

Faculty Candidate Seminar

Scalable and Secure Cloud Service in Big Data Systems

Yuzhe (Richard) Tang, Georgia Institute of Technology

Monday, March 31st 11:00 to 12:00pm

Venue: CS205

(Refreshments will be served at 10:45 a.m.)

Abstract – Cloud computing and big data technology continue to revolutionize how computing and data analysis are delivered today and in the future. To store and process the fast-changing big data, various scalable systems (e.g. key-value stores and MapReduce) have recently emerged in industry. However, there is a huge gap between what these open-source software systems can offer and what the real-world enterprise applications demand. First, scalable key-value stores are designed for simple data access methods, which limit their use in advanced database applications. Second, the key-value stores, usually hosted at a third party public cloud, require security mechanisms to protect cloud customers' data in terms of data authenticity and confidentiality. Third, the demand continues to grow for privacy preserving search over big data from multiple autonomous data providers, as exemplified by the Healthcare Information networks.

In this talk, I will present my research work towards bridging these gaps. First, I will describe a suite of privacy preserving index algorithms that allow data sharing among unknown parties and yet maintaining a desired level of content privacy. To securely construct the index, it makes a very efficient use of the traditional expensive multi-party computations. Second, I will talk about the HIndex work for high performance real-time indexing on top of write-optimized key-value stores (e.g. HBase and Cassandra). HIndex has been implemented and is currently being transferred to an IBM big data product. Finally, I will briefly overview my other work on providing efficient authentication mechanisms on key-value stores and system performance optimizations of streaming applications on multi-core machines.

Bio - Yuzhe Tang is a Ph.D. student in the College of Computing at Georgia Institute of Technology. His research interests are broadly in distributed systems and cyber security. Specifically, his Ph.D. work has focused on building scalable secure systems for big data management in the cloud; he has lead research projects on a variety of topics including big data indexing, secure key-value stores, efficient privacy preserving computations, performance optimizations of stream applications and indexing in peer-to-peer networks. Most often, his research begins with a real-world problem, and ends with a practical solution that is implemented. He has done internships at IBM Research, NEC Labs and Microsoft Research. He has received his BSc degree and MSc degree in computer science and engineering from Fudan University, Shanghai, China, in 2006 and 2009 respectively. More information can be found at his school web page, <http://www.cc.gatech.edu/~ytang36/>.