Statement of Research

H.M.N. Dilum Bandara

With the advent of Microelectromechanical Systems (MEMS) sensors and actuators have become miniature and low cost enabling widespread deployments. Such sensors enhance the perception of our surroundings by sensing the physical world at a far greater spatial and temporal granularity than have been hitherto possible. Value of the sensor data grows exponentially when multiple and different forms of sensors are connected together. For example, smart-city applications rely on a multimodal set of sensors such as CCTV cameras, smart traffic cameras, ambient sensors, smart meters, water level and quality sensors, wearables, and even social media feeds by citizens. These disparate sensors have different power, computing, bandwidth, and latency characteristics; hence, require a mix of infrastructure based and ad-hoc networks for interconnection. While sensor mote-level communication primarily relies on data-centric, low power, low bandwidth, and multi-hop routing, gateways and data aggregators are typically based on TCP/IP. Once delivered, sensor data need to be timely processed to generate useful insights while dealing with four Vs of big data, namely high volume, velocity, variety, and veracity. Thus, it is imperative to provision on-demand and scalable computing resources and Machine-Learning (ML) algorithms to integrate, clean, and process data. Therefore, our realization of the grand visions such as smart cities and Industry 4.0 lies at our ability to seamlessly integrate and timely perform sensor data *generation*, *transmission*, and *processing*. My research addresses the challenges across these three aspects of the Internet of Things (IoT) and Cyber-Physical Systems (CPS) where I have extensively worked towards addressing interdisciplinary problems in weather/flood monitoring and fleet management.

Graduate Research

During my Masters' thesis, I addressed inter-node data transfer problems in a Wireless Sensor Network (WSN) based subsurface, chemical plume monitoring system. To enable message and power efficient communication in large WSNs, I adopted the idea of Virtual Sensor networks (VSNs), an emerging concept that supports dynamic and collaborative WSNs. I developed a configurable, top-down cluster and Cluster Tree (CT) formation algorithm that enables the formation of a CT with controlled breadth and depth, uniform cluster size, and hexagonal-like clusters. To power efficiently detect and track the movement of a chemical plume, I extended the algorithm to dynamically group sensors that detect the same chemical plume. Then a CT-based routing strategy was developed to facilitate node-to-sink/node/group communication. This is the first group formation and communication solution for VSNs. The algorithm was further extended to support the formation of a secure backbone that enables secure upper-layer functions and dynamic distribution of cryptographic keys. The efficacy of the solution was demonstrated by simulating the subsurface, chemical plume monitoring system, where I worked closely with two water resource scientists from the Colorado Schools of Mines, USA to understand the application requirements and fluid dynamics aspects.

My Doctoral work focused on Distributed Collaborative Adaptive Sensing (DCAS) systems where distributed and timely data processing is essential among resource and data-rich sensors such as radars. I proposed to use Peer-to-Peer (P2P) communities to enable nodes with diverse capabilities to collaborate in achieving greater sensing, communicating, and computing tasks promptly. Through an Exploratory Data Analysis (EDA) of four real-world systems , I demonstrated that the characteristics of real-world resources and queries are very different compared to widely used assumptions in related work. I then developed a tool to generate large synthetic traces of multi-attribute resources and range queries using statistical and timeseries behavior learned from the EDA and techniques such empirical Copula. Using those traces, simulation, and a set of equations derived to capture the cost of multi-attribute Resource Discovery (RD), I demonstrated that the related work is ill-equipped to address the real-world cases. Consequently, I developed a resource and query aware P2P-based multi-attribute RD solution that is both efficient and load balanced. Further, I developed a distributed caching solution that exploits P2P communities to improve the community-wide and system-wide lookup performance. I also proved that the distributed caching problem is NP-complete, and a satisfactory solution could be derived by combining a sub-overlay formation scheme and a local-knowledge-based distributed caching algorithm. Moreover, I demonstrated the applicability of Named Data Networking (NDN) for DCAS systems by developing a distributed multi-user, multi-application, and multi-sensor data fusion solution. The proposed solutions and the analysis apply to a wide variety of contexts such as DCAS systems, P2P, grid/cloud/fog computing, job schedulers, and GENI (Global Environment for Network Innovations). This work was carried out and evaluated in the context of the Collaborative Adaptive Sensing of the Atmosphere (CASA) project, a multidisciplinary partnership among four universities, industry, and government to engineer revolutionary weather-sensing networks in the USA.

