My research interests are in the areas of networks and computer systems. Over the past five years, I have addressed some key research problems spanning the control, service, and management planes of networked systems. To solve these problems, I have employed techniques in Operations Research, Computational Complexity Theory, Meta-Heuristics, and Algorithm design.

Past Research: Supporting services hosted in Cloud-based Data Center Networks

The spectrum of online services that we use on a daily-basis have surpassed being a mere subsidiary, to becoming a necessity. Services continuously enhance and facilitate virtually every aspect of our daily lives. My long-term research goal is to support current, emerging, and future online services by improving the underlying infrastructure and systems. To this end, in my doctoral studies I focused on improving the support of services hosted in Cloud Data Center networks. Cloud Computing [1] offers many salient features, e.g., lowering CAPEX and OPEX, dynamic resources provisioning, support for elasticity and geo-distribution, among others. Furthermore, this paradigm lowers the barrier to entry, and enables start-ups and Over-the-Top service providers to quickly market their services without the need for upfront investment in infrastructure hardware. For instance, AirBnB [2] is one example of an Over-the-Top service provider that leveraged the benefits of the Cloud by migrating its service to the Amazon EC2 platform. Moreover, Cloud Computing enables service providers to scale their infrastructure needs to fit their business growth over time. As this transition unfolds, multiple challenges emerged. Namely, with the rapid rise and adoption of Cloud Computing, Data Centers have grown at an unprecedented scale [1]. Further, the traffic pattern in Cloud-based Data Centers shifted from a predominantly north-south traffic to east-west traffic [3]. This is problematic since Data Centers topologies were mainly designed with north-south traffic in mind. Additionally, services hosted in the Cloud exhibit stringent Service Level Agreements (SLAs) in terms of compute, network resources, and reliability guarantees (among others). Indeed, any violation of SLAs incurs severe monetary penalties for Cloud providers. Moreover, service providers typically demand value-added network functions to monitor [4] and/or enhance the performance [5] and security of their services. These problems of scale, reliability, traffic management, and resource provisioning are fundamental to efficiently support services running in the Cloud. During my doctoral studies, I have addressed the following key research questions:

- How to achieve scalable Traffic Engineering in Cloud Data Center networks?
- How to efficiently provision resources for multi-tenant services hosted in Cloud Data Center networks?
- How to guarantee reliability of Cloud Services in failure-prone Data Center networks?
- How to provision software-based network functions to enhance the performance and/or security of Cloud services?

I devised mathematical models to optimally solve each of these problems. I truly enjoy modeling networking problems mathematically as it allows to capture the intricate relationship between the different interplaying variables and constraints. These models further constituted the foundation to develop optimization- and algorithmic- based techniques to provide scalable and efficient solutions. Further, in several occasions I employed computational complexity theory to prove the intractability of the problems at hand. Learning and employing these techniques to practical networking problems has provided me with strong and solid skills to take on future research endeavors.

Ongoing Research: Machine Learning for Networking

Machine learning is perched on the 2016 Gartner Hype Cycle [6] for emerging technologies, and will likely remain at top of the list for the upcoming 2 to 5 years. Along with a team of PhDs, postdoctoral researchers, and research professors at the University of Waterloo, I am investigating the impact of this technology on networking. Specifically, we are currently devising a thorough study on existing work that used machine learning for networking. We aim not only to shed light on related literature, but also on the potential, as well as the obstacles and limitations, of incorporating intelligence in networking across its different planes.

In addition, I am currently investigating the role that Machine Learning can play to enhance the security of online services. Traditional security measures employed today rely on signature-based detection to find patterns of known attacks. However, this renders the network vulnerable to "Zero-Day" attacks. The main research question here is: *How can we detect something that we have not previously seen?*. Previous research efforts have revealed that security threats exhibit distinguished behavior from normal communication patterns. Hence, if one can detect these anomalies with high accuracy we may be able to detect unknown attacks. This however is a daunting task given the high complexity and heterogeneity of the network. Some behavior that is abnormal in one instance can also be a completely normal one under different circumstances; e.g., a spike in traffic on a weekday can be perfectly normal, while the same spike on a weekend can be indication of a Denial of Service (DoS) attack. Machine Learning has been proven to understand the complex structure of underlying raw data. This has motivated my current work on employing Machine Learning for Early Signs of Malware Detection. This work is in collaboration with the Royal Bank of Canada.

Future Research: Supporting Services Running in 5th Generation Networks

A natural extension to my previous work is to consider the challenges related to supporting services running in the next generation of mobile networks (5G). This research area is particularly pertinent to my background and expertise given the envisioned architecture for 5G networks. Mobile operators are now facing an explosive growth of mobile data traffic. With spectrum shortage, mobile operators turn towards dense mobile infrastructure deployment to support this growth; which yields significant increase in CAPEX and OPEX. As the average revenue per mobile user is at a stand-still, mobile operators are looking to keep pace with the increase in traffic volume with minimal additional cost. Network resources sharing is thus emerging as a key business model for reducing future costs [7]. Mobile operators are thus embracing the concepts of Cloud Computing, and Network "Softwarization", and aiming to redesign the next generation of mobile networks with inherit support for multi-tenancy [8]. Mobile services will be running as end-to-end network slices; where each slice consists of virtualized resources and network functions extending from the antennas to the network core. These slices will coexist atop the same infrastructure and share its resources.

Network slicing is fundamental to the realization of 5G. While this problem bares some similarities to the resource allocation problem in the Cloud, it also exhibits notable differences in terms of the underlying infrastructure, the characteristics of mobile services, and their differentiated QoS requirements. The latter is critical given the various classes of services that 5G aims to support [8]; e.g., bandwidth-intensive services (e.g., streaming HD video), low-latency and high-reliability services (e.g., self driving cars). I have thus planned my research agenda for the next five years around future anticipated developments in 5G. I aim to answer the following open research questions:

- How to provision network slices each tailored to the different class (type) of mobile services?
- How to achieve statistical multiplexing without violating the QoS requirements of mobile services?
- How to support mobile services elasticity?
- How to circumvent failures and congestions in this multi-tenancy environment.

References

[1] Qi Zhang, Lu Cheng, and Raouf Boutaba. Cloud computing: state-of-the-art and research challenges. *Journal of internet services and applications*, 1(1):7–18, 2010.

[2] Airbnb case study. https://aws.amazon.com/solutions/case-studies/airbnb/. Accessed: 2017-03-08.

[3] HP Networking Blog effects of virtualization and cloud computing on data center networks. http://www.ics.uci.edu/ cs237/reading/Download.pdf. Accessed: 2017-03-08.

[4] Amazon Web Services LLC amazon cloudwatch. Technical report. Accessed: 2017-03-08.

[5] Amazon Web Services LLC elastic load balancer. Technical report. Accessed: 2017-03-08.

[6] Gartner hype cycle. http://www.gartner.com/newsroom/id/3412017. Accessed: 2017-03-08.

[7] Jeffrey G Andrews, Stefano Buzzi, Wan Choi, Stephen V Hanly, Angel Lozano, Anthony CK Soong, and Jianzhong Charlie Zhang. What will 5g be? *IEEE Journal on selected areas in communications*, 32(6):1065–1082, 2014.

[8] NGMN Alliance. 5g white paper. *Next generation mobile networks, white paper*, 2015.