THE UNIVERSITY OF SYDNEY

FACULTY OF ENGINEERING

Engineering Vacation Research Internship Program

COMPUTER SCIENCE RESEARCH PROJECTS FOR WINTER 2022

COMPUTER SCIENCE PROJECTS

CS22/1 Evaluating the Performance-Durability Trade-off of MongoDB

Supervisors: Prof Alan Fekete and A/Prof Uwe Roehm

Eligibility: Strong academic background in Computer Science. Particularly useful would be, good knowledge of Database and/or Operating Systems. Experience in benchmarking or tuning systems would be desirable.

Project Description:
MongoDB is a popular NoSQL database that offers database like functionality to store and access collections of JSON data. The data in MongoDB is usually replicated across several nodes, and for performance reasons, the system provides alternative ways to update the data; these choices differ in how thoroughly the change is protected against loss from system crashes. The goal of this project is to develop some guidance for programmers to help them decide which MongoDB configuration is best suited for their applications. This project will extend previous work (from an Honours thesis) to measure the properties of these choices; the focus will be on identifying conditions which make the system more vulnerable to data loss, and to evaluate the trade-of between update scalability and write durability. This winter internship is offered in collaboration with MongoDB and will allow a close insight into the performance trade-offs with modern distributed NoSQL databases.

Requirement to be on campus: No

CS22/2 Effective Guidance for the MongoDB Query Optimiser

Supervisors: Prof Alan Fekete and A/Prof Uwe Roehm

Eligibility: Strong academic background in Computer Science. Particularly useful would be, good knowledge of Database and/or Operating Systems. Experience in benchmarking or tuning systems would be desirable.

Project Description:
MongoDB is a popular NoSQL database that offers database like functionality to store and access collections of JSON data. The platform chooses a way to execute each query to obtain good performance; the platforms approach to this query optimisation activity is quite different from other databases.

The goal of this project is to identify situations where the optimiser makes good choices, and any situations where it chooses a poor execution; for the latter, we seek to improve the optimiser. The project will extend previous work (from an Honours thesis) which measured the success of the optimiser for simple queries on simple documents; the focus will now be on looking a nested document structure. This winter internship is offered in collaboration with MongoDB and will allow a close insight into the performance issues with modern NoSQL databases.

Requirement to be on campus: No

CS22/3 Evaluating the Performance of different logical designs in MongoDB

Supervisors: Prof Alan Fekete and A/Prof Uwe Roehm

Eligibility Criteria: Strong academic background in Computer Science. Particularly useful would be, good knowledge of Database and/or Operating Systems. Experience in benchmarking or tuning systems would be desirable.

Project Description:
MongoDB is a popular NoSQL database that offers database like functionality to store and access collections of JSON data. There are different ways to arrange logically equivalent information in documents: for example, connections between two kinds of item may be stored within one item, within the other, or externally to both. The goal of this project is to develop some guidance for programmers to help them decide which MongoDB document structure to use, to get the best performance (depending on the pattern of operations needed). This project will extend previous work (from an Honours thesis) to measure the performance of a wider variety of structural designs for the data, and then in particular, to take account of enforcing integrity constraints.

Requirement to be on campus: No

CS22/4 Indoor positioning data analysis

Supervisor: Prof Athman Bouguettaya

Eligibility:

- Experience in programming Android apps (Java or Kotlin)
- Experience with backend developments (preferably NodeJS)
- Working knowledge of database technologies (e.g., mysql)
- Basic knowledge of WiFi access points

Project Description:
Indoor positioning refers to the process of finding out the location of people/objects indoors, e.g., inside malls or offices. One way to determine the position of an object is to leverage WiFi access points. Specifically, the strength of the received signal can reveal information about how far the access point is. This project aims to use the signals from multiple access points to determine the position of objects. The project involves collecting indoor position data using an existing app. The data then will be used to model and estimate the actual position.

