

Stefan Saroiu

CONTACT INFORMATION

Box 352350
Dept. of Computer Science & Engineering
University of Washington
Seattle, WA 98195-2350 USA

Cell: (425) 985-9486
Fax: (206) 543-2969
E-mail: tzoompy@cs.washington.edu
<http://www.cs.washington.edu/homes/tzoompy>

RESEARCH INTERESTS

My research interests lie at the intersection of distributed systems, operating systems and networks. They encompass Internet content delivery systems, peer-to-peer systems, security, scalability and robustness of Internet services, wireless, and operating systems and networking in general.

EDUCATION

University of Washington, Seattle, WA, USA

❑ Ph.D. Candidate, Computer Science, Expected: August 2004

Dissertation Topic: "Measurement and Analysis of Internet Content Delivery Systems"
Advisors: Steven D. Gribble and Henry M. Levy

❑ M.S., Computer Science, December 2000

University of Waterloo, Waterloo, ON, Canada

❑ Double B.Math., Computer Science and Combinatorics & Optimization, August 1999

HONORS AND AWARDS

- Most cited Computer Science article published in 2002, as reported by Citeseer: "A Measurement Study of Peer-to-Peer File Sharing Systems".
- Best Paper Award: USENIX Symposium on Internet Technologies (USITS) 2003.
- Best Student Paper Award: Operating Systems Design and Implementation (OSDI) 2002.
- Best Student Paper Award: ACM SIGCOMM Internet Measurement Workshop (IMW) 2002.
- Best Paper Award: Multimedia Computing and Networking (MMCN) 2002.
- Reviewer for Infocom (2004, 2003), Usenix (2003), Sigmod (2003), VLDB (2002), IEEE Magazine (2002), SOSP (2001).

DISSERTATION

Title: Measurement and Analysis of Internet Content Delivery Systems

Advisors: Steven D. Gribble and Henry M. Levy

My thesis develops a measurement-based characterization of content delivery systems in the modern Internet. Currently, the Internet is experiencing an astronomical increase in the use of specialized content delivery systems, such as peer-to-peer file sharing systems (e.g. Kazaa, Napster, Gnutella) and content delivery networks (e.g. Akamai). These systems are rapidly changing the nature of Internet traffic; therefore, an understanding of the modern Internet requires a detailed understanding of these new mechanisms and the data they serve. At the same time, the surge in popularity of peer-to-peer file sharing has resulted in a flurry of research activity into novel peer-to-peer designs. Surprisingly little is known about the performance, behavior and workload of such systems in practice. For example, the characteristics of peers that choose to participate in a file sharing network must be understood and taken into account. A carefully designed system would avoid placing a large burden on peers with transient lifetimes or low-bandwidth Internet connections.

My dissertation bridges the gap between the architecture proposals of new content delivery systems and the design trade-offs, workloads and requirements they must accommodate in practice. In this context, my work focused on two parts of this problem. First, I performed a detailed characterization of the network behavior and properties of peers participating in modern peer-to-peer systems, such as Napster and Gnutella. This characterization includes measuring bottleneck bandwidths between peers and the Internet at large, IP-level latencies to send packets to these hosts, and correlations between these characteristics. These measurements show that there is significant heterogeneity and lack of cooperation across peers participating in these systems. Second, I examined

Internet content delivery by focusing on four popular content delivery systems: the World Wide Web, the Akamai content delivery network and the Kazaa and Gnutella peer-to-peer file sharing systems. My findings quantify the extent to which peer-to-peer traffic has overwhelmed Web traffic as the leading consumer of Internet bandwidth and the dramatic differences in the characteristics of objects being transferred as a result. All these findings have influenced the design of newer content delivery systems, including new P2P sharing systems.

**OTHER RESEARCH
EXPERIENCE** **University of Washington, Seattle, WA USA**

A New Internet Security Threat: Spyware **June, 2003 - Present**

In recent years, a relatively new computing threat has gained momentum: the spread of *spyware*. Though most people are aware of spyware, the research community has to date spent little effort understanding the nature and extent of the spyware problem. My research in this direction is an initial attempt to do so. Using the trace infrastructure I developed to monitor the traffic exchanged between the University of Washington and the rest of the Internet, I performed a quantitative study of spyware, characterizing the spread of spyware within the University.

