaptation behavior actually can be very fault-prone, for two reasons. First, interdependencies between rules accumulate very quickly, even for a small number of rules, leading to situations where there is a nondeterministic choice among rules for a particular sensed context value, or a dead rule superseded by higher priority rules, or cascading chains of rules that can be triggered for an arbitrary length of time by the same sensed context values. Such situations manifest themselves to users as anomalous and unexpected outputs or effects that in the best case are simply annoying and in the worst case possibly dangerous to the user. Second, the fact that context changes are sensed asynchronously with respect to application behavior, plus the fact that the sensors operate independently and at varying sampling rates, can cause transient instability in the adaptation behavior, further adding to the anomalous and unexpected effects experienced by users.

Previous research produced a suite of novel model checking algorithms for automatically identifying many classes of potential faults in rule-based, context-aware adaptive applications. For one simple but representative application involving 16 adaptation rules and nine different application states, the algorithms detected hundreds of distinct instances of the fault classes.

More recent work has investigated the treatment of unmanned aerial vehicle (UAV) missions and other forms of robot systems as context-aware ubiquitous computing systems, with a focus on the message-passing behavior of the systems. Certain message-passing scenarios can produce faulty behavior in system components. This research produced a novel approach for systematically analyzing message sequences and message frequencies that can lead to component failures, due to message queues filling up and messages being dropped.

Representative Publication

Charles Lucas, Sebastian G. Elbaum and David S. Rosenblum: Detecting problematic message sequences and frequencies in distributed systems, OOPSLA, 2012.

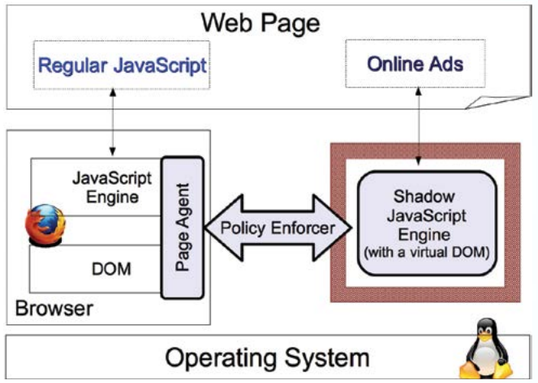
The faculty members in software engineering research are:

- Dong Jin Song
- Stanislaw Jarzabek
- Damith Chatura Rajapakse
- David Rosenblum
- Bimlesh Wadhwa

COMPUTER SYSTEM AND INFORMATION SECURITY

Cyber-security has been recognized as an increasingly critical area of computing by academia, industry and governments. Our research in this domain covers many aspects of security and privacy, ranging from foundational to practical issues, and classical to emerging topics. Research by our faculty members have had significant impact on areas like web security, mobile security, OS security, software vulnerability detection, biometrics, privacy protection and quantum cryptography. In the following pages, we highlight some of the recent research projects undertaken by our colleagues.

Web Security



The web platform has become a cornerstone of our computing infrastructure. Thanks to its convenience and scalability, the web platform has been attracting important financial services, personal data management applications and even online government services. Our research addresses security issues in web applications in two directions. First, we develop new security primitives in the web platform to enforce mandatory security properties. For example, JavaScript-based online advertisements pose threats to the embedding web applications as well as users operating systems. We develop solutions to provide strong and flexible confinement techniques to isolate JavaScript-based advertisements. Second, we develop several large-scale analysis tools to analyze web applications in an end-to-end manner, covering both client side and server side. For instance, we have developed infrastructure to analyze dynamic information flow properties in PHP and Microsoft .NET web applications. We have also developed analysis tools based on taint-tracking and symbolic execution on client-side JavaScript code. These techniques are generally applicable to program transformation, inferring isolation compartments and so on. We have been successful in scaling them up to real-world programs, which have enabled us to find a variety of web vulnerabilities such as DOM-based cross-site scripting and sanitization failures in web applications.

Representative Publications

AUTHSCAN: Automatic Extraction of Web Authentication .Protocols from Implementations, Guangdong Bai, Jike Lei, Guozhu Meng, Sai Sathyanarayan Venkatraman, Prateek Saxena, Jun Sun, Yang Liu, and Jin Song Dong, NDSS, 2013.

Adsentry: Comprehensive and flexible confinement of JavasSript-based advertisements, Xinshu Dong, Minh Tran, Zhenkai Liang, and Xuxian Jiang, ACSAC, 2011.

Towards fine-grained access control in JavaScript contexts, Kailas Patil, Xinshu Dong, Xiaolei Li, Zhenkai Liang, and Xuxian Jiang, ICDCS, 2011.

Library Isolation in Desktop Operating Systems

We rethink the security mechanisms for traditional operating systems. In such systems, e.g., Windows and UNIX, a library usually carries the same privilege as the program using it. A bug in the library can lead to compromise of the whole program. In addition, attackers can plan backdoors in the libraries. To ensure least privilege, we designed security mechanisms to isolate and confine libraries.

Authentication systems have long been threatened by malicious developers who can plant backdoors to bypass normal authentication, which is often seen in insider-related incidents. A malicious developer can plant backdoors by hiding logic in source code, by planting delicate vulnerabilities, or even by using weak cryptographic algorithms. Because of the common usage of cryptographic techniques and code protection in authentication modules, it is very difficult to detect and eliminate backdoors from login systems. We propose a framework to ensure that the authentication process is not affected by backdoors. The key component of our approach is NaPu, a native sandbox to ensure pure functions, which protects the complex and backdoor-prone part of a login module. We guarantee that the component secured by our approach has no backdoors that can be used practically.

Existing native sandbox solutions, such as Google NativeClient used by NaPu, require source code of the isolated libraries and assume limited memory interactions between the library and the program using it. To overcome such limitations, we develop a new library isolation mechanism Codejail, which supports isolation of libraries that have tight memory interactions with the main program.

Representative Publications

A Framework to Eliminate Backdoors from Response Computable Authentication, Shuaifu Dai, Tao Wei, Chao Zhang, Tielei Wang, Yu Ding, Zhenkai Liang, and Wei Zou. IEEE Symp. on Security and Privacy, 2012.

Codejail: Application-transparent Isolation of Libraries with Tight Program Interactions, Yongzheng Wu, Sai Sathyanarayan, Roland Yap, and Zhenkai Liang. ESORICS, 2012.

Operating System Security

Operating Systems such as Windows have a large attack surface. Malware make use of numerous features in the normal interaction and normal functionality in the operating system possibly combined with vulnerabilities to attack the system. One direction of the research is to understand and protect against vulnerabilities from binaries. Many of the works focus on Windows which is where most of the security problems arise.

We propose novel visualizations which can be used to understand the behavior of Windows and also the interactions between components and binaries in Windows. Many common attacks on Windows can be prevented by ensuring only trusted binaries are used, we propose a simple security model for binaries which is easy to manage while providing security.

Browser components, in particular, ActiveX controls in Internet Explorer can also be exploited. Such attacks are dangerous since they exploit features rather than vulnerabilities. We show how to detect such dangerous functionality in ActiveX controls and provide a mechanism to prevent their use.

Once a host machine has been infected with malware, it can be difficult to detect malicious behavior since the host can be compromised. We propose to make use of external mechanisms and sensors to detect compromised hosts. As such the external detector can work with the host but need not trust the host.

Representative Publications

Detecting and Preventing ActiveX API-Misuse Vulnerabilities in Internet Explorer, T. Dai, S. Sathyanarayan, R.H.C. Yap, Z. Liang, ICICS, 2012.

Enhancing host security using external environment sensors, E.-C. Chang, L. Lu, Y. Wu, R.H.C. Yap and J. Yu, Intl. J. of Information Security, 10(5), 2011.

Towards a binary integrity system for windows, Y. Wu and R.H.C. Yap, ASIACCS, 2011.

Analysis and Classification of Malware and Attacks

Security companies usually receive a large number of daily reports of malware and attacks. Efficiently analyzing such attacks is a challenging task. We develop techniques to automatically classify and diagnose malware and attacks.

A large family of remote exploits works by corrupting memory of the victim process to execute malicious code. Security companies and research labs usually collect a huge number of attacks samples daily, but they do not have enough manpower to analyze and classify them. The samples need to be prioritized so that new vulnerabilities and attack techniques can be discovered in a timely manner. To quickly respond to these attacks, it is critical to automatically diagnose such exploits to find out how they circumvent existing defence mechanisms. Because of the complexity of the victim programs and sophistication of recent exploits, existing analysis techniques fall short: they either miss important attack steps or report too much irrelevant information. We propose a novel solution, PointerScope, to use type inference on binary execution to detect the pointer misuses induced by an exploit. These pointer misuses highlight the important attack steps of the exploit, and therefore convey valuable information about the exploit mechanisms. Our approach complements dependency-based solutions to perform more comprehensive diagnosis of sophisticated memory exploits. PointerScope is efficient and effective on real-world exploit samples, e.g., identifying three key instructions from an attack trace containing 64 million instructions. Based on the graph representing key attack steps, PointerScope automatically classify attacks according to attack techniques, which can help security companies to prioritize the tasks of diagnosing attack samples.

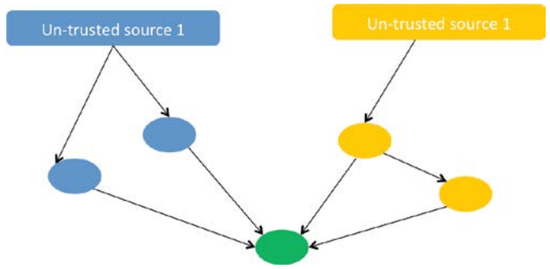
The classification of malware is not a well-defined problem. We investigate whether or not it is possible to classify malware using visualization. By using a form of dotplots, we found that it is possible to see similarities between the code and data of different malware. Such visualization is complementary to other visualization techniques.

Representative Publications

Identifying and Analyzing Pointer Misuses for Sophisticated Memory-corruption Exploit Diagnosis. Mingwei Zhang, Aravind Prakash, Xiaolei Li, Zhenkai Liang, and Heng Yin. NDSS, 2012.

Experiments with Malware Visualization,Y. Wu Yongzheng and R.H. C. Yap, DIMVA, 2012.

Symbolic Taint Analysis



Much of the functionality in our daily lives are software controlled and hence protecting our software against security vulnerabilities is of extreme importance. A common source of vulnerabilities comes from the input to the software, which may not be checked within the application. These vulnerabilities take different forms and names such as cross-site scripting, SQL injection and so on. In particular cross-site scripting allows attackers to pass unchecked input in the form of problematic scripts which may then get executed on the site of a non-malicious user. SQL injection refers to unchecked program input being used to construct database queries (which may then be exploited by an attacker to reveal confidential information such as user passwords). In this project, we develop and employ analysis methods for detecting the impact of program inputs on (parts of) an application. The main purpose is to detect or explain potential software attacks – thereby enhancing software security. One of the innovative outputs of the project will be to use software analysis and symbolic execution methods for generating and explaining potential attack scenarios, without actually encountering the attacks. Our infrastructure is geared to find out and summarize the input dependent parts of an application. It may suggest mechanisms to the programmer for making their applications more robust by inserting more checks at the appropriate places in their program. More interestingly, our analysis infrastructure can potentially reveal attack scenarios prior to the deployment of an application. This is achieved by a judicious mix of symbolic execution and program dependency analysis methods.

As part of our work, we observed that often a piece of software may come with a reference implementation. For example, web-servers implement the HTTP protocol - so for testing and analyzing a specific web-server --- we can use a well-known and well-tested web-server such as Apache as the reference implementation. Based on this observation, our project has developed several test generation, debugging and analysis methods, as evidenced by the relevant publications.

Representative Publications

Modeling Software Execution Environment, Dawei Qi, William Sumner, Feng Qin, Mai Zheng, Xiangyu Zhang, A Roychoudhury, WCRE 2012.

DARWIN: An Approach for Debugging Evolving Programs, Dawei Qi, Abhik Roychoudhury, Zhenkai Liang, Kapil Vaswani, ACM Transactions on Software Engineering and Methodology, 21(3), 2012.
Path Exploration based on Symbolic Output, Dawei Qi, Hoang D.T. Nguyen, Abhik Roychoudhury, SIGSOFT FSE, 2011.

Analysis and Test Generation for Evolving Software

Programmers do not write programs entirely from scratch. Over time, a program gradually evolves from one program version to another. However, as software evolves from one version to another - testing/debugging to ensure that the program continues to meet its intended functionality is difficult. Validation of such evolving programs (say, to address possible bugs introduced via program changes) remains a huge problem in software development. This adds to the cost for software maintenance, which is much larger than the initial software development cost. The cost of maintaining a software and managing its evolution is said to account for more than 90% of the total cost of a software project, prompting authors to call it the “legacy crisis”. In this project, we seek to develop testing and validation methods for evolving software to bring down such costs. The main innovation of the proposed work lies in making testing/debugging methods change-aware. A commonly used terminology in software engineering is “regression testing” where a changed program version is tested to check for possible regressions (some functionality which worked earlier but is broken due to the changes). In practice, regression testing often amounts to re-testing, or testing a selected sub-set of tests. Ideally, as changes are made in the program, the test-suite should also evolve with the program - an aspect that we study in this project. Akin to regression testing, we also study regression debugging, where we root-cause the reason for any detected software regression. The innovative aspect of our proposed work in regression debugging lies in using symbolic execution based semantic analysis of the failed test case as opposed to enumerating the changes.

Representative Publications

Partition-based Regression Verification, Marcel Böhme, Bruno C.d.S. Oliveira, Abhik Roychoudhury, ACM/IEEE ICSE 2013.

SemFix: Program Repair via Semantic Analysis, Hoang D.T. Nguyen, Dawei Qi, Abhik Roychoudhury, Satish Chandra, ACM/IEEE International Conference on Software Engineering (ICSE) 2013.

Software Change Contracts, Dawei Qi, Jooyong Yi, Abhik Roychoudhury, ACM SIGSOFT FSE 2012, NIER Track.

Modeling Software Execution Environment, Dawei Qi, William Sumner, Feng Qin, Mai Zheng, Xiangyu Zhang, Abhik Roychoudhury, IEEE WCRE, 2012.

Public Key Infrastructure

Public Key Infrastructure (PKI) is the critical infrastructure for securing communications and transactions over insecure public networks such as the Internet. It relies on certificates which may need to be revoked. There are a number of existing certificate revocation mechanisms but they either require increased network bandwidth or are not scalable. Furthermore, there is increasing need that revocation is more timely, possibly, being able to revoke a valid certificate in real-time. We propose some mechanisms which are more scalable with less resource requirements and deal with real-time. These mechanisms are suitable for mobile devices since less network bandwidth is required while being more scalable on the Certificate Authority (CA) computational resources.

Representative Publications

Trusted Principal-Hosted Certificate Revocation, Sufatrio and R.H.C. Yap, IFIPTM: IFIP WG 11.11 International Conference on Trust Management, 2011.

Quantifying the Effects of More Timely Certificate Revocation on Lightweight Mobile Devices, Sufatrio and R.H.C. Yap, MetriSec: Intl. Workshop on Security Measurements and Metrics, 2011.

Quantum Cryptography Made Real - The World’s First Implementation of Quantum Bit Commitment

In the past year, we have - among numerous other works - performed the world’s first experimental implementation of quantum bit commitment in the noisy storage model. This demonstrates that such protocols can in principle be performed with today’s quantum optics technologies. We have also introduced a novel concept to quantum cryptography called the quantum to classical randomness extractor which can be used to obtain secure randomness (key) in two-party protocols where it is impossible to estimate the min entropy of the raw key such as in quantum key distribution. We expect this concept to have applications in many areas of quantum cryptography. Finally, we have derived a fundamental result on the foundations of quantum mechanics. More precisely, we have shown that a violation of the uncertainty principle would allow for a violation of the second law of thermodynamics, providing us with a strong physical motivation of uncertainty relations.

Representative Publications

Experimental implementation of bit commitment in the noisy-storage mode, N. Ng, S. Joshi, C. Chia, C. Kurtsiefer and S. Wehner, Nature Communications 3, 1326, 2012.

Quantum to classical randomness extractors, M. Berta, O. Fawzi and S. Wehner, CRYPTO, 2012.

Secure Aggregation in Wireless Sensor Networks

In wireless sensor networks, obtaining aggregate information (e.g., average temperature sensed by all the sensors) is often of critical importance. Classic approaches (e.g., simple tree-based aggregation) for obtaining such aggregate information are extremely vulnerable to malicious sensors in the system -- even a single malicious sensor can completely corrupt the final aggregation result. The goal of secure aggregation is to thwart such attacks. The need for secure aggregation has been widely acknowledged by the research community. Nevertheless, due to the inherent difficulty, the problem is far from being completely solved.

Specifically, imagine that we deploy a sensor network for battlefield monitoring and the user wants to know where the enemy troops are moving. Prior approaches on secure aggregation for general sensor networks can only detect whether the aggregation result is corrupted by an attacker. On the other hand, the attacker can keep corrupting the result, so that the system can never generate a correct result. Clearly, we need to tolerate the attacker in order to make the system useful.

Within such context, we have designed two novel and complementary algorithms that can tolerate the attacker. The first tree-sampling algorithm is based on a novel concept of set sampling, which can provably generate a correct aggregation result despite adversarial interference. Our second VMAT algorithm incurs smaller overhead than tree sampling when there is no attacker. When under attack, the VMAT algorithm can pinpoint the attacker and revoke some capability held by attacker. Clearly, after revoking all the capabilities of the attacker, the algorithm will be able to produce a correct answer.

Representative Publications

Secure Aggregation with Malicious Node Revocation in Sensor Networks, Binbin Chen and Haifeng Yu, ICDCS, 2011.

Secure and Highly-Available Aggregation Queries in Large-Scale Sensor Networks via Set Sampling, Haifeng Yu, Distributed Computing, Volume 23, Issue 5, pp. 373-394, April 2011.

Defending Against Rational Sybil Attacks in Overlay Multicast Systems

Overlay multicast systems are representative of many today’s p2p systems, where users need to contributed to the system (e.g., by uploading data to other users). With sybil attacks, one malicious user pretends to have many fake identities. Sybil defense in overlay multicast systems are tricky because individual users are often selfish, and may not want to spend their bandwidth to participate in whatever defense protocol we design. This quickly leads to the tragedy of the commons, which is a key and well-known challenge in designing these systems. In such context, we have proposed a novel and practical DCast protocol that can properly defense against sybil attacks. This is the very first practical protocol that can deal with sybil attacks in this setting. DCast achieves its strong functionality by carefully addressing a number of key challenges. For example, one key challenge addressed is the design of an effective punishment mechanism to punish those users who fail to contribute to the system, despite the presence of collusion among the sybil identities.

Representative Publication

DCast: Sustaining Collaboration in Overlay Multicast despite Rational Collusion, Haifeng Yu, Phillip B. Gibbons, and Chenwei Shi, PACM CCS, 2012.

Data Privacy

Organizations such as ministries or hospitals regularly release microdata (e.g., census data or medical records) for purposes that serve the common good through the advancement of knowledge. Still, the same data can inadvertently reveal sensitive personal information to malicious adversaries – even anonymous personal records can be re-identified using third-party data sources. Our research focuses on a variety of different data types, including traditional structured data, streaming data, transaction data and set-valued data. We have studied various concepts of privacy guarantee that an anonymized data set should satisfy before publication (e.g., k-anonymity, t-closeness) and developed new privacy notions such as r-uncertainty for transaction data and ks-anonymity and ls-diversity for streaming data. We have also developed comprehensive and systematic schemes that offer insights into the tradeoffs between privacy and information loss.

Representative Publications

Anonymizing Set-Valued Data by Nonreciprocal Recoding, M. Xue, P. Karras, C. Raissi, J. Vaidya, K.L. Tan, ACM SIGKDD, 2012.

Adaptive Differentially Private Histogram of Low-Dimensional Data, C.F. Fang and E.C. Chang, PETS, 2012.

CASTLE: Continuously Anonymizing Data Streams, J. Cao, B. Carminati, E. Ferrari, K.L. Tan, IEEE Transactions on Dependable and Secure Computing, IEEE CS, 2011.

SABRE: A Sensitive Attribute Bucketization and REdistribution Framework for t-closeness, J. Cao, P. Karras, P. Kalnis, K.L. Tan, The VLDB Journal, Springer, Vol. 20, No. 1, pp. 59-82.

Privacy in Social Networks

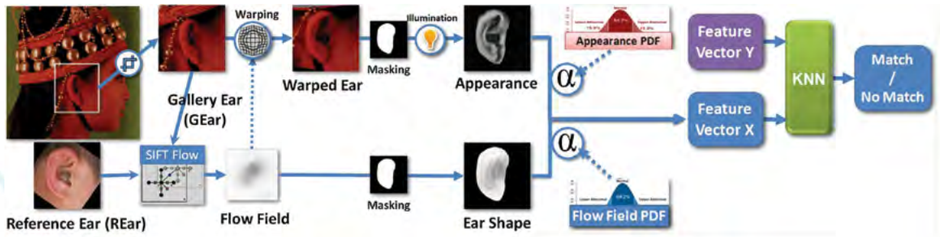
Social networks are quite ubiquitous nowadays. In many cases, social networks can reveal a lot of information about the users. This leads to the question whether are social networks too public? We investigate privacy policies which balance between two tradeoffs, utility of publically available information from the perspective of the network as a whole or the network provider, versus how much privacy is available to the users. It turns out that there is a range of policies which either make more links or relationships private or making the profile of a user private. By changing the edges or profiles which are made private, one can balance between more privacy or more utility. It turns out that making links private is more effective in this balance than making nodes private. Some social networks are intrinsically private in that information is only made available to users and their uses in the neighbourhood. One attack is the link privacy attack which reveals private graph data either by compromising a user or obtaining the information by bribing the user. We show that the privacy leak from the link privacy attack is significant. By only attacking a small fraction of the nodes, it is possible to obtain information about a large part of the graph. We explain why it is difficult to defend against the attack and give some strategies to limit the privacy leak.

Representative Publications

Revisiting Link Privacy in Social Networks, S. Effendy, R.H.C. Yap and F. Halim, ACM CODASPY, 2012.

Partial Social Network Disclosure and Crawlers, S. Effendy, F. Halim and R.H.C. Yap, SCA, 2011.

The Biometrics of Identical Twins



Identical Twins pose a significant challenge to existing biometric systems. For example, face recognition algorithms currently cannot distinguish between such twins. And while it is known that Identical Twins have identical DNA, it is not known if their other biometrics, such as voice patterns, are similar as well. We are exploring this problem by analyzing a set of different biometrics from about 40 pairs of identical twins, including speech recordings, facial and head movements, and ear images. Our hypothesis is that even though twins may look alike, they will behave differently. In other words, their behavioral biometrics are different. Our investigations into speech patterns, facial expressions, and head gestures support this hypothesis. Moreover, these biometrics are discriminative for the general, non-twin population as well. Interestingly, we also found that some physical biometrics, such as ear lobes, may be used to distinguish between twins. In light of our findings, we have proposed a new way to augment existing face recognition systems so that they are robust against imposter attacks by identical twins.

Representative Publications

A Talking Profile to Distinguish Identical Twins, L. Zhang, N. Hossein, C. L. Foo, K. T. Ma, T. Sim and D. Guo. IEEE Automatic Face and Gesture Conference, 2013.

Wonder Ears: Identification of Identical Twins from Ear Images, N. Hossein, L. Zhang, T. Sim, E. M. Martinez and D. Guo. IAPR International Conference on Pattern Recognition, 2012.

The faculty members of the Security Group are:

- Chang Ee-Chien
- Liang Zhenkai
- Rahul Jain
- Abhik Roychoudhury
- Prateek Saxena
- Sim Mong Cheng, Terence
- Tan Kian Lee
- Stephanie Wehner
- Yap Hock Chuan, Roland
- Yu Haifeng

SYSTEMS AND NETWORKING

The Systems and Networking group focuses on the architecting and understanding of complex computer systems. The research primarily is on the analysis, architecture and algorithms for complex embedded, parallel and distributed systems, often employing the next-generation technologies. The group has an excellent track record in funding, publication and impact. Below is a summary of some of the latest projects by the group members:

Timing Analysis of Embedded Software

Real-time embedded software are ubiquitious and control the functionality of many devices used in our everyday lives. The focus of the proposed project will be on compositional timing analysis methods which efficiently derive Worst Case Execution Time (WCET) estimates of large real-life embedded software. WCET analysis usually involves program path analysis as well as micro-architectural analysis. We aim for achieving compositionality both at the program level, as well in the construction of timing models for the underlying micro-architecture. At the program level, we generate timing summaries of the individual program fragments such as functions / procedures. Consequently, if a library or a small part of a program is changed – we do not need to re-do the entire analysis. At the micro-architecture level, our focus is on building timing models to capture the timing effects of multi-cores. Processor manufacturers have increasingly moved towards multi-cores, and this has impacted the embedded processor market as well. However, when we build a software timing analyzer for multi-core platforms, we want the analyzer to be configurable with respect to small changes, such as changes in the number of cores. Such configurability may be achieved by compositional analysis of the timing effects of the individual cores, that is, we want to analyze the timing effects of each processor core individually and compose them to get the timing effects of a multi-core platform. This removes the need to re-do the complete analysis in case of small changes in the platform. The main deliverable of the project is a scalable WCET analysis tool that can generate tight execution time bounds for software (possibly running via a real-time operating system RTOS) on multi-core platforms.