Requirement to be on campus: Yes *dependent on government's health advice

CS22/5 Multi-Package Delivery Using Single Drone in a Skyway Network

Supervisor: Prof Athman Bouguettaya

Eligibility: Having Python programming and preferably data science knowledge for data cleaning

Project Description:
Drones offer cost-effective and fast delivery services. The potential utilization of drones is limited by payload capacity and battery consumption constraints. Drones may need multiple times of recharge for persistent delivery operation. The drone delivery environment is highly constrained because the availability of recharging stations is not guaranteed. We leverage the service paradigm to address the key challenges in delivery by drones. The functional and non-functional properties of drones are abstracted as Drone Services. The drone services operate in a skyway network where each node represents a recharging station or a delivery target. We collect a trajectory dataset for a single drone delivering multiple packages to multiple destinations in a single trip. The dataset will be used for predicting the drone's arrival at certain stations considering battery limitations and energy consumption trends. In addition, the dataset will help in computing the best skyway paths leading to the destinations faster.

Requirement to be on campus: Yes *dependent on government's health advice

CS22/6 Drones Reordering in Swarm-based Services

Supervisor: Prof Athman Bouguettaya

Eligibility: Python knowledge

Project Description:
Drone swarms used for delivery may fly in different formations for energy conservation purposes. Depending on the wind direction, a swarm may undertake different fly formations to reduce drag forces that consume energy. The reordering of swarm members is a common behaviour that occurs in nature (eg. bird flocks) to conserve energy. A swarm member's energy consumption is typically determined by its position in a formation. We propose reordering a drone swarm formation to reduce the load on the most energy-consuming positions and efficiently distribute energy use. In this project, we use real drones to learn the effect of reordering a formation on the energy consumption through data collection and analysis. Machine learning algorithms would be implemented to predict the optimal reordering mechanism given the different intrinsic (eg. payload and battery capacity) and extrinsic constraints (eg. wind conditions).

Requirement to be on campus: Yes *dependent on government's health advice

CS22/7 Machine Learning-based Approach to Estimate Occupancy from Thermal Image Data

Supervisor: Prof Athman Bouguettaya

Eligibility:
Required Skills:

- Experience in programming in Python or Matlab.
- Experience in image processing libraries such as OpenCV or PyTorch.
- Able to implement image processing algorithms to separate background infra-red radiation from human.
- Able to implement machine learning classification algorithms, including K-Nearest Neighbors (KNN), Multilayer Perceptron Artificial Neural Network (MLP), and Linear Regression.

Good to Have (But Not Essential):

- Knowledge of Github

Project Description:
Occupancy estimation is important for smart building applications such as controlling heating, ventilation, lighting, air conditioning systems and determining occupants' behaviours. This project aims to detect and estimate (i.e., count) occupancy in a room from thermal imaging data using state-of-the-art machine learning algorithms. Existing approaches either use intrusive sensors that do not ensure people's privacy when obtaining high accuracy estimations, such as RGB cameras, or use non-intrusive sensors with lower accuracy, such as context sensing devices. In this regard, we use thermal cameras to preserve people's privacy. In addition, thermal cameras have some advantages over the RGB cameras such as they are not light-dependent and can work in dark environments.

Students who decide to work on this project will have the opportunity to implement image processing algorithms to extract features from the thermal image data and implement machine learning algorithms to build the occupancy detection and estimation model.

Requirement to be on campus: Yes *dependent on government's health advice

CS22/8 Understanding the Assimilation of Electronic Medication Management System in an Australian Hospital

Supervisor: Dr Kevin Kuan

Eligibility: Basic proficiency in R for statistical modelling.

Project Description:
Electronic Medication Management (eMeds) systems support the improved quality, safety and effectiveness of medication management with NSW hospitals. This includes providing support for doctors, nurses, and pharmacists to prescribe, order, check, reconcile, dispense and record the administration of medicines. In the project, students will have the opportunity to review literature in health IT management, and to perform statistical analyses on a survey data using techniques such as Partial Least Squares Structural Equation Modelling (PLS-SEM) and fuzzy set Qualitative Comparative Analysis (fsQCA).

Requirement to be on campus: No

CS22/9 Supporting Consumer Decision Using Online Reviews

Supervisor: Dr Kevin Kuan

Eligibility: Basic proficiency in Python for text and data mining

Project Description:
Consumers are increasingly relying on online reviews in their everyday lives, including shopping, dining, traveling etc. This project aims to better support consumer decisions using online reviews by understanding how different numeric and text features of online reviews affect consumers in their decision-making and behaviour. In the project, students will have the opportunity to review literature in disciplines such as business, psychology, computer science, etc., and to analyse a large data set of online reviews containing both number and text.

