

Xiaohui Helen Gu

CONTACT INFORMATION

Xiaohui Gu
IBM T.J. Watson Research Center
19 Skyline Drive
Hawthorne, New York 10532

Voice: (914) 784-6092 (O), (914) 299-1153 (C)
Fax: (914) 784-7455
E-mail: xiaohui.helen.gu@gmail.com
Homepage: www.research.ibm.com/people/x/xgu

RESEARCH INTERESTS

My general research interests are in the areas of computer systems, with a current focus on information processing systems, large-scale networked systems, and wireless mobile systems.

EDUCATION

University of Illinois at Urbana-Champaign, Urbana, IL, USA
PhD in Computer Science **October, 2004**
Dissertation: "SpiderNet: A Quality-Aware Service Composition Middleware"
Advisor: Prof. Klara Nahrstedt

University of Illinois at Urbana-Champaign, Urbana, IL, USA
MS in Computer Science **January, 2001**
Thesis: "Visual Quality of Service Programming Environment for Distributed Heterogeneous Systems", **awarded David J. Kuck Best Master Thesis Award**
Advisor: Prof. Klara Nahrstedt

Peking University, Beijing, China
BS in Computer Science **July, 1999**
Graduated with Honor

SELECTED HONORS AND AWARDS

- First Invention Plateau Award, 2006, IBM T.J. Watson Research
- Invention Achievement Award, 2004, IBM T.J. Watson Research
- Saburo Muroga Fellowship to outstanding graduate students, 2003 - 2004, UIUC
- ILLIAC Fellowship to best PhD candidates, 1999 - 2002, UIUC
- David J. Kuck Best Master Thesis Award, 2001, UIUC
- Student Scholarship to OSDI 2002, USENIX
- Motorola Scholarship to best undergraduate students, 1998, Peking University
- Outstanding undergraduate scholarships, 1995 - 1998, Peking University

RESEARCH EXPERIENCE

IBM T. J. Watson Research Center, Hawthorne, NY
Position Held: *Research Staff Member* **August, 2004 - December, 2007**

Adaptive load diffusion for windowed stream joins. Designed and implemented an *adaptive load diffusion* operator to enable scalable processing of Multiway Windowed Stream Joins (MWSJs) using a cluster system. The load diffusion is achieved by a set of novel semantics-preserving tuple routing algorithms. The load diffusion operator can (1) preserve the MWSJ semantics while spreading tuples to different hosts for parallel join processing; (2) achieve fine-grained load balancing among distributed hosts; and (3) perform semantics-preserving on-line adaptations to maintain optimal performance in dynamic stream environments. The load diffusion system has been deployed at the IBM System S stream processing cluster.

Distributed stream processing system. Designed and implemented a distributed stream processing system that combines distributed composition probing with coarse-grained global state management for instantiating query networks. Guided by the coarse-grained global state information, the system selectively probes a subset of candidate components to compose best query networks that can satisfy the user's QoS/resource requirements and achieve load balancing in the stream processing system. To adapt to dynamic stream environments, the system can adaptively adjust the number of probes to maintain a specified composition performance target (i.e., composition success rate) with minimum overhead. In addition, the system allows safe reuse of processing components by performing QoS impact projection to examine whether shared stream processing can cause QoS violations on existing running applications. Moreover, the system can dynamically discover and reuse available result streams as much as possible to avoid redundant computations. The system has been implemented and tested on the PlanetLab wide-area network testbed.

Stream dissemination overlay networks. Designed and implemented *BridgeNet*, an *adaptive multi-source stream dissemination overlay network*, which can efficiently collect streaming information from distributed locations and disseminate aggregated information to different stream consumers. BridgeNet provides a new *distributed cell tree* structure for multi-source stream aggregations, which can adaptively expand or contract itself in response to workload changes. In particular, BridgeNet performs *stream-pattern-based* cell tree adaptations, stream clustering, and overlay topology adaptations to deliver efficient stream dissemination without losing system stability. The system has been deployed at the PlanetLab wide-area network testbed.