Representative Publications

Precise Micro-architectural modeling for WCET analysis via AI+SAT. Abhijeet Banerjee, Sudipta Chattopadhyay, Abhik Roychoudhury. IEEE RTAS, 2013.

A Unified WCET Analysis Framework for Multi-core Platforms. Sudipta Chattopadhyay, Chong Lee Kee, Abhik Roychoudhury, Timon Kelter, Peter Marwedel and Heiko Falk, IEEE RTAS, 2012.

Performance Debugging of Esterel Specifications. Lei Ju, Bach Khoa Huynh, Abhik Roychoudhury, Samarjit Chakraborty. Real-time Systems Journal, 48(5), 2012.

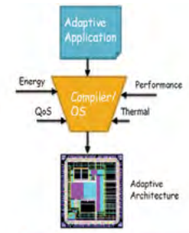
Timing Analysis of Concurrent Programs running on Shared Cache Multi-cores. Yun Liang, Huping Ding, Tulika Mitra, Abhik Roychoudhury, Yan Li, Vivy Suhendra. Real-time Systems Journal, 48(6), 2012.

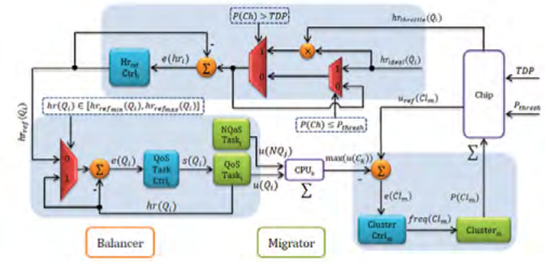
Scalable and Precise Refinement of Cache Timing Analysis via Model Checking. Sudipta Chattopadhyay and Abhik Roychoudhury. IEEE RTSS, 2011.

Scope-aware Data Cache Analysis for WCET Estimation. Bach Khoa Huynh, Lei Ju and Abhik Roychoudhury. IEEE RTAS, 2011.

Energy-Efficient Computing

This project aims to develop a framework, called Monsoon, for energy and thermal efficient computing in high-performance embedded systems. The key concept in our approach is to leverage on the inherent characteristics of many modern applications (e.g., media processing, recognition and data mining, machine learning) to tolerate imprecise and approximate solutions. We exploit this approximate nature of contemporary and future applications through context-dependent synergistic adaptation of the both the application and the architecture to achieve energy savings and improve thermal profile while meeting the quality of service (QoS) guarantees.





Representative Publication

Hierarchical Power Management for Asymmetric Multi-Core in Dark Silicon Era. Thannirmalai Somu Muthukaruppan, Mihai Pricopi, V Vanchinathan, Tulika Mitra, Sanjay Vishin, ACM/IEEE DAC, 2013.

Polymorphic Heterogeneous Multi-Core for the Dark Silicon Era

Computing systems have made an irreversible transition towards parallel architectures with multi-cores and many cores. However, power and thermal limits are rapidly bringing the computing community to another crossroad where a chip can have many cores but a significant fraction of them are left unpowered --- or dark --- at any point in time. This phenomenon, known as dark silicon, is the driving force behind our polymorphic heterogeneous multi-core architecture called Bahurupi. The architecture is fabricated as a homogeneous multi-core system consisting of multiple identical (simple) cores and some amount of reconfigurable fabric on chip. The main novelty of Bahurupi lies in its highly flexible architecture. Post-fabrication, software can compose together primitive cores as well as exploit the re-configurable fabric for hardware accelerators to create a customized multi-core system consisting of cores with diverse functionality-power-performance characteristics enabling better match between application requirements and computation capabilities. This software-directed re-configurability allows us to enjoy substantially improved energy-efficiency without paying the hefty price of design and fabrication for customization.

Representative Publications

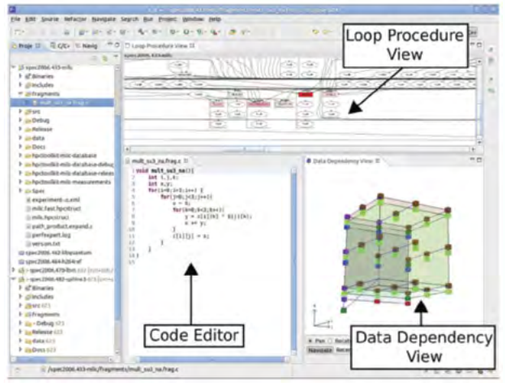
Bahurupi: A Polymorphic Heterogeneous Multi-core Architecture. Mihai Pricopi, Tulika Mitra. ACM Transactions on Architecture and Code Optimization, 8(4), January 2012

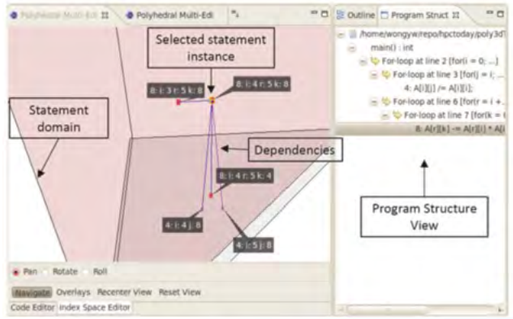
Graph Minor Approach for Application Mapping on CGRAs. Liang Chen, Tulika Mitra. Int. Conf. on FPT, 2012.

Shared Reconfigurable Fabric for Multi-Core Customization. Liang Chen, Tulika Mitra. ACM/IEEE DAC, 2011.

High Performance Computing with Heterogeneous Platforms

In this project, we are developing tools and algorithms to enable the running of legacy applications on modern heterogeneous platforms. In particular, we have focused on the GPU. Unlike previous approaches that tried to be fully automatic, we have built our tools and algorithms with the assumption that the human is in the loop. Thus one of our goals is to assist the human software developer in understanding the massive amount of information that comes with program analysis and application characterization. To this end, we have developed Tulipse, a plugin to the Eclipse Integrated Software Development Environment, the displays key analysis information of a program, allowing the developer to zoom and lock in to areas in the code that are bottlenecks and/or most promising for parallelization. Around Tulipse, we are developing a slew of techniques for utilizing highly tuned libraries that were developed by third parties or vendors. We have also developed highly optimized libraries of our own for stencil and sparse matrix computation.





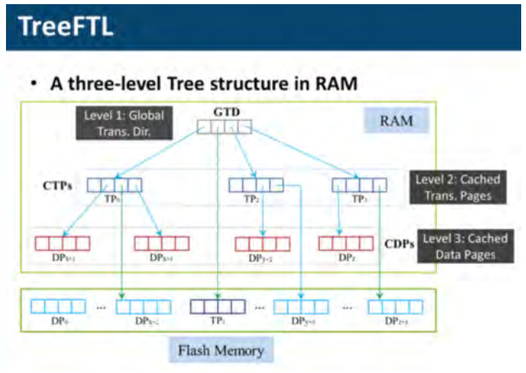
Representative Publications

W.T. Tang, W.J. Tan, R. Krishnamoorthy, Y.W. Wong, S-H. Kuo, R.S.M. Goh, S.J. Turner, and W.F. Wong, “Optimizing and Auto-Tuning Iterative Stencil Loops for GPUs with the In-Plane Method”. IEEE IPDPS, 2013.

W.T. Tang, Y. Wong, W.J. Tan, T. Dubrownik, R.S.M. Goh, S. Kuo, R. Duan, S. Turner, and W.F. Wong. “Tulipse: A Visualization Framework for User-Guided Parallelization”. EUROPAR, 2012.

Memory Hierarchy for the Next-Generation Technologies

We are at an important cusp in technology when several key memory technologies that are non-volatile in nature are near their commercialization phase. These include the Spin-Transfer-Torque RAM, Phase Change Memory, Memristors, and the most mature of them all Flash memory. While Flash is now ubiquitous, the former presents a new feature for future computer systems to exploit – namely, non-volatility. In addition, these memory technologies have their own idiosyncrasies that need to be handled at the system level. In this project, we have been working how to incorporate these technologies into all levels of the memory hierarchy, from level 1 processor caches, last level caches, main memory, down to solid state disks. We have new hardware designs and algorithms that provide better efficiency as well as workarounds for the problems that these memory technologies face.



Representative Publications

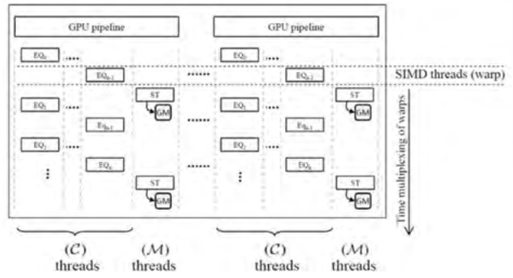
C. Wang, and W.F. Wong, “Observational Wear Leveling: An Efficient Algorithm for Flash Memory Management”. DAC, 2012.

C. Wang, and W.F. Wong, “ADAPT: Efficient Workload-sensitive Flash Management Based on Adaptation, Prediction and Aggregation”. IEEE MSST, 2012.

Z. Sun, X. Bi, H. Li, W.F. Wong, Z.L. Ong, X. Zhu, and Wenqing Wu, “Multi-Retention Level STT-RAM Cache Designs with a Dynamic Refresh Scheme.” IEEE/ACM MICRO, 2011.

Using GPUs for Computational Systems Biology

Biopathways are often modeled as systems of ordinary differential equations (ODEs). Such systems will usually have many unknown parameters and hence will be difficult to calibrate. Since the data available for calibration will have limited precision, an approximate representation of the ODEs dynamics should suffice. One must, however, be able to efficiently construct such approximations for large models and perform model calibration and subsequent analysis. In this project, we innovated a graphical processing unit (GPU) based scheme by which a system of ODEs is approximated as a dynamic Bayesian network (DBN). A model checking procedure for DBNs based on a simple probabilistic linear time temporal logic was then constructed. The GPU implementation considerably extends the reach of a previous PC-cluster-based implementation. Further, the key components of our algorithm can serve as the GPU kernel for other Monte Carlo simulations-based analysis of biopathway dynamics. Similarly, our model checking framework is a generic one and can be applied in other systems biology settings.



Strategy for our GPU implementation of probabilistic biopathway dynamics simulation

Representative Publications

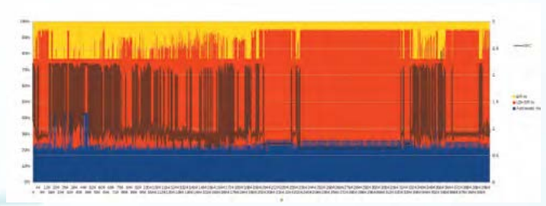
B. Liu, A. Hagiescu, S. Palaniappan, B. Chattopadhyay, Z. Cui, W.F. Wong, and P.S. Thiagarajan, “Approximate Probabilistic Analysis of Biopathway Dynamics”. Bioinformatics. Vol. 28, No. 11. pp. 1508-1516. 2012.

Online Multithreaded Program Behavior Monitoring through Dynamic Instrumentation and Hardware Counter

Dynamic instrumentation allows us to monitor any binary under execution. It provides a promising means of understanding an application. Through understand an application’s behaviours such as its cache and branch behaviours, program developers as well as compiler writings can then better optimize their code.

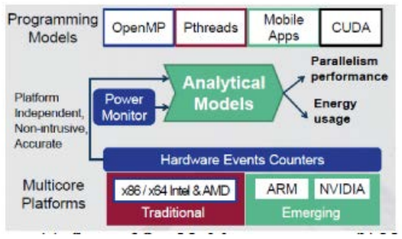
The project was started in August 2012. We now have a working tool built on a binary rewriting system called DynamoRio that works with the Performance API (PAPI) interface. Using this tool, we have collected the execution data on a number of SPEC 2006 benchmarks. Currently, we are looking at that data analysis in a summative (report and “explain” program behavior), as well as a predictive (categorize program and predict performance).

Below is a graph plot on the information collected for the 453.PovRay SPEC benchmark. The instructions composition and Instruction-Per-Cycle (IPC) are plotted on the same graph.

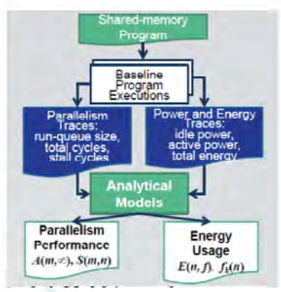


Parallelism and Energy Performance of Multicore Systems

Multicore systems are increasingly adopted across many application domains. Consequently, understanding their performance is becoming an important issue for a growing number of users. However, performance analysis of parallel programs on multicore systems is still challenging, especially for large programs or applications developed different programming languages. We have developed an analytical modeling approach for studying the parallelism and energy performance of shared-memory programs on multicore systems. The proposed model derives the parallelism and parallelism loss from data dependency and memory overhead in traditional UMA and NUMA multicore systems, and emerging platforms such as ARM multicores. Using only widely available inputs derived from the trace of the operating system run-queue and hardware events counters, the proposed model achieves high practicality and generality across many types of shared-memory programs running on different multicore platforms. Applications of the model include understanding achieved speedup and parallelism loss, and prediction of optimal core and memory configuration, where the optimality criteria is minimum execution time, minimum energy usage or a tradeoff between these two.



(a) Scope of Our Model



(b) Measurement-Analytic Model Approach

Representative Publications

B.M. Tudor and Y.M. Teo, On Understanding the Energy Consumption of ARM-based Multicore Servers, ACM SIGMETRICS, 2013.

B.M. Tudor and Y.M. Teo, Towards Modeling Parallelism and Energy Performance of Multicore Systems, IEEE IPDPS, 2012.

B.M. Tudor, Y.M. Teo and S. See, Understanding Off-chip Contention of Parallel Programs in Chip Multiprocessors, Int. Conf. on Parallel Processing, 2011.

B.M. Tudor and Y.M. Teo, A Practical Approach for Performance Analysis of Shared Memory Programs, IEEE IPDPS, 2011.

Augmented Stability in Unmanned Aerial Vehicles

Unmanned aerial vehicles, particularly those based on commercial-off-the-shelf (COTS) platforms meant originally for radio-control use, are very light, weighing in at between just under 1 kg to 10 kg. As such they are particularly susceptible to cross-winds, especially at the common flight altitude of 300 feet in an open space.The cross-winds make it difficult for a UAV to accurately follow a flight path. This can affect swarms of UAVs flying in formation at the same altitude as a cross-wind can push a lighter UAV into the path of a heavier one, endangering both aircraft.

In this project a specialized form of radial basis function neural networks called “Dynamic Cell Structure” or DCS networks adapt to varying wind conditions and supply corrective roll commands to the UAV’s autopilots to help the UAV maintain an accurate flight path. As the corrective commands are based on deviations from the intended flight path rather than from actual cross-winds, there is no need to measure wind conditions affecting the aircraft, which can be difficult as measurements will tend to be affected by the geometry of the airframe and its forward movement.

Experiments show an improvement in tracking accuracy of between 20% and 60%. Experiments were conducted using a Multiplex Mentor platform with brushless motors weighing 1 kg, Ardupilot 1.0 autopilot module with IMU sensor shield, and GS407 Helical U-Blox GPS receiver and antenna assembly unit. Figures 2 and 3 show improvements in tracking accuracy on test flights near the Raffles Marina, Tuas, Singapore.

Figure 1. Multiplex Mentor platform

Figure 2. DCS tracking (yellow) and stock autopilot tracking (blue) on racetrack flight path.

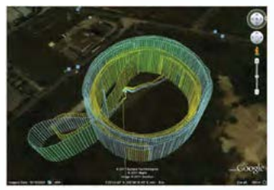


Figure 3. DCS tracking (yellow) and stock autopilot tracking (blue) on circular flight path.

Representative Publication

Achudhan Sivakumar, UAV Swarm Coordination and Control Under Realistic Weather and Network Conditions for Establishing A Wireless Communication Backbone, Dissertation submitted in partial fulfilment for the degree of Doctor of Philosophy, NUS, 2011.

Sensor Networks

Advances in integrated circuit technology, micro electro mechanical systems (MEMS) and wireless technology have enabled construction of low power and low cost devices that support sensing, computation and communication capabilities. While the capabilities of these devices are limited, by utilizing a large number of these sensing devices, it is possible to create a wireless sensor network (WSN) that allows the deployment of new and exciting types of applications in a wide range of areas including surveillance, transportation, tracking, health-care and wildlife monitoring.

We have looked at a wide spectrum of problems covering many aspects of sensor network algorithms. A key requirement of sensor network is resource efficiency, in particular, energy efficiency. We look at algorithms that are related to and based on network connectivity, including algorithms for efficient neighbor table management, connectivity monitoring, coverage control, routing and hole detection.

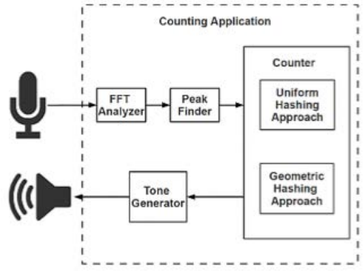
There is a strong emphasis on system implementation in this work. Most of the above mentioned protocols have been implemented in TinyOS and evaluated on testbeds consisting of MICA2, MICAz and/or TelsoB motes. We have gone through 3 generations of sensor testbeds. Currently, we have deployed INDRIYA, a 140 nodes sensor testbed. Besides being one of the largest testbeds in academic institutions, sensors in INDRIYA are also deployed over multiple floors, thus providing a realistic 3D sensor platform. More information of INDRIYA can be found at http://indriya.ddns.comp.nus.edu.sg.

Highlight: Low Cost Crowd Counting using Audio Tones

We have developed a crowd counting solution based on audio tones, leveraging the speakerphones that are commonly available in most phones. To the best of our knowledge, this is the first use of audio tones as a networking mechanism for multi-hop, large-scale counting. We have implemented the tone-counting system in Android based smartphones.

By using audio tones which are barely audible to humans, and not requiring any user inputs, our solution has minimal intrusion to the mobile user experience. It also preserves anonymity, privacy, and security, since only (randomly generated) tones are sent by the phones with no additional exchange of link or MAC layer information.

The technique could be used for fast, low-cost estimations of the number of passengers who board a bus or subway can greatly aid public transport planning. Crowd counts localized to specific subway cabins can be leveraged for crowd control. Counts of taxi queues can guide better deployment of taxis towards high demand areas. A counting solution can also be used in conducting surveys in real time. An effective, low-cost solution like ours can substantially expand the reach of transportation surveys, which in the past; can only be done rarely to mitigate costs.



The maximum number of nodes that can be counted increases exponentially with the number of frequencies available for use. Our current implementation, with 98 usable frequencies for counting, supports more than 800 devices. With the use of multi-hop communication, the devices can cover a much larger geographical area than the range of an audio transmission. Assuming an audio transmission range of 5m, 800 or more devices forming a 28x28 grid pattern can cover an area of about the size of 3.5 football fields. We have experimentally demonstrated counting in a network of up to 7 hops.

We have conducted two experiments counting the crowd at a bus stop as well as aboard a moving bus. The technique developed is not limited for use in transportation but also in event planning, visitor survey and proximity marketing etc.

Representative Publications

M Doddavenkatappa, MC Chan and Ben Leong, “Splash: Fast Data Dissemination with Constructive Interference in Wireless Sensor Networks,” USENIX NSDI, 2013.

PG Kannan, SP Venkatagiri, MC Chan, AL Ananda, LS Peh, “Low-Cost Crowd Counting using Audio Tones,” ACM Sensys, 2012.

M Doddavenkatappa, MC Chan and AL Ananda, “Indriya: A Low-Cost, 3D Wireless SensorNetwork Testbed”,TRIDENTCOM,2011.

M Doddavenkatappa, MC Chan and Ben Leong, “Improving Link Quality by Exploiting Channel Diversity in Wireless Sensor Networks,” RTSS 2011.

Heterogeneous Delay Tolerant Networks

In DTN, due to the sparseness of the network, mobile and/or static nodes can only communicate when they are in communication range and these contacts are intermittent and unpredictable. We have studied various aspects of DTN, including routing protocol efficiency, routing protocol robustness and incentive mechanism.

Highlight: Taxi Advisory System (TADS)
We have developed a distributed algorithm for monitoring and providing guidance to taxis. Typically, map is partitioned into many regions and the supply/demand for each region is monitored in a distributed manner. As our taxi communication is operating over a DTN, we propose techniques to minimize over/under counting of supply/demand. We performed evaluation of our proposed system based on traces obtained from the given taxi traces with more than 15,000 taxis. Our simulation results show that even when only 10% of the clients participated in using TADS, TADS can still reduce the long client waiting time noticeably. With 50% of the clients participated in using TADS, TADS can reduce the number of clients with wait times longer than 60 minutes by 27%.

Representative Publications

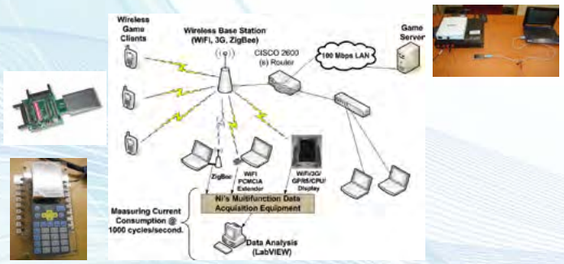
FC Choo, MC Chan, AL Ananda, LS Peh, “A Distributed Taxi Advisory System”, ITST, 2012

H Wang, L-S Peh, E Koukoumidis, S Tao and MC Chan, “Meteor Shower: A Reliable Stream Processing System for Commodity Data Centers,” IEEE IPDPS, 2012

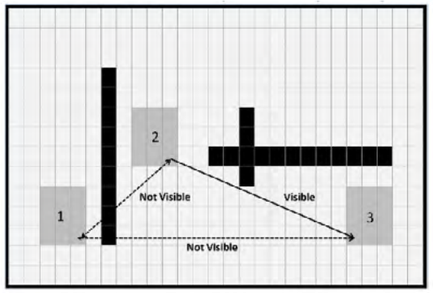
FC Choo, PV Seshadri, MC Chan, “Application-Aware Disruption Tolerant Network,” IEEE MASS, 2012

Energy Efficient Network Protocols

Multiplayer mobile games are an increasingly important class of mobile application. Supporting multiplayer games on mobile clients (e.g. PDAs and mobile phones) that communicate over a wireless network poses a number of interesting technical problems including power saving mechanism and light weight energy aware network protocols. While device features and application quality are rapidly growing, the battery technologies are not growing at the same pace. Battery lifetime is one of the key factors that hinder the usability of the mobile devices for resource-intensive applications. In this work, we design a framework for power management that adapts its behavior to the intent of the user and the game, the characteristics of the network and the network interface. Our system is being designed to reduce the overall device power usage without sacrificing the end-user game experience.



Highlight: Dynamic Lookahead Mechanism for Conserving Power in Multi-Player Mobile Games
We have developed a technique called dynamic AoV lookahead for reducing wireless interface power consumption upto 50% while playing a popular, yet resource intensive, mobile multiplayer games and demonstrated the same to show that the quality is not compromised. The work is continuation of the work reported in the previous year and published in Mobisys 2011 under the title, “Adaptive display power management for mobile games”.



We have already shown that how tone mapping techniques can be used to dynamically increase the image brightness, thus allowing the LCD backlight levels to be reduced in display systems particularly in smartphones. This saves significant power as the majority of the LCD’s display power is consumed by its backlight. The Gamma function (or equivalent) can be efficiently implemented in smartphones with minimal resource cost. We describe how we overcame the Gamma function’s non-linear nature by using adaptive thresholds to apply different Gamma values to images with differing brightness levels. These adaptive thresholds allow us to save significant amounts of power while preserving the image quality. We implemented our solution on laptops and Android smartphones. We have also conducted user study to show that we can save up to 68% of the display power without significantly affecting the perceived gameplay quality. Finally, we are in the process of integrating the various power savings mechanisms for display, network and processing into a single integrated framework.

Representative Publications

B Anand, LK Chong, EC Chang, MC Chan, AL Ananda and WT Ooi,“El-pincel: A Painter Cloud Service for Greener Web Pages,” ACM Multimedia, 2012.

K Thirugnanam, B Anand,, J Sebastian, PG Kannan, AL Ananda, RK Balan, and MC Chan, “Dynamic Lookahead Mechanism for Conserving Power in Multi-Player Mobile Games,” IEEE INFOCOM 2012.

B Anand, AL Ananda, MC Chan and RK Balan, “ARIVU: Making Networked Mobile Games Green - A Scalable Power-Aware Middleware”, Mobile Networks and Applications, Feb 2012

B Anand, K Thirugnanam, J Sebastian, PG Kannan, AL Ananda, MC Chan, RK Balan, “Adaptive Display Power Management for Mobile Games,” ACM Mobisys, 2011.