Current Research

While Usage Based Insurance (UBI) is introduced in several countries, it primarily relies on a one-time snapshot of a driver and basic metrics such as the number of hard breaks. We developed a cloud-based, vehicular data acquisition and analytics system for real-time and long-term driver behavior monitoring, trip analysis, and vehicle diagnostics to support advanced UBI and vehicle self-care applications. The proposed system consists of an On-Board Diagnostics (OBD) port to Bluetooth dongle, a mobile app named Kampana (available on Google Play), and a cloud-based backend. Based on OBD data, Complex Event Processors (CEPs) at both the smartphone and backend detect and notify unsafe and anomalous events in real time. We calculate a driver's relative skill, aggression, comfort, and safety scores, as well as predict vehicle service times and impending sensor failures using statistical and ML techniques such as Regression Analysis, Kurtosis, Hidden Markov Model, k-Means Clustering, and AdaBoost algorithm. The system was tested on multiple vehicles and results confirmed that the computing, bandwidth, and power consumption of the app are reasonable. This work leads to a start-up company that focus on vehicular data analytics.

Noticing our work on data analytics a vehicle delivery company in UK requested us to develop a driver-scheduling solution for their vehicle delivery business. Similarly, Nimbus Venture (Pvt) Ltd., Sri Lanka, a fleet management company, reached us to conduct an EDA of customer data, predict fuel consumption, and develop vehicle and driver scheduling solutions for Ready Mix Concrete (RMC) and heavy goods delivery. We prosed a Random-Forrest based solution to predict the fuel consumption with 95% accuracy and to classify fuel-inefficient driving based on both the driver dependent parameters and external factors such as traffic, road topography, and weather. A rule checker and a scheduler based solution was proposed for scheduling problems. Rule checker enforces constraints such as deadlines, driver skills, driver and truck availability, and working and resting hours. Scheduler used Simulated Annealing (SA) to assign jobs while maximizing the job coverage and minimizing the overall cost. Using workloads derived from customer data, we demonstrated that the proposed SA-based solutions outperform Genetic and Hill Climbing algorithms. We recently developed a crowdsourced, smartphone-based road condition and fuel consumption monitoring and mapping solution targeting the Sri Lankan road network. We are currently calibrating and finetuning the signal processing and ML techniques to accurately detect potholes and estimate International Roughness Index (IRI).

While applications such as weather predictions have been mostly running on HPC clusters, a better cost-performance ratio could be gained by running them on Public Cloud. Therefore, we proposed to build the computing infrastructure of the Center for Urban Water (CUrW), Sri Lanka (i.e., flood control and water management system for Metro Colombo) on Public Cloud. Today, CUrW uses a fully containerized Cloud platform (first of its kind) connected to a flexible workflow engine. We achieved this by containerizing some of the difficult models such as WRF, HEC-HMS, and Flo-2D, as well be by extending Apache Airflow to support hierarchical and loop-based workflows. We are also working on a weather data integration and assimilation system for CUrW that supports multimodal weather data.

Other recent work includes addressing proactive, workload, resource, policy, and cost-aware auto-scaling challenges in IaaS and PaaS Clouds. This is achieved by developing an ensemble workload prediction model, cost model, and smart killing. We further developed a multi-cloud library that could be integrated into the SaaS application to provide automatic failover while alleviating vendor lock-in and single point of failover problem. The utility of these solutions was demonstrated using both simulations and implementations using AWS and CloudStack. A European client uses the multi-cloud library since early 2017. Another area of work includes CEP, which is essential for real-time stream analytics in IoT. I worked closely with WSO2 Inc. (developers of Siddhi CEP engine) to improve Siddhi's throughput and latency using GPUs, automate CEP query generation, and developed a CEP engine for Arduino-based edge devices. Moreover, we demonstrated that a low-end weather monitoring framework could be developed using CEP and basic ML. Furthermore, we developed a Software Defined Networking (SDN) based OpenFlow enabled smart-home gateway for IoT applications. A heuristic-based clustering solution to visualize and rationalize Police patrol beats based on a city's crime data was also developed.

Future Directions

My future research would be an outgrowth of the challenges experienced while working on data generation, transmission, and processing in IoT and CPS. I plan to work on essential elements of smart cities such as disaster prediction and mitigation, intelligent transportation systems, industry 4.0, and open data. In fact, very recently, I got an internal grant to develop a data integration platform for smart cities. I plan to enable seamless movement of multimodal data across agencies/users while keeping the data distributed with the data owners and enabling them to define finer authorization levels. I believe this would reduce the reluctance to open up data due to issues such as ownership, recognition, multi-party integration, and data quality.