Requirement to be on campus: No

CS22/10 Efficient Algorithms for Large Scale Graph Processing

Supervisor: Dr Lijun Chang

Eligibility: Competent in algorithm design and implementation

Project Description:
Graph model has been widely used to represent the relationships among entities in a wide spectrum of applications such as social networks, communication networks, and biological networks. We are nowadays facing a tremendous amount of large-scale real-world graphs with millions or billions of edges. Thus, there is a need of designing efficient algorithms for processing large graphs. In this project, our aim is to design efficient algorithms to speed up graph processing on the ever-growing large graph datasets. The problems that we will be investigating can be (1) dense subgraph (e.g., clique, k-plex) computation over a large sparse graph which finds one dense subgraph of the maximum size, or (2) dense subgraph enumeration which enumerates all maximal dense subgraphs.

Requirement to be on campus: No

CS22/11 Best-Possible Wiretap Coding via Cryptography

Supervisor: Dr Qiang Tang

Eligibility: Strong math/theory background

Project Description:
A wiretap coding scheme (Wyner, Bell Syst. Tech. J. 1975) enables Alice to reliably communicate a message m to an honest Bob by sending an encoding c over a noisy channel ChB, while at the same time hiding m from Eve who receives c over another noisy channel

ChE. Wiretap coding is clearly impossible when ChB is a degraded version of ChE, in the sense that the output of ChB can be simulated using only the output of ChE.

Recent progress **https://eprint.iacr.org/2022/343.pdf** have circumvented this impossibility via cryptography, i.e., relying on computational assumptions (despite very strong). We will explore whether we can provide alternative assumptions.

Requirement to be on campus: No

CS22/12 Analysing the Security of Consensus Protocols in Popular Blockchain Projects

Supervisor: Dr Qiang Tang

Eligibility: Familiar with blockchain, distributed computing, or cryptography

Project Description:
There are many blockchain projects with very high market values (e.g., billions USD), however, many of their protocols, including the fundamental layer of consensus protocols are designed in an ad hoc way, and have not been rigorously analysed. This is in sharp contrast with the "provable security" paradigm of modern cryptography.

We will analyse the consensus protocol of some well-known blockchain projects, trying to identify its vulnerability or a rigorous security proof.

Requirement to be on campus: No

CS22/13 Configuration Analysis: Augmentation of set-theoretics with statistical analyses

Supervisor: A/Prof Simon Poon

Eligibility: Good knowledge in data science techniques like Association Rule Mining, Clustering, and SVM. Programming skilling and understanding in statistics and algebra. Students interested in honours and postgraduate research studies are encouraged to apply.

Project Description:
Assessing effects from complex interactions of multiple study factors are common in empirical studies. Changing one factor may have little effect on study outcome if other factors remain unchanged. Furthermore, such interactions may extend to different configurations with complex synergistic relationships amongst many seemingly unrelated factors. From the view of statistical association analysis, especially correlation analysis with the well-known correlation coefficients has been a useful technique for identifying relationships from observational data. However, conventional statistical methods cannot account for situations in which only specific combinations of variables reveal their impact on the outcome (conjunctural causation) or all paths that lead to an outcome need to be simultaneously uncovered (equifinality). These methods also fall short in explaining situations in which a given combination of variables contributes to the presence of an outcome but at the same time is irrelevant for the absence of that outcome (causal asymmetry).

In this project, we address these issues by integrating set-theoretic approaches in conjunction with statistical approaches (like regressions) for discovering meaning configurations from observational data with limited diversity.

Requirement to be on campus: No

CS22/14 Apply process mining in electronic health records

Supervisor: A/Prof Simon Poon

Eligibility:

- Strong interest in medical and health informatics, and process-oriented data science
- Good knowledge in data mining and machine learning, and programming skills. Students interested in honours and postgraduate research studies are encouraged to apply.