Improving Performance for 3G/4G Mobile Data Networks

Mobile cellular data networks are an important medium for wireless connectivity on-the-go. We are currently investigating the performance of local 3G/4G mobile cellular networks in order to identify and address potential performance issues.
We conducted a comprehensive measurement study involving all three of our local ISPs where we found that the performance of downloads in a 3G/HSPA mobile network can be significantly degraded by a concurrent upload that saturates the uplink buffer on the mobile device. To address this problem, we have developed two algorithms: Receiver-side Flow Control (RSFC) and TCP Receiver-Rate Estimation (TCP-RRE). These two algorithms address this problem at the receivers and senders of data respectively and can improve download speeds in the presence of a concurrent upload without requiring any modification to existing mobile devices.

We also developed and deployed on the Android AppStore a mobile application that can allow users to measure the performance of 3G/4G mobile networks called ISP Check (https://play.google.com/store/apps/details?id=com.ispcheck). Because data can be rather expensive, we are currently investigating techniques to allow us to measure mobile cellular networks accurately with minimum data transfer, with a long-term goal of developing a crowd-sourcing-based technique to measure network performance and diagnose network faults.

Representative Publication

Yin Xu, Wai Kay Leong, Ben Leong and Ali Razeen. “Dynamic Regulation of Mobile 3G/HSPA Uplink Buffer with Receiver-Side Flow Control”. IEEE ICNP, 2012.

Modelling and Understanding Peer-to-Peer File Sharing Algorithms

As the amount of data and the size of media files distributed over the Internet continue to grow, file distribution will continue to be a predominant activity on the Internet. Server-client systems are inherently less scalable and peer-assisted file distribution systems are used to support file distribution.
We developed performance models for peer-assisted file distribution systems. Based on a measurement study of BitTorrent on PlanetLab, we made two key observations: (i) there is a fixed pattern in the utilization of the available bandwidth over the course of a download, and (ii) peers enjoy an amount of service that is commensurate with their contribution. Building on these insights, our model can estimate the download time for each class of peers in a well-provisioned peer-assisted file distribution system based on BitTorrent. Our model accurately predicts the download time and achieves an average error rate of 16.5% for heterogeneous swarms up to 150 nodes in size and can be used to estimate the server capacity for achieving a specific quality of service in a large heterogeneous swarm.

We also investigated a linear optimization approach for studying download finish times in peer-to-peer networks that allow but do not require coding. We demonstrate that using the network coding framework simplifies analysis even in scenarios where the optimal solution does not require coding. We use this framework to study the effect of requiring reciprocity, a typical feature of incentive-compatible protocols and show that for a dynamically changing network scenario, coding can provide a robust and optimal solution that outperforms routing.

Representative Publication

Cristina Carbunaru, Yong Meng Teo, and Ben Leong. “A Performance Study of Peer-assisted File Distribution with Heterogeneous Swarms”. IEEE LCN, 2011.

Effectively Exploiting Multiple Paths on the Internet

It is well known that there are many possible paths between two end-points on the Internet. It was however not clear if it was possible to use additional detour paths to improve the end-to-end throughput between a pair of source and destination nodes. We have developed a new massively-multipath (mPath) source routing algorithm that can improve end-to-end throughput for high-volume data transfers. We demonstrate that our algorithm is practical by implementing a system that employs a set of proxies to establish one-hop detour paths between the source and destination nodes.
Our algorithm can fully utilize the available access link bandwidth when good proxied paths are available, without sacrificing TCP friendliness, and achieves throughput comparable to TCP when such paths cannot be found. For 40% of our test cases on PlanetLab, mPath achieved significant improvements in throughput. Among these, 50% achieved a throughput of more than twice that of TCP.

Research on Fault-Tolerant Communication Complexity

Communication complexity is a vital subarea of theoretical computer science. It is concerned with the minimum amount of communication needed, to compute a certain function whose inputs are separately held by distributed players connected by a certain topology. (The simplistic case would be two players connected by a single edge.) By definition, this involves distributed computation. On the other hand, a key focus of distributed computing research, since the very beginning, has been to tolerate failures. It is thus natural to ask “If we want to compute a certain function in a fault-tolerant way (i.e., while tolerating player crash failures), what will the communication complexity be?” Let us call such communication complexity as fault-tolerant communication complexity.

Whether fault-tolerant communication complexity is interesting to study largely depends on how big a difference that failures make. In this work, we prove that the impact of failures is significant, at least for the SUM function: As our central result, we prove the existence of an exponential gap between the non-fault-tolerant and fault-tolerant communication complexity of SUM. Since SUM can be reduced to/from many other functions, the gap carries over to those functions too. We also show that for some other functions such as the MAX function, there is no large (asymptotic) gap. Our exponential gap result attests that fault-tolerant communication complexity needs to be studied separately from all the existing research on “non-fault-tolerant” communication complexity.

This work has resulted in the following publication:
Binbin Chen, Haifeng Yu, Yuda Zhao, and Phillip B. Gibbons, “The Cost of Fault Tolerance in Multi-Party Communication Complexity.” ACM PODC, 2012.

Research on Error Estimating Coding

Error correcting codes have long been playing fundamental roles in communication systems. Behind decades of research on error correcting codes, the implicit philosophy has always been correcting the errors as soon as the errors are first detected. Doing so provides a clean abstraction, so that the upper layer will only see fully correct data. In recent years, however, many emerging designs in wireless networking start to leverage packets with errors. These packets are processed and relayed (by the routers) inside the network, and the errors in them are often corrected only at the final destination and at a later time. These emerging designs invalidate the previous philosophy, and lead to the following natural question: Should we start looking beyond error correcting codes?

We introduced the novel concept of Error Estimating Coding (EEC). EEC enables the receiver of a packet to estimate the number of errors in the packet, without correcting the errors. Compared to error correcting codes, EEC offers weaker functionality while incurring substantially smaller overhead. We show via real-world experiments that EEC significantly benefits many modern designs in wireless networking.

Representative Publication

Binbin Chen, Ziling Zhou, Yuda Zhao, and Haifeng Yu, "Efficient Error Estimating Coding: Feasibility and Applications." IEEE/ACM Transactions on Networking (ToN), Volume 20, Issue 1, pp. 29-44, February 2012.

MultiZoom: Fast and Reliable Algorithms for Multichannel Radio Networks

Today, we are entering the age of open airwaves: researchers and engineers are developing exciting new wireless applications, and these applications rely on the open airwaves. These unlicensed bands of spectrum are freely available for everyone to use, fostering innovation and new ideas. Yet the very fact that the airwaves are open to all poses its own set of challenges, and these challenges risk derailing the ongoing wireless revolution. First, the rapid growth of wireless networking is leading to crowding on the airwaves, creating inadvertent interference among users. This contention for bandwidth leads to poor performance and complicates the design of wireless applications. Second, the open and shared spectrum is particularly vulnerable to noise and signal interference, leading to corrupted messages and unpredictable network delays. Thus, applications that rely on the open airwaves suffer from unreliable performance and unexpected delays.

The primary goal of this project is to address the challenges of the open airwaves by exploiting the power of dynamic spectrum access. We focus primarily on ad hoc networks, that is, networks of devices that are deployed in an ad hoc fashion and communicate directly amongst each other in a peer-to-peer fashion. Moreover, modern wireless radios have the capacity to use a wide range of spectrum which can be accessed dynamically. In this project, we focus on how to leverage such a multi-channel environment to achieve fast and reliable protocols, even in the presence of interference and noise.

Representative Publications

Sebastian Daum, Seth Gilbert, Fabian Kuhn, Calvin C. Newport: Leader election in shared spectrum radio networks. PODC 2012: 215-224

Mohsen Ghaffari, Seth Gilbert, Calvin C. Newport, Henry Tan: Optimal Broadcast in Shared Spectrum Radio Networks. OPODIS 2012: 181-195

Keren Censor-Hillel, Seth Gilbert, Fabian Kuhn, Nancy A. Lynch, Calvin C. Newport: Structuring unreliable radio networks. PODC 2011: 79-88

Shlomi Dolev, Seth Gilbert, Majid Khabbazian, Calvin C. Newport: Leveraging Channel Diversity to Gain Efficiency and Robustness for Wireless Broadcast. DISC 2011: 252-267

Computation in Highly Dynamic Mobile Networks

Consider the problem of studying a city, exploring its evolution over an average week. One might wonder: How do populations move through the city? How do weather and air quality correlated with these movements? Are there patterns to how people dress? And so on. To answer these types of questions, imagine equipping a large number of taxicabs with sensors that record everything in their nearby vicinity. The resulting mobile (vehicular) sensor network provides the perfect platform for monitoring and collecting data. Taxis move throughout the city, providing high levels of coverage at a relatively low cost.

In this project, we are exploring several algorithmic issues that arise in collecting and aggregating data. How efficiently can we aggregate data? How do we cope with the constant changes in the vehicular network? Can we leverage hidden patterns in the underlying vehicular movement to improve the efficiency of aggregation? How do we trade-off different communication mechanisms, e.g., 3G vs. 802.11?

Representative Publication

Alejandro Cornejo, Seth Gilbert, Calvin C. Newport: Aggregation in dynamic networks. PODC 2012: 195-204

Resource-competitive Algorithms

It has long been known that attacking is often more expensive that defending. This has been true in both the real world (see, e.g., ancient fortified cities), and the on-line world. In the latter context, this approach has been proposed to resolve denial-of-service attacks (see, e.g., computational puzzles), reduce spam (e.g., micropayments), and is perhaps at the basis of all cryptography. In this project, we have two (related) goals. The first is to apply this idea to the domain of wireless networking; the second is to develop a more precise theory of resource-competitive algorithms that will allow us to formalize how competitive a given algorithm is, and hence provide precise performance guarantees in the presence of adversarial attacks. We have been working on several different resource-competitive algorithms for wireless networks that focus on energy as the key resource: both attackers and defenders must spend energy, and we can design algorithms where attackers must spend a large amount of energy to disrupt the protocol, while the honest nodes can proceed cheaply. We are applying this idea to basic communication problems, backoff protocols, sybil attacks, and robot coordination.

Representative Publications

Seth Gilbert, Maxwell Young: Making evildoers pay: resource-competitive broadcast in sensor networks. PODC 2012: 145-154

Seth Gilbert, Jared Saia, Valerie King, Maxwell Young: Resource-competitive analysis: a new perspective on attack-resistant distributed computing. FOMC 2012: 1

Faster Multiprocessor Algorithms: Cooperation in the Cloud

This project has focused on problems in the area of high-performance parallel computing. We have developed new and faster algorithms for enabling a set of processors/cores/threads to cooperatively work on a shared problem. For example, consider two complementary problems: task allocation (i.e., how do processes cooperate to share work amongst themselves) and mutual exclusion (i.e., how do processes cooperate to avoid accessing the same shared resource). Both problems are questions of long-standing interest, with research dating back to the origins of computer science. Over the last two years, we have developed new algorithms for each of these problems that are significantly faster than the best previously known techniques. In the case of mutual exclusion, the new algorithm is an exponential improvement over existing results, and has led to significant new developments in very fast (sub-logarithmic) protocols for multiprocessor coordination. In the case of task allocation, the new algorithm has the potential to reduce the coordination overhead for large cloud computing workloads, improving the efficiency of big data analysis.

Representative Publications

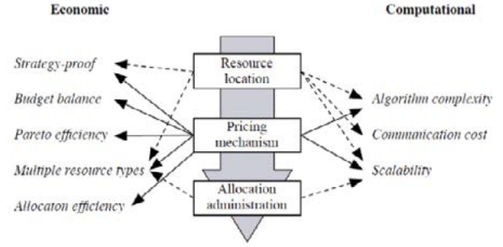
Dan Alistarh, Michael A. Bender, Seth Gilbert, Rachid Guerraoui: How to Allocate Tasks Asynchronously. FOCS 2012: 331-340

Michael A. Bender, Seth Gilbert: Mutual Exclusion with O(log^2 Log n) Amortized Work. FOCS 2011: 728-737

Dan Alistarh, James Aspnes, Seth Gilbert, Rachid Guerraoui: The Complexity of Renaming. FOCS 2011: 718-727

Strategy-proof Resource Pricing on Multiple Clouds

Resource sharing on the Internet is becoming increasingly pervasive. Recently, there is growing interest in federated systems where globally distributed and commoditized resources are traded, such as in peer-to-peer computing, grids, and cloud computing. An important issue in the allocation of shared resources in the context of large federated systems, such as multi-clouds, is that users are rational and attempt to maximize their self-interest. Our work investigates the use of resource pricing with financial incentives to allocate shared resources when users are rational. Using mechanism design, we developed a strategy-proof, VCG-based resource pricing scheme, with provable economic properties that include individual rationality, incentive compatibility and budget balance. We show that multiple resource types per consumer request can be allocated to multiple resource providers with polynomial time complexity. Although we trade-off Pareto efficiency and obtain 80% optimal economic efficiency, we achieve strategy-proof, budget balance and computational efficiency, and at the same time increase the number of allocations by about half under different market conditions. To improve scalability, we developed a new distributed auction scheme that leverages on a peer-to-peer DHT overlay network for resource lookups and distribute pricing by resource type. As an application of our pricing scheme, we have prototyped SkyBoxz, a multi-cloud platform for cloud consumers to utilize and cloud providers to manage virtual machines across multiple public and private clouds.



Strategy-proof Resource Pricing – Economic and Computational Efficiencies

Representative Publication

M. Mihailescu and Y.M. Teo, The Impact of User Rationality in Federated Clouds, IEEE/ACM Int. Sym.on Cluster, Cloud and Grid Computing, 2012.

Understanding Emergent Properties in Complex Systems

In complex systems, simple component interactions can lead to unexpected behaviors and new properties. These properties, called emergent properties, represent a significant challenge in engineering and understanding the behaviors of complex systems. Examples of emergence include trends in big data analytics such as connection patterns in social networks, Ethernet capture effect, and load-balancer failures in a multi-tiered distributed system. Emergence can be a positive addition to the system properties of a product because these properties can be adapted to support new tasks that the designers have never originally intended. Though there is a lot of interest in emergence, this is still not well understood. Our formal approach centers on detecting weak emergence from a computer science perspective by factoring emergence in system design and models. We have developed a grammar-based set-theoretic approach to identify and to quantify the set of emergent property states in a given system, and demonstrates how known emergence can be detected. Ongoing work includes applying our approach to study emergence in concurrent program verification, self-organizing properties in AI and multi-agent systems, and big data such as in social networks.

Representative Publications

Y.M. Teo, B.L. Luong, C. Szabo, Formalization of Emergence in Multi-agent Systems, ACM SIGSIM Conf. on Principles of Advanced Discrete Simulation, 2013.

C. Szabo, Y.M. Teo, Post-mortem Analysis of Emergent Behavior in Complex Simulation Models, ACM SIGSIM Conf. on Principles of Advanced Discrete Simulation, 2013.

C. Szabo and Y.M. Teo, An Integrated Approach for the Validation of Emergence in Component-based Simulation Models, Winter Simulation Conference, 2012. IEEE Computer Society Press, Berlin, Germany, Dec 2012.

C. Szabo and Y.M. Teo, An Objective-based Approach for Semantic Validation of Emergence in Component-based Simulation Models, ACM/IEEE/SCS Workshop on Principles of Advanced and Distributed Simulation,2012.

Y.M. Teo and C. Szabo, Semantic Validation of Component-based Models with Emergent Properties, Book Chapter in Ontology, Epistemology, and Teleology of Modeling and Simulation – Philosophical Foundations for Intelligent M&S Applications, edited by Andreas Tolk, Springer-Verlag, 2012.

Traffic Simulation and Symbiotic Simulation

Traffic Management and Control has been receiving wide attention as the number of car in cities increases. We are collaborating with SMART (Singapore MIT Alliance for Research and Technology) to design and implement SimMobility, the proposed simulation platform of the Future Mobility project that aims to integrate and link together various mobility-sensitive behavioural models with state-of-the-art simulators to predict impacts of mobility demands on transportation networks, intelligent transportation services and vehicular emissions. This will thereby enable the simulation of the effects of a portfolio of technology, policy and investment options under alternative future scenarios. In short, SimMobility encompasses the modeling of millions of agents, from pedestrians to drivers, from phones, traffic lights to GPS, from cars to buses and trains, from second-by-second to year-by-year simulations, across entire countries. Potential users go beyond SMART researchers, including transportation planners, intelligent transportation system designers, urban planners, etc. Figure 1 shows a high-level overview of SimMobility.

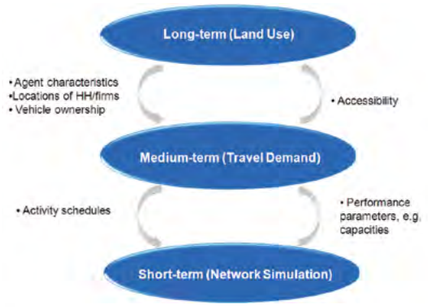


Figure 1: Overview of SimMobility

We are also working with SMART to introduce symbiotic simulation into a Real-Time Traffic Simulator, called DynaMIT. Symbiotic simulation systems have been proposed by the Parallel/Distributed Simulation Work group at the Dagstuhl Seminar on Grand Challenges for Modeling and Simulation as one of the grand challenges for Modeling and Simulation. A symbiotic simulation system is defined as one that interacts with the physical system in a mutually beneficial way; it not only performs "what-if" experiments that are used to control the physical system, but also accepts and responds to data from the physical system. Symbiotic simulation has so far been applied to socio-economic systems and manufacturing systems. Modern traffic management systems, with their deployment of cameras, sensors, tracking systems, etc., lend themselves very well to symbiotic systems. When a major incident happens, our symbiotic simulation will kick in to perform 'what if' alternative strategy simulation and decide on what the best solution is, and this will be fed back into the real system.

Representative Publications

Y.Xu and G.Tan, "Offline Road Network Partitioning in Distributed Transportation Simulation", Workshop on Principles of Advanced and Distributed Simulation, 2012.

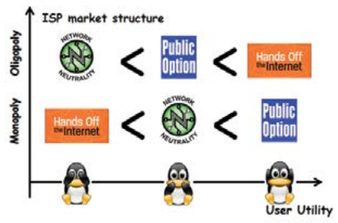
S.Hetu, and G.Tan, "Perennial Simulation of a Legacy Traffic Model: Implementation, Considerations and Ramifications", Winter Simulation Conference, 2011.

S.Hetu, and G.Tan, "Application of Symbiotic Decision Support to Manage Evacuation Studies using a Perennial Framework", Asia Simulation Conference, 2011.

Internet Regulation and the Network Neutrality Debate

The network neutrality principle bars Internet Service Providers (ISPs) from offering premium services to customers willing to pay higher fees. While potential network neutrality legislation is still hotly debated around the world, we have found a less intrusive and more effective alternative: the introduction of a Public Option ISP.

Because the Internet only offers a single class of "best-effort" service, it has been a real challenge to enable seamless end-to-end communications and guarantee quality of service for applications like real-time immersive telepresence. While ISPs could introduce premium services to help solve this problem, the fear is that non-premium users would suffer as a result; hence the push for network neutrality legislation. Our new approach introduces a Public Option ISP. Whether run by the government or a private organization, the Public Option ISP implements a network-neutral policy that does not prioritize or discriminate against particular types of traffic. Other ISPs can provide premium services if desired. The resulting system provides higher overall utility (a measure of whether users get the network throughput they need) for users than could be obtained by legislating network neutrality.



The Public Option approach can reshape the Internet landscape, creating more competitive and efficient ISP markets and supporting diversity in the industry. Its new economic incentives will enable new business models for ISPs to provide novel services and higher user utility. Through competition, more innovative applications can emerge and be supported by the future Internet. Most importantly, the Public Option ISP will always act as a safety net for users through its own network-neutral policy.

Representative Publications

Richard T. B. Ma and Vishal Misra. Congestion and Its Role in Network Equilibrium. IEEE Journal on Selected Areas in Communications, Volume 30, Issue 11, December, 2012.

Richard T. B. Ma and Vishal Misra. The Public Option: a Non-regulatory Alternative to Network Neutrality. ACM CoNEXT 2011.

The faculty members of the Systems and Networking research area are:

- Akkihebbal L. Ananda
- Chan Mun Choon
- Seth Gilbert
- Leong Wing Lup, Ben
- Richard Tianbai Ma
- Tulika Mitra
- Abhik Roychoudhury
- Soo Yuen Jien
- Tan Soon Huat, Gary
- Tan Keng Yan, Colin
- Teo Yong Meng
- Wong Weng Fai
- Yu Haifeng

Department of **Information Systems**

HEAD'S OVERVIEW

The Department of Information Systems has attained international research leadership position in recent years with high quality and impactful research publications. Based on the publications listed in the top two Information Systems journals (Information Systems Research and MIS Quarterly), the Department was ranked eight (see http://www.vvenkatesh.com/ISranking /RankingsMISQISRbyUni.asp?RYear=3) and several of our faculty members' work have garnered significant number of citations and won international research awards. Our faculty members are actively serving on the editorial boards of top-tier journals (e.g., MIS Quarterly, Information Systems Research, Management Science, Journal of the AIS, Journal of Management Information Systems, and IEEE Transactions on Engineering Management) and the program committees of top-tier conferences (e.g., International Conference on Information Systems). Indeed, we are one of the few information systems departments worldwide which have faculty members serving on the editorial board of every top-tier journal in the field of information systems.

To scale new heights in our research, we have embarked on research projects that will deliver thought leadership and significant impact to the local industry and the international community. For example, the research group on Information Systems and Management in Healthcare seeks to assess the impact of healthcare systems on different dimensions and provide guidance on how IT can be designed, used and managed for better health outcomes and a safer and more efficient healthcare system. In connection with this, the Department has set up the Centre of Health Informatics to develop key human capital and engage in intensive research in health informatics with the support from InfoComm Development Authority of Singapore in January 2012. The research groups on Social Computing and Business Intelligence are also developing tools and best practices to exploit the vast amount of information that exists on technology-enabled platforms to improve business decisions and firms' competitiveness. As services will be the next driver of economic growth for Singapore, the service innovation group will be looking at the interplay of people, technology, and practices to create and deliver innovative and transformative services for the firms. These projects have received very positive feedback. On another dimension, as we seek to establish ourselves as an important node of knowledge and expertise on information systems issues in Asia, we have begun to collaborate closely with several top Chinese universities to investigate interesting electronic commerce phenomena in China. Locally, we also look forward to collaborating with industry and government partners in these projects.

The doctoral programme in the Department of Information Systems has been widely acknowledged for the quality of its graduates. The programme has been continuously refined to include comprehensive coursework and research components as well as a rigorous examination process. Graduates of the programme have received good placements in industry and academia. In academia, our doctoral graduates have earned placements in good universities in the United States, Canada, the United Kingdom, France, Australia, New Zealand, India, China Mainland, China Hong Kong, Switzerland and Singapore. A key objective of our doctoral programme is to develop a global community of doctoral alumni.

To maintain international leadership position in research, the Department of Information Systems will continue to subject its research initiatives and education programme to rigorous review by international scholars. We owe much of our success to our Visiting Committee and support from our Deanery and senior management. We will continue to invite distinguished scholars in the field of information systems to visit us, critique our research and collaborate with us in our projects. In so doing, we are continuing a practice that has proven to raise our international profile in research significantly.

The Department of Information Systems aims to increase its research vibrancy by adding to the many research collaboration efforts that are already ongoing. We welcome prospective partners from academia, industry and government to work with us in our venture in knowledge creation and application.

Teo Hock Hai
Professor and Head
Department of Information Systems

IT-ENABLED SERVICE INNOVATION

The Research Group on IT-Enabled Service Innovation focuses on the interplay of people, information technology, and organizational practices to deliver enhanced services. As the global economy turns increasingly service oriented, the role of IT in enabling and innovating services requires systematic investigation and understanding. Particularly, increasing numbers and varieties of services are provided through IT platforms e.g., SaaS and Cloud services. Furthermore, IT enables the creation of new services in diverse industries such as the financial and public sectors. The group's research encompasses salient topics in this area. The objectives are to develop models and frameworks that enhance theoretical understanding as well as provide insights to practitioners on how to leverage and manage IT to create, deliver, and experience services. Key themes include:

- Open Service Innovation
- Sector Studies of e-Service Innovation and Impacts
- IT-enabled Service Outsourcing
- Cloud Computing Management

Open Service Innovation

Service innovation is increasingly being recognized as an important strategy for firms to differentiate themselves and sustain competitive advantage. The connectivity and interactivity afforded through IT has transformed how services are designed. However, there are untested claims about whether IT-enablement actually improves service performance. Further, customers are being considered as important value co-creators and increasingly included in organizational innovation efforts, e.g., users are major innovators of mobile data services. Yet, it still remains unclear how user participation affects the quality and performance of these services. This is particularly salient with the explosive growth of mobile services, yet the success of very few among them. Our research on mobile service innovation examines the drivers behind the successful creation of such user-generated services.

