

## Computer Science Seminar Series, 2010

### National Capital Region

# Data Exploration and Privacy Preservation over Hidden Web Databases

**Speaker: Prof. Nan Zhang**  
Department of Computer Science  
George Washington University

**Friday, September 10, 2010**  
**1:00PM-2:00PM, NVC 325**

#### Abstract

A large number of online databases are hidden behind form-like web interfaces which allow users to execute search queries by specifying desired (ranges of) attribute values of the sought-after tuple(s). We consider the problem of approximate query processing over hidden databases, and propose novel sampling-based techniques which use a small number of queries through the web interface of a hidden database to produce unbiased estimates with small variance. We also explain the threats posed by such sampling-based techniques on sensitive aggregate information of hidden databases. The protection of sensitive aggregates stands in contrast to the traditional privacy problem where individual tuples must be protected while ensuring access to aggregating information. We propose privacy-preserving techniques to thwart bots from sampling the hidden database to infer aggregate information.

\*Collaborative work with Xin Jin of George Washington University, Arjun Dasgupta, Bradley Jewell, Anirban Maiti, and Dr. Gautam Das of University of Texas at Arlington, and Dr. Surajit Chaudhuri of Microsoft Research.

#### Biography

Dr. Nan Zhang is an Assistant Professor of Computer Science at the George Washington University, Washington, DC, USA. Prior to joining GWU, he was an assistant professor of Computer Science and Engineering at the University of Texas at Arlington from 2006 to 2008. 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 span security and privacy issues in databases, data mining, and computer networks, including privacy and anonymity in data collection, publishing, and sharing, privacy-preserving data mining, and wireless network security and privacy. He received the NSF CAREER award in 2008.