Project Description:
In recent years, process mining has been increasingly applied to healthcare to gain insights on complex, multidisciplinary clinical pathways. It has been used to inform efforts to improve patient quality of care, improve coordination between departments, and optimize the use of resources [1] It bridges the gap between clinical pathways prescribed by guidelines and the actual processes patients are undertaking. This project aims to apply process mining (PM) methods to generate process models for clinical pathways from electronic health records (eMR). These process models will allow us to study causes of pathway constraints and improvement strategies based on data-driven performance analysis. Process mining algorithms (like Heuristics, inductive and Fuzzy miners) would be used to extract process models from the event logs extracted from eMR. Different data-driven process models will be systematically evaluated to terms of accuracy and fitness in representing clinical pathways.

[1] – Munoz-Gama, J. et al. (2022), Process mining for healthcare: Characteristics and challenges, Journal of Biomedical Informatics 127, 103994.

Requirement to be on campus: No

CS22/15 Evaluation of AI Imaging Diagnostic Medical Decision Support Systems in Radiology

Supervisors: A/Prof Simon Poon, Dr Neysa Petrina

Eligibility

- Good knowledge of Health Information
- System implementation & evaluation
- Interest in usability research of technology in clinical context
- Good knowledge in mixed method research

Project Description:
The implementation of artificial intelligence (AI) technology in radiology systems as support tools for clinical decision-making promises to substantially expedite the medical screening process, reduce workload burdens, and improve the quality of care provided. However, the utility of AI in a real-world radiology context is significantly impacted by human factors related to daily clinical practice which go beyond diagnostic accuracy such as user acceptability and system usability. Neglect of these in the adoption of past healthcare technologies (e.g. electronic health records, computer-aided detection software) has resulted in suboptimal outcomes such as increased clinician distrust and frustration, and degraded decision-making quality even where the technology was not functionally flawed.

Prospective student will engage in a short user experience study that evaluates Radiologists behavioural responses to AI technology in the diagnostic decision-making process. We will systematically investigate factors contributing to Radiologist's acceptance or rejection of AI.

This is a health information systems project. Student enrolled in combined social/health science and engineering/computing degrees and planning to pursue honours or masters by research are encouraged to apply.

Requirement to be on campus: No

CS22/16 Random Features in Deep Neural Networks

Supervisor: Dr Chang Xu

Eligibility: Experience with Deep Learning programming.

Project Description:
Popular algorithms towards efficient neural networks include the training of redundant networks. For example, pruning techniques require pretrained networks, and neural architecture search algorithms require the training of supernet, since it is hard to fetch prior knowledge of current networks or architectures without the training of their massive weight parameters. However, some modules in the network could bring some hints. Batch normalization (BN) layers normalize activations and then subsequently applies a learned affine transform. Recent studies investigate the performance of randomly initialized CNNs when only the parameters in BN layers are trained. Surprisingly, a relative high performance can be achieved considering the limited training budget. This performance could come from the activation sparsity learning via affine parameters in BN layers. Those properties provide some insights into the expressive power of random features, which could become the potential indicators for further model compression techniques or NAS frameworks without a time-consuming training of redundant networks.
Requirement to be on campus:

Requirement to be on campus: No

CS22/17 Randomness in Learning and Testing Algorithms

Supervisor: Dr Clement Canonne

Eligibility: WAM>75 and Undergraduate candidates must have already completed at least 96 credit points towards their undergraduate degree at the time of application.

Project Description:
Fast and theoretically accurate algorithms for testing or learning properties of data are known, under various types of "constraints" (such as communication constraints, or privacy requirements). Many of those algorithms, however, are *randomised*, which can be an issue in settings where generating random seeds is expensive, impossible, or could lead to vulnerabilities.

The aim of this project is to
1. Investigate whether one can optimise the amount of randomness required by these algorithms [theory+empirical]
2. Implement and evaluate in practice their performance (time- and accuracy-wise) [empirical]
3. Assess whether the type of randomness introduced can lead to security issues or attacks. (i.e., can tampering with the random seed make the algorithm "fail" catastrophically?) [theory and/or empirical]

[The applicant can choose to focus on a subset of the above 3 goals, depending on their background and preferences.]