Predictive failure management for large-scale cluster systems. Designed and implemented a *predictive failure management framework* that employs *failure predictions* to achieve cost-efficient, fault-tolerant continuous query processing. Timely failure warnings allow the system to minimize failure penalty for running queries while reducing fault tolerance cost by taking preventive actions on *abnormal* hosts or operators only. The predictive failure management framework offers (1) an adaptive log stream sampling technique that can dynamically adjust log data sampling rate based on the state of the monitored object, and employ reservoir sampling to incrementally maintain a desired size of training data set; (2) a self-evolving failure prediction model that uses stream-based learning methods to continuously capture the evolving system behavior under dynamic stream workloads; and (3) an iterative inspection algorithm to ensure smooth query processing in the presence of failures. The system has been deployed at the IBM System S stream processing cluster.

Self-configuring information management for overlay networks. Designed *InfoEye*, a model-based self-configuring information management system that consists of a set of monitoring sensors deployed on different overlay nodes. InfoEye can dynamically configure the operations of different sensors based on statistical application query patterns and system attribute distributions. InfoEye can answer multi-attribute range queries about dynamic information of overlay networks with much lower cost than previous solutions. The system has been deployed and tested at the PlanetLab wide-area network testbed.

Peer-to-Peer (P2P) voice-over-IP (VoIP) system. Designed and implemented *peerTalk*, a P2P VoIP system that can provide multi-party VoIP services in a fully distributed and self-organizing fashion. Compared to traditional approaches such as centralized voice mixing, peerTalk achieves better scalability and failure resilience by dynamically distributing stream processing workload among different peers. Moreover, peerTalk decouples the multi-party VoIP service delivery into two phases: mixing phase and distribution phase. The decoupled model allows us to explore the asymmetric property of multi-party VoIP services (e.g., distinct speaking/listening activities, unequal inbound/out-bound bandwidths) so that the system can better adapt to distinct stream mixing and distribution requirements. The system has been deployed and tested at the PlanetLab wide-area network testbed.

Department of Computer Science, University of Illinois, Urbana, IL

Position Held: Research Assistant

January, 2000 - July, 2004

Quality-aware distributed service composition (PhD dissertation research). Designed and implemented *SpiderNet*, a quality-aware distributed service composition framework for wireless mobile computing environments and large-scale networked systems. SpiderNet allows application services to be automatically assembled from distributed service components based on user- and application- specified requirements. SpiderNet outperforms previous approaches in terms of flexibility, scalability, resource efficiency, and quality-of-service (QoS) provisioning. To address heterogeneity and resource constraints in wireless mobile computing environments, SpiderNet provides ordered coordination to ensure QoS consistencies among interacting service components. SpiderNet then provides optimal service graph partitioning, which decides how to bind selected service components onto various devices based on their capabilities and capacities. To achieve scalability in large-scale networked systems, SpiderNet proposes a new probing-assisted decentralized approach to QoS-aware and load-sensitive service composition. During runtime, SpiderNet provides integrated proactive and reactive failure recovery to maintain the QoS of running applications. SpiderNet is implemented as a distributed middleware infrastructure that is deployed and evaluated on both UIUC smart room testbed and PlanetLab wide-area network testbed.

Visual QoS Programming Environment (MS thesis work): Designed and developed *QoS*Talk, a visual QoS programming environment to facilitate QoS-aware application programming. Proposed a novel XML-based specification language HQML and implemented a visual programming environment for automatically generating HQML specifications from visual specifications. The software is licensed by University of Illinois, which has been downloaded by many researchers from universities and companies of different countries.

IBM T.J. Watson Research Center, Hawthorne, NY

Position Held: Research Intern

May, 2003 - August, 2003

Voice-over-IP Conferencing System: Designed a distributed VoIP conferencing system using the QoS-assured service composition technique, which outperforms the conventional VoIP conferencing systems in terms of multi-constrained QoS provisioning and resource efficiency.

IBM T.J. Watson Research Center, Hawthorne, NY

Position Held: Research Intern

May, 2002 - August, 2002

Contract-based Service Composition: Designed a QoS-aware service composition framework for enterprise service overlay networks. Introduced service level agreement (SLA) into the service composition framework to govern the commitment of QoS provisioning in a heterogeneous environment involving third-party service providers.