Such involvement of external sources e.g., customers, is being recognized as a new paradigm in innovation, termed as open innovation. This approach can take on various forms including crowdsourcing and open networks with knowledge intermediaries. Our research in this area explores various forms of IT-enabled open innovation and suggests optimal designs in terms of incentives, task specification, and contest elements that can lead to greater reach and better quality of participation, both from seekers (firms posing innovation problems) and solvers (external parties providing solutions). This in turn can ensure that seekers obtain both efficiency and effectiveness benefits from such open innovation platforms.

Task Type and Crowdsourcing Approach Framework

Task type	Crowdsourcing approach	Requirement specificity	Motivations for participation
Simple task with low outcome variety	Open call for participation, Open call for solution	Specific	Extrinsic: Micro-paid Intrinsic: Enjoyment, Sense of achievement, Solvers' need fulfillment
Simple task with high outcome variety	Open call for solution, Open call for participation	General	Extrinsic: Financial rewards, Visibility in the job market Intrinsic: Skill enhancement, Enjoyment in solving novel tasks
Complex task with low outcome variety	Open call for candidates, Open call for solution	Specific	Extrinsic: Financial rewards, Peer reputation Intrinsic: Solvers' need fulfillment and autonomy
Complex task with high outcome variety	Open call for candidates	Specific at a higher level, general at a lower level	Extrinsic: Financial rewards Intrinsic: Enjoyment in solving challenges, Sense of achievement

Representative Publications

Hua Ye, J., and Kankanhalli, A. "Leveraging Crowdsourcing for Value Co-Creation", Communications of AIS

Hua Ye, J., and Kankanhalli, A. "Exploring Innovation through Open Networks: A Review and Initial Research Questions", IIM Bangalore Management Review, 25th Anniversary Volume, Invited Paper.

Hua Ye, J., Kankanhalli, A., and Yang, Z. (2012) "Knowledge Brokering for Open Innovation: A Case Study of Innovation Intermediaries," International Conference on Information Systems, Orlando.

Hua Ye, J., Kankanhalli, A., Goh, K.Y., and Sun, J. (2011) "Investigating Value Co-Creation in Innovation of IT-enabled Services: An Empirical Study of Mobile Data Services", International Conference on Information Systems, Shanghai.

Hua Ye, J., and Kankanhalli, A. (2011) "Investigating Antecedents of IT-enabled Service Innovation", Academy of Management Meeting, San Antonio.

Xu, H., Teo, H.H., Tan, B.C.Y. and Agarwal, R. (2010) "Role of Push-Pull Technology in Privacy Calculus: The Case of Location-Based Services," Journal of Management Information Systems, 26(3), 137-176.

Sector Studies of e-Service Innovation and Impacts

Digital technologies are being embedded into an ever-increasing range of services, thereby expanding the role and relevance of IT in their innovation. Concurrently, the continued pressure to reduce the time and cost of innovation has driven many companies to adopt standardized innovation structures, tools, processes, and metrics enabled through IT. Yet, there is limited evaluation and assessment of the impacts of e-service innovation, which are likely to be sector specific. With the above motivation, this research stream aims to develop models of e-service innovation and performance in key service sectors in Singapore e.g., financial and government. The performance assessment consists of a set of dimensions that describe the development of e-services and their impacts on salient organizational outcomes such as customer satisfaction and sales. This provides a way to measure organizations' ability to improve their e-service processes, as well as innovate and scale up e-services, thus enhancing consumer adoption and satisfaction.

Representative Publications

Yang, Y., and Kankanhalli, A. (2013) "Investigating the Influence of IT and Other Resources on Service Innovation in Banking", Pacific Asia Conference on Information Systems, Jeju, Korea.

Yang, Z., and Kankanhalli, A. (2013) "Innovation in Government Services: The Case of Open Data", IFIP WG 8.6 Conference, Bangalore.

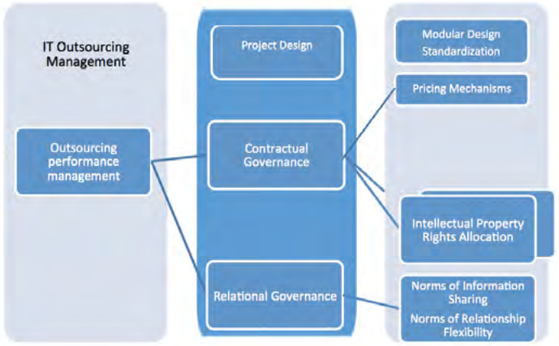
Hua Ye, J., Kankanhalli, A., Huber, M., Blohm, I., Bretschneider, U., Goswami, S., Leimeister, J. M., Krcmar, H. (2012) "Collaboration and the Quality of User Generated Ideas in Online Innovation Communities," Academy of Management Meeting, Boston.

Phang, C.W. and Kankanhalli, A. (2011) "A Service Systems

Perspective of E-government Development", Service Systems Implementation in the Service Science: Research and Innovations in the Service Economy Series, Demirkan, H., Spohrer, J.C., Krishna, V. (Eds.), Springer e

Kankanhalli, A., Lee, O.K., and Lim, K. (2011) "Knowledge Reuse through Electronic Repositories: A Study in the Context of Customer Service Support", Information and Management, 48(2-3), 106-113.

IT Enabled Service Outsourcing



Information technology (IT) enabled service outsourcing has become an important global strategy for many organizations today. The economies of scale and scope coupled with labor arbitrage provide a convincing business case, especially for organizations in the developed world, to shift IT-intensive functions to offshore vendors located in foreign countries. In service outsourcing processes, firms transfer valuable assets such as intellectual property and confidential information to vendors. The separation of ownership and control of these valuable and intangible assets can unleash a variety of problems and risks, such as moral hazard problems, vendor hold-up, security risks, and loss of intellectual property. Outsourcing companies need to understand the transaction risks they will face when transferring intellectual property and information across countries in order to make appropriate decisions to protect themselves proactively and retroactively. Our research explored effective risk-mitigating mechanisms to circumvent outsourcing risks. In particular, we examined three types of risk-mitigating mechanisms: contract design, relational governance, and project design (e.g., modularity of projects). We developed a model contract as a template for IT outsourcing contract negotiation. Outsourcing companies can use this contract template to decide whether they need to bargain for more controls rights or select appropriate payment structure and project milestones to protect their intellectual property and valuable assets from misappropriation. We also examined the effects of non-contractual mechanisms (i.e., relational governance and project design) and found that relational governance and project design function as good substitutes for contract design in specific circumstances. These findings provide practical guidelines for IT professionals and business executives in IT outsourcing management. Our research have been published in several top IS journals and conferences.

Representative Publications

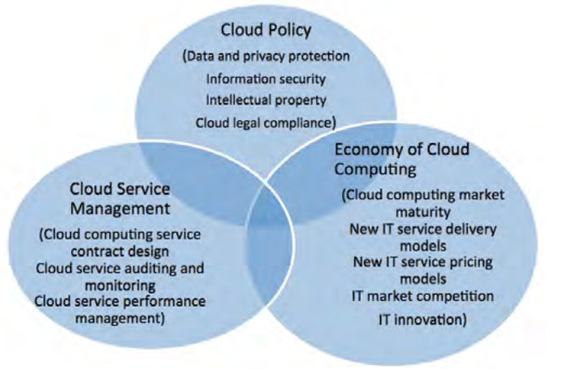
Mathew, S., and Chen, Y.Y. (2013) "Achieving Offshore Software Development Success: An Empirical Analysis of Risk Mitigation through Relational Norms," Journal of Strategic Information Systems.
Chen, Y.Y., and Bharadwaj, A. (2013) "An Empirical Analysis of Intellectual Property Sharing Arrangements in Software Development Outsourcing," International Conference of Outsourcing Information Systems (ICOIS), Mannheim.

Chen, Y.Y. and Heng C.S. (2012) "Contract Renegotiation and Bargaining Power: Evidence from IT-related Outsourcing Agreements", International Conference on Electronic Commerce, Singapore, Best Theme Paper Award.

Png, I.P.L., and Chen, Y.Y. (2011) "Installment Payments and Contractual Opportunism: Empirical Evidence from Technology Development", International Conference of Information Systems, Shanghai.

Leong, S.H.S., and Chen Y.Y. (2010) "The Rights of Communication in Singapore", Singapore Academy of Law Journal, 22(2), 602-631.

Cloud Computing Management



Cloud computing has emerged as a new paradigm in IT services provisioning providing alternate mechanisms for service modeling and deployment. Although the industry adoption of this new model shows a growing trend it is not very clear from usage data how this approach in IT services is impacting the IT market landscape in terms of performance of traditional IT service and product providers and how it further affects IT market restructuring. Our research seeks to understand the profound effects of cloud computing on the economy, measured in three different ways: productivity, diffusion of new technologies, and ongoing investments in information technologies. We are also constructing empirical research models examining impacts of cloud computing on firm performance from the service provider perspective. Specifically, our research answers the following two strategic questions: (1) how does cloud computing influence the transition to a new economy and ongoing investment in information technologies? and (2) what does the shift towards the Cloud mean to different kinds of IT suppliers such as a traditional software company, IT outsourcing vendors, and IT hardware companies? Knowing how the Cloud influences any of these two areas is useful for policy and decision making on technology investment. For example, our study about the impacts of the Cloud on economic growth and fluctuations is useful for governments in providing the best infrastructure of innovation for IT entrepreneurs. Our research about the effects of the Cloud on IT providers' market structure and competition has important managerial implications for IT professionals and business executives. The research findings have been published in top international management conferences and have garnered invitations to serve as panel discussant in several cloud computing technology development conferences.

Representative Publications

Chen, Y.Y., and Zhan, J. (2013) "An Empirical Study of the Cannibalization Effects of SaaS on on-Premise Software Firm Performance", Academy of Management Annual Meeting, Orlando.

Zou, X., and Huang, K.W. (2013) "Optimal Pricing and Capacity Planning for Information Technology Service Provider under Service-Level Agreement," Frontiers in Service Conference, Taipei.

Zou, X., and Huang, K.W. (2012) "A Model on Location-Based Service as Infomediary," ICEC, Singapore. (Best Student Paper Nominee)

Clemons, E.K., and Chen, Y.Y. (2011) "Making the Decision to Contract for Cloud Services: Managing the Risk of an Extreme Form of IT Outsourcing", Hawaii International Conference on System Sciences.

Ge, C., and Huang, K.W. (2011) "R&D and Catch-Up Effect Among Software-as-a-Service Firms: A Stochastic Frontier Approach," ICIS, Shanghai.

Saya, S., Pee, L.G., and Kankanhalli, A. (2010) "The Impact of Institutional Influences on Perceived Technological Characteristics and Real Options in Cloud Computing Adoption", International Conference on Information Systems, St. Louis.

The faculty members involved in IT-enabled service innovation research are:

- Chen Yuanyuan
- Goh Khim Yong
- Heng Cheng Suang
- Huang Ke-Wei
- Atreyi Kankanhalli
- Tan Cheng Yian, Bernard
- Teo Hock Hai

ELECTRONIC AND SOCIAL COMMERCE

The Research Group on Electronic and Social Commerce seeks to understand and develop information technologies to influence social and economic behaviors of individuals, teams, organizations and online communities. In recent years, the Internet has greatly facilitated the creation and exchange of user-generated content within and across web sites, making such related media sites and channels increasingly "social". Indeed, social media can turn traditionally fragmented user inputs into all kinds of value co-creation. As a phenomenal way to leverage social media and to realize its business value, more and more e-commerce platforms support large-scale consumer interactions, aggregate consumer opinions and contributions, and optimize online purchases that cannot be achieved through isolated individual consumers' participation in conventional e-commerce. Hence, it is particularly important to understand how to best deploy technologies platforms to facilitate peer-to-peer consumer interactions and harness the wisdom of crowds for informed product search and decision.

Collaborative Shopping

Shopping is often a social process, in which a shopper is accompanied by friends or family members. Tauber (1972) has argued that one of the prime motives for shopping is the desire to communicate with others who have similar interests, to share ideas about particular products with them, to seek their feedback, and to enjoy leisure time with friends

and family. Despite the apparent importance of the social nature of shopping, most e-commerce literature has focused on consumers' shopping on their own and very few studies have investigated the phenomenon of collaboration among online shoppers, presumably because it is difficult to implement collaboration due to technological constraints and difficult to measure the performance of collaboration. Hence, in this proposed study we will investigate this emerging and important online buying mode.

We use the term collaborative online shopping to describe the activity in which a consumer shops at an online store concurrently with one or more remotely located shopping partners. We aim to achieve three goals: 1) to explore this new frontier of e-commerce and develop a new theory about collaborative online shopping, which integrates collaborative technologies with coordination performance, mutual understanding, social presence, and shopping enjoyment; 2) to identify the fundamental characteristics of the technologies that can facilitate collaborative online shopping; 3) to design a series of experiments to empirically test the effects of different technologies on collaborative shoppers' behaviour and to shed light on the design of such technologies.

Induced Product Learning and Search from Consumer-Generated Data

Consumers are increasingly sharing product preferences and experiences with others on websites. For example, consumers can "tag" products using their own words and these "product tags" are then aggregated and shared with other online consumers who seek information. In addition, highly experienced and influential information contributors on websites are often highlighted as "featured users" and serve as direct information sources. This study examines the effects of these two distinct design mechanisms, i.e., product tags and featured users. While the former facilitates attribute-based product search, the latter facilitates people-based search. We constructed experimental websites using actual data from one of the largest social-network-based product-search websites in China. The results show that information seekers perceive their product search experience as highly diagnostic when connected product tags are provided. The presence of featured users leads information seekers to perceive the product search experience as serendipitous, but this is evident when connected tags are also provided.

Representative Publications

Tan B.C.C., Pan, S L., Lu X H, and Huang L H (2014), "The Role of IT Capabilities in the Development of Multi-Sided Digital Platforms: Insights from the Platform Development Strategy of Alibaba.com". Journal of AIS.

Goh, K.Y., Heng, C.S., and Lin, Z., "Social Media Brand Community and Consumer Behavior: Quantifying the Relative Impact of User- and Marketer-Generated Content", Information Systems Research, 24(1), 2013, pp. 88-107.

Kim, H. W., Chan, H. C. and Kankanhalli, A., "What Motivates People to Purchase Digital Items on Virtual Community Websites? The Desire for Online Self-Presentation," in Information Systems Research.

Yi, C., Jiang, Z. and Benbasat, I. "Enticing Consumers via Incomplete Product Experience: An Investigation of Online Product Interactivity Designs," Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI), Vancouver, Canada, 2011.

Zhu, L., Benbasat, I. and Jiang, Z. (equal contribution) "Let's Shop Online Together: An Empirical Investigation of Collaborative Online Shopping Support," Information Systems Research, Vol. 21, No. 4, pp. 872-891, 2010.

Tan, C. H., Teo, H. H., Benbasat, I., "Assessing Screening and Evaluation Decision Support Systems: A Resource-Matching Approach," Information Systems Research, Volume 21, Issue 2, 2010.

Yi, C., Jiang, Z. and Benbasat, I. "Towards Organized Search and Unexpected Discoveries: Investigating the Impacts of Product Tags and Featured Users in Online Product Search," the International Conference on Information Systems (ICIS), Saint Louis, US, 2010.

The faculty members involved in e-commerce/social commerce research are:

- Chan Hock Chuan
- Goh Khim Yong
- Jiang Zhenhui, Jack
- Atreyi Kankanhalli
- Pan Shan Ling
- Tan Cheng Yian, Bernard
- Teo Hock Hai

HEALTHCARE INFORMATION SYSTEMS

The Research Group on Healthcare Information Systems seeks to assess the impact of healthcare information systems, providing guidance on how they can be designed, used and managed for better and more efficient outcomes. Plagued by issues of quality, changing demographics and rising healthcare costs, information technology (IT) has been recognized as an important enabling tool to provide a safer, more efficient and patient-centric healthcare delivery system. Globally, investments in health IT have grown in recent years, creating immense opportunities and need for IS research in this area. In light of these developments, our research focuses particularly on both IS applications for healthcare providers and health consumers. The main areas of focus are: 1) Adoption and Use, 2) Impact, 3) Consumer health informatics, and 4) Analytics.

Adoption and Use

In the area of adoption and use, we have conducted studies in hospital settings as well as in physician practices. These studies are aligned with the National Electronic Health Record (NEHR) initiative in Singapore to develop lifetime electronic health records of their citizens. As the adoption of institution-based Electronic Medical Records is essential for the success of the NEHR, there is a need to understand the challenges that inhibit adoption so appropriate technologies and policies can be designed to facilitate adoption by providers. To this end, we have conducted a study that evaluates the EMR features that physician practitioners view as empowering their work and how that in turn affects their likelihood of adopting the EMR. In another study, we have evaluated the enabling factors of hospitals' decision in adoption of HIT.

Compared to the outpatient setting, IT adoption in Singapore hospitals is much more advanced. Out of 8 public hospitals in Singapore, 4 have achieved stage 6 of the EMRAM (EMR Adoption Model) Award given by HIMSS Analytics. At the same time, hospitals in Singapore are actively exploring and implementing new systems to improve their healthcare processes and to deliver better patient care. To this end, we have conducted studies to examine and evaluate the implementation and use of medication management systems, physician order entry systems (CPOE), Radiology & Picture Archiving & Communication systems (RIS/PACS), and Messaging systems. Our work has been published in premium conferences (e.g. ICIS, PACIS) and journals (e.g. DSS, IJHCS), and some working papers are concurrently under review.

Representative Publications

Ng, B.Y., A. Kankanhalli, and J.W.L., Yip (2011), "Use of Healthcare IS by Multiple User Groups: An Empirical Study of a Medication Management System", International Conference on Information Systems, Shanghai.

Tong, Y., Goonawardene, N., Tan, S.S.L., Teo, H.H. and Low, C.H. "The Influence of Job Rotation on Physicians' System Use: A Situated Learning Perspective", Proceedings of the 32nd Annual International Conference on Information Systems, Shanghai, China, Dec 2011.

Xu, L.L., Tan, S.S.L., Chan, H.C., Teo, H.H., and Goh, L.G. "Empowering Physicians with Electronic Health Records: System Capabilities to Adoption Intention," Academy of Management Annual Meeting, Montréal, Canada, August 2010.

Yang, Z., Ng, B.Y., Kankanhalli, A. and Lim, J., "Analyzing the Enabling Factors of the Organizational Decision to Adopt Healthcare Information Systems", Decision Support Systems.

Yang, Z., Ng, B.Y. and Kankanhalli, A. "Workarounds in the use of IS in healthcare: A Case Study of an Electronic Medication Administration System", International Journal of Human Computer Studies, 70(1): 43-65, 2012.

Impact

To understand the value of IT investments in healthcare, it is important to examine the outcomes and impact of the technologies that have been implemented. To date, research results in this area are still equivocal with some studies showing positive outcomes and some showing negative outcomes. Our research in this area examines the impact of health IT in hospital settings and e-health in general. For instance, one of our studies examined how upgrades in information technology enabled physician referral systems affect the efficiency of healthcare-related processes by facilitating or inhibiting the learning process of hospital employees.

Representative Publications

Karimi, F., Poo, C.C.D., and Tan, Y.M. "Clinicians' Satisfaction with Clinical Information Systems: A Disconfirmation Paradigm Perspective" Academy of Management Annual Meeting, Boston, Massachusetts, United States, August 2012.

Mukhopadhyay, T., Singh, P.V. and Kim S.H. "Learning Curves of Agents with Diverse Skills in Information Technology Enabled Physician Referral Systems", Information Systems Research, 22(3): 586-605, 2011.

Piette, J.D., Lun, K.C., Moura Jr, L.A., Fraser, H.S.F., Mechael, P.N., Powellf, J. and Khoja, S.R. "Impacts of e-health on the outcomes of care in low- and middle-income countries: where do we go from here?" Bulletin of the World Health Organization, 90: 365–372, 2012.

Consumer Health Informatics (CHI)

There is growing interest in developing e-Health solutions and applications for healthcare consumers. Today, there is a burgeoning of websites and mobile apps that target the health consumer. Our research in this area focuses on understanding how consumers use and interact with CHI applications and the impact these applications have on their health behaviors and outcomes. For example in one study, we looked at how online health information seeking can affect patient's self-management behavior. In another study, we explore how technical features of microblogging platforms can influence users' use of microblog health messages.

Representative Publications

Goonawardene, N., Jiang, J., Tan S.S.L. and Jiang, Z. "Online health information seeking and adolescents' intention towards health self-management", Proceedings of the 17th Pacific Asia Conference on Information Systems, Jeju Island, Korea, June 2013.

Jiang, J., Tong, Y. and Tan S.S.L. "Do you Re-Tweet health advice on microblogging platforms? The Effects of Health Topic and Website Design on Credibility Assessment" Proceedings of the 33rd Annual International Conference on Information Systems, Orlando, Florida, Dec 2012.

Healthcare Analytics

With the increasing use of IT in healthcare, consolidating the large volumes of data generated daily from a healthcare institution's array of information systems is increasingly posing a massive challenge to administrators, care providers and researchers. Leveraging on such data is vital to healthcare institutions in terms of gathering business intelligence, streamlining workflows, keeping operating costs down and having the means to track and improve operations so as to offer the highest quality of customer service and patient care possible. Analytics research also holds great promise in enabling predictive and personalized healthcare. In view of the strategic importance of analytics in healthcare, we have taken on several initiatives to position the Department, as well as its Centre for Health Informatics, as a centre for research excellence in healthcare analytics. We are currently working on an analytics project with a hospital on their patient loyalty program as well as organizing the first International Conference on Big Data and Analytics in Healthcare from 8-10 July 2013.

Representative Publications

Lun, K. C. "Why Health Informatics needs Health Analytics" Proceedings of the Asia Pacific Association for Medical Informatics (APAMI 2012) Conference (invited Plenary Paper), Beijing, PRC, 22-25 October, 2012

Lun, K. C. "Is Healthcare ready for the Big Data Tsunami?" Proceedings of the 4th Middle East Health Informatics (MEAHI) Conference (invited Plenary Paper), Dubai, UAE, 22-24 April 2013.

The faculty members in healthcare information systems research are:

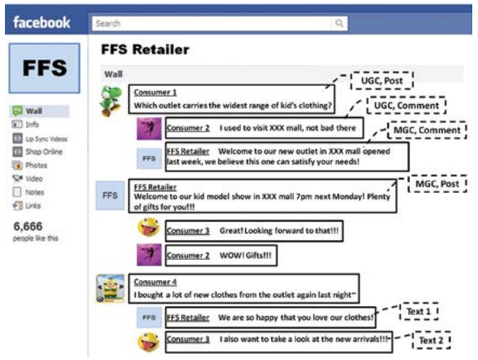
- Chan Hock Chuan
- Goh Khim Yong
- Atreyi Kankanhalli
- Lun Kwok Chan
- Poo Chiang Choon, Danny
- Tan Swee Lin, Sharon
- Teo Hock Hai

BUSINESS INTELLIGENCE

The Research Group on Business Intelligence and Analytics seeks to develop systems and capabilities to leverage the data, knowledge and information from businesses and organizations for intelligent decision making and competitive advantage. Rapid developments in information and communication technologies, the widespread availability and accumulation of data assets, and the need for organizations to better utilize their intellectual capital and data have led to increased interest in business intelligence (BI) domains. With massive amounts of information being added to corporate databases and the Internet every day, effective and efficient knowledge discovery and its management have become imperative. Diverse BI technologies and applications that we have researched on include intra-organizational social media applications, social computing tools, mobile and web analytics applications, credit scoring applications, revenue optimization systems, customer data analytics platforms, and advanced web mining systems. Broadly, BI research can be categorized into the quantitative and economic aspects of BI and the technical aspects of BI.

Quantitative and Economic Aspects

Our group's research on the quantitative and economic aspects of BI broadly involves the extraction, categorization, organization and analysis of the vast amounts of individual consumer and firm level business data that are particularly challenging in terms of high volume, velocity and variety, across a wide spectrum of business domains such as finance, marketing, supply chain, and e-commerce. We use methods and techniques such as econometric analysis, event history analysis, agent-based simulation models, neural network models, social network analysis, text mining and classification, and analytical modeling to evaluate research issues that can uncover deep and fresh insights into areas such as business operations, firm strategies, online and offline consumer behaviors, information content production and consumption behaviors.



Representative Publications

Goh, K.Y., Heng, C.S., and Lin, Z. (2013), "Social Media Brand Community and Consumer Behavior: Quantifying the Relative Impact of User- and Marketer-Generated Content", Information Systems Research, 24(1), pp. 88-107.

Goh, K.Y., Hui, K.L., and Png, I.P.L. (2011), "Newspaper Reports and Consumer Choice: Evidence from the Do Not Call Registry", Management Science, 57(9), pp. 1640-1654.

Huang, K.W. and Sundararajan, A. (2011), "Pricing Digital Goods: Discontinuous Costs and Shared Infrastructure", Information Systems Research, 22(4), pp. 721-738.

Kim, S.H. and Mukhopadhyay, T. (2011), "Determining Optimal CRM Implementation Strategies", Information Systems Research, 22(3), pp. 624-639.

Lehman, D.W. and Hahn, J. (2012), "Momentum and Organizational Risk Taking: Evidence from the National Football League", Management Science, 59(4), pp. 852-868.