Requirement to be on campus: No

CS22/18 Apply Natural Language Processing to analyse breastfeeding promotion and sentiments in social media

Supervisor: Dr Sue Chng

Eligibility: Good knowledge in text mining and Natural Language Processing including NLP toolkits. Interest in social media analysis

Project Description:
Promoting healthy infant feeding practices such as breastfeeding has a range of profound short-term and long-term health benefits to both mothers and their children. The World Health Organization (WHO) recommends that all infants should be exclusively breastfed for the first six months of life, after which, families should introduce nutritious complementary foods with continued breastfeeding for up to 2 years of age or beyond. Breastfeeding communities also flourish on social media platforms, including Facebook, Twitter, and Instagram. The communication channels and patterns are different on various platforms.

In this project, the student will apply text analytics tools to explore community sentiment related to breastfeeding on social media and examine the alignment between the information shared on social media. General study directions:

- To identify and analyse marketing approaches that are used on social media to promote BMS in public forums
    - Identify marketing approaches used in traditional media to promote BMS
    - Review BMS social media platforms and public pages/groups on social media to extract BMS marketing contents
    - Categorize contents extracted from social media based on identified marketing approaches to determine marketing approaches commonly used on social media.

- To examine the factors that are associated with sharing and favourable commenting on breastfeeding in public pages and groups on social media;
    - Extract posts/comments/videos/images related to BMS on public pages and groups on social media
    - Perform sentiment analysis on extracted media to determine correlation between sentiment and BF marketing approaches.

Requirement to be on campus: No

CS22/19 Attacks Against Graph Neural Networks

Supervisor: Dr. Suranga Seneviratne

Eligibility: Background in Machine Learning and Python programming experience

Project Description:
Graph Neural Networks (GNNs) have become the state-of-the-art for graph-based data mining, surpassing traditional concepts on graph modelling and walk-based algorithms. However, similar to other deep learning models, graph neural networks are also vulnerable to adversarial examples and data poisoning attacks. This project explores such attacks on the node classification problem in social network graphs. You will work on several Facebook and Twitter social network graph datasets and conduct node feature perturbation attacks and compare the success rates.

Requirement to be on campus: No

CS22/20 Blockchain virtual machine comparison

Supervisors: Vincent Gramoli and Deepal Tennakoon

Eligibility:

Project Description:
With the recent advances in consensus optimisations, modern blockchains [1] can scale to large number of machines, delivering unprecedented performance.
The main drawback is that these scalable and secure blockchains are limited by the expressiveness they offer to users due to the scripting language that they can execute.

More recent blockchains feature a virtual machine that can execute program written in a Turing programming language (e.g., Solidity). Unfortunately, they tend to fork and the assets they manage are frequently lost or stolen.

In a recent effort, the Concurrent Systems Research Group has combined the security and efficiency of the new blockchain with a dedicated virtual machine that is quite efficient. The virtual machines, which stems from an improved version of the go Ethereum Virtual Machine appears to be the bottleneck that caps its performance to few thousands of transactions per second.

The goal of this project is to optimize such a virtual machine to reach performance closer to tens of thousands of transactions per second. Approaches include changing some of its data structures, or testing a more recent version of the EVM, written in WASM.

[1] *Crain, Natoli, Gramoli. Red Belly: A Secure, Fair and Open Blockchain. IEEE Security and Privacy 2021.*

Requirement to be on campus: No

CS22/21 Deep neural network enhancement for real-time video applications

Supervisor: Wei Bao

Eligibility:
- Knowledge in computer networks (especially in UDP and HTTP DASH).
- Knowledge in deep neural networks.
- Experience in Python programming.

Project Description:
Real-time videos (video conferencing, online teaching, etc.) have experienced tremendous growth during the pandemic. UDP is adopted as the underlying transport-layer protocol to avoid delay, but it suffers from packet loss. Another issue is the limited bandwidth in the network. The sender sends low-resolution/low-frame-rate video when the network bandwidth is low, which drastically lowers users' Quality of Experience (QoE).

With the advances of deep neural networks (DNNs), a variety of DNN-based enhancement schemes have been designed. (1) Super-resolution DNNs recover high-resolution videos from low-resolution ones; (2) loss-recovery DNNs can fix distorted frames; (3) interpolation DNNs will recover high-frame rate from low-frame-rate videos; (4) video denoising DNNs can clear noise generated during transmission. Without the need of investing on the improvement in network infrastructure, we can still drastically improve the QoE of real-time videos by DNNs.

