Abstract:

Nowadays geographically distributed data centers are common supporting infrastructure for cloud systems to provide customers with better reliability and quality of services. The computing capacity and scale of such data centers keep increasing to meet the soaring demand from cloud applications and services. Each of such data centers can consume up to megawatts of electricity. The hunger for power brings large electricity bills to cloud providers (e.g., Google and Microsoft) and causes significant negative impact to the environment. The associated energy cost and emissions have become major concerns of these cloud providers.

In this talk, I will present my study on designing energy-efficient, low-cost, and environment-friendly operation for cloud data centers. First, I will introduce Geographical Load Balancing (GLB), which is defined as spatial workload placement among geographically distributed datacenters. GLB explores the spatial flexibilities of workloads and the spatial diversities of parameters such as electricity prices and emission factors in different regions, in order to realize cost reduction and sustainability improvement for data centers. Then, I will discuss a case study, which studies applying GLB to cloud-based video streaming systems (e.g., YouTube). For this case study, I develop an on-line distributed algorithm and leverage competitive analysis to demonstrate its performance guarantee. Finally, I will highlight my research plan on cloud computing and data centers.

Short Bio:

Fanxin Kong is currently a Ph.D. candidate in the School of Computer Science at McGill University. He received his Bachelor and Master Degrees both in Computer Science and Technology from Northeastern University, China. He was awarded AMD Scholarship by Advanced Micro Devices, China. He is the recipient of Graduate Excellence Award at McGill University.

His research interests include computer systems, computer networks, and power systems, and specially span the areas of cloud computing and big data, cyber-physical and real-time systems, and smart grid and energy systems. He has published over 20 research papers in major peer-reviewed international journals and conference proceedings in these areas.