



HITC/IGB Joint Seminar



Large-Scale Privacy-Preserving Mapping of Human Genomic Sequences on Hybrid Clouds

XiaoFeng Wang

Associate Professor

School of Informatics and Computing, Indiana University

Thursday, February 9, 2012, 4:00 PM

Siebel Center for Computer Science, Room 2405

One of the most important analyses on human DNA sequences is read mapping, which aligns a large number of short DNA sequences (called reads) produced by sequencers to a reference human genome. The analysis involves intensive computation (calculating edit distances over millions upon billions of sequences) and therefore needs to be outsourced to low-cost commercial clouds. This asks for scalable privacy preserving techniques to protect the sensitive information sequencing reads contain. Such a demand cannot be met by the existing techniques, which are either too heavyweight to sustain data-intensive computations or vulnerable to re-identification attacks. In this talk, I describe a new technique that makes an important step towards secure and scalable read mapping on the hybrid cloud, which includes both the public commercial cloud and the private cloud within an organization. Inspired by the famous “seed-and-extend” method, our approach strategically splits a mapping task: the public cloud seeks exact matches between the keyed hash values of short read substrings (called seeds) and those of reference sequences to roughly position reads on the genome; the private cloud extends the seeds from these positions to find right alignments. Our novel seed-combination technique further moves most workload of this task to the public cloud. The new approach is found to work effectively against known inference attacks, and also easily scale to millions of reads.

Dr. XiaoFeng Wang is an associate professor in the School of Informatics and Computing at Indiana University. He received his Ph.D. in ECE from CMU in 2004, and has since been a faculty member at IU. Dr. Wang is a recognized researcher on system and network security. His current work focuses on privacy issues in processing and dissemination of human genome data and security/privacy issues in Cloud Computing. He is a recipient of 2011 PET Award and the Best Practical Paper Award at the 32nd IEEE Symposium on Security and Privacy. His research is supported by the NSF, Department of Homeland Security, the Air Force and Microsoft Research. He served as the acting director for the Security Informatics program including the Master Program in Security at IU in 2010.

www.healthit.illinois.edu