You are expected to design a new DNN model to realise video enhancement. The model should be adaptive to the system environment (limited bandwidth and computational capacity), with small data size and small inference delay.

Requirement to be on campus: No

CS22/22 Optimal Schedule for Discretely Controllable Processing Tasks in Edge Computing

Supervisor: Wei Bao

Eligibility: Familiar with basic scheduling algorithms.

Project Description:
With the advances of edge computing, computational tasks can directly be run in the edge devices, facilitating a wide range of smart applications. However, due to the limited computational capacity and energy, the devices may not be able to complete computational-intensive tasks. There are two evident solutions: (1) We reduce the processing time/workload by sacrificing processing quality (i.e., controllable processing time). (2) We offload the tasks to a nearby resource-rich computing node (edge server).

In this project, we aim study the scheduling problem in this edge computing environment. There are several low-speed machine edge devices and one high-speed edge server. The tasks arrive at the devices, which can be processed at the device locally or offloaded to the edge server. Each task has multiple levels of controllable processing times: If a task is fully completed, it will obtain a full utility; If a task is completed with a lower processing level, a partial utility will be obtained. Our objective is to maximize the overall accumulated utility. We aim to propose a schedule algorithm to solve the problem with theoretical analysis on its optimality.

Requirement to be on campus: No

CS22/23 Solving Network Traffic Scheduling Problems via Reinforcement Learning

Supervisor: Dr Xi Wu

Eligibility: Have basic knowledge of Deep Learning, Reinforcement Learning, and their implementations

Project Description:
Time-triggered (TT) communication has long been studied in various industrial domains, whose challenging task is to find a feasible schedule table. Network changes are inevitable due to the topology dynamics, varying data transmission requirements, etc. Once changes occur, the schedule table needs to be re-calculated in a timely manner. Solver-based methods and heuristic-based methods were proposed to solve this problem. However, solver-based methods employ integer linear programming (ILP) or satisfiability modulo theories (SMT) which have high computational complexity. On the other hand, heuristic-based methods are fast, but they need to be handcrafted based on the application characteristics. Thus, these methods are not general enough to work in complex scenarios especially in large networks. In this project, we are going to use Reinforcement Learning to solve Network Traffic Scheduling problems.

Requirement to be on campus: No

CS22/24 Preventing Vulnerabilities in Object-Oriented Programming Language via Language-based Security Approach

Supervisor: Dr Xi Wu

Eligibility: Strong Programming Skill in Java and Interested in Language-Based Security

Project Description:
Java is an object-oriented programming language. It has an access control model for guarding access to resources which relies on programmer discipline to insert security checks, which are then performed at runtime. It makes use of a capability-like notion for access to some resources. The goal of our work is providing access control at the programming-

language level using a capability-based approach. This is challenging because of the use of shared memory and pointers, and the level of interaction between trusted and untrusted code. In this context, language-based security approaches may be used to prevent vulnerabilities that are not addressed by process-based access control at the operating system level. Our project includes building a prototype implementation of the type of system and validating its usability by applying it to real-world case studies.

Requirement to be on campus: No

CS22/25 Implementation and performance evaluation of multi-join algorithm in CUDA using Julia/C++ across a large-scale distributed setting with GPU accelerations

Supervisors: Prof Albert Zomaya and Dr M. Reza Hoseiny

Eligibility:

- Good knowledge of CUDA and Julia or C++
- Experience working with Git, Linux, SSH, command line tools
- Understanding of the multi-threading programming paradigm, low-latency and performant applications
- Work closely with the rest of team to solve problems, and transfer knowledge
- Writing reusable, testable, and efficient code
- Good knowledge of Performance evaluation metrics for computer systems
- Knowledge of Design patterns, fundamental data structures and algorithms