Microsoft Research, Redmond, WA USA

Making Overlay Networks Practical **June - September, 2002**

Part of the Herald project, I was one of the members who designed and implemented SkipNet, a new overlay network protocol. SkipNet addresses key practical problems in current distributed hash table designs. SkipNet provides controlled data placement and routing locality guarantees by organizing data primarily by lexicographic key ordering. Another useful consequence of SkipNet's locality properties is that peers maintain overlay connectivity in the face of network partitions, unlike other overlay networks.

Microsoft Research, Redmond, WA USA

Bringing Predictable Real-Time Execution to Commodity Operating Systems **June - September, 2000**

I worked on the Consumer Real-Time project. This project attempts to understand the real-time requirements of applications running by home-users on commodity operating systems, like Windows. In this context, I performed a study of practical performance characteristics and requirements of a popular soft modem. My work characterized and quantified the benefits that a real-time scheduler would bring to the performance and reliability of a commodity operating system.

PUBLICATIONS All of the following publications may be downloaded from:

<http://www.cs.washington.edu/homes/tzootpy/publications.html>

Stefan Saroiu, Steven D. Gribble, and Henry M. Levy, "Measurement and Analysis of Spyware Infections in a University Environment". Proceedings of the 1st USENIX/ACM Symposium on Networked Systems Design and Implementation, NSDI, San Francisco, CA, 2004.

Krishna P. Gummadi, Richard J. Dunn, Stefan Saroiu, Steve D. Gribble, Henry M. Levy, and John Zahorjan, "Measurement, Modeling, and Analysis of a Peer-to-Peer File-Sharing Workload". Proceedings of the 19th ACM Symposium on Operating Systems Principles, Bolton Landing, NY, 2003.

Nicholas J. A. Harvey, Michael B. Jones, Stefan Saroiu, Marvin Theimer, and Alec Wolman, "SkipNet: A Scalable Overlay Network with Practical Locality Properties". Proceedings of the 4th USENIX Symposium on Internet Technologies and Systems, USITS, Seattle, WA, 2003. Received Best Paper Award. I gave the conference presentation.

Stefan Saroiu, Krishna P. Gummadi, and Steven D. Gribble, "Measuring and Analyzing the Characteristics of Napster and Gnutella Hosts". Multimedia Systems Journal, Volume 9, Number 2, pp. 170-184, 2003, Springer-Verlag.

Stefan Saroiu, Krishna P. Gummadi, Richard J. Dunn, Steven D. Gribble, and Henry M. Levy, "An Analysis of Internet Content Delivery Systems". Proceedings of the 5th Symposium on Operating Systems Design and Implementation, OSDI, Boston, MA, 2002. Received Best Student Paper Award. I gave the conference presentation.

Krishna P. Gummadi, Stefan Saroiu, and Steven D. Gribble, "King: Estimating Latency between Arbitrary Internet End Hosts". Proceedings of the 2nd ACM SIGCOMM Internet Measurement Workshop, IMW, Marseille, France, 2002. Received Best Student Paper Award.

Atar Baer, Stefan Saroiu, and Laura A. Koutsky, "Obtaining Sensitive Data Through the Web: An Example of Design and Methods". Epidemiology, Volume 13, Issue 6, pp: 640-645, 2002.

Stefan Saroiu, P. Krishna Gummadi, and Steven D. Gribble, "Exploring the Design Space of Distributed and Peer-to-Peer Systems: Comparing the Web, TRIAD, and Chord/CFS. Proceedings of the 1st International Workshop on Peer-to-Peer Systems, IPTPS, Cambridge, MA, 2002. Also appeared in Lecture Notes for Computer Science, Vol. 2429, pp. 214-224, 2002. I gave the conference presentation.

Jared Saia, Amos Fiat, Steven D. Gribble, Anna R. Karlin, and Stefan Saroiu, "Dynamically Fault-Tolerant Content Addressable Networks". Proceedings of the 1st International Workshop on Peer-to-Peer Systems, IPTPS, Cambridge, MA, 2002. Also appeared in Lecture Notes for Computer Science, Vol. 2429, pp. 270-280, 2002.

Stefan Saroiu, P. Krishna Gummadi, and Steven D. Gribble, "A Measurement Study of Peer-to-Peer File Sharing Systems". Proceedings of Multimedia Computing and Networking, MMCN, San Jose, CA, 2002. Received Best Paper Award. I gave the conference presentation.

