Bringing Performance to the Cloud

**SPEAKER**  Dr Dongsu HAN
Assistant Professor
Department of Electrical Engineering
Korea Advanced Institute of Science and Technology
Korea

**DATE**  20 April 2015 (Monday)
**TIME**  11:00 am - 12:00 noon
**VENUE**  CS Seminar Room, Y6405, 6th Floor
           Yellow Zone, Academic 1
           City University of Hong Kong
           83 Tat Chee Avenue
           Kowloon Tong

**ABSTRACT**

Cloud computing has been tremendously successful in providing the scalability needed for popular Internet-based services such as Facebook, Google, and Netflix. Popular Internet-based services that we use every day rely on a public or private cloud platform to provide their services world-wide. To scale out the system performance, these services utilize hundreds of thousands to millions of servers that cost billions of dollars. Effectively utilizing individual resources in the cloud becomes crucial in reducing the cost (and the energy consumption) at this scale.

In this talk, we look at how to improve the performance of a single machine within the cloud. One of the key problems that limits the performance is that applications and networking stacks are not design to scale well in multicore environments. Modern machines have tens of cores and multiple 10Gbps Ethernet links. However, existing designs cannot effectively utilize these resources. In this talk, we focus on improving the performance of two essential building blocks, an in-memory key-value store and the TCP stack, that Internet-based services commonly rely on. By employing multi-core aware designs and by-passing the OS kernel to eliminate its overhead, we show that we can dramatically increase their performances by up to 13.5x and 3x respectively.

**BIOGRAPHY**

Dongsu Han is an assistant professor at KAIST in the Department of Electrical Engineering and the Graduate School of Information Security. He received his Ph.D. from the Computer Science Department at Carnegie Mellon University in 2012. His research interests include Internet architectures, cloud and distributed systems, and Internet content delivery. As part of this thesis work, he worked on a new Internet architecture called XIA, one of the major, still on-going future Internet architecture projects funded by the NSF.

All are welcome!