Project Description:
Modern NVIDIA® GPU devices can now parse petabytes of data orders of magnitude faster than they could using traditional CPUs, in applications ranging from energy exploration to deep learning that can deliver the highest performance for modern applications. In this project, we try to investigate the capabilities of modern GPU for processing the real-time streaming data and performing Incremental and Iterative computation and delta computation
commonly to be running across a large-scale deployment of (possibly several) GPU devices. Evaluating the relational join is perhaps the most centric algorithm in the database systems.
While there are many ways to perform join over multiple relationships (such as hash-join), Ngo, Re, and Rudra [1] proposed a worst-case optimal join algorithm based on iterative stateful approach for computing relational joins over multiple relationships with common attributes (aka Generic-join). The approach adds in common attributes one at a time to the result of the join from the previous round. In each round of computation, each relation needs to propose some extensions, and then ask the other relations to validate them. In this project, our aim is to provide a distributed implementation of Generic-join algorithm over multiple relationships when the data enters the system in a continuous streaming manner. The aim is to improve the run-time performance by exploiting the parallel computation, pipelining the functions, minimizing the loss in bandwidth, and reusing the optimal local on-chip memory embedded in modern GPUs.

We can improve the performance of computations by extending a strategy to pipeline records through relations of a computing session in a set of never-ending data streams. Student(s) may start by understanding the detail of Generic-Join mechanism over multiple streaming data. The student/group can continue by implementing the algorithm using CUDA C++ ro Julia standard libraries. Next step is to evaluate the run-time performance of Generic-Join approach over other implementations of multi-join such as hash-based join used in conventional database systems (e.g. MySQL), and streaming engine such as Apache Spark/Storm/Flink. Finally, possible performance improvement of implementations (such as parallel computation, pipelining the kernel functions, and local on-chip memory) can be further explored.

Expected outcomes/deliverables:

1. A set of Julia/C++ CUDA libraries/scripts for implementation and testing of Generic-Join approach on NVIDIA's accelerators.
2. A detailed report that explains the internal mechanism of the developed system and interactions among different components as well as is expected.
3. a detailed performance evaluation report

Reading material:

- Julia/C++ CUDA SDK documentation for NVIDIA's accelerators
- Timely Dataflow architecture and its Progress Tracking documentation

References:

[1] Hung Q. Ngo and Christopher Re and Atri Rudra, ``Skew Strikes Back: New Developments
in the Theory of Join Algorithms``, https://arxiv.org/abs/1310.3314, 2013
[2] Frank McSherry, Worst-case optimal joins, in dataflow,
http://www.frankmcsherry.org/dataflow/relational/join/2015/04/11/genericjoin.html

Requirement to be on campus: No

CS22/26 Neural Interfaces for Visually Impaired as Tactile Enabling Technology

Supervisor: Dr Anusha Withana

Eligibility: You will work with the supervisor and a PhD student, and we expect you are a fast learner. Excellent skills in programming, skills in embedded systems, machine learning, knowledge in design and fabrication, and human computer interaction are added benefits.

Project Description:
Modern computers frequently use visual and auditory interfaces. For example, we can see the visual information on our screens and hear the audio computers generate. However, one important modality is less explored, that is, what we feel through our skin. We feel vibrations, temperature, and pressure through sensory receptors in our skin. This modality is particularly important to create enabling interfaces, for instance computer interfaces used by visually impaired people. In this project, we will explore how we can create enabling technologies using novel tactile interfaces, particularly an interface directly communicates with our neural system.

Requirement to be on campus: Yes *dependent on government's health advice

CS22/27 Shoe-Sense: Non-optical methods for tracking in-shoe foot motion

Supervisors: Dr Anusha Withana, Dr Alycia Fong Yan and Prof Alistair McEwan

Eligibility: You will work with the supervisor and a PhD student, and we You will work with three supervisors (computer science/health and medicine/bio med), and we expect you are a fast learner. Excellent skills in programming, skills in embedded systems, machine learning, knowledge in design and fabrication, and human computer interaction are added benefits.

Project Description:
Understanding the effect of footwear on foot motion has many health benefits for all populations. It is also important to inform footwear designers and manufacturers. The current gold standard optical method of motion capture uses retro-reflective or active markers captured with multiple infrared cameras. This requires cutting holes in the shoe to place the markers and cannot be used outside lab settings. In this project we plan to develop a novel, magnet based wearable sensor to track foot movements inside the shoe without any modification to the shoe. Currently we have developed initial prototypes using sock worn magnets. The goal of this project is to improve the sensor and collect data and analyse them.
Requirement to be on campus: Yes *dependent on government's health advice