Department of Computer Sciences
The University of Texas at Austin
1 University Station C0500
Austin, TX 78712

Phone: (512) 789 6858
Fax: (512) 232 7886
Email: nav@cs.utexas.edu
http://www.cs.utexas.edu/users/nav

Navendu Jain

Research Interests

Data management and networked computer systems: large-scale distributed systems, database systems, operating systems, and security.

Education

University of Texas, Austin, TX
Ph.D. candidate in Computer Science, Spring 2003 – Present (degree expected May 2008)
Dissertation title: "Precision-Integrated Scalable Monitoring"
Advisor: Prof. Mike Dahlin

Indian Institute of Technology (IIT), Delhi, India
M.Tech. in Computer Science and Engineering (Ist in class), May 2002
B.Tech. in Computer Science and Engineering (2nd in class), May 2001

Awards and Honors

§ IBM Ph.D. Fellowship (2007-2008)
§ Certificate of Appreciation, University of Texas Research Policy Committee (2005)
§ Microsoft Graduate Merit Scholarship (2001-2002)
§ Institute Merit Award and Scholarship, IIT Delhi (2000-2002)
§ Outstanding Project Award of the Year, IBM India Research Lab (2000)
§ 7^{th} rank in International Code Design Contest, Veritas (1999)
§ Top 0.1% IIT-JEE examination among more than 100,000 students (1997)
§ Top 0.1% All-India AISSE Merit Award, Government of India (1995)

Research Experience

Graduate Research Assistant — **University of Texas, Austin**
Laboratory of Advanced Systems Research — **June 2003 – Present**

Lead a research project on designing and building a scalable, high performance, and robust monitoring framework for large-scale networked systems [*SOSP 2005 WIP*].

- Designed and co-implemented STAR, a scalable monitoring system that delivers bounded approximate answers to continuous aggregate queries while minimizing communication costs under high-volume workloads [*VLDB 2007*].
- Designed and implemented SMART, another scalable monitoring system that maximizes query precision subject to constraints on the available bandwidth resources [*In Review*].
- Developed and co-implemented a new consistency metric to quantify the correctness of query results under node failures and network disruptions, a common case in large-scale distributed systems [*In Review*].

Poster and Demo Presentations: [*NSDI 2006, WORLDS 2006*].
Additional details on the project are available at: `http://www.cs.utexas.edu/users/nav/prism`.

Research Intern — **IBM T. J. Watson Research Center**
Hawthorne, New York — **Summer 2007, Fall 2005**

- Co-developed a new resource allocation framework to achieve high availability for partial-fault tolerant applications [*INFOCOM 2008, Filed a patent application*].
- Designed and implemented the Linear Road benchmark for distributed stream data management systems, and compared performance of SPC, Aurora, and STREAM stream processing systems [*SIGMOD 2006*].
- Co-developed an adaptive resource manager for stream data processing systems [*ICDCS 2006*].

Summer Research Intern **IBM Almaden Research Center**
San Jose, California **Summer 2004**

- Designed and implemented a replica synchronization protocol for data replication [*FAST 2005*].
- Designed and implemented a Bloom filter based similarity detection technique [*WebDB 2005*].

Extreme Blue Intern **IBM Zurich Research Lab**
Ruschlikon, Switzerland **July 2002 – Sep. 2002**
Co-developed a scalable middleware architecture for data processing in sensor networks and built a prototype implementation for telematics applications [*CCNC 2004*].
The system prototype was showcased at the IBM Exposition in London, UK 2002.

Teaching Experience

Teaching Assistant, CS313E **Elements of Software Design**
UT Austin **Fall 2004**
Tutored 30+ students with lab assignments, held office hours, and helped grade quizzes and exams.

Teaching Assistant, CS 380L **Advanced Operating Systems**
UT Austin **Fall 2003, Fall 2004**
Graded and gave feedback on 25+ students' critiques of research papers; guided students in multi-threaded programming projects.

Teaching Assistant, CS303E **Elements of Computer and Programming**
UT Austin **Spring 2003**
Taught two weekly sections (25+ each) of undergraduates – introduced new material, reviewed concepts taught by the instructor, supervised homework graders, held office hours, and helped prepare exam questions and course grading.
Review: "The instructor encourage questions very well and was readily prepare which was great."
Review: "...but our prof's lecture format is too generalized and so I get my needed information from the TA – and he delivered – good job."