Michael B. Jones and Stefan Saroiu, "Predictability Requirements of a Soft Modem". Proceedings of ACM SIGMETRICS 2001, Cambridge, MA, 2001. I gave the conference presentation.

Michael B. Jones, John Regehr, and Stefan Saroiu, "Two Case Studies in Predictable Application Scheduling Using Rialto/NT". Proceedings of the 7th Real-Time Technology and Applications Symposium, RTAS, Taipei, Taiwan, 2001.

INDUSTRY EXPERIENCE

Microsoft Corp., Redmond, WA USA

Software Design Engineer

January - April, 1999 and January - April, 1998

I worked on the Microsoft SQL Server product ver. 7.0, 7.1 and 8.0. I was a member of the Query Optimizer and Execution group. I worked on the design and implementation of a variety of hash-based databases operators. I also extended the algorithm for detecting functional dependencies. Finally, I was in charge of researching and designing a solution to the problem of estimating the number of distinct values of an attribute based on sampling – a classical problem of traditional databases.

Microsoft Corp., Redmond, WA USA

Software Design Engineer

May - August, 1997

I worked on the Microsoft Bookshelf project. I was a member of the group that designed and implemented a Web version of the Microsoft Bookshelf dictionary.

Nortel Wireless, Calgary, AB Canada

Webmaster

September - December, 1996

I worked as a Webmaster, database administrator and Unix sysadmin in one of Nortel's centers producing CDMA/TDMA equipment.

Bell Northern Research, Ottawa, ON Canada

Programmer

January - April, 1996

I worked on a Web interface to DDME: Nortel's largest manufacturing database containing comprehensive information about any item that Nortel has ever manufactured. Within a couple of months, this interface was used by several hundreds of clients.

**TEACHING AND
MENTORING**

University of Washington, Seattle, WA USA

Graduate Seminar on Wireless

2003

I organized a UW graduate seminar on wireless networks research over a Summer quarter. This involved the setup of a typical readings graduate seminar, including selecting a syllabus and assigning papers. The seminar was suggestively called CSE 802.11. Several wireless researchers from Microsoft Research attended the seminar and gave invited talks.

Undergraduate Mentor

2001

I advised undergraduate student, Adam Prewett, during his independent research project titled "The Effects of Packet Timestamping Accuracy and Transport Protocol on Network Latency".

Teaching Assistant

1999 - 2000

- CSE 544 - Graduate Course in Databases. Spring 2000.
- CSE 594 - Databases for Professional Masters Program. Autumn 1999.
- CSE 444 - Undergraduate Course in Databases. Autumn 2000.

Graduate Admissions

2002

I served as a graduate admissions committee member.

University of Waterloo, Waterloo, ON USA

Teaching Assistant

1997

- MATH 127 - Calculus I for Honors Physics and Chemistry. Winter 1997.
- MATH 115 - Linear Algebra for Engineering. Autumn 1997.

**PERSONAL
INFORMATION**

Date of Birth: April 6th, 1975
Place of Birth: Bucharest, Romania
Citizenship: Canadian
Languages: English, Romanian
Visa Status: Currently on F-1 as a Canadian student

REFERENCES

- **Dr. Steven D. Gribble.** Professor, Dept. of Computer Science and Engineering, University of Washington. Office: 578 Paul G. Allen Center, Box 352350, University of Washington, Seattle, WA, USA. Tel: (206) 685-1958. E-mail: gribble@cs.washington.edu
- **Dr. Henry M. Levy.** Professor, Dept. of Computer Science and Engineering, University of Washington. Office: 596 Paul G. Allen Center, Box 352350, University of Washington, Seattle, WA, USA. Tel: (206) 543-9204. E-mail: levy@cs.washington.edu
- **Dr. Michael B. Jones.** Researcher, Systems and Networking Research Group, Microsoft Research, One Microsoft Way, Bldg. 113/2178, Redmond, WA, USA. Tel: (425) 706-8846. E-mail: mbj@microsoft.com
- **Dr. John Zahorjan.** Professor, Dept. of Computer Science and Engineering, University of Washington. Office: 534 Paul G. Allen Center, Box 352350, University of Washington, Seattle, WA, USA. Tel: (206) 543-0101. E-mail: zahorjan@cs.washington.edu