Ma, X., Kim, S.H., and Kim, S.S. (2013), "Online Gambling Behavior: The Impacts of Cumulative Outcomes, Recent Outcomes, and Prior Use", conditionally accepted at Information Systems Research.

Zhang, C., Hahn, J., and De, P. (2013), "Continued Participation in Online Innovation Communities: Impact of the Member-Innovation Involvement Stage", in Information Systems Research.

Technical Aspects

Research in the technical aspects of BI broadly involves designing techniques to process the vast amounts of information generated on the Internet. In particular, we are interested in new media information sources, such as blogs and social media platforms such as Facebook and Twitter. These new media outlets have become mega-producers of useful information that can be leveraged by a whole host of real-world application areas such as retail, finance and publishing. We develop techniques and systems to process this vast information base and to capture the real-time raw data, store it efficiently and finally extract useful information from it that can be consumed by specific business application domains. We collect and study consumption patterns of this data and note that increasingly, consumers interact with this information through mobile devices such as Apple and Android smartphones. A particular outcome of this research is Mobilewalla, a company that pioneered the search, discovery and rating of mobile apps for mobile audience measurement and deep analytics.

Representative Publications

Datta, A., Dutta, K., Liang, Q., and VanderMeer, D. (2011), "SOA Performance Enhancement through XML Fragment Caching", Information Systems Research, 23(2), pp. 505-535.

Huang, K.W. and Li, Z. (2011), "A Multi-Label Text Classification Algorithm for Labeling Risk Factors in SEC Form 10-K", ACM Transactions on Management Information Systems, 2(3), pp. 18:1-18:19.

Pervin, N., Fang, F., Datta, A., Dutta, K., and Vandermeer, D. (2013), "Fast, Scalable, and Context-Sensitive Detection of Trending Topics in Microblog Post Streams", ACM Transactions on Management Information Systems, 3(4), Article No. 19.

Setiono, R., Baesens, B., and Mues, C. (2011), "Rule Extraction from Minimal Neural Networks for Credit Card Screening", International Journal of Neural Systems, 21(4), pp. 265-276.

Shaw, R., Datta, A., VanderMeer, D., and Dutta, K. (2013), "Building a Scalable Database-Driven Reverse Dictionary", IEEE Transactions on Knowledge and Data Engineering, 25(3), pp. 528-540.

The faculty members in business intelligence research are:

- Anindya Datta
- Kaushik Dutta
- Chen Yuanyuan
- Goh Khim Yong
- Jungpil Hahn
- Huang Ke-Wei
- Kim Seung Hyun
- Phan Tuan Quang
- Poo Chiang Choon, Danny
- Rudy Setiono

Innovation and Entrepreneurship

INNOVATION AND ENTREPRENEURSHIP

SoC actively nurtures innovation and entrepreneurship. Various steps have been taken to promote innovation and to support faculty and students who wish to embark on an entrepreneurial track. These include:

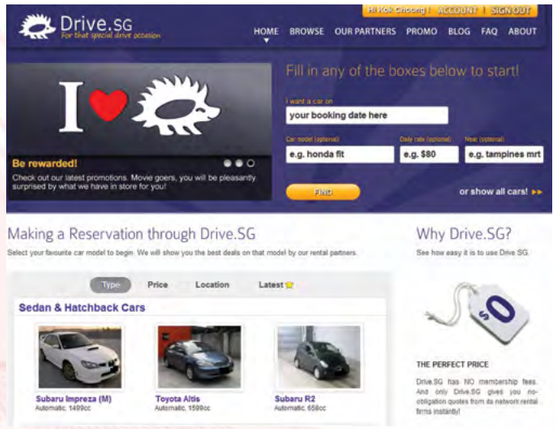
1. A set of courses to equip the aspiring innovator with appropriate tools and methods, and entrepreneurial skills
2. Incubation facilities and support, including guidance and networking
3. Exposure to industry specialists and entrepreneurs, particularly those working in regional and global markets
4. Targeted inclusion of adjunct faculty with entrepreneurial track records
5. Awards for undergraduates with outstanding innovations

School of Computing (SoC) Incubation Centre

The SoC Incubation Centre offers support to help students and alumni bring ideas through development to commercial fruition. The goal is to nurture young businesses and help them to survive and grow during the start-up period when they are most vulnerable.

The Centre provides physical space in lockable rooms and infrastructure such as network connectivity and telephone lines, to successful applicants for a period of one to two years (reviewed half-yearly). There is access to expert guidance from SoC faculty. Basic advisory services can also be made available, including advice on finance, legal, marketing and management matters.

Exciting Developments at SoC Incubatee Companies



Drive.SG, Singapore's first online direct sell platform for car rental companies, has announced that they have closed a S$200,000 seed round with private investor John Tan. John is also an investor in Chope, Impulseflyer, 8 Villages and Payroll Hero. Drive.SG was part of Echelon 2012's Startup Marketplace and managed to catch the attention of John, among another four to five other investors. According to John, what drew him to Drive.SG was the premium domain that they had. Adrian and his team took about five months after that before managing to woo John to invest in their company. <e27>

Mozat, a leading mobile social platform for operators and service providers worldwide. Mozat serves 16 million mobile users worldwide, through 15 telco partners. It is building the largest, most vibrant mobile social communication, file sharing and casual gaming platform worldwide with Vodafone. With this partnership, Mozat further underscores its position as a leading mobile social networking solutions provider for global operators. <WebWire>

tenCube, a five-year-old company which developed a security program for mobile phones and laptops was acquired by anti-virus giant McAfee. The software, named WaveSecure, can lock and wipe clean a phone's memory remotely, use the Web to back up and restore phone data, and provide a location and SIM card tracking service. <SPH>

Tiny Whale developed a Facebook application for Apple's iPads that became the top-seller in iTunes App Store only four months after its launch. The application, named Pica, priced at US$3.98, has been downloaded about 150,000 times since its launch. <SPH>

Beating the Odds

Students are invited to present their business plan or breakthrough idea to a panel of experienced Entrepreneurs and Venture Capitals so as to sharpen their business proposals and cultivating new entrepreneurs.

Innovation Path

SoC will introduce a Special Focus Area on Innovation to guide students in designing innovative, scalable and commercially viable products and services. A primary aim is to integrate technical skills and design thinking in for digital products and services.

Series of Courses

SoC has launched CP2201 (Journey of the Innovator), CS3882 (Developing Breakthrough Ideas) and CS4880 (Digital Entrepreneurship) to guide students through the process of digital innovation, and to encourage them to embark on a personal journey of creativity and challenge.

VaSCo Grant

VaSCo (Validating Startup Concept) was launched in 2012 to encourage students who have innovative and commercialisable ideas for digital markets to apply for seed funding of up to S$10,000 to develop their ideas. The grant is awarded on a project-basis and can be from multidisciplinary student teams.

SoC Final Year Project (FYP) Innovation Award

The FYP Innovation Award was launched in July 2009 to encourage undergraduate students to develop innovative, practical and commercialisable ideas that substantial potential for real world impact.

Workshops

Workshops are held to promote interaction with entrepreneurs and to impart relevant knowledge. Examples of workshops held are China workshop and Branding for Startups.

Companies currently hosted at the SoC Incubation Centre:

- 4D Earpiece
- Drive.SG <https://www.drive.sg/>
- ILOVEDEALS Pte Ltd <http://www.ilovedeals.sg/deal>
- KWIX <http://kwix.com.sg/>
- MokoMomo Interactive Pte Ltd <http://www.mokomomo.com/>
- STOB Pte Ltd
- Singapore Carpentry <http://carpentry.sg/>
- Social@Work <http://socialatwork.com/>
- SqrDots Pte Ltd <http://50.56.206.127/info/about>
- Streetgaga <https://streetgaga.com/>
- Tiny Whale <http://tinywhale.net/>
- Visenze Pte Ltd <http://www.visenze.com/>

Companies previously hosted at the SoC Incubation Centre:

- 51tuofu <http://www.51tuofu.com/>
- Audiary Pte Ltd <http://www.audiary.com/>
- BigDeal Pte Ltd <http://plus.bigdeal.sg/>
- Daylight Studios Pte Ltd <http://www.day-lightstudios.com/>
- Decision-Ware Simulations & Games Ltd < http://www.decisionwaresim.com/>
- Dream Axis Pte Ltd <http://www.dreamaxis.com.sg/>
- FlickEvents <http://www.flickevents.com/>
- FriarTuck Pte Ltd <http://www.friartuck.net/>
- HuluTech Pte Ltd <http://www.hulutech.com/>
- Innov Systems Pte Ltd <http://www.innvo.com/>
- LD-Tech Pte Ltd < http://www.ld-tech.com/>
- mChron Pte Ltd < http://www.mchron.com/>
- Nineo.com Pte Ltd <http://www.nineo.com/>
- Novatap Private Ltd <http://www.novatap.com/>
- Orango Pte Ltd <http://www.giftail.com/>
- PrivyTV <http://www.privytv.com/>
- SGEye Pte Ltd <http://www.sgeye.com/>
- Smoov Pte Ltd <http://corporate.smoov.me/index.php>
- tenCube Pte Ltd <https://www.wavesecure.com/>
- Thenewlead.com Pte Ltd <http://www.thenewlead.com/>
- Thothe Technologies Pte Ltd <http://www.thothe.com/index.html>
- Vinova Pte Ltd <http://vinova.sg/>

Appendices

FACULTY MEMBERS

(Department of **Computer Science**)

Faculty Members	Academic Qualifications	Current Research Interests
AKKIHEBBAL L. Ananda Professor of Practice Email: ananda@comp.nus.edu.sg http://www.comp.nus.edu.sg/~ananda	PhD (University of Manchester, UK) MSc (University of Manchester, UK) MTech (IIT Kanpur, India) BE (Bangalore University, India)	Performance of transport protocols, wireless and sensor networks, and IPv4 to IPv6 transition mechanisms.
ANDERSON Norman Hugh Associate Professor Email: hugh@comp.nus.edu.sg http://www.comp.nus.edu.sg/~hugh	PhD (National University of Singapore) MSc (University of the South Pacific, Fiji Islands) BSc (Victoria University of Wellington, New Zealand)	Formal methods, refinement, tool support, network administration, distributed systems, encryption
ANG Chuan Heng Senior Fellow Email: angch@comp.nus.edu.sg http://www.comp.nus.edu.sg/~angch	PhD (University of Maryland, USA) MSc (Nanyang University, Singapore) BSc (Nanyang University, Singapore)	Quadtree related data structures and algorithms, iconic image representation and retrieval
BRESSAN Stephane Associate Professor Email: steph@comp.nus.edu.sg http://www.comp.nus.edu.sg/~steph	PhD (University of Lille, France) MSc (University of Lille, France) Engineering Degree (Ecole Universitaire d'Ingenieur de Lille, France) BSc (University of Lille, France)	Integration and management of information from heterogeneous, distributed and autonomous information sources
BROWN Michael Scott Associate Professor Email: brown@comp.nus.edu.sg http://www.comp.nus.edu.sg/~brown	PhD (University of Kentucky, USA) BSc (University of Kentucky, USA)	Interactive computer-vision, camera-projector display systems, document imaging and processing
CHAN Chee Yong Associate Professor Email: chancy@comp.nus.edu.sg http://www.comp.nus.edu.sg/~chancy	PhD (University of Wisconsin at Madison, USA) MSc (National University of Singapore) BSc (Hons) (National University of Singapore)	Database usability, XML databases, query processing and optimization
CHAN Mun Choon Associate Professor Email: chanmc@comp.nus.edu.sg http://www.comp.nus.edu.sg/~chanmc	PhD (Columbia University, USA) MPhil (Columbia University, USA) MS (Columbia University, USA) BSc (Purdue University, USA)	Wireless networking, sensor networks, disruption tolerant networks, mobile computing
CHANG Ee Chien Associate Professor Email: changec@comp.nus.edu.sg http://www.comp.nus.edu.sg/~changec	PhD (Courant Institute of Mathematical Sciences, New York University, USA) MSc (National University of Singapore) BSc (Hons) (National University of Singapore)	Network security, applied cryptography, multimedia/ biometric security and multimedia delivery
CHIN Wei Ngan Associate Professor Email: chinwn@comp.nus.edu.sg http://www.comp.nus.edu.sg/~chinwn	PhD (Imperial College, University of London, UK) MSc (University of Manchester, UK) BSc (Hons) (University of Manchester, UK)	Automated program verification, secure and dependable software, type systems and program analyses, software specification and models, programming language design
CHUA Tat Seng Professor Email: chuats@comp.nus.edu.sg http://www.comp.nus.edu.sg/~chuats	PhD (University of Leeds, UK) BSc (University of Leeds, UK)	Multimedia information retrieval and social media analysis, research focus in extraction, retrieval and question-answering (QA) of text, video and live media arising from the web and social networks
DONG Jin Song Associate Professor Email: dongjs@comp.nus.edu.sg http://www.comp.nus.edu.sg/~dongjs	PhD (University of Queensland, Australia) BInfTech (Hons) (University of Queensland, Australia)	Design analysis and verification, context awareness and pervasive computing, real-time concurrent system specification, web semantics, services, agent and reasoning, formal methods and safety critical systems, object, component, and language semantics
GILBERT Seth Dean's Chair Assistant Professor Email: gilbert@comp.nus.edu.sg http://www.comp.nus.edu.sg/~gilbert	PhD (Massachusetts Institute of Technology, USA) MSc (Massachusetts Institute of Technology, USA) BSc (Yale University, USA)	Algorithms for highly dynamic, distributed environments, such as mobile ad hoc networks, wireless sensor networks, and peer-to-peer systems.
HENZ Martin Associate Professor Email: henz@comp.nus.edu.sg http://www.comp.nus.edu.sg/~henz	Dr. rer. nat. (Universitat des Saarlandes, Germany) MSc (State University of New York at Stony Brook, USA)	Combinatorial optimization, defining evolution, client-centered web-based computing

Faculty Members	Academic Qualifications	Current Research Interests
HSU David Associate Professor Email: dyhsu@comp.nus.edu.sg http://www.comp.nus.edu.sg/~dyhsu	PhD (Stanford University, USA) BSc (University of British Columbia, Canada)	Robotics, artificial intelligence and computational biology
HSU Wynne Professor Email: whsu@comp.nus.edu.sg http://www.comp.nus.edu.sg/~whsu	PhD (Purdue University, USA) MSc (Purdue University, USA) BSc (National University of Singapore)	Spatial, temporal, spatio-temporal data mining, recommender systems, medical image analysis, machine learning
JAFFAR Joxan Shaw Chair Professor Email: joxan@comp.nus.edu.sg http://www.comp.nus.edu.sg/~joxan	PhD (Monash University, Australia) MSc (University of Melbourne, Australia) BSc (Hons) (University of Melbourne, Australia)	Programming languages and applications, with emphasis on the logic and constraint programming paradigms
JAIN Rahul Assistant Professor, Joint Appt (CQT) Email: rahul@comp.nus.edu.sg http://www.comp.nus.edu.sg/~rahul	PhD (Tata Institute of Fundamental Research, India) B.Tech (Indian Institute of Technology, Mumbai, India)	Information theory, quantum computation, communication complexity, complexity theory, cryptography
JAIN Sanjay Provost's Chair Professor Email: sanjay@comp.nus.edu.sg http://www.comp.nus.edu.sg/~sanjay	PhD (University of Rochester, USA) MS (University of Rochester, USA) BTech (IIT Kharagpur, India)	Inductive inference, computational learning theory, structural complexity, recursion theory and parallel algorithms
JARZABEK Stanislaw Associate Professor Email: stan@comp.nus.edu.sg http://www.comp.nus.edu.sg/~stan	PhD (Warsaw University, Poland) MSc (Warsaw University, Poland)	Software engineering, software reuse and evolution, software clone detection
KAN Min Yen Associate Professor Email: kanmy@comp.nus.edu.sg http://www.comp.nus.edu.sg/~kanmy	PhD (Columbia University, USA) MS (Columbia University, USA) BS (Columbia University, USA)	Digital libraries, natural language processing and information retrieval
KANKANHALLI Mohan Professor Email: mohan@comp.nus.edu.sg http://www.comp.nus.edu.sg/~mohan	PhD (Rensselaer Polytechnic Institute, USA) MS (Rensselaer Polytechnic Institute, USA) BTech (IIT Kharagpur, India)	Multimedia information systems, information security and surveillance, digital video processing
KHOO Siau Cheng Associate Professor Email: khoosc@comp.nus.edu.sg http://www.comp.nus.edu.sg/~khoosc	PhD (Yale University, USA) MPhil (Yale University, USA) MSc (National University of Singapore) BSc (National University of Singapore)	Specification mining, static and dynamic program analysis, program transformation, functional programming, domain-specific languages, aspect-oriented programming
LEE Mong Li, Janice Associate Professor Email: leeml@comp.nus.edu.sg http://www.comp.nus.edu.sg/~leeml	PhD (National University of Singapore) MSc (National University of Singapore) BSc (Hons) (National University of Singapore)	Cleaning and integration of heterogeneous and semi-structured data, database performance issues in dynamic environments, medical informatics.
LEE Wee Sun Associate Professor Email: leews@comp.nus.edu.sg http://www.comp.nus.edu.sg/~leews	PhD (Australian National University, Australia) BEng (Hons) (University of Queensland, Australia)	Machine learning, planning under uncertainty
LEONG Hon Wai Associate Professor Email: leonghw@comp.nus.edu.sg http://www.comp.nus.edu.sg/~leonghw	PhD (University of Illinois at Urbana-Champaign, USA) BSc (University of Malaya, Malaysia)	Design and analysis of efficient algorithms for optimization problems, transportation logistics, multimedia video processing, bioinformatics and computational biology
LEONG Tze Yun Associate Professor Email: leongty@comp.nus.edu.sg http://www.comp.nus.edu.sg/~leongty	PhD (Massachusetts Institute of Technology, USA) SM (Massachusetts Institute of Technology, USA) SB (Massachusetts Institute of Technology, USA)	Decision-theoretic artificial intelligence, temporal probabilistic reasoning, machine learning, cognitive modeling, adaptive computing, and biomedical informatics
LEONG Wing Lup, Ben Assistant Professor Email: bleong@comp.nus.edu.sg http://www.comp.nus.edu.sg/~bleong/	PhD (Massachusetts Institute of Technology, USA) MEng (Massachusetts Institute of Technology, USA) SB (Massachusetts Institute of Technology, USA)	Computer networking, network architecture, overlay networks, peer-to-peer networking, routing algorithms, network coding, distributed systems, organization, architecture and programming models, networked games, algorithms for self-organization, novel data structures

Faculty Members	Academic Qualifications	Current Research Interests
LEOW Wee Kheng Associate Professor Email: leowwk@comp.nus.edu.sg http://www.comp.nus.edu.sg/~leowwk	PhD (University of Texas at Austin, USA) MSc (National University of Singapore) BSc (National University of Singapore)	Computer vision, medical image analysis, and protein docking
LIANG Zhenkai Assistant Professor Email: liangzk@comp.nus.edu.sg http://www.comp.nus.edu.sg/~liangzk	PhD (Stony Brook University, USA) MSc (Stony Brook University, USA) BSc (Peking University, China)	System and software security, web security, program analysis, and software engineering
LING Tok Wang Professor Email: lingtw@comp.nus.edu.sg http://www.comp.nus.edu.sg/~lingtw	PhD (Waterloo University, Canada) MMath (Waterloo University, Canada) BSc (Hons) (Nanyang University, Singapore)	Data modeling, entity-relationship approach, object-oriented data model, normalization theory, logic and database, integrity constraint checking, semi-structured data model, XML twig pattern query processing, and XML keyword query processing
LOW Kian Hsiang, Bryan Assistant Professor Email: lowkh@comp.nus.edu.sg http://www.comp.nus.edu.sg/~lowkh	PhD (Carnegie Mellon University, USA) MSc (National University of Singapore) BSc (Hons) (National University of Singapore)	Adaptive sampling and active learning, multi-agent systems(i.e., multi-agent coordination, planning, and learning), statistical machine learning, optimization, game theory
LOW Kok Lim Assistant Professor Email: lowkl@comp.nus.edu.sg http://www.comp.nus.edu.sg/~lowkl/	PhD (University of North Carolina at Chapel Hill, USA) MSc (University of North Carolina at Chapel Hill, USA) BSc (National University of Singapore)	Computer graphics with focus in computational art, computational photography/Image manipulation, real-time 3D rendering
Ma Tianbai, Richard Assistant Professor, Joint Appt (ADSC) Email: tbma@comp.nus.edu.sg http://www.comp.nus.edu.sg/~tbma/	PhD (Columbia University, USA) Mphil. (Chinese University of Hong Kong, Hong Kong) BSC. (Chinese University of Hong Kong, Hong Kong)	Economics of the Internet, information centric networking, performance evaluation
MITRA Tulika Associate Professor Email: tulika@comp.nus.edu.sg http://www.comp.nus.edu.sg/~tulika	PhD (State University of New York at Stony Brook, USA) MEng (Indian Institute of Science at Bangalore, India) BE (Jadavpur University, India)	Compilers and architectures for embedded systems, application-specific processors design, software timing, power, thermal analysis, and optimization.
MOTIWALLA Juzar Professor of Practice Email: juzar@comp.nus.edu.sg http://www.comp.nus.edu.sg/~juzar	PhD (University of Wisconsin, USA) MSc (University of Wisconsin, USA) BBA (University of Singapore)	Business strategy in digital markets, innovation, digital entrepreneurship
NG Hwee Tou Professor Email: nght@comp.nus.edu.sg http://www.comp.nus.edu.sg/~nght	PhD (University of Texas at Austin, USA) MS (Stanford University, USA) BSc (University of Toronto, Canada)	Natural language processing, information retrieval
NG Teck Khim Associate Professor of Practice Email: ngtk@comp.nus.edu.sg http://www.comp.nus.edu.sg/~ngtk/	PhD (Carnegie Mellon University, USA) MSc (National University of Singapore) BSc (National University of Singapore)	Geometrical computer vision such as 3D reconstruction from images, as well as audio processing, and applications on mobile devices
OOI Beng Chin Professor Email: ooibc@comp.nus.edu.sg http://www.comp.nus.edu.sg/~ooibc	PhD (Monash University, Australia) BSc (Hons) (Monash University, Australia)	Database system architectures, performance issues, indexing techniques and query processing, multimedia and spatio-temporal, distributed, parallel, P2P, and cloud database systems and applications
OOI Wei Tsang Associate Professor Email: ooiwt@comp.nus.edu.sg http://www.comp.nus.edu.sg/~ooiwt	PhD (Cornell University, USA) BSc (Hons) (National University of Singapore)	Systems support for multimedia applications: video streaming, graphics streaming, networked virtual environment.
PUNG Hung Keng Associate Professorial Fellow Email: punghk@comp.nus.edu.sg http://www.comp.nus.edu.sg/~punghk	PhD (University of Kent at Canterbury, UK) BSc (University of Kent at Canterbury, UK)	Mobile and pervasive computing, service oriented systems architecture, context aware pervasive computing
ROSENBLUM, David S. Professor Email: david@comp.nus.edu.sg http://www.comp.nus.edu.sg/~david/	PhD (Standford University, USA) MS (Stanford Unviersity, USA) MS (North Texas University USA) BS *North Texas State University, USA)	Ubiquitous computing, scalability of software systems, publish/subscribe communication, probabilistic modeling of distributed systems, internet-scale software technologies, software architecture, software testing

