Abstract:
A large number of online databases are hidden in the "deep web" and only accessible through restrictive search or browsing web interfaces. We consider third-party data analytics over these hidden databases, specifically the problems of crawling, sampling, and aggregate estimations. We also explain how the recent advancements of such data analytics techniques pose significant privacy threats to certain sensitive aggregate information over hidden databases. The protection of sensitive aggregates stands in sharp contrast to the traditional privacy problem where individual tuples must be protected while ensuring access to aggregating information. We propose privacy-preserving techniques to suppress the inference of aggregate information from hidden databases.

Bio:
Dr. Nan Zhang is an Associate Professor of Computer Science at the George Washington University, Washington, DC, USA. He received the B.S. degree from Peking University in 2001 and the Ph.D. degree from Texas A&M University in 2006, both in computer science. His current research interests include databases and information security/privacy. He received the NSF CAREER award in 2008.