Hewlett-Packard Research Labs, Palo Alto, CA

Position Held: Research intern

May, 2001 - August, 2001

Adaptive offloading for wireless mobile devices: Designed and implemented an adaptive offloading system that can alleviate the resource constraints (e.g., memory, CPU speed) of a mobile device by offloading part of application execution and data to a nearby surrogate (e.g., application server).

TEACHING EXPERIENCE

University of Illinois at Urbana-Champaign, Urbana, IL

Position Held: Teaching Assistant - CS 323: Operating Systems Design

August, 1999 - January, 2000

Developed and graded programming assignments, helped students during office hours, prepared and graded exams, gave substitute lectures to about 130 students, led lab sessions and review sessions.

University of Illinois at Urbana-Champaign, Urbana, IL

Position Held: Teaching Assistant - CS 314: Multimedia Systems

January, 2000 - May, 2000

Gave substitute lectures to about 50 students, designed and graded the class group project on implementing a distributed video conferencing system, prepared and graded homework and exams.

Peking University, Beijing

Position Held: Teaching Assistant - Operating Systems Implementation **February, 1999 - June, 1999**

Gave lectures to about 150 students, designed and graded the class group project on implementing a mini operating system.

PUBLICATIONS

REFEREED JOURNAL/MAGAZINE/BOOK CHAPTER PAPERS

[J6] Xiaohui Gu, Zhen Wen, Philip S. Yu, Zon-Yin Shae, "peerTalk: A Peer-to-Peer Voice-Over-IP System", *IEEE Transactions on Parallel and Distributed Systems (TPDS)*. (to appear)

[J5] Xiaohui Gu, Klara Nahrstedt, "On Composing Stream Applications in Cooperative Peer-to-Peer Environments", *IEEE Transactions on Parallel and Distributed Systems (TPDS)*, 17(8): pp. 824-837, 2006.

[J4] Xiaohui Gu, Klara Nahrstedt, "Distributed Multimedia Service Composition with Statistical QoS Assurances", *IEEE Transactions on Multimedia (TMM)*, 8(1): pp. 141-151, 2006.

[J3] Xiaohui Gu, Alan Messer, Ira Greenberg, Dejan Milojicic, Klara Nahrstedt, "Adaptive Offloading for Pervasive Computing", in *IEEE Pervasive Computing Magazine*, Vol. 3, No. 3, July, 2004. (Best papers from PerCom 2003).

[J2] Xiaohui Gu, Klara Nahrstedt, "QoS-Aware Service Composition for Large-Scale Peer-to-Peer Systems", Book Chapter in *Grid Resource Management: State of the Art and Future Trends*, Chapter 24, pp. 395-410, Kluwer Academic Press, October, 2003.

[J1] Xiaohui Gu, Klara Nahrstedt, Wanghong Yuan, Duangdao Wichadakul, Dongyan Xu, "An XML-based Quality of Service Enabling Language for the Web", *Journal of Visual Language and Computing (JVLC)*, special issue on multimedia languages for the Web, volume 13, No.1, pp. 61-95, Academic Press, February, 2002.

REFEREED CONFERENCE/WORKSHOP PAPERS

[C23] Chen Chen, Xifeng Yan, Philip S. Yu, Jiawei Han, Dongqing Zhang, Xiaohui Gu, "Toward Graph Containment Search and Indexing", In Proc. of Very Large Data Bases Conference, VLDB 2007.

[C22] Kun-Lung Wu, Philip S. Yu, Bugra Gedik, Kirsten W. Hildrum, Charu C. Aggarwal, Eric Bouillet, Wei Fan, David A. George, Xiaohui Gu, Gang Luo, Haixun Wang, "Challenges and Experience in Prototyping a Multi-Modal Stream Analytic and Monitoring Application on System S", In Proc. of Very Large Data Bases Conference, VLDB 2007.

[C21] Xiaohui Gu, Philip S. Yu, "Toward Self-Managed Media Stream Processing Service Overlays", In Proc. of IEEE International Conference on Multimedia and Expo (ICME), Beijing, China, July, 2007.

