Guozhang WANG

Department of Computer Science
4107A Upson Hall
Cornell University, Ithaca, NY

Phone: (607) 339-8352
E-mail: guoz@cs.cornell.edu

RESEARCH INTERESTS

Large-scale data driven systems with particular interests in cloud computing infrastructure.

EDUCATION

- **September 2008 – present** — **Cornell University, Ithaca, NY, USA**
  Ph.D., Computer Science
  Minor: Statistics
- **July 2006 – November 2006** — **National University of Singapore, Singapore**
  Exchange Student, Computer Science
- **September 2004 – June 2008** — **Fudan University, Shanghai, China**
  B.S., Science and Technology

HONORS AND AWARDS

- **Cornell University Computer Science Department TA Excellence Award**, May.2013
- **Cornell University Computer Science Department TA Excellence Award**, May.2009
- **Tung Orient Overseas Container Line Scholarship** (top 5/140 of the department), Oct.2007
- **Chun-Tsung Scholar of China** (supported by Nobel Prize Laureate: Lee, T.D), May.2007
- **Fudan University First-class Renmin Scholarship**, Oct.2005, Oct.2006
- **Chinese Mathematical Olympiad First-class Award** (totally 33 students of the state), Feb.2004

PUBLICATIONS

Journal Articles

- Xiaokui Xiao, *Guozhang Wang*, Johannes Gehrke, "Differential Privacy via Wavelet Transforms", In *TKDE special issue of "Best Papers of ICDE 2010", 2010*.
- Michaela Götz, Ashwin Machanavajjhala, *Guozhang Wang*, Xiaokui Xiao, Johannes Gehrke, "Publishing Search Logs – A Comparative Study of Privacy Guarantees", In *TKDE, 2009*.

Conference Proceedings

- *Guozhang Wang*, Wenlei Xie, Alan Demers, Johannes Gehrke, "Asynchronous Large-Scale Graph Processing Made Easy", In *Proc. CIDR 2013*.
- Tao Zou, *Guozhang Wang*, Marcos Vaz Salles, David Bindel, Alan Demers, Johannes Gehrke,

Walker White, "Making Time-stepped Applications Tick in the Cloud", In *Proc. SOCC 2011*.

- *Guozhang Wang*, Marcos Vaz Salles, Benjamin Sowell, Xun Wang, Tuan Cao, Alan Demers, Johannes Gehrke, Walker White, "Behavioral Simulations in MapReduce", In *Proc. VLDB 2010*.
- Xiaokui Xiao, *Guozhang Wang*, Johannes Gehrke, "Differential Privacy via Wavelet Transforms", In *Proc. ICDE 2010*.
- Xiaokui Xiao, *Guozhang Wang*, Johannes Gehrke, "Interactive Anonymization of Sensitive Data", In *Proc. ACM SIGMOD 2009*. (Demo Paper)

Patents

- Sanjay Agrawal, *Guozhang Wang*, "Allocation of Tenants to Database Services", US Patent 12, 982,899, Dec 31, 2010.

TEACHING EXPERIENCE

- **2012 Fall**

CS 5320: "Introduction to Database", Teaching Assistant ***TA Excellence Award***

- **2012 Fall**

CS 5321: "Practicum in Database Systems", Teaching Assistant

- **2010 Spring**

CS 5300: "The Architecture of Large-Scale Information Systems", Teaching Assistant

- **2008 Fall**

CS 4320: "Introduction to Database", Teaching Assistant ***TA Excellence Award***

- **2008 Fall**

CS 4321: "Practicum in Database Systems", Teaching Assistant

RESEARCH AND WORKING EXPERIENCE

- **Sep. 2008 – Present**

Research Assistant, Cornell University

Title 1: Automatic Scaling Out Iterative Computations

Worked on building large scale frameworks for iterative computations, including simulations, graph processing, and machine learning applications. Leveraged natural properties of iterative computation patterns such as data locality and update order independence for efficient data partitioning and synchronization. Proposed a programming framework and the underlying parallel runtime for iterative graph processing applications to enable both programming simplicity and asynchronous executions by separating algorithmic logic with scheduling policies.

Title 2: Privacy-Preserving Data Publishing

Worked on mechanisms of publishing search logs including frequent keywords, queries and clicks under relaxed ε-differential privacy guarantees. Developed a data publishing framework applying wavelet transforms to provide accurate answers for range-count queries. Implemented a data anonymization toolkit with intuitive

interface that can interactively guide users through the privacy-preserving data publishing process.

- **May. 2012 – Aug. 2012**

Engineering Intern, LinkedIn Data Infrastructure Team
Title: Consumer Redesign in Apache Kafka
Worked on Apache Kafka, a distributed publish-subscribe messaging system. Focused on the consumer redesign project, which aims at migrating the rebalance logic to a centralized coordinator at the server side and removing functional dependencies such as ZooKeeper for communication and synchronization purposes.

- **May. 2011 – Aug. 2011**

Research Intern, Microsoft Jim Gray Systems Lab
Title: Implementing Queues in Main-memory Databases
Worked on main-memory component of the Microsoft SQL Server. Compared the performance of different lock-free data structures and algorithms for supporting queue services under various workloads. Studies related transactional semantics of queuing operations under the multi-version concurrency control mechanism.

- **May. 2010 – Aug. 2010**

Research Intern, Microsoft Research Redmond
Title: Multi-tenant Architecture for Conference Management Tools
Worked on Microsoft Conference Management Tool (CMT). Explored challenges of transmitting the CMT servers to a shared-table multi-tenant data infrastructure. Studied a related work of allocating applications to multiple database instances in the server.

- **Jul. 2007 – Nov. 2007**

Research Intern, Microsoft Research Asia
Title: Unsupervised Table Entity Retrieval for Non-Template Web Pages
Worked on techniques of Web table retrieval. Applied SVD for feature extraction using both web structure and text content information to detect table regions from web. Studied techniques for name entity recognition using character level n-gram Semi-CRF models.

- **May. 2007 – May. 2008**

Research Assistant, Shanghai (International) Database Research Center
Title: Data Stream Anomaly Detection
Worked on data stream processing, with special focus on anomaly detection. Applied Principle Component Analysis (PCA) to study the activity of different hosts from the same cluster.

SKILLS

- **Programming**

C/C++, Java, Matlab, SQL, Python, Scala

- **Operating Systems**

Linux (Ubuntu, Red Hat), Windows

REFERENCES

Johannes Gehrke, Professor
Dept. of Computer Science
Cornell University, USA

Ping Li, Assistant Professor
Dept. of Statistical Science
Cornell University, USA

Donghui Zhang, Senior Software Engineer
Development Team
Paradigm4, USA

Sanjay Agrawal, Software Engineer
Ads Backend Team
Google, USA

David DeWitt, Technical Fellow
Jim Gray Systems Lab
Microsoft, USA