Faculty Members	Academic Qualifications	Current Research Interests
ROYCHOUDHURY Abhik Associate Professor Email: abhik@comp.nus.edu.sg http://www.comp.nus.edu.sg/~abhik	PhD (State University of New York at Stony Brook, USA) MS (State University of New York at Stony Brook, USA) BE (Jadavpur University, India)	Software testing and analysis, design and analysis of embedded software and systems, trustworthy computing
SAXENA, Prateek Dean's Chair Assistant Professor Email: prateeks@comp.nus.edu.sg http://www.comp.nus.edu.sg/~prateeks/	PhD (University of California, USA) MS (Stony Brook University, USA) BE(University of Pune, India)	Web security and binary analysis, computer security, intersection with programming languages, verification, compilers and operating systems
SIM Khe Chai Assistant Professor Email: simkc@comp.nus.edu.sg http://www.comp.nus.edu.sg/~simkc	PhD (University of Cambridge, UK) MEng (University of Cambridge, UK) MPhil (University of Cambridge, UK) BA (University of Cambridge, UK)	Statistical pattern classification and acoustic modelling for automatic speech recognition
SIM Mong Cheng, Terence Associate Professor Email: tsim@comp.nus.edu.sg http://www.comp.nus.edu.sg/~tsim	PhD (Carnegie Mellon University, USA) MSc (Stanford University, USA) SB (Massachusetts Institute of Technology, USA)	Facial analysis, recognition and verification. Modeling, animation, rendering. Biometrics. Pattern recognition. Computational photography. Computer vision. Image processing
STEPHAN Frank Christian Professor, Joint Appt (Math) Email: fstephan@comp.nus.edu.sg http://www.comp.nus.edu.sg/~fstephan/	PhD (University of Karlsruhe, Germany) BSc (University of Karlsruhe, Germany)	Recursion theory, learning theory, computational complexity, automata and formal languages
SUNG Wing Kin, Ken Associate Professor Email: ksung@comp.nus.edu.sg http://www.comp.nus.edu.sg/~ksung	PhD (University of Hong Kong, Hong Kong) BSc (University of Hong Kong, Hong Kong)	Computational biology, algorithm
TAN Chew Lim Professor Email: tancl@comp.nus.edu.sg http://www.comp.nus.edu.sg/~tancl	PhD (University of Virginia, USA) MSc (University of Surrey, UK) BSc (Hons) (University of Singapore)	Document image analysis, text mining, natural language processing
TAN Kian Lee Provost's Chair Professor Email: tankl@comp.nus.edu.sg http://www.comp.nus.edu.sg/~tankl	PhD (National University of Singapore) MSc (National University of Singapore) BSc (Hons) (National University of Singapore)	Multimedia information retrieval, query processing and optimization in multiprocessor, distributed systems and P2P systems, database performance, and database security and privacy
TAN Soon Huat, Gary Associate Professor Email: gtan@comp.nus.edu.sg http://www.comp.nus.edu.sg/~gtan	PhD (University of Manchester, UK) MSc (University of Manchester, UK) BSc (Hons) (National University of Singapore)	Parallel and distributed computing, scheduling and load balancing, declarative multiprocessors, parallel and distributed (interactive) simulation and high level architecture
TAN Sun Teck Associate Professor Email: tanst@comp.nus.edu.sg http://www.comp.nus.edu.sg/~tanst	PhD (Essex University, UK) MSc (Essex University, UK) BSc (Nanyang University, Singapore)	Multimedia computer-based tutoring system, artificial intelligence, Chinese computing
TAN Tiow Seng Associate Professor Email: tants@comp.nus.edu.sg http://www.comp.nus.edu.sg/~tants	PhD (University of Illinois at Urbana-Champaign, USA) MSc (National University of Singapore) BSc (National University of Singapore)	Computational geometry; algorithms and data structures, GPU algorithm
TAY Yong Chiang Professor, Joint Appt (Math) Email: tayyc@comp.nus.edu.sg http://www.comp.nus.edu.sg/~tayyc	PhD (Harvard University, USA) BSc (Hons) (University of Singapore)	Performance modeling, distributed computing, database systems.
TEO Yong Meng Associate Professor Email: teoym@comp.nus.edu.sg http://www.comp.nus.edu.sg/~teoym	PhD (University of Manchester, UK) CEng (Engineering Council, UK) MSc (University of Manchester, UK) BTech (Hons) (University of Bradford, UK)	Parallel and distributed computing including parallel & distributed simulation, grid computing, parallel computer architecture, and performance evaluation.
THIAGARAJAN P.S. Professor Email: thiagu@comp.nus.edu.sg http://www.comp.nus.edu.sg/~thiagu	PhD (Rice University, USA) MS (Rice University, USA) BTech (IIT Madras, India)	System-level design and analysis methods for real time embedded systems and computational systems biology

Faculty Members	Academic Qualifications	Current Research Interests
TUNG Kum Hoe, Anthony Associate Professor Email: atung@comp.nus.edu.sg http://www.comp.nus.edu.sg/~atung	PhD (Simon Fraser University, Canada) MSc (National University of Singapore) BSc (Hons) (National University of Singapore)	Databases, data mining, big data processing, data analytics and visual analytics
WANG Ye Associate Professor Email: wangye@comp.nus.edu.sg http://www.comp.nus.edu.sg/~wangye	PhD (Tampere University of Technology, Finland) Diplom-Ingenieur (Technische Universität Braunschweig, Germany) BSc (Southern China University of Technology, China)	Mobile computing, sound and music computing, music information retrieval and eHealth
WEHNER Stephanie Assistant Professor, Joint Appt (CQT) Email: wehner@comp.nus.edu.sg http://www.comp.nus.edu.sg/~wehner	PhD (University of Amsterdam, Netherlands) MSc (University of Amsterdam, Netherlands) BSc (University of Amsterdam, Netherlands)	Quantum cryptography, quantum information theory and complexity theory
WONG Lim Soon Provost's Chair Professor Email: wongls@comp.nus.edu.sg http://www.comp.nus.edu.sg/~wongls/	PhD (University of Pennsylvania, USA) MSc (University of Pennsylvania, USA) BSc (Imperial College London, UK)	Query language theory, finite model theory, bioinformatics, knowledge discovery
WONG Weng Fai Associate Professor Email: wongwf@comp.nus.edu.sg http://www.comp.nus.edu.sg/~wongwf	DrEngSc (University of Tsukuba, Japan) MSc (National University of Singapore) BSc (Hons) (National University of Singapore)	Embedded systems, compilers, computer architecture, GPGPU and parallel processing
YAP Hock Chuan, Roland Associate Professor Email: ryap@comp.nus.edu.sg http://www.comp.nus.edu.sg/~ryap	PhD (Monash University, Australia) MSc (Monash University, Australia) BSc (Hons) (Monash University, Australia)	Programming languages, applications, computational molecular biology, computational finance, medical informatics
YIN Kang Kang Assistant Professor Email: kkyin@comp.nus.edu.sg http://ww.comp.nus.edu.sg/~kkyin	PhD (University of British Columbia, Canada) MSc (Zhejiang University, China) BSc (Zhejiang University, China)	Computer animation and simulation, especially responsive character animation and real-time humanoid motion control
YU Haifeng Associate Professor Email: yuhf@comp.nus.edu.sg http://www.comp.nus.edu.sg/~yuhf/	PhD (Duke University, USA) MSc (Duke University, USA) BEng (Shanghai Jiao Tong University, China)	Distributed computing, distributed algorithms, communication complexity, applied algorithms in networking, and distributed systems security
ZHAO Shengdong Assistant Professor Email: zhaosd@comp.nus.edu.sg http://www.comp.nus.edu.sg/~zhaosd/	PhD (University of Toronto, Canada) MSc (University of California at Berkley, USA) BSc (Linfield College, USA)	Human-computer interaction
ZIMMERMANN Roger Associate Professor Email: rogerz@comp.nus.edu.sg http://www.comp.nus.edu.sg/~rogerz/	PhD (University of Southern California, USA) MSc (University of Southern California, USA	Distributed and peer-to-peer systems, collaborative environments, streaming media architectures, georeferenced video search, and mobile location-based services

FACULTY MEMBERS

(Department of **Information Systems**)

Faculty Members	Academic Qualifications	Current Research Interests
CHAN Hock Chuan Associate Professor Email: chanhc@comp.nus.edu.sg http://www.comp.nus.edu.sg/is/bio/chanhc.html	PhD (University of British Columbia, Canada) MA (University of Cambridge, UK) BA (Hons) (University of Cambridge, UK)	User-database interaction, spreadsheet visualization, information systems acceptance.
CHANG Ting Ting Assistant Professor Email: changtt@comp.nus.edu.sg http://www.comp.nus.edu.sg/is/bio/changtt.html	PhD (Carnegie-Mellon University, USA) MSc (Carnegie-Mellon University, USA) MSc (National University of Singapore) BSc (Hons) (National University of Singapore)	Psychological contracts in knowledge sharing, virtual teams, social computing technologies
CHEN Yuanyuan Assistant Professor Email: chenyy@comp.nus.edu.sg http://www.comp.nus.edu.sg/is/bio/chenyy.html	PhD (Emory University, USA) LL.M (Emory University, USA) LL.M (National Huaqiao University, China)	Global sourcing, pricing and piracy of digital goods, Information poaching and data security, economics of technology and law
DATTA, Anindya Associate Professor Email: datta@comp.nus.edu.sg	PhD (Maryland University, USA) MS (Maryland University, USA) BTech (Indian Institute of Technology, India)	Big data, data mining, mobile systems, social networks
DUTTA, Kaushik Associate Professor Email: dutta@comp.nus.edu.sg http://www.comp.nus.edu.sg/is/bio/dutta.html	PhD (Georgia Institute of Technology, USA) BS Electrical Engineering (Jadavpur University, India) MS Computer Science (Indian Statistical Institute, India)	Energy efficiency of IT infrastructure and applications, data mining and big data management, mega-scale IT infrastructure
FAIK, Isam Assistant Professor Email: faik@comp.nus.edu.sg	PhD (University of Cambridge, UK) 2011 M.Eng (McGill University, Canada) 2005 B.Eng (McGill University, Canada) 2002	Practice-based and critical approaches to the study of the dynamics between organisational and technological change
GOH Khim Yong Assistant Professor Email: gohky@comp.nus.edu.sg http://www.comp.nus.edu.sg/is/bio/gohky.html	PhD (University of Chicago, USA) MSc (National University of Singapore) BSc (Hons) (National University of Singapore)	Direct marketing and internet marketing, competitive pricing and advertising, empirical models of Information technology and industrial organization, econometric and Markov Chain Monte Carlo methods
HAHN Jungpil Associate Professor Email: jungpil@comp.nus.edu.sg http://www.comp.nus.edu.sg/is/bio/jungpil.html	PhD (University of Minnesota, USA) MBA (Yonsei University, South Korea) BBA (Yonsei University, South Korea)	Software development projects and processes, organizational learning, knowledge management systems, open innovation and collaborative problem solving, computational organizational theory
HENG Cheng Suang Assistant Professor Email: hengcs@comp.nus.edu.sg http://www.comp.nus.edu.sg/is/bio/hengcs.html	PhD (Stanford University, USA) MSc (National University of Singapore) BSc (Hons) (National University of Singapore)	Organization strategies & decision making , organizations & social media , IT entrepreneurship, IT outsourcing & offshoring
HUANG Ke-Wei Assistant Professor Email: huangkw@comp.nus.edu.sg http://www.comp.nus.edu.sg/is/bio/huangkw.html	PhD (New York University, USA) M.Phil (New York University, USA) MSc (New York University, USA) MBA (National Taiwan University) BSc (National Taiwan University)	Economics of information systems, e-commerce, pricing ,theoretical models of industrial organization
JIANG Zhenhui, Jack Associate Professor Email: jiang@comp.nus.edu.sg http://www.comp.nus.edu.sg/is/bio/jiang.html	PhD (University of British Columbia, Canada) MBA (Tsinghua University, China) BEcon (Tsinghua University, China) BEng (Tsinghua University, China)	Electronic commerce, human-computer interaction, recommendation agents, information privacy
KANKANHALLI Atreyi Associate Professor Email: atreyi@comp.nus.edu.sg http://www.comp.nus.edu.sg/is/bio/atreyi.html	PhD (National University of Singapore) MSc (Rensselaer Polytechnic Institute, USA) BTech (IIT Delhi, India)	Knowledge management, IT innovation adoption and change management, IT-enabled innovation in service sectors (e.g., Government, Healthcare)
KIM Seung Hyun Assistant Professor Email: kimsh@comp.nus.edu.sg http://www.comp.nus.edu.sg/is/bio/kimsh.html	PhD (Carnegie-Mellon University, USA) MSc (Carnegie-Mellon University, USA)	Economics of information systems, business value of IT, knowledge management, customer relationship management, information security

Faculty Members	Academic Qualifications	Current Research Interests
LIM Lai Huat, John Associate Professor Email: jlim@comp.nus.edu.sg http://www.comp.nus.edu.sg/is/bio/jlim.html	PhD (University of British Columbia, Canada) MSc (National University of Singapore) BEng (Hons) (National University of Singapore)	Multi-lingual and cross-cultural aspects of negotiation support
LUN Kwok Chan Professorial Fellow Email: lunkc@comp.nus.edu.sg http://www.comp.nus.edu.sg/is/bio/lunkc.html	PhD (University of Birmingham, UK) MSc (London School of Hygiene & Tropical Medicine, UK) BSc (Hons) (University of Singapore)	Biomedical informatics with special interest in telehealth, biostatistics with special interest in clinical trials, medical demography with special interest in population ageing
PAN Shan Ling Associate Professor Email: pansl@comp.nus.edu.sg http://www.comp.nus.edu.sg/is/bio/pansl.html	PhD (University of Warwick, UK) MBA (University of Texas at San Antonio, USA) MA (University of London, UK) BSBA (Southeast Missouri State University, USA)	IT Strategies in China, eco-business system of IT industry, strategic enterprise systems
PHAN Tuan Quang Assistant Professor Email: phantq@comp.nus.edu.sg http://www.tuanqphan.us	DBA (Harvard Business School, USA) BSc (Massachusetts Institute of Technology, USA)	Social networks, social media, Big Data, product diffusion, word-of-mouth, and web and mobile commerce
POO Chiang Choon, Danny Associate Professor Email: dpoo@comp.nus.edu.sg http://www.comp.nus.edu.sg/is/bio/dpoo.html	PhD (University of Manchester, UK) MSc (University of Manchester, UK) BSc (Hons) (University of Manchester, UK)	Software engineering, information management, knowledge management
SETIONO Rudy Associate Professor Email: rudys@comp.nus.edu.sg	PhD (University of Wisconsin-Madison, USA) MSc (Eastern Michigan University, USA) BSc (Eastern Michigan University, USA)	Neural networks, data mining, nonlinear optimization
TAN Cheng Yian, Bernard Professor Email: btan@comp.nus.edu.sg http://www.comp.nus.edu.sg/is/bio/btan.html	PhD (National University of Singapore) MSc (National University of Singapore) BSc (Hons) (National University of Singapore)	Cross-cultural issues in information systems, computer-mediated communication, knowledge management, information privacy, software project management
TAN Swee Lin, Sharon Assistant Professor Email: tansl@comp.nus.edu.sg http://www.comp.nus.edu.sg/is/bio/tansl.html	PhD (Carnegie Mellon University, USA) MSc (Carnegie Mellon University, USA) MSc (National University of Singapore) BSc (National University of Singapore)	Knowledge management and work interruption, and healthcare information technology
TEO Hock Hai Associate Professor Email: teohh@comp.nus.edu.sg http://www.comp.nus.edu.sg/is/bio/teohh.html	PhD (National University Singapore) MSc (National University Singapore) BSc (Hons) (National University Singapore)	Electronic commerce, IT innovation adoption, information privacy, user-database interactions
WOON Mei Yen, Irene Associate Professor Email: iwoon@comp.nus.edu.sg http://www.comp.nus.edu.sg/is/bio/iwoon.html	PhD (Aston University, UK) MSc (Aston University, UK) BBA (University of Singapore)	Organizational and human factors in software project management

GRADUATE RESEARCH DEGREES AWARDED

(Department of Computer Science)

Name	Degree	Supervisor	Thesis Title
Achudhan Sivakumar	PhD	Tan Keng Yan, Colin	UAV Swarm Coordination and Control for Establishing Wireless Connectivity
Bao Zhifeng	PhD	Ling Tok Wang	Towards an Effective Processing of XML Keyword Query
Bhojan Anand	PhD	Ananda A.L.	Energy Efficient Algorithms and Techniques for Wireless Mobile Clients
Cao Jianneng	PhD	Tan Kian Lee	Privacy-Preserving Data Publication for Static and Streaming Data
Cao Yu	PhD	Tan Kian Lee Chan Chee Yong	Optimizing Complex Queries with Multiple Relational Instances
Chen Binbin	PhD	Chan Mun Choon	Cooperative Internet Access Using Heterogeneous Wireless Networks
Chen Ding	PhD	Chan Chee Yong	XML Query Optimization
Chen Su	PhD	Ooi Beng Chin	On Optimizing Moving Object Databases
Cheng Wei	PhD	Ooi Wei Tsang	Streaming of High-resolution Progressive Meshes Over the Internet
Chia Wai Kit, Henry	PhD	Tan Chew Lim	Cognitive-inspired Approaches to Neural Logic Network Learning
Chiang Tsung Han	PhD	David Hsu	Markov Dynamic Models for Long-timescale Protein Motion
Chidansh Amitkumar Bhatt	PhD	Mohan Kankanhalli	Probabilistic Temporal Multimedia Datamining
Chong Ket Fah	PhD	Leong Hon Wai	New Models and Algorithm for De Novo Peptide Sequencing of Multi-charge MS/MS Spectra
Claudia Szabo	PhD	Teo Yong Meng	Composable Simulation Models and Their Formal Validation
Dai Bingtian	PhD	Tung Kum Hoe, Anthony	On Column Heterogeneity
David Cristina Mariana	PhD	Chin Wei Ngan	Enhanced Specification Expressivity for Verification with Separation Logic
Deepak Gangadharan	PhD	Roger Zimmermann	Quality-aware Performance Analysis for Multimedia MPSoC Platforms
Ding Feng	PhD	Leow Wee Kheng	3D Segmentation of Soft Tissues by Flipping-free Mesh Deformation
Dong Difeng	PhD	Wong Lim Soon	Relapse Prediction in Childhood Acute Lymphoblastic Leukemia by Time-series Gene Expression Profiling
Ehsan Rehman	PhD	David Hsu	Bounded Uncertainty Roadmaps
Felix Halim	PhD	Yap Hock Chuan, Roland	Solving Big Data Problems from Sequences to Tables and Graphs
FENG YUZHANG	PhD	Dong Jinsong	Reasoning About Complex Agent Knowledge - Ontologies, Uncertainty, Rules and Beyond
Gherghina Cristian Andrei	PhD	Chin Wei Ngan	Efficiently Verifying Programs with Rich Control Flows
Gong Tianxia	PhD	Tan Tiow Seng	Automatic Annotation, Classification and Retrieval of Traumatic Brain Injury CT Images
GUO DONG	PhD	Sim Mong Cheng, Terence	A Collection of Digital Photo Editing Methods
Hagiescu Miriste Andrei Mihai	PhD	Wong Weng Fai	A Computing Origami: Optimized Code Generation for Emerging Parallel Platforms
Harish Katti	PhD	Mohan Kankanhalli Chua Tat Seng	Human Visual Perception, Study and Applications to Understanding Images and Videos
Hugo Willy	PhD	Sung Wing Kin, Ken Ng See Kiong	On Interaction Motif Inference from Biomolecular Interactions: Riding the Growth of the High Throughput Sequential and Structural Data
Ju Lei	PhD	Abhik Roychoudhury	Model-driven Timing Analysis of Embedded Software
Lam Soon Lee, Edmund	PhD	Dong Jinsong	Parallel Execution of Constraint Handling Rules - Theory, Implementation and Application
Lau Qiangfeng Peter	PhD	Wynne Hsu	Effective Reinforcement Learning for Collaborative Multi-agent Domains
Lee Wah Heng Charlie	PhD	Sung Wing Kin, Ken	Bioinformatic Applications for Virology Research
Li Hao	PhD	Leow Wee Kheng	Predictive Surgical Simulation for Preoperative Planning of Complex Cardioviscular Surgeries
Liang Ke	PhD	Roger Zimmermann	Peer-to-peer Interactive 3D Media Dissemination in Networked Virtual Environments

Name	Degree	Supervisor	Thesis Title
Liang Yun	PhD	Tulika Mitra	Instruction Cache Optimizations for Embedded Systems
Lin Ziheng	PhD	Kan Min Yen Ng Hwee Tou	Discourse Parsing: Inferring Discourse Structure, Modeling Coherence, and Its Applications
Liu Ruizhe	PhD	Tan Chew Lim	Automatic Quantification of Brain Midline Shift in CT Images
Liu Yang	PhD	Dong Jinsong Rudy Setiono	Model Checking Concurrent and Real-time Systems: The PAT Approach
Lu Haiyun	PhD	Leow Wee Kheng	Knowledge-guided Docking of Flexible Ligands to Protein Domains
Lu Liming	PhD	Chan Mun Choon Chang Ee-Chien	Traffic Monitoring and Analysis for Source Identification
Lu Zheng	PhD	Michael Brown	High-resolution Imaging for E-Heritage
Marian Mihailescu	PhD	Teo Yong Meng	Strategy-proof Resource Pricing in Dederated Systems
Mukesh Kumar Saini	PhD	Mohan Kankanhalli	Privacy Aware Surveillance System Design
Ng Hoong Kee	PhD	Leong Hon Wai	Algorithms for Multi-point Range Query and Reverse Nearest Neighbour Search
Patel Dhavalkumar Chaturbhai	PhD	Wynne Hsu	Mining Patterns in Complex Data
Pavel Korshunov	PhD	Ooi Wei Tsang	Video Quality for Video Analysis
Peter Wittek	PhD	Tan Chew Lim	Compactly Supported Basis Functions as Support Vector Kernels: Capturing Feature Interdependence in the Embedding Space
Raman Balaji	PhD	Ooi Wei Tsang	Application-specific Workload Shaping in Resource-constrained Media Players
Ramkumar Jayaseelan	PhD	Tulika Mitra	Application-specific Thermal Management of Computer Systems
Sadegh Heyrani Nobari	PhD	Stephane Bressan	Scalable Data-Parallel Graph Algorithms from Generation to Management
Sandeep Kumar	PhD	Khoo Siau Cheng Abhik Roychoudhury	Mining Behavioral Specifications of Distributed Systems
Satish Kumar Verma	PhD	Ooi Wei Tsang	Techniques for Improving Predictability and Message Efficiency of Gossip Protocols
Saurabh Garg	PhD	Leow Wee Kheng	Modeling of Leaves and Constrained Leaf Morphing in Leaf Space
Shao Tao	PhD	Ananda A.L. Chan Mun Choon	Geographic Routing for Point to Point Data Delivery in Wireless Sensor Network
Sheng Chang	PhD	Wynne Hsu	Pattern Mining in Spatiotemporal Database
Sim Joon Edward	PhD	Wong Weng Fai	Hardware-software Codesign for Run-time Reconfigurable FPGA-based Systems
Sriganesh Maniganahalli Srihari	PhD	Leong Hon Wai	Integrating Biological Insights with Topological Characteristics for Improved Complex Prediction from Protein Interaction Networks
Steven Halim	PhD	Yap Hock Chuan, Roland	An Integrated White+Black Box Approach for Designing and Tuning Stochastic Local Search Algorithms
Su Bolan	PhD	Tan Chew Lim Lu Shijian	Document Image Enhancement
Sudipta Chattopadhyay	PhD	Abhik Roychoudhury	Time-predictable Execution of Embedded Software on Multi-core Platforms
Sun Jun	PhD	Tan Chew Lim	Kernel Engineering on Parse Trees
Tan Chek Tien	PhD	Cheng Holun, Alan	Adaptive Agent Architectures in Modern Virtual Games
Tan Hwee Xian	PhD	Chan Mun Choon	Communication Protocols for Energy Constrained Networks
Tan Yee Fan	PhD	Kan Min Yen	Cost-sensitive Web-based Information Acquisition for Record Matching
Tran Quoc Trung	PhD	Chan Chee Yong	Query by Output
Vo Hoang Tam	PhD	Ooi Beng Chin	Design of Efficient and Elastic Storage in the Cloud
Wang Jie	PhD	Tan Chew Lim	Image Registration: Features and Applications