[C20] Jin Liang, Xiaohui Gu, Klara Nahrstedt, "Self-Configuring Information Management for Large-Scale Service Overlays", In Proc. of *IEEE INFOCOM*, Alaska, May, 2007. (acceptance rate:18%)

[C19] Xiaohui Gu, Zhen Wen, Philip S. Yu, "BridgeNet: An Adaptive Multi-Source Stream Dissem-

ination Service Overlay”, In Proc. of *IEEE INFOCOM Mini-Symposium*, Anchorage, AL, May, 2007. (acceptance rate: 25%)

[C18] Xiaohui Gu, Philip S. Yu, Haixun Wang, “Adaptive Load Diffusion for Multiway Windowed Stream Joins”, In Proc. of *IEEE International Conference on Data Engineering (ICDE)*, Istanbul, Turkey, April, 2007. (full paper, acceptance rate:18%)

[C17] Xiaohui Gu, Zhen Wen, Ching-Yung Lin, Philip S. Yu, “ViCo: An Adaptive Distributed Video Correlation System”, In Proc. of *ACM Multimedia (SIGMM)*, Santa Barbara, CA, October, 2006. (full paper, acceptance rate: 17%)

[C16] Thomas Repantis, Xiaohui Gu, Vana Kalogeraki, “Synergy: Sharing-Aware Component Composition for Distributed Stream Processing Systems”, In Proc. of *the 7th ACM/IFIP/USENIX International Middleware Conference (Middleware)*, Melbourne, Australia, November 2006 (acceptance rate: 17%)

[C15] Xiaohui Gu, Philip S. Yu, Klara Nahrstedt, “Optimal Component Composition for Scalable Stream Processing”, In Proc. of *25th IEEE International Conference on Distributed Computing Systems (ICDCS 2005)*, Columbus, Ohio, June 6-10, 2005. (acceptance rate: 14%)

[C14] Xiaohui Gu, Philip S. Yu, “Adaptive Load Diffusion for Stream Joins”, In Proceedings of *ACM/IFIP/USENIX International Middleware Conference (Middleware)*, Grenoble, France, 2005. (acceptance rate: 21%)

[C13] Xiaohui Gu, Zhen Wen, Philip S. Yu, Zon-Yin Shae, “Supporting Multi-Party Voice-Over-IP Services with Peer-to-Peer Stream Processing”, In Proc. of *ACM Multimedia (SIGMM)*, Singapore, November, 2005. (short paper)

[C12] Xiaohui Gu, Klara Nahrstedt, Bin Yu, “An Integrated Peer-to-Peer Service Composition Framework”, In Proc. of *IEEE International Symposium on High-Performance Distributed Computing (HPDC-13)*, Honolulu, Hawaii, June 4-6, 2004. (acceptance rate: 15%)

[C11] Xiaohui Gu, Klara Nahrstedt, Rong N. Chang, Zon-Yin Shae, “An Overlay Based QoS-Aware Voice-Over-IP Conferencing System”, in Proc. of *IEEE International Conference on Multimedia and Expo 2004 (ICME 2004)*. (acceptance rate: 30%)

[C10] Xiaohui Gu, Klara Nahrstedt, Rong N. Chang, Christopher Ward, “QoS-Assured Service Composition in Managed Service Overlay Networks”, in Proc. of *23rd IEEE International Conference on Distributed Computing Systems (ICDCS 2003)*, Providence, Rhode Island, May 19-22, 2003. (acceptance rate: 17%)

[C9] Xiaohui Gu, Klara Nahrstedt, Alan Messer, Ira Greenberg, Dejan Milojevic, “Adaptive Offloading Inference for Delivering Applications in Pervasive Computing Environment”, in Proc. of *IEEE International Conference on Pervasive Computing and Communications (PerCom 2003)*, Dallas-Fort Worth, Texas, March 23-26, 2003. (acceptance rate: 30%)