Teaching Assistant, CS130, CS391 **Data Structures, Algorithms**
IIT Delhi **Fall 2001, Spring 2002**
Lead a weekly discussion class (40+ students) – introduced new material, worked in-class practice examples, graded software assignments, and guided student projects.
Received recognition letters from both course instructors.

Manuscripts Under Review

N. Jain, D. Kit, P. Mahajan, P. Yalagandula, M. Dahlin, and Y. Zhang, "Known Unknowns in Large-Scale System Monitoring," *In double-blind review.*

N. Jain, P. Yalagandula, M. Dahlin, and Y. Zhang, "SMART: Adaptive Precision Setting for Aggregation Queries over Distributed Data Streams," *In double-blind review.*

N. Jain, M. Dahlin, and Y. Zhang, "WebView: Scalable Information Monitoring for Data-Intensive Web Applications," *In double-blind review.*

Refereed Publications

N. Bansal, R. Bhagwan, **N. Jain**, Y. Park, D. Turaga, and C. Venkatramani, "Towards Optimal Operator Placement in Partial-Fault Tolerant Applications," Accepted to 27th IEEE International Conference on Computer Communications (**INFOCOM** 2008), Phoenix, AZ. Acceptance rate: 21%.

N. Jain, D. Kit, P. Mahajan, P. Yalagandula, M. Dahlin, and Y. Zhang, "STAR: Self-Tuning Aggregation for Scalable Monitoring," Proceedings of 33rd International Conference on Very Large Databases (**VLDB** 2007), Vienna, Austria. Acceptance rate: 16%.

N. Jain, L. Amini, H. Andrade, R. King, Y. Park, P. Selo, and C. Venkatramani, "Design, Implementation, and Evaluation of the Linear Road Benchmark on the Stream Processing Core," Proceedings of 25th ACM International Conference on Management of Data (**SIGMOD** 2006), Chicago, IL. Acceptance rate: 13%.

L. Amini, **N. Jain**, A. Sehgal, J. Silber, and O. Verscheure, "Adaptive Control of Extreme-Scale Stream Processing Systems," Proceedings of 26th International Conference on Distributed Computing Systems (**ICDCS** 2006), Lisboa, Portugal. Acceptance rate: 14%.

N. Jain, M. Dahlin, and R. Tewari, "TAPER: Tiered Approach for Eliminating Redundancy in Replica Synchronization," Proceedings of 4th USENIX Conference on File and Storage Technologies (**FAST** 2005), San Francisco, CA. Acceptance rate: 20%.

N. Jain, P. Yalagandula, M. Dahlin, and Y. Zhang. "INSIGHT: A Distributed Monitoring System for Tracking Continuous Queries," Work-in-Progress Session at 20th ACM Symposium on Operating Systems Principles (**SOSP** 2005 WIP), Brighton, United Kingdom.

N. Jain, M. Dahlin and R. Tewari, "Using Bloom Filters to Refine Web Search Results," Proceedings of 8th International Workshop on the Web and Databases (**WebDB** 2005), Baltimore, MD. Acceptance rate: 27%.

A. Chen, **N. Jain**, T. Pietraszek, A. Perniola, S. Rooney, and P. Scotton, "Scaling Real-Time Telematics Applications using Programmable Middleboxes," Proceedings of IEEE Consumer Communications and Networking (**CCNC** 2004), Las Vegas, NV. Acceptance rate: 30%.

N. Pabuwal, **N. Jain**, and B. N. Jain, "An Architectural Framework to deploy Scatternet-based Applications over Bluetooth," Proceedings of IEEE International Conference on Communications (**ICC** 2003), Anchorage, Alaska. Acceptance rate: 37%.

P. Niebert, M. Mahfoudh, E. Asarin, M. Bozga, **N. Jain**, and O. Maler, "Verification of Timed Automata via Satisfiability Checking," Proceedings of 7th International Symposium on Formal Techniques in Real-Time and Fault Tolerant Systems (**FTRTFT** 2002), Oldenburg, Germany. Acceptance rate: 23%.