Name	Degree	Supervisor	Thesis Title
Wang Kai	PhD	Chua Tat Seng	Retrieving Questions and Answers in Community-based Question Answering Services
Wang Nan	PhD	Tung Kum Hoe, Anthony	Dense Graph Pattern Mining and Visualization
Wang Xiangyu	PhD	Mohan Kankanahalli	Multimedia Decision Fusion
Wu Huayu	PhD	Ling Tok Wang	Using Semantics in XML Query Processing
Wu Ji	PhD	Tan Kian Lee	Exploring Time Related Issues in Data Stream Processing
Wu Sai	PhD	Ooi Beng Chin	Query Processing in Peer-to-peer Based Data Management System
Wu Xiuchao	PhD	Ananda A.L. Chan Mun Choon	Improving TCP Performance in the Mobile, High Speed, Heterogeneous and Evolving Internet
Wu Yongzheng	PhD	Yap Hock Chuan, Roland	Operating System Auditing and Monitoring
Xiang Shili	PhD	Tan Kian Lee	Multiple Query Optimization in Wireless Sensor Networks
Xu Jia	PhD	Chang Ee-Chien	Towards Efficient Proofs of Storage and Verifiable Outsourced Database in Cloud Computing
Xu Liang	PhD	Ling Tok Wang	Labeling Dynamic XML Documents: An Order-centric Approach
Xue Mingqiang	PhD	Pung Hung Keng	Privacy Protection via Anonymization for Publishing Multi-type Data
Xue Yinxing	PhD	Stanislaw Jarzabek	Reengineering Legacy Software Products into Software Product Line
Yang Xiaoyan	PhD	Ooi Beng Chin	Extracting Schemas for Database Exploration
Ye Ning	PhD	Sim Mong Cheng, Terence	Personal Identification from Facial Expression
Yin Hongli	PhD	Leong Tze Yun	A Model Driven Approach to Imbalanced Data Learning
Yu Jianxing	PhD	Chua Tat Seng	Hierarchical Organization of Consumer Reviews for Products and Its Applications
Zhang Bingjun	PhD	Wang Ye	Adaptive Multimodal Fusion Based Similarity Measures in Music Information Retrieval
Zhang Dongxiang	PhD	Tung Kum Hoe, Anthony	Efficient Location-based Spatial Keyword Query Processing
Zhang Mingze	PhD	Chan Mun Choon Ananda A.L.	Coverage and Connectivity Management in Wireless Sensor Networks
Zhang Xian	PhD	Dong Jinsong	Verification of Timed Process Algebra and Beyond
Zhang Xiaopeng	PhD	Sim Mong Cheng, Terence	Gradient Variation: A Key to Enhancing Photographs Across Illumination
Zhang Zhenjie	PhD	Tung Kum Hoe, Anthony	Local Bounding Technique and Its Applications to Uncertain Clustering
Zhang Zhiyong Melvin	PhD	Leong Hon Wai	Conserved Gene Cluster Discovery and Applications in Comparative Genomics
Zhao Shanheng	PhD	Ng Hwee Tou	Coreference Resolution: Maximum Metric Score Training, Domain Adaptation, and Zero Pronoun Resolution
Zhu Jian	PhD	Pung Hung Keng	Provision, Discovery and Development of Ubiquitous Services and Applications
Aisha Siddiqa Azim	MSc	Sim Khe Chai	Combining Speech with Textual Methods for Arabic Diacritization
Cao Nannan	MSc	Low Kian Hsian, Bryan	Information-theoretic Multi-robot Path Planning
Chanaka Aruna Munasinghe	MSc	Ooi Wei Tsang	Texture Caching in Networked Virtual Environments
Cheng Weiwei	MSc	Tan Kian Lee	Query Authentication and Processing on Outsourced Databases
Cui Weiwei	MSc	Roger Zimmermann	An Experimental Study of Video Uploading from Mobile Devices with HTTP Streaming
Fang Shunkai	MSc	Roger Zimmermann	EnAcq: Energy-efficient Location Data Acquisition based on Improved Map Matching
Fu Qiang	MSc	Wynne Hsu	Augmented Nomogram with Dependent Feature Pairs
Gong Bozhao	MSc	Tay Yong Chiang	HBM: A Hybrid Buffer Management Scheme for Solid State Disks
Gong Jian	MSc	Leong Wing Lup, Ben	Chimera: Large-scale Data Collection and Processing
Guo Fei	MSc	Ooi Beng Chin	SHAPE: Scalable Hadoop-based Analytical Processing Environment
Guo Jiayan	MSc	Low Kok Lim	Video Inpainting for Non-repetitive Motion
He Lian	MSc	Wang Ye	Obstructive Sleep Apnea Diagnosis with Apnea Event Detection in Snoring Sound Using a Conditional Random Field

Name	Degree	Supervisor	Thesis Title
Hoang Cong Duy Vu	MSc	Kan Min Yen	Towards Automated Related Work Summarization
Hou Song	MSc	Tay Yong Chiang	Systems and Networking
Hu Junfeng	MSc	Chan Chee Yong	Solving Empty Result Problem in Keyword Search over Relational Databases
Huang Wei	MSc	Hsu Wynne	Mining Non-contiguous Mutation Chain in Biological Sequences Based on 3D-Structure
Hui Mei	MSc	Ooi Beng Chin	Efficiently Indexing Sparse Wide Tables in Community Systems
Huynh Bach Khoa	MSc	Abhik Roychoudhury	Scope-aware Data Cache Analysis for WCET Estimation
Kuang Xiaole	MSc	Zhao Shengdong	A Multilevel Analysis of Commercial Software Online Help Forums
Leow Su Jun	MSc	Low Kok Lim	Automatic Paper Pop-up Design
Li Jianran	MSc	Sim Mong Cheng, Terence	Supervised Extraction of Face Subspaces Based on Multimodal Discriminant Analysis
Li Mo	MSc	Ashraf Golam	Sketch-based Character Prototyping by Deformation
Li Ruoru	MSc	Low Kok Lim	Multi-view Image Refocusing
Li Yan	MSc	Tulika Mitra	Timing Analysis of Concurrent Programs Running on Shared Cache Multi-cores
Lin Yuting	MSc	Ooi Beng Chin	Llama: Leveraging Columnar Storage for Scalable Join Processing in Mapreduce
Liu Bin	MSc	Chan Chee Yong	Efficient Indexing for Skyline Queries with Partially Ordered Domains
Liu Shuaicheng	MSc	Michael Brown	Digital Image Super Resolution
Liu Xiaomin	MSc	Roger Zimmermann	Performance Evaluation of Speech Quality for VOIP on the Internet
Liu Yugang	MSc	Stephane Bressan	Step: Set of T-uples Expansion Using the Web
Lu Han	MSc	Sim Mong Cheng, Terence	Separation of Reflected Images Using WFLD
Luo Yan	MSc	Ooi Wei Tsang	HTTP Live Streaming for Zoomable Video
Luu Anh Tuan	MSc	Dong Jinsong	Seve: Automatic Tool for Verification of Security Protocols
Mai Dang Quang Hung	MSc	Yap Hock Chuan, Roland	Distributed SAT Solving Engine
Mao Lei	MSc	Tan Tiow Seng	Nagivation Mesh Construction and Path-Finding for Architecture Environment
Meduri Venkata Vamsikrishna	MSc	Tan Kian Lee	Exhaustive Reuse of Subquery Plans to Stretch Iterative Dynamic Programming for Complex Query Optimization
Mohammad Oliya	MSc	Pung Hung Keng	Incremental Query Answering Over Semantic Contextual Information
Nguyen Le Nguyen	MSc	Chua Tat Seng	Applying Semantic Analysis to Finding Similar Questions in Community Question Answering Systems
Pramila Nuwantha Ariyaratne	MSc	Sung Wing Kin, Ken	De Novo Genome Assembly Using Paired-end Short Reads
Pratibha Sundar Sundaramoorthy	MSc	Leong Wing Lup, Ben	Practical 3D Geographic Routing for Wireless Sensor Network
Qian Kun	MSc	Zhao Shengdong	Extending Input Range through Clutching: Analysis, Design, Evaluation and Case Study
Sai Sathyanarayan Venkatraman	MSc	Liang Zhenkai	Systematically Enhancing Black-box Web Vulnerability Scanners
Shen Zhong	MSc	Tay Yong Chiang	Using Map-reduce to Scale an Empirical Database
Situ Liangji	MSc	Tan Chew Lim	Text Localization in Web Images Using Probabilistic Candidate Selection Model
Song Zhiyuan	MSc	Leow Wee Kheng	Model-based Cardiac CT Segmentation and 3D Heart Reconstruction
Su Zhan	MSc	Tung Kum Hoe, Anthony	Supporting Top-K item Exchange Recommendations in Large Online Communities
Sun Yang	MSc	Tay Yong Chiang	Data Modeling and Query Processing for Online Social Networking Services
Suraj Pathak	MSc	Tay Yong Chiang	Status of Phase Change Memory in Memory Hierarchy and Its Impact on Relational Database
Tan Huiyi	MSc	Leong Hon Wai	A Parent Mass Filter Algorithm for Peptide Sequencing from Tandem Mass Spectra

Name	Degree	Supervisor	Thesis Title
Truong Duc Thang	MSc	Ashraf Golam	Application of Semi-Lagrangian Method for Visualization of Near-body Hydrodynamics and Enhancement of Amorphous Effect
Wang Tao	MSc	Tay Yong Chiang	Data Storage and Retrieval for Social Network Services
Wang Yumei	MSc	Leow Wee Kheng Cheng Holun, Alan	Tracking of Coronary Arteries in Angiogram Sequence by Structural Matching of Junctions
Xiong Yuanting	MSc	Sim Khe Chai	Modeling of Non-native Speech Automatic Speech Recognition
Xu Jinyu	MSc	Ooi Wei Tsang	Multipath Routing over Wireless Mesh Networks
Yang Xin	MSc	Zhao Shengdong	Robots in My Contact List: Using Social Media Platforms for Human-robot Interaction in Domestic Environment
Yi Bo	MSc	Zhao Shengdong	Exploring Eyes-free User Motivation and Predicting Mental Workload in Mobile HCI
Yi Yu	MSc	Wang Ye	A Tempo-based Music Search Engine with Multimodal Query
Zhang Jiexin	MSc	Dong Jinsong	Formalizing and Verifying Software Architectures
Zhang Lingyan	MSc	Roger Zimmermann	Presentation of Multiple GEO-referenced Videos
Zhao Wei	MSc	Wang Ye	Utilizing EEG Signal in Music Information Retrieval
Zhao Zhen Dong	MSc	Wang Ye	Large Scale Music Information Retrieval by Semantic Tags
Zhou Shaoping	MSc	Zhao Shengdong	ActiveCite: An Interactive System for Automatic Citation Suggestion
Zhou Yuan	MSc	Ooi Beng Chin	Supporting Database Applications as a Service
Zhou Zenan	MSc	Wynne Hsu	MaxFirst: an Efficient Method for Finding Optimal Regions

GRADUATE RESEARCH DEGREES AWARDED

(Department of Information Systems)

Name	Degree	Supervisor	Thesis Title
Achudhan Sivakumar	PhD	Tan Keng Yan, Colin	UAV Swarm Coordination and Control for Establishing Wireless Connectivity
Bao Zhifeng	PhD	Ling Tok Wang	Towards an Effective Processing of XML Keyword Query
Bhojan Anand	PhD	Ananda A.L.	Energy Efficient Algorithms and Techniques for Wireless Mobile Clients
Cao Jianneng	PhD	Tan Kian Lee	Privacy-Preserving Data Publication for Static and Streaming Data
Cao Yu	PhD	Tan Kian Lee Chan Chee Yong	Optimizing Complex Queries with Multiple Relational Instances
Chen Binbin	PhD	Chan Mun Choon	Cooperative Internet Access Using Heterogeneous Wireless Networks
Chen Ding	PhD	Chan Chee Yong	XML Query Optimization
Chen Su	PhD	Ooi Beng Chin	On Optimizing Moving Object Databases
Cheng Wei	PhD	Ooi Wei Tsang	Streaming of High-resolution Progressive Meshes Over the Internet
Chia Wai Kit, Henry	PhD	Tan Chew Lim	Cognitive-inspired Approaches to Neural Logic Network Learning
Chiang Tsung Han	PhD	David Hsu	Markov Dynamic Models for Long-timescale Protein Motion
Chidansh Amitkumar Bhatt	PhD	Mohan Kankanhalli	Probabilistic Temporal Multimedia Datamining
Chong Ket Fah	PhD	Leong Hon Wai	New Models and Algorithm for De Novo Peptide Sequencing of Multi-charge MS/MS Spectra
Claudia Szabo	PhD	Teo Yong Meng	Composable Simulation Models and Their Formal Validation
Dai Bingtian	PhD	Tung Kum Hoe, Anthony	On Column Heterogeneity
David Cristina Mariana	PhD	Chin Wei Ngan	Enhanced Specification Expressivity for Verification with Separation Logic
Deepak Gangadharan	PhD	Roger Zimmermann	Quality-aware Performance Analysis for Multimedia MPSoC Platforms
Ding Feng	PhD	Leow Wee Kheng	3D Segmentation of Soft Tissues by Flipping-free Mesh Deformation
Dong Difeng	PhD	Wong Lim Soon	Relapse Prediction in Childhood Acute Lymphoblastic Leukemia by Time-series Gene Expression Profiling
Ehsan Rehman	PhD	David Hsu	Bounded Uncertainty Roadmaps
Felix Halim	PhD	Yap Hock Chuan, Roland	Solving Big Data Problems from Sequences to Tables and Graphs
FENG YUZHANG	PhD	Dong Jinsong	Reasoning About Complex Agent Knowledge - Ontologies, Uncertainty, Rules and Beyond
Gherghina Cristian Andrei	PhD	Chin Wei Ngan	Efficiently Verifying Programs with Rich Control Flows
Gong Tianxia	PhD	Tan Tiow Seng	Automatic Annotation, Classification and Retrieval of Traumatic Brain Injury CT Images
GUO DONG	PhD	Sim Mong Cheng, Terence	A Collection of Digital Photo Editing Methods
Hagiescu Miriste Andrei Mihai	PhD	Wong Weng Fai	A Computing Origami: Optimized Code Generation for Emerging Parallel Platforms
Harish Katti	PhD	Mohan Kankanhalli Chua Tat Seng	Human Visual Perception, Study and Applications to Understanding Images and Videos
Hugo Willy	PhD	Sung Wing Kin, Ken Ng See Kiong	On Interaction Motif Inference from Biomolecular Interactions: Riding the Growth of the High Throughput Sequential and Structural Data

RESEARCH PROJECT

(Department of **Computer Science**)

Principal Investigator	Project Title	Commencement Date	Completion Date	Total Project Value (S$)
AKKIHEBBAL L. Ananda SMART	Application-guided Network Design	1-Jul-10	30-Jun-11	$202,400
AKKIHEBBAL L. Ananda SMART	Application-guided Network Design	1-Jul-11	30-Jun-12	$334,097
AKKIHEBBAL L. Ananda SMU	Power and Network-Aware Software Infrastructure for Multiplayer Mobile Games	15-May-09	31-May-12	$453,212
BROWN, Michael Scott A*STAR	Applications of Spectral Imaging for Consumer Cameras	6-Feb-12	5-Feb-15	$652,600
BROWN, Michael Scott EADS	Data Distribution Strategies for Dynamically Connected Embedded Nodes	1-Feb-09	28-Feb-11	$31,200
BROWN, Michael Scott FRC	User-assisted Extraction of 3D Scene Properties from Images and Videos	1-Apr-10	31-Mar-13	$50,000
BROWN, Michael Scott FRC	Next Generation Medical Imaging Reporting	1-Aug-12	31-Jul-15	$56,500
BROWN, Michael Scott MSRA	Linearizing the Xbox Kinect for Computer Vision Applications	1-May-12	30-Apr-15	$43,671
BROWN, Michael Scott ODPRT	Perceptual Fidelity and Seamlessness in Projector-Based Displays	1-Jan-08	31-Dec-10	$149,137
BROWN, Michael Scott ODPRT	Hyperspectral Analysis of Old and Damaged Manuscripts	9-Feb-09	31-Dec-12	$499,745
CHAN Chee Yong FRC	Enhanced Querying and Browsing of Database	1-Apr-11	30-Sep-13	$82,748
CHAN Chee Yong FRC	Query Acquisition Techniques for Querying Ad Hoc Databases	1-Apr-13	31-Mar-16	$146,500
CHAN Mun Choon FRC	A Framework for Mobile Wireless Access from Vehicles	1-Sep-08	31-Aug-11	$78,923
CHAN Mun Choon SMART	Application-Guided Network Design	1-Jul-12	30-Jun-13	$209,520
CHANG Ee Chien FRC	Managing Information Leakage in Data Outsourcing and Cloud Computing Applications	1-Jan-13	31-Dec-15	$93,200
CHANG Ee Chien TDSI	Security of Information Sharing and Collaborative Authoring	1-Dec-09	30-Nov-12	$150,000
CHENG Holun, Alan FRC	Auto-robotic Reconfiguration Design	1-Jan-11	31-Dec-13	$38,100
CHENG Holun, Alan FRC	Dynamic Quality Volumetric Homeomorphic Mesh for Deforming Surfaces	1-Apr-09	30-Jun-12	$31,500

Principal Investigator	Project Title	Commencement Date	Completion Date	Total Project Value (S$)
CHENG Holun, Alan FRC	Discovery and Visualization of Macromolecular Structures in 3D Density Map	1-Jan-10	31-Dec-12	$71,400
CHIN Wei Ngan ARC	Specification and Verification for Future Programmers	11-Dec-09	31-May-13	$858,519
CHIN Wei Ngan FRC	A Permission Calculus for Automated Reasoning of Shared Resources	1-Jan-12	31-Dec-14	$65,200
CHIN Wei Ngan FRC	A Core Calculus to Support Exception Handling	1-Nov-08	31-Oct-11	$66,000
CHIN Wei Ngan IBM	Verifying Real-time Programs with Separation Logic	15-Mar-09	14-Sep-10	$14,188
CHUA Tat Seng A*STAR	Geographical Information Retrieval via Spatial Annotation of Web Media	1-Aug-10	31-Jul-13	$600,890
CHUA Tat Seng COMCAST	Event Detection in Movie	1-Oct-08	30-Sep-10	$579,600
CHUA Tat Seng MDA	Interactive Media Search	1-Nov-07	31-Oct-10	$1,979,500
CHUA Tat Seng NEXT	NExT Search Centre	1-May-10	30-Apr-15	$10,000,000
CHUA Tat Seng SPH	Intelligent Media Search for Local Information	30-Jul-08	29-Jul-11	$2,172,450
DONG Jin Song A*STAR	Formal Verification Systems for Simulink and Stateflow	6-Feb-12	5-Feb-15	$542,000
DONG Jin Song A*STAR	Activity Monitoring and UI Plasticity for Supporting Ageing with Mild Dementia at Home (AMUPADH)	1-Mar-10	31-Aug-12	$132,490
DONG Jin Song ARC	Advanced Model Checking Systems	1-Nov-09	31-Jan-13	$918,019
DONG Jin Song MDA	Systematic Design Methods and Tools for Developing Location Aware, Mobile and Pervasive Computing Systems	1-Nov-07	31-Oct-10	$361,500
DONG Jin Song ODPRT	Image and Pervasive Access Lab (IPAL)	22-Jan-12	21-Jan-15	$180,000
DONG Jin Song TDSI	Model Checking System of Systems	15-Apr-11	14-Apr-14	$252,867
DONG Jin Song TDSI	Timed Dynamic Planning System with Liveness and Resources	1-May-08	30-Apr-11	$191,100
GILBERT, Seth Lewis ODPRT	MultiZOOM: Fast Algorithms for Multichannel Wireless Networks	1-Sep-10	31-Aug-13	$150,000

Principal Investigator	Project Title	Commencement Date	Completion Date	Total Project Value (S$)
HENZ Martin FRC	TexiCard - Towards An Operating System for Web Pages	1-Apr-09	31-Mar-12	$81,500
HOBOR Aquinas Adam FRC	Program Verification for Concurrency, Compilers and Self-Modifying Code	1-Aug-09	31-Jul-12	$67,800
HSU David ARC	Large-Scale Uncertainty Planning in Partially Observable Environments	1-Jul-11	30-Jun-14	$668,400
HSU David ARC	Human-Centered Adaptive Computing: A Probabilistic Approach	1-Dec-07	31-Mar-11	$769,720
HSU David FRC	Uncertainty Modeling and Planning for Systems with Large State Spaces	1-Apr-10	31-Jan-12	$49,700
HSU David SMART	Autonomy in Mobility-on-Demand Systems	1-Apr-12	31-Mar-13	$127,131
HSU David SMART	Autonomy in Mobility-on-Demand Systems	1-Aug-10	30-Sep-11	$116,000
HSU Wynne Exploit Technologies	Ocular Imaging	18-Oct-10	31-Dec-15	$645,605
HSU Wynne FRC	Mining and Visualization of Infectious Diseases Spread Patterns	1-Apr-09	30-Jun-12	$89,700
HSU Wynne SERI	Retinal Image Analysis	1-Apr-09	31-Mar-14	$654,277
HSU Wynne SERI	Singapore Retinal Image Archival and Analysis Network (SIRIAN) for Disease Prediction	31-Oct-07	30-Oct-10	$165,000
HSU Wynne SSHSPH	GeoVisualization of Spatio-Temporal Patterns and Disease Trends in Singapore	1-May-12	15-Feb-14	$226,000
JAFFAR Joxan ARC	Symbolic Tracing	1-Jun-09	31-Aug-12	$623,000
JAFFAR Joxan FRC	Symbolic Tracing II (Stopgap)	1-Aug-12	31-Jul-13	$137,469
JAFFAR Joxan FRC	A Compiler for Symbolic Tracing	1-May-08	30-Apr-11	$105,000
JAIN Sanjay FRC	Complexity Issues in Inductive Inference	10-Aug-07	9-Aug-10	$45,000
JARZABEK Stanislaw FRC	Towards a Model for Comparing and Evaluating Generative Techniques	15-Mar-08	14-Jul-11	$69,500
JARZABEK Stanislaw FRC	New Tools for XVCL Workbench and Exploring Non-Software Application Domains for XVCL	1-Apr-09	31-Mar-11	$60,000
JARZABEK Stanislaw FRC	Flexible Generators: Applying Generators in Software Reuse and Evolution	1-Apr-10	31-Mar-12	$31,744
KAN Min Yen FRC	Data Mining for Supporting Critical Reviews in Evidence Based Nursing	15-Sep-10	14-Mar-13	$98,000
KANKANHALLI Mohan A*STAR	Characterizing and Exploiting Human Visual Attention for Automated Image Understanding and Description	1-Aug-10	31-Jul-13	$300,200
KANKANHALLI Mohan FRC	Real-time Best View Selection in Cyberphysical Environments	1-Apr-10	30-Sep-13	$107,000
KHOO Siau Cheng DSTA	Software Dependability through Specification Mining	21-Aug-07	20-Oct-11	$347,334
KHOO Siau Cheng FRC	Malware Detection via Mining	1-Nov-09	30-Apr-13	$77,444
KHOO Siau Cheng FRC	Bug Signature Discovery	1-Apr-12	31-Mar-15	$65,096
LEE Mong Li, Janice A*STAR	Gaze Tracking for Visual Field Testing in Glaucoma	1-Aug-12	31-Jul-14	$257,202
LEE Wee Sun AOARD	Combining Offline and Online Computation for Solving Partially Observable Markov Decision Process	7-Mar-12	6-Mar-15	$189,000
LEE Wee Sun DSO	Active Learning for Information Extraction	1-Sep-11	31-Mar-13	$120,547
LEE Wee Sun DSO	Transfer Learning for Adaptive Relation Extraction	1-Nov-09	31-Aug-11	$180,000
LEE Wee Sun Singapore-MIT GAMBIT Game Lab	Tools for Creating Intelligent Game Agents	1-Jun-09	31-Jul-12	$403,248
LEONG Hon Wai FRC	Algorithmic Methods for Enhancement of Metabolic Models	1-Sep-11	31-Aug-14	$47,500
LEONG Hon Wai FRC	Algorithms for Comparative Genomics Using Gene Team Trees	15-Sep-08	14-Sep-11	$80,000
LEONG Tze Yun FRC	Representing and Reasoning with Change in Human-Centered Computing	1-Jan-13	31-Dec-15	$133,000
LEONG Tze Yun FRC	Representation Discovery in Uncertain and Dynamic Environments	1-Sep-10	31-Dec-12	$51,100
LEONG Wing Lup, Ben ARC	Achieving High-Bandwidth Peer-to-Peer Data Transfers with Network Coding and the Right Incentives in Heterogeneous Networking Environments	7-Apr-08	6-Jul-11	$812,020
LEONG Wing Lup, Ben FRC	Towards Mobile Aerial Meshes for Interactive Streaming	1-Jan-13	31-Dec-15	$123,830
LEONG Wing Lup, Ben FRC	OpenNAT: Enabling Universal Connectivity in an Increasingly Heterogeneous World	1-Jan-11	31-Dec-12	$136,188
LEONG Wing Lup, Ben NUS	Accessible Fault-Tolerant Computing on Commercial Hardware	2-Jan-12	31-Dec-13	$49,800

Principal Investigator	Project Title	Commencement Date	Completion Date	Total Project Value (S$)
LEOW Wee Kheng A*STAR	MMedWeb - Multimedia Medical Conceptual Web for Intelligent Information Access	1-Jul-07	30-Jun-10	$530,840
LIANG Zhenkai FRC	Securing the Android Mobile Platform against Browser-based Attacks	15-Aug-11	14-Aug-14	$152,000
LIANG Zhenkai ODPRT	Improving Security Incident Response by Automatic Vulnerability Diagnosis	1-Oct-08	30-Sep-11	$167,880
LIANG Zhenkai ODPRT	A Framework for General Security Support in Web Browsers	9-Feb-09	8-Aug-12	$477,050
LING Tok Wang FRC	Semantic Keyword Search on XML Databases	1-Apr-11	31-Mar-14	$59,800
LING Tok Wang FRC	Semantic Query Optimization on XML Databases	1-Apr-07	31-Mar-11	$45,720
LOW Kian Hsiang ODPRT	Active Robotic Exploration and Mapping for Environmental Sensing Applications	1-Apr-10	31-Mar-13	$165,377
LOW Kian Hsiang SMART	Spatiotemporal Modeling and Prediction of Traffic Patterns	1-Oct-12	30-Sep-13	$108,656
LOW Kian Hsiang SMART	Spatiotemporal Modeling and Prediction of Traffic Patterns	1-Oct-11	30-Sep-12	$66,000
LOW Kok Lim FRC	Computational Paper Pop-Up Design	1-Sep-11	31-Aug-14	$92,250
LOW Kok Lim FRC	Computational Photographic Aesthetics	1-Oct-08	30-Sep-11	$69,800
MA Tianbai, Richard ODPRT	Resource Provisioning and Optimization for Enterprise Cloud Users	15-Jan-11	14-Jan-14	$178,000
MITRA Tulika ARC	Bahurupi: Polymorphic Heterogeneous Multi-Core Systems	1-Nov-09	30-Apr-13	$806,020
MITRA Tulika ARC	Monsoon: Energy and Thermally Efficient Adaptation of Embedded Computing Systems	1-Jan-13	31-Dec-15	$642,808
MITRA Tulika CSR	Dynamic Thermal Management for Heterogeneous MPSoCs	1-Nov-11	31-Dec-15	$215,200
MITRA Tulika FRC	Computation Model for 3D Multi-core Systems	1-Apr-12	31-Mar-15	$71,740
MITRA Tulika FRC	Probabilistic Timing Analysis of Embedded Software	1-Apr-09	30-Sep-11	$30,000
NG Hwee Tou DSTA	Machine Translation for Resource-Poor Languages	26-Feb-08	25-Nov-11	$388,000
NG Hwee Tou MDA	Research Collaboration between NUS and CSIDM: Phase 1	1-Jun-08	31-Dec-11	$2,657,000