[C8] Duangdao Wichadakul, Xiaohui Gu, Klara Nahrstedt, “A Programming Framework for Ubiquitous Multimedia Applications”, in Proc. of *ACM Multimedia 2002 (SIGMM 2002)*, Juan Les Pins, France, December 1-6, 2002. (full paper, acceptance rate: 15%)

[C7] Xiaohui Gu, Klara Nahrstedt, “A Scalable QoS-Aware Service Aggregation Model for Peer-to-Peer Computing Grids” in Proc. of *11th IEEE International Symposium on High Performance Distributed Computing (HPDC 2002)*, Edinburgh, Scotland, July 24-26, 2002. (acceptance rate: 30%)

[C6] Xiaohui Gu, Klara Nahrstedt, "Dynamic QoS-Aware Multimedia Service Configuration in Ubiquitous Computing Environments", in Proc. of *22nd IEEE International Conference on Distributed Computing Systems (ICDCS 2002)*, Vienna, Austria, July 2-5, 2002. (acceptance rate: 18%)

[C5] A. Messer, I. Greenberg, P. Bernadat, D. Milojevic, D. Chen, T.J. Guili, X. Gu, "Towards a Distributed Platform for Resource-Constrained Devices" in Proc. of *22nd IEEE International Conference on Distributed Computing Systems (ICDCS 2002)*, Vienna, Austria, July 2-5, 2002. (acceptance rate: 18%)

[C4] Duangdao Wichadakul, Klara Nahrstedt, Xiaohui Gu, Dongyan Xu, "2KQ+: An Integrated Approach of QoS Compilation and Reconfigurable, Component-based Runtime Middleware for the Unified QoS Management Framework", in Proc. of *IFIP/ACM International Conference on Distributed Systems Platforms (Middleware 2001)*, Heidelberg, Germany, Nov. 2001 (acceptance rate: 17%)

[C3] Xiaohui Gu, Klara Nahrstedt, "An Event-Driven, User-Centric, QoS-aware Middleware Framework for Ubiquitous Multimedia Applications", in Proc. of *9th ACM Multimedia (Multimedia Middleware Workshop)*, Ottawa, Canada, September 30 - October 5, 2001.

[C2] Wanghong Yuan, Klara Nahrstedt, Xiaohui Gu, "Coordinating Energy-Aware Adaptation of Multimedia Applications and Hardware Resource," in Proc. of *9th ACM Multimedia (Multimedia Middleware Workshop)*, Ottawa, Canada, September 30 - October 5, 2001.

[C1] Xiaohui Gu, Duangdao Wichadakul, Klara Nahrstedt, "Visual QoS Programming Environment for Ubiquitous Multimedia Services", in Proc. of *IEEE International Conference on Multimedia and Expo 2001(ICME 2001)*, Tokyo, Japan, August 22 - August 25, 2001. (acceptance rate: 30%)

REFEREED POSTERS AND WORK-IN-PROGRESS PRESENTATION

[W3] Xiaohui Gu, Duangdao Wichadakul, Chui Ong, Klara Nahrstedt, "A Unified Framework for QoS and Security aware Software Development and Execution", *First International Space Mission Challenges for Information Technology Conference (Organized by NASA and JPL)*, Pasadena, CA, July 2003.

[W2] Xiaohui Gu, Klara Nahrstedt, "Scalable Constrained Routing in Overlay Networks", *Work In Progress Session in USENIX Operating Systems Design and Implementation (OSDI 2002)*, Boston, MA, December 2002.

[W1] Xiaohui Gu, Klara Nahrstedt, "QoS-Talk: QoS Programming Environment for Ubiquitous Multimedia Applications", *Poster Session at ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI 2001)*, Snowbird, UT, June 2001.

PATENTS

[P8] Xiaohui Gu, Spyridon Papadimitriou, Shu-Ping Chang, Philip Yu, System and Methods for Predictive Failure Management, Filed with the US patent office, 2007.

[P7] Haixun Wang, Xiaohui Gu, Lipyeow Lim, Min Wang, System and Method for Indexable XML Data Serialization, Filed with the US patent office, 2007.

[P6] Xiaohui Gu, Philip S. Yu, Shu-Ping Chang, Model-based Self-Optimizing Distributed Information Management, Filed with the US patent office, 2006.