I plan to extend my blockchain work beyond food traceability and spectrum allocation in cognitive and citizens broadband radio services to other smart-city related applications such as Self-Driving Vehicles (SDV), ride and car sharing, and citizen-generated renewable energy. While IoT, CPS, and WSN envisioned rich actuators, most systems to date are open-loop (except in home automation) where processed data is usually visualized on a dashboard enabling human experts to initiate some action. Whereas to realize the full benefits of these systems we need to run them in closed-loop operation, as it enables accurate, reliable, fast, and dynamic systems. I believe it is high time that we realize the true potential of IoT by fully exploiting closed-loop operation by addressing challenges like automated decision making, finer control, feedback, stability, redundancy, and failover.

While the advent of SDVs could make my work on driver behavior modeling less relevant, plenty of work can be done in the areas of vehicle-to-x communication, vehicle safety, and vehicle self-care. This is essential as the driver/passengers will no longer have a sense of minor issues such as steering difficulties, wheel alignment, noises, and engine misses which could be early warnings of much bigger problems. While modern vehicles are well equipped with sensors, our challenges lie in the ability to collect and transmit high volumes of sensor data while on the road and process them fast enough to detect impending issues. Thus, I plan to work on both in-vehicle and in-network data processing, efficient content dissemination and retention, and real-time and long-term data analytics. Moreover, rich monitoring and modeling of SDV behavior are essential to quickly adapt our road infrastructure which is developed based on over 100-years of human-centered driving.

One of the challenges I faced in fleet scheduling is the accommodation of last minutes jobs and cancellations while still covering already scheduled and confirmed jobs. Among the possible optimization algorithms, there seems to be no easy solution than re-running the entire solution.

This is not acceptable in cases where the complexity of the optimization algorithm is too high, or vehicles are already on the field and have covered some of the jobs. Therefore, I plan to explore low-complexity algorithms that still provides optimized job coverage and cost.

While ML has become an everyday buzzword, the widening gap in data literacy and access to data is poised to make it the next global differentiator. Through major Public Cloud providers now offer ML as a service, choosing a fitting ML technique, identifying a good feature set, and interpreting the data remain magic to most. With our pursuit for more accurate algorithms that could unearth even the most obscure patterns in rich data, we are missing out on many applications with low or no data yet good enough predictions and recommendations are equally important. While rich datasets give established business every opportunity to squeeze in existing and potential customers, new business struggle to stay competitive due to the lack of data. Nevertheless, customers expect the same level of customizations and relevance from new businesses and quickly revert to established players when new ones are unable to keep up. Therefore, it is imperative to work on bootstrapping and low-data problems where businesses that are either new or with modest volumes of data could still give meaningful predictions and recommendations. While education could bridge the data literacy gap, algorithms for automatically choosing ML models, feature identification, and parameter tuning are also of interest. While simple models and techniques are the need of the former, latter requires more complex ML techniques. Moreover, while we collect and process data at massive scale, our ability to safeguard the privacy is questionable. Thus, data anonymization techniques that preserve data utility, provide provable properties, and applicable across different datasets are essential to bridge the access to data gap. Therefore, I believe further work in these areas would establish a level playing field for data analytics.

Achievements

My work has been published at IEEE Tran. on Parallel and Distributed Sys., J. of Network and Computer App., Computer Networks, ACM Tran. on Internet Technology, and IEEE Cloud, ICC, and GLOBECOM. My students and I have won Student Paper Award (Merit) at IEEE SOLI 2018, Best Poster at ERU Symposium 2018, Best Poster at ISTec Student Research Poster Contest 2008, and Outstanding Research Performances at the University of Moratuwa in 2015 and 2016 (2017 results pending). I have been able to secure three grants from the Senate Research Committee and student stipends from the industry (worth USD 30,750) and 4-year, USD 675,000 World Bank grant for CUrW computational infrastructure R&D in collaboration with Prof. Srikantha Herath. 13 MSc, 12 MBA in IT, and 59 Undergraduate students have graduated under my research supervision, and six MSc students are finalizing their thesis. I currently supervise three PhD, seven MSc, and four MBA in IT students. My interest to promote research at the University prompted me to take up research administration positions such as the Director, Engineering Research Unit of the Faculty; committee member roles at the Faculty of Graduate Studies, Senate Research Committee, Faculty Higher Degrees Committee, and Post Graduate Board of Studies; and President, Student Leadership Council of the NSF Engineering Research Center for CASA.