

Fall 2017 Seminar Series

Reading between the lines of datacenter logs

November 22nd 2017

Time 10:30am-11:30am – HEC 450

How can rigorous data analysis, based on various logs collected at large-scale datacenters, help us improve the resilience and performance of these systems and the applications they run? In this talk, I will first show how trace-driven analysis helped us uncover various interesting patterns in the behaviour of jobs running on large clusters, using data from Google, CMU, and Los Alamos National Lab. Our analysis revealed properties that distinguish unsuccessful jobs from others, including certain configuration settings and resource consumption patterns. Using these insights, we designed a machine learning-based framework for predicting job terminations with high precision and recall, and demonstrated how these predictions can be used to mitigate the effect of unsuccessful executions in datacenters. In the second half of this talk, I will dig deeper into examining the memory behaviour of applications, since memory accesses limit the performance and scalability of numerous workloads in real systems. I will show how careful cache and memory profiling can help us better understand the way programs interact with the memory hierarchy, while discussing and demonstrating practical use cases for real-world systems.

Nosayba El-Sayed
Postdoctoral Associate at CSAIL, MIT

Nosayba El-Sayed is a Postdoctoral Associate at CSAIL, MIT. Her research focuses on designing and implementing data-driven techniques that exploit the wealth of data generated in modern platforms to improve the reliability and performance of large-scale datacenters. She completed her PhD at the University of Toronto, during which time she interned at Amazon's Datacenter Global Services division to work on server-outage analysis and prediction. More recently, Nosayba has focused on investigating how new features available in modern hardware (e.g. cache partitioning) can be used to improve datacenter utilization. Nosayba's work was published in conferences such as SIGMETRICS, DSN, ICDCS, SC, and HPCA (coming soon). Her work on datacenter reliability received a SIGMETRICS best paper award and was featured in ;login! Usenix Magazine, Data Center Knowledge, and Communications of the ACM.

Hosted by: Dr. Aziz Mohaisen