[P5] Xiaohui Gu, Philip S. Yu, Haixun Wang, System and Method for Scalable Processing of Multi-way Data Stream Correlations, Filed with the US patent office, 2006.

[P4] Xiaohui Gu, Zhen Wen, Philip S. Yu, Zon-Yin Shae, System and Method for Peer-to-Peer Multi-Party Voice-Over-IP Services, Filed with the US patent office, 2006.

[P3] Xiaohui Gu, Philip S. Yu, Method and Apparatus for Providing Load Diffusion in Data Stream Correlations, Filed with the US patent office, 2005.

[P2] Xiaohui Gu, Philip S. Yu, System and Methods for Optimal Component Composition in a Stream Processing System, Filed with the US patent office, 2005.

[P1] Xiaohui Gu, Rong N. Chang, Zon-Yin Shae, System and Method for Providing a Communication Session, Filed with US, European, Chinese patent office, 2004.

INVITED TALKS /CONFERENCE TALKS

[T10] "System Support for Large-Scale On-line Data Analysis", Invited Talk, Boston University, March, 2007.

[T9] "Building Service Marketplace for Pervasive Computing", Invited Talk, State University of New York at Buffalo, Feb., 2007.

[T8] "System Support for Large-Scale Data Stream Processing ", Invited Talk, University of California at Riverside, October, 2006.

[T7] "System Support for Large-Scale Data Stream Processing ", Invited Talk, Michigan State University, October, 2006.

[T6] "ViCo: An Adaptive Distributed Video Correlation System", ACM Multimedia (SIGMM), October, 2006.

[T5] "Adaptive Load Diffusion for Stream Joins", ACM/IFIP/USENIX International Middleware Conference (Middleware), December, 2005.

[T4] "Optimal Component Composition for Scalable Stream Processing", *25th IEEE International Conference on Distributed Computing Systems (ICDCS 2005)*, June, 2005.

[T3] "An Integrated Peer-to-Peer Service Composition Framework ", *IEEE International Symposium on High-Performance Distributed Computing (HPDC-13)*, June, 2004.

[T2] "QoS-Assured Service Composition in Managed Service Overlay Networks", *23rd IEEE International Conference on Distributed Computing Systems (ICDCS)*, May, 2003.

[T1] "Adaptive Offloading Inference for Delivering Applications in Pervasive Computing Environment", *IEEE International Conference on Pervasive Computing and Communications (PerCom)*, March, 2003.

PROFESSIONAL SERVICE

- TPC Member, ACM/IFIP/USENIX 8th International Middleware Conference (Middleware), Doctoral Symposium, Newport Beach, CA, 2007.
- TPC Member, The 13th International Conference on Parallel and Distributed Systems (ICPDS), Hsinchu, Taiwan, 2007.
- TPC Member, Work-in-Progress Co-Chair, IEEE International Conference on Pervasive Computing and Communications (PerCom), New York, 2007.
- TPC Member, International Workshop on Ambient Intelligence, Media, and Sensing, in conjunction with ICDE, 2007

- TPC Member, ACM Multimedia Modelling Conference (MMM) 2007.
- TPC Member, The 27th IEEE Real-Time Systems Symposium (RTSS) 2006.
- TPC Member, IEEE International Conference on Pervasive Computing and Communications (Per-Com) 2006.
- TPC Member, ACM Multimedia Service Composition Workshop (in conjunction with ACM Multimedia) 2005.
- TPC Member, Publicity Chair, IEEE International Conference on Pervasive Services (ICPS), 2005.
- Reviewer for IEEE Transactions on Parallel and Distributed Systems, IEEE Transactions on Networking, IEEE Transactions on Mobile Computing, IEEE Transactions on Multimedia, ACM Transactions on Multimedia, Journal of Grid Computing, ACM SIGMETRICS, ACM Multimedia, ACM SIGMOD, VLDB, IEEE ICDE, ACM/IFIP/USENIX Middleware, IEEE ICDCS, IEEE HPDC, IEEE INFOCOM, IEEE ICC, etc.

MISCELLANEOUS

US Permanent Resident