T.V. Ashwin, **N. Jain**, and S. Ghosal, "Improving Image Retrieval Performance using Negative Relevance Feedback," Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (**ICASSP** 2001), Salt Lake City, UT.

N. Jain, S. Bansal and S. Kapoor, "Efficient Object Ellipsoid Trees," Proceedings of Indian Conference on Computer Vision, Graphics, and Image Processing (**ICVGIP** 2000), Bangalore, India.

Technical Reports

N. Jain, D. Kit, P. Mahajan, P. Yalagandula, M. Dahlin, and Y. Zhang, "PRISM: Precision-Integrated Scalable Monitoring," Technical Report TR-06-22, UT Austin Department of Computer Sciences, 2006.

N. Jain, N. Pabuwal, and B. N. Jain, "Distributed, Robust, and Self-organizing Bluetooth Scatternet Formation," Technical Report TR04-41, UT Austin Department of Computer Sciences, 2004.

N. Jain and J. Chaw, "The GPU as a Scientific Computing Engine: Performance and Analysis," Technical Report TR04-40, UT Austin Department of Computer Sciences, 2004.

T.V. Ashwin, S. Ghosal, A. Sarkar, **N. Jain**, and S. Sarkar, "Robust Non-parametric Relevance Feedback for Image Retrieval," IRL Technical Report, IBM Research, 2001.

Talks

- "STAR: Self-Tuning Aggregation for Scalable Monitoring," conference talk at VLDB 2007.
- "iMon: A Scalable Information Monitoring System," invited talk at IBM Research, NY 2007.
- "Design, Implementation, and Evaluation of the Linear Road Benchmark on the Stream Processing Core," conference talk at SIGMOD 2006.
- "TAPER: Tiered Approach for Eliminating Redundancy in Replica Synchronization," conference talk at FAST 2005.
- "Using Bloom Filters to Refine Web Search Results," workshop talk at WebDB 2005.
- "A Booster Box for Floating Car Data Application," invited talk at IBM Research, Zurich 2002.
- "A Booster Box for Floating Car Data Application," invited talk at IBM Exposition, UK 2002.

Patent Applications

"System and Method for Resource Allocation in Partial Fault-Tolerant Applications" by Navendu Jain, Yoonho Park, Deepak Turaga, and Chitra Venkatramani. IBM YOR920070525US1, 2008.

"A Method for Redundancy Elimination for Replicating File System Data" by Navendu Jain and Renu Tewari. IBM ARC-820040139, 2004.

"Two New Content-based Techniques for Similarity Detection" by Navendu Jain and Renu Tewari. IBM ARC-820040156, 2004.

Software Artifacts

PRISM/SDIMS distributed monitoring system. Used by researchers at UCSD, Univ. of Arizona, and UIUC. Public release planned. (With a UTCS alumnus and two junior graduate students.)

Linear Road benchmark: prototype design and implementation of the benchmark in SPC distributed stream processing system. Part of IBM Research System S project.

TAPER: prototype of replica synchronization protocol for data replication. Source code released.

Service

- Scribed NSDI 2008 PC meeting; wrote summaries of the PC discussion for selected papers.
- External Reviewer: NSDI 2007, FAST 2007, and HPDC 2005.
- Graduate Student Faculty Recruiting Committee, Department of Computer Sciences, 2005, 2007.
- Graduate Student Representative, University of Texas Research Policy Committee, 2005.
- Computer Science Representative, Graduate Student Assembly, 2003-2005.

References

Prof. Mike Dahlin
UT Austin
1 University Station, CO500
Austin, TX 78712
(512) 471-9549
dahlin@cs.utexas.edu

Prof. Joe Hellerstein
UC Berkeley
685 Soda Hall
Berkeley, CA 94720
(510) 643-4011
hellerstein@cs.berkeley.edu

Prof. Yin Zhang
UT Austin
1 University Station, C0500
Austin, TX 78712
(512) 232-7496
yzhang@cs.utexas.edu

Dr. Lisa Amini
IBM T. J. Watson Research Center
19 Skyline drive
Hawthorne, NY 10532
(914) 784-7366
aminil@us.ibm.com