Principal Investigator	Project Title	Commencement Date	Completion Date	Total Project Value (S$)
NG Hwee Tou MDA	Research Collaboration between NUS and CSIDM: Phase 2	1-Jun-11	10-Jul-14	$1,343,000
OOI Beng Chin A*STAR	Logbase: A Scalable Log-Structured Distributed Data Platform	1-Feb-13	31-Jan-16	$594,000
OOI Beng Chin A*STAR	S3: A Scalable, Sharable and Secure P2P Based Data Management System	1-Aug-07	31-Jul-10	$546,640
OOI Beng Chin ARC	epiC - An Elastic Power-Aware Data Intensive Cloud System	1-Apr-11	31-Mar-14	$638,403
OOI Beng Chin ARC	Utab: A High Performance Community Web Management System	1-Apr-09	30-Sep-12	$1,094,120
OOI Beng Chin FRC	CDAS: A Crowdsourcing Data Analytics Systems	1-Jan-13	31-Dec-15	$106,000
OOI Beng Chin FRC	Indexing in Cloud Computing Platforms	1-Apr-09	31-Mar-11	$45,000
OOI Beng Chin FRC	Efficiently Querying Uncertain Data in Distributed Devices based on Probing	1-Oct-09	30-Sep-11	$30,000
OOI Beng Chin MDA	Structure-Aware Data and Query Modeling for Effective Information Retrieval over Heterogeneous Data Sources in Co-Space	15-Dec-08	14-Dec-12	$1,812,000
OOI Beng Chin MSRA	Managing Probabilistic Data with Crowd Intelligence	1-Jul-12	30-Jun-15	$43,671
OOI Beng Chin NRF	Database Support for Research on Megacities	1-Aug-12	31-Jul-17	$4,484,000
OOI Beng Chin NRF	Design and Development of a Comprehensive Information Technology Infrastructure for Data-Intensive Applications and Analysis	1-Dec-12	30-Nov-17	$6,461,500
OOI Wei Tsang FRC	Secure and Efficient Remote 3D Rendering	1-Jan-13	31-Dec-15	$109,500
OOI Wei Tsang French Embassy and NUS	Remote Visualization of Large Collections of 3D Data	1-Sep-08	31-Aug-10	$10,300
OOI Wei Tsang Singapore-MIT GAMBIT Game Lab	Fast Rendering of 3D Objects on Mobile Devices	1-Jan-10	30-Jun-11	$88,320
OOI Wei Tsang URC	Magnifiable Media Stream	1-Dec-08	31-May-11	$297,600
P.S. Thiagarajan ARC	A Computational Framework for the Study of Multi-mode Biopathways	1-Jun-12	31-May-15	$656,954
P.S. Thiagarajan ARC	Decomposition and Composition of Large Signalling Pathway Models with Emphasis on Parameter Estimation	15-May-08	14-May-11	$487,020
P.S. Thiagarajan FRC	Approximate Analysis Methods for Networks of Dynamical Systems	1-Jan-12	31-Dec-14	$62,000

Principal Investigator	Project Title	Commencement Date	Completion Date	Total Project Value (S$)
P.S. Thiagarajan FRC	Scenario-based Study of Distributed Systems and Supervisory Control	1-Apr-09	31-Mar-12	$30,708
ROSENBLUM, David S. ODPRT	Advances in Probabilistic Modeling and Analysis	1-Jun-11	31-May-14	$300,000
ROSENBLUM, David S. PVO and SoC	Institute on Felicitous Computing	1-Jan-12	31-Dec-14	$2,000,000
ROYCHOUDHURY Abhik A*STAR	Scalable Timing Analysis Methods for Embedded Software	6-Feb-12	5-Feb-15	$492,000
ROYCHOUDHURY Abhik ARC	Analysis and Test Generation for Evolving Software	16-May-11	15-May-14	$830,500
ROYCHOUDHURY Abhik DSTA	Symbolic Taint Analysis	27-Feb-09	26-Aug-12	$397,290
ROYCHOUDHURY Abhik FRC	Validation and Comprehension Methods for Large-scale Systems with Complex Interactions	1-Mar-09	31-May-12	$66,268
ROYCHOUDHURY Abhik FRC	Multi-Tenant Software Engineering for Cloud Computing	1-Apr-10	30-Sep-12	$82,600
ROYCHOUDHURY Abhik URC	Platform-aware Timing Analysis of Behavioural System Models	15-Nov-07	14-May-11	$250,000
SAXENA, Prateek ODPRT	WEBINSPECT: A Security Architecture for Web Applications with Auditability Guarantees	12-Jul-12	11-Jul-15	$300,000
SIM Khe Chai FRC	Haptic Voice Recognition: Perfecting Voice Input with a Magic Touch	1-Apr-12	31-Mar-15	$87,370
SIM Khe Chai ODPRT	Model-based Approaches to Content-Aware Voice Extraction (CAVE)	1-Jan-09	31-Dec-11	$175,000
SIM Mong Cheng, Terence FRC	Sketch Me If You Can	1-Mar-11	31-Aug-13	$105,700
SIM Mong Cheng, Terence FRC	Collaborative Computational Photography	1-Apr-09	31-Mar-12	$102,068
SIM Mong Cheng, Terence FRC	Wonder Twins	15-Sep-09	14-Mar-11	$73,900
SOO Yuen Jien FRC	Online Multithreaded Program Behaviour Monitoring through Dynamic Instrumentation and Hardware Counter	1-Apr-12	31-Mar-15	$96,500
STEPHAN, Frank Christian FRC	Hypothesis Spaces and Complexity of Learning	1-Apr-10	30-Sep-13	$41,000
SUNG Wing Kin, Ken ARC	Extracting Biological Signals from the Second Generation Sequencing	1-Oct-10	30-Sep-13	$888,477
SUNG Wing Kin, Ken URC	Motif Discovery for DNA and Protein	16-Nov-07	14-May-11	$259,800

Principal Investigator	Project Title	Commencement Date	Completion Date	Total Project Value (S$)
TAN Chew Lim A*STAR	Video Text Recognition Research	1-Sep-09	28-Feb-13	$576,588
TAN Chew Lim ARC	Clinical Feature Quantification on Large TBI CT Database for Automatic Prognosis	1-Apr-12	31-Mar-15	$340,200
TAN Chew Lim ARC	Content-based Retrieval and Automated Interpretation of Brain CT Scan Images	7-Apr-08	6-Oct-11	$621,020
TAN Chew Lim FRC	Accurate Modelling for Statistical Machine Translation based on Sophisticated Exploration of Bilingual Knowledge Resource	1-Apr-11	31-Mar-12	$76,600
TAN Keng Yan, Colin Institute for Infocomm Research	Project Finch	10-Oct-10	31-May-13	$152,082
TAN Kian Lee A*STAR	Hippocratic Data Stream Cloud for Secure, Privacy-preserving Data Analytics Services	1-Aug-10	31-Jul-13	$404,085
TAN Kian Lee FRC	A MapReduce-based System Data Cube Materialization and View Maintenance	1-Aug-12	31-Jul-15	$116,250
TAN Kian Lee TDSI	Mobile Sensor Database Synchronization and Query Processing in Dynamic Environments	1-Apr-08	31-Dec-10	$175,000
TAN Kian Lee URC	Efficient Methods for Preserving the Privacy of Individuals in Publicly-Released Data	10-Dec-07	9-Jun-11	$271,493
TAN Soon Huat, Gary FRC	A Framework for Symbiotic Simulation of Crisis Management	1-Sep-07	31-Aug-10	$44,523
TAN Soon Huat, Gary SMART	Parallel, Scalable Transportation Simulation	1-Nov-12	31-Oct-13	$92,700
TAN Soon Huat, Gary SMART	Parallel, Scalable Transportation Simulation	1-May-11	31-Oct-12	$176,160
TAN Tiow Seng FRC	G2: Geometry and Graphics	1-Apr-08	31-Mar-11	$63,500
TEO Yong Meng FRC	Modelling and Validation of Emergent Properties in Complex Systems	1-Dec-11	30-Nov-14	$63,090
TEO Yong Meng FRC	DiRAM: Decentralized Market-based Resource Pricing and Allocation	1-Feb-08	31-Jan-11	$36,830
TEO Yong Meng FRC	Live Process Relocation in Large Distributed Systems	15-Sep-09	14-Sep-12	$48,200
TEO Yong Meng SMART and NUS	A Testing and Verification System for Today's Multicore and Tomorrow's Multiprocessing Systems	1-Jan-10	31-Dec-10	$49,119
TEO Yong Meng Sun Microsystems	Performance Analysis of Petascale Systems	15-Jan-07	14-Jul-12	$352,974.15
TUNG Kum Hoe, Anthony FRC	Interactive Visual Analytics using Tag-Clouds	1-Apr-12	31-Mar-15	$65,000

Principal Investigator	Project Title	Commencement Date	Completion Date	Total Project Value (S$)
TUNG Kum Hoe, A nthony FRC	Visual Exploration and Mining of Cohesive Subgraphs in Complex Relational Graph	15-Nov-08	14-Nov-11	$94,100
TUNG Kum Hoe, Anthony URC	GEMINI: Gene Expression Mining	1-Nov-06	30-Jun-10	$250,000
WANG Ye A*STAR	A Cloud-based Therapy Delivery System That Uses Music To Enhance Limb Function and Speech Production in Patients With Neurological Impairments	1-Nov-12	30-Apr-14	$152,365
WANG Ye FRC	Sensor-Based Gait Analysis: A First Step Towards a RAS Gait Training System	1-Sep-11	31-Aug-14	$118,200
WANG Ye FRC	The Development of a Cloud-based Physiological Data Centre for Predictive Analytics	1-Jan-13	31-Dec-15	$134,650
WANG Ye FRC	MagicMirror - A Learning Companion for Singing	1-Apr-08	31-Mar-11	$68,000
WANG Ye NUS VISA	Automatic Music Playlist Generation for Treating Insomnia of the Elderly	1-Aug-10	31-Jul-12	$20,000
WONG Lim Soon A*STAR	Supporting Diagnostic Data Mining via Exploratory Hypothesis Testing and Analysis	1-Aug-10	31-Jan-14	$655,340
WONG Lim Soon A*STAR	Pattern Spaces: Theory, Techniques, and Applications	1-Aug-07	31-Dec-10	$485,840
WONG Lim Soon ARC	Construction of Reliable Protein Interactomes for Infectious Diseases	1-Apr-10	30-Jun-13	$856,020
WONG Lim Soon ARC	Proteomic Profile Analysis Based on Biological Networks	12-Oct-12	11-Oct-15	$545,000
WONG Lim Soon ARC	Intensional Expressive Power of Query Languages	1-Oct-12	30-Sep-15	$119,000
WONG Lim Soon NRF	Informatics and Search Algorithms	1-Feb-08	31-Jan-13	$600,000
WONG Weng Fai A*STAR	Software Optimization of High Performance Computing Code for Hetreogeneous Multicore Platforms	1-Aug-10	31-Jan-14	$451,060
WONG Weng Fai A*STAR	Wearable Integrated BSN Hardware and Software Platform	1-Mar-06	30-Nov-10	$95,000
WONG Weng Fai ARC	Processor Memory Hierarchy Built using the Next-generation Technologies	1-Oct-10	31-Mar-14	$850,164
WONG Weng Fai FRC	Accelerating Systems Biology Computations using Graphics Processors	1-Apr-10	31-Mar-13	$93,531
YAP Hock Chuan, Roland ARC	Unifying Constraint Representations, Ad-hoc Constraints and Global Constraints	1-Jan-13	31-Dec-15	$536,289
YAP Hock Chuan, Roland FRC	Effective Crawling of Microblogs under Resource and API Limitations	1-Oct-12	30-Sep-14	$54,860

Principal Investigator	Project Title	Commencement Date	Completion Date	Total Project Value (S$)
YAP Hock Chuan, Roland FRC	Maximum Flow for Internet Scale Graphs	1-Jan-11	31-Dec-12	$97,518
YAP Hock Chuan, Roland Microsoft	Understanding the Windows Scheduler	1-Oct-08	31-Mar-11	$20,456
YAP Hock Chuan, Roland URC	Programming and Solving Ad-hoc and Problem Specific Constraints	1-Jun-07	30-Nov-10	$250,000
YIN Kang Kang ARC	Building Digital Doubles from Multimedia Datasets	1-Aug-12	31-Jul-15	$328,500
YIN Kang Kang ODPRT	Data-driven Character Animation	10-Aug-10	9-Aug-13	$400,000
YU Haifeng ARC	Fault-tolerant Communication Complexity in Wireless Networks	1-Apr-12	31-Mar-15	$577,000
YU Haifeng FRC	Combating Selfish and Malicious Behaviour in Large-Scale Decentralized Distributed Systems	1-Oct-09	30-Sep-12	$69,460
YU Haifeng ODPRT	Towards Secure and Highly-Available Aggregation Queries in Large-Scale Sensor Networks	7-Jan-08	6-Jul-11	$426,602
ZHAO Shengdong Autodesk	Question-Answer Social Forum Analysis Project	1-Aug-10	31-Jul-13	$26,830
ZHAO Shengdong C.K. Tang Limited	Virtual Dressing Room	1-Oct-11	30-Sep-13	$5,000
ZHAO Shengdong FRC	AutoComPaste and Eye-Copy: Enabling Efficient Information Gathering via Innovative Interaction Techniques	1-Jan-12	31-Dec-14	$126,600
ZHAO Shengdong ODPRT	Multi-Dimensional Adaptive Interface for the Aged	1-Mar-10	31-Aug-13	$366,000
ZHAO Shengdong ODPRT	Alternative Interaction Techniques for Mobile and Ubiquitous Environments	1-Feb-09	31-Jan-12	$179,590
ZIMMERMANN Roger A*STAR	Location-adaptive Peer-to-Peer Audio Streaming for Networked Virtual Environments	1-Aug-08	31-Jul-11	$495,780
ZIMMERMANN Roger FRC	Location-Based Services in Support of Social Media Applications	1-Apr-10	31-Mar-12	$79,200

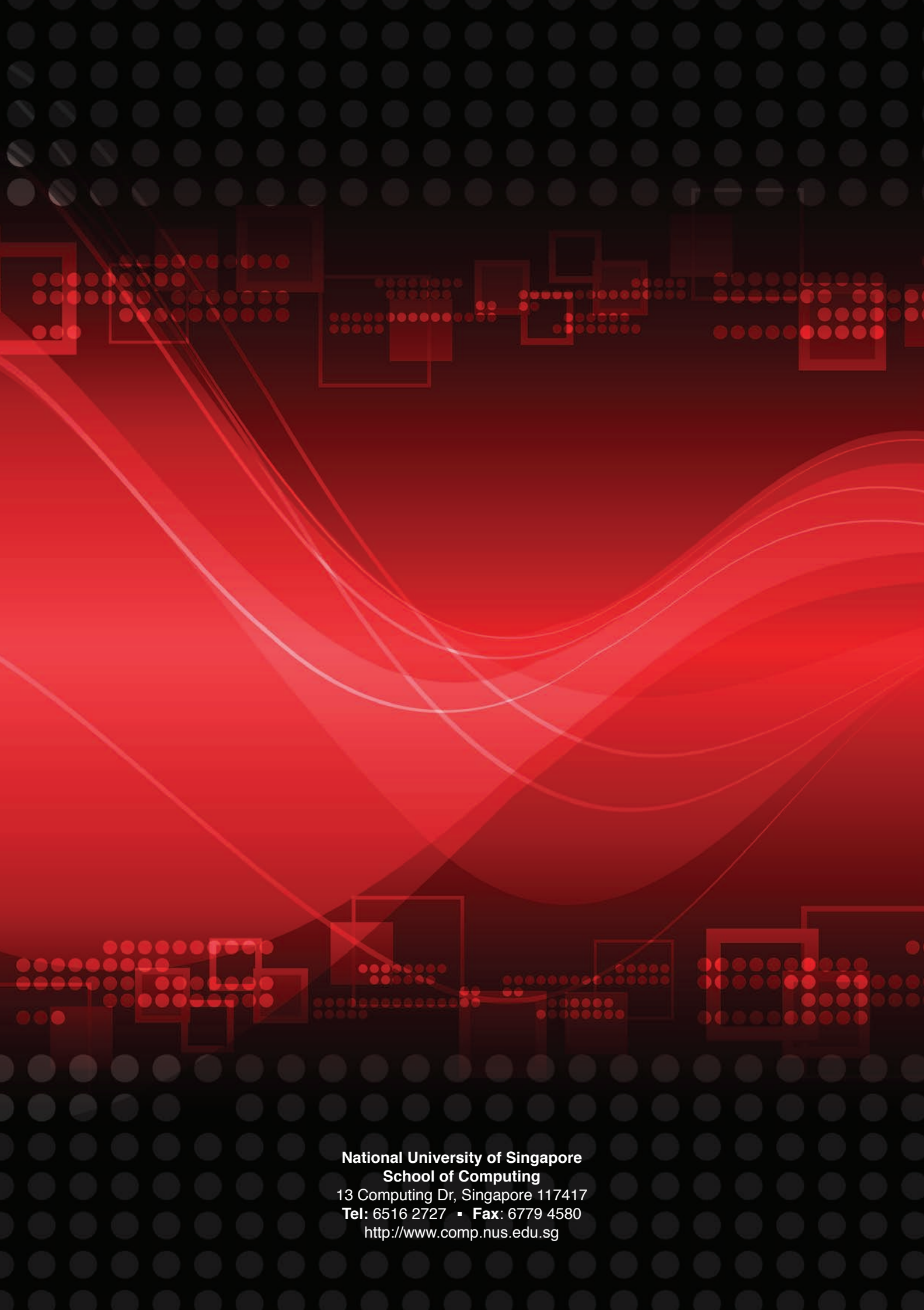
RESEARCH PROJECT

(Department of Information Systems)

Principal Investigator	Project Title	Commencement Date	Completion Date	Total Project Value (S$)
CHANG Ting Ting Klarissa ODPRT	Leveraging Social Computing Technologies for Competitive Advantage	1-Nov-07	30-Apr-10	$59,800
CHANG Ting Ting Klarissa URC	The Role of New Information and Communication Technologies on Medicine 2.0	1-Jun-10	30-Nov-13	$187,500
CHEN Yuanyuan ODPRT	A Global Policy Framework for Cloud Computing	1-Feb-11	31-Jan-14	$359,000
CHEN Yuanyuan ODPRT	Assessment of Intellectual Property and Security Risks in Information Technology (IT) Enabled Service Sourcing	1-Oct-08	30-Sep-11	$60,000
DATTA Anindya FRC	Rated Aspect Summarization of Online Reviews through Constrained Topic Modeling	1-Feb-12	31-Jan-14	$47,000
DATTA Anindya FRC	Enabling Document Driven Search	15-Sep-10	31-Dec-11	$73,600
DATTA Anindya NRF-POC	Creating a Comprehensive Lexical Index of Documents from the World Wide Web (WWW)	1-May-10	31-Jul-11	$248,000
DATTA Anindya ODPRT	The Design and Implementation of a Search Engine for Blogs	1-May-09	30-Apr-12	$172,800
DUTTA Kaushik FRC	Estimating Energy Consumption of an Application	1-Apr-13	31-Mar-16	$129,445
DUTTA Kaushik ODPRT	Reduce, Reuse, Recycle: Building Greener Software	1-Oct-11	30-Sep-14	$150,000
GOH Khim Yong FRC	Consumer Responses and Firm Strategies for Social-Mobile Advertising and Location-Based Marketing	1-Aug-12	31-Jul-15	$138,300
GOH Khim Yong FRC	Investigating Advertising and Promotional Response in Digital Entertainment and Media Environments	1-Apr-09	31-Mar-12	$55,000
HAHN Jungpil ODPRT	Toward a Behavioral Theory of Software Development: Software Development as Design Problem Solving	1-Aug-12	31-Jul-15	$150,000
HENG Cheng Suang FRC	IT Outsourcing: Contracts & Innovation	1-Feb-11	31-Jan-14	$179,000
HENG Cheng Suang FRC	IT Outsourcing: An Investigation of Multiple Vendors	1-Feb-09	31-Jan-12	$84,600
HUANG Ke-Wei FRC	Examining the Impacts of Enterprise Software Innovations by Quantifying Technology News	1-May-10	30-Apr-13	$95,600

Principal Investigator	Project Title	Commencement Date	Completion Date	Total Project Value (S$)
JIANG Zhenhui, Jack ARC	The New Frontier of ECommerce: An Investigation of Collaborative Online Shopping	1-Dec-09	31-May-13	$750,020
JIANG Zhenhui, Jack FRC	From Index Search to Social Sensemaking: An Investigation of Internet Product Search Mechanism	15-Jan-08	14-Jan-11	$49,001
KANKANHALLI Atreyi FRC	Open Innovation through Virtual Knowledge Intermediaries	1-Apr-12	31-Mar-15	$65,000
KANKANHALLI Atreyi IBM	Service Innovation	1-Nov-10	30-Apr-14	$51,856
KANKANHALLI Atreyi URC	Role of Information & Communication Technologies in Governance	4-Jul-08	3-Jul-12	$181,600
KIM Seung Hyun FRC	Impact of EHR on Privacy and Information Security in the Health Care Industry	1-Dec-11	30-Nov-14	$62,500
KIM Seung Hyun ODPRT	An Empirical Assessment of Knowledge Management Systems	1-Oct-08	30-Sep-11	$60,000
LIM Lai Huat, John FRC	IT-Enabled Learning, Work and Socialization	1-Apr-11	31-Mar-14	$82,075
PAN Shan Ling FRC	Agility in Asian IT Management: Resource Management and Capability Development	1-Apr-09	30-Sep-12	$57,500
PAN Shan Ling ISS	eGovernment and Service Transformation in Singapore	1-Dec-07	30-Nov-10	$50,000
PHAN Tuan Quang FRC	Social and Media Webcrawling and Analytics for Financial Valuations and Consumer Index	1-Aug-12	31-Jul-14	$64,000
PHAN Tuan Quang ODPRT	Investigation on Consumer Behaviour in Social Networks through Theoretical, Empirical, and Experimental Methods	10-Oct-11	9-Oct-14	$240,000
POO Chiang Choon, Danny FRC	Classification-based Approach to Sentiment Analysis	1-Jan-11	30-Jun-13	$42,552
POO Chiang Choon, Danny FRC	Quantitative Analysis of Health Information Technology Impact on Patient Safety and Healthcare Quality and Productivity	1-Oct-09	30-Sep-12	$32,100
POO Chiang Choon, Danny FRC	Universal Automated Single Polymorphism Identification	1-Apr-10	31-Mar-12	$39,680
TAN Swee Lin, Sharon FRC	A Study of the Use and Impact of Online Health in Singapore	1-Mar-12	28-Feb-15	$68,131

Principal Investigator	Project Title	Commencement Date	Completion Date	Total Project Value (S$)
TAN Swee Lin, Sharon FRC	Designing and Evaluating a Web and Mobile-based eHealth Behaviour Intervention Program	1-Apr-13	31-Mar-16	$141,531
TAN Swee Lin, Sharon FRC	Developing a Model of EHR Implementation: A Case Study of Singapore EHR Initiative	1-Apr-09	31-Mar-12	$55,000
TEO Hock Hai ARC	Effects of Secondary Web Stimuli on the Processing of Primary Online Tasks	1-Apr-12	31-Mar-15	$424,558
TEO Hock Hai FRC	Effects of Web Ads on Online Consumers Shopping Behavior	1-Jan-12	31-Dec-14	$89,545
TEO Hock Hai FRC	Next Generation Cloud Computing-based Electronic Commerce Platform	1-May-12	30-Apr-14	$48,000
TEO Hock Hai URC	Information Technology, Value Network Transparency and Supply Chain Performance	24-Jun-08	23-Dec-10	$220,000
ZIMMERMANN Roger MDA	Center of Social Media Innovations for Communities	1-Sep-10	31-Aug-15	$4,400,000



National University of Singapore
School of Computing
13 Computing Dr, Singapore 117417
Tel: 6516 2727 ▪ **Fax**: 6779 4580
http://www.comp.nus.edu.sg