Research Statement and Agenda

Stefan Saroiu
tzoompy@cs.washington.edu

My interests lie at the intersection of distributed systems, operating systems and networks. My thesis work has focused on providing much-needed insight into the behavior of peer-to-peer (P2P) systems and content delivery networks, through a combination of measurement and trace-driven analysis. My most recent work uses measurement and analysis to provide initial insight into a recent Internet security threat: spyware. I have also worked on designing and building practical and scalable overlay networks, such as my work on the SkipNet project. My early work includes a real-time characterization and modeling of the requirements of "soft devices" (such as software modems) on commodity operating systems.

As a by-product of my dissertation work, I have made a collection of P2P traces available to the community, enabling at least 38 research projects from 22 different institutions to run their own experiments. One of my publications in this area has become the most cited computer science article published in 2002, according to Citeseer [1].

In the remainder of this research statement, I will drill down into a selected subset of my research accomplishments and then describe my future research interests. Of my publications, two have won best paper awards and two have won best student paper awards.

Measuring and Analyzing Content Delivery Systems in the Modern Internet

In recent years, the Internet has experienced an astronomical increase in the use of specialized content delivery systems, such as peer-to-peer file sharing systems (e.g., Kazaa, Napster, or Gnutella) and content delivery networks (e.g. Akamai). The sudden popularity of peer-to-peer file sharing systems has resulted in a flurry of research activity into novel peer-to-peer system designs. Because these systems (1) are fully distributed, without any infrastructure that can directly be measured and monitored; (2) have novel distributed designs requiring new and highly scalable crawling techniques; and (3) use proprietary naming and encoding mechanisms and protocols, surprisingly little is known about the performance, behavior, and workload of such systems in practice. My dissertation provides a remedy to this situation, by measuring and analyzing the behavior and workloads of these systems.

My Ph.D. thesis bridges the gap between architectures proposed for new content delivery systems and the design trade-offs, workloads and requirements they must accommodate in practice. 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. I have also examined content delivery workloads in the modern Internet 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 first work [2] in this direction was an extensive study of Napster and Gnutella peers. Measuring these systems was challenging because of their large scale, the novelty of their protocols and the uncooperative nature of their constituent peers. To carry out the study, I developed a suite of highly scalable cluster-based tools, including a crawler for the Gnutella P2P network and fast bandwidth, latency, and peer-lifetime measurement tools. These tools allowed us to monitor, measure and characterize millions of peers in these networks. I learned several lessons from this measurement study. First, there is a significant amount of heterogeneity in such systems; bandwidth, latency, availability, and the degree of sharing vary between three and five orders of magnitude across the peers in the system. This implies that any similar peer-to-peer system must be deliberate and careful about delegating responsibilities across peers. Second, even though these systems were designed with symmetry of responsibilities in mind, there is clear evidence of client-like or server-like behavior in a significant fraction of the systems' populations. Third, peers tend to deliberately misreport information if there is an incentive to do so. Because effective delegation of responsibility depends on accurate information, this implies that future systems must either have built-in incentives for peers to tell the truth or systems must be able to directly measure and verify reported information.

My next step in analyzing Internet content delivery was a full characterization of the workload of four content delivery systems: HTTP Web traffic, the Akamai content delivery network, and Kazaa and Gnutella peer-to-peer file sharing traffic. For this, I ported and extended a network monitoring infrastructure to identify, capture and analyze Kazaa and Gnutella peer-to-peer traffic, Akamai CDN traffic and regular Web traffic. Using this infrastructure, I collected year-long traces of incoming and outgoing Web, Akamai and P2P traffic at the University of Washington, a large university with over 60,000 students, faculty, and staff. This work uncovered several important results [3]. First, it quantified the rapidly increasing importance of new content delivery systems, particularly peer-to-peer networks. Second, it showed how the mixture of object types delivered over the Internet has changed, with video and audio traffic dominating the more traditional Web traffic. Third, it described how a small number of peers downloading a small number of objects consume a disproportionately high

amount of bandwidth in peer-to-peer systems. This finding points to serious limitations about the scalability of currently deployed peer-to-peer systems. Finally, this analysis showed that proxy caching would have a large effect on a wide-scale peer-to-peer system, potentially reducing wide-area bandwidths dramatically.

The final step in the analysis of content delivery involved understanding the forces that drive peer-to-peer file-sharing workloads in order to anticipate the potential impacts of future change [4]. For this, I captured a trace of over 200 consecutive days of Kazaa traffic exchanged between our University and the rest of the Internet. Using this trace, my colleagues and I developed a model of multimedia workloads present in file-sharing systems. This model helped to predict and explain fundamental differences of peer-to-peer workloads when compared to Web workloads. For example, a Kazaa-like workload does not exhibit the Zipf-like popularity distribution present in Web workloads. Ultimately, these findings showed that while the Web is ultimately driven by the rate of change of documents, the performance of file-sharing systems is only limited by the birth rates of objects and clients.

A by-product of this research is the development of several tools that measure network characteristics in a fast, scalable and unintrustive manner, such as *SProbe* [5] for bandwidth measurement and *King* [6] for latency measurement. These tools also work in an *uncooperative environment*, i.e. one in which measurement software is only deployed locally on the measurement host.

A New Internet Security Threat: Spyware

A relatively new computing threat has recently gained momentum: the spread of *spyware*. Although the definition of spyware is debatable, spyware typically refers to software that gathers information surreptitiously about a user or a computer and relays that information back to a third party. Though most people are aware of spyware, the research community has spent little effort understanding the nature and extent of the spyware problem. Yet, without quantifying the scope of the spyware problem, it is hard to understand whether spyware is an important problem to the security of the Internet at large.

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. My findings [7] show that four "popular" spyware programs infect nearly 10% of active University hosts, and that hosts infected with this spyware tend to have more than one spyware program running. Additionally, a majority of organizations within the University contain at least one spyware-infected host, suggesting that existing organization-specific security policies and mechanisms (such as perimeter firewalls) are not effective at preventing spyware infections. While the representativeness of measurements gathered at only one site is difficult to characterize, my results confirm that spyware is indeed a significant problem for the Internet in general.

Other Research

Scalable overlay networks have recently emerged as flexible infrastructure for building a large peer-to-peer system. In practice, two disadvantages of these systems are: (1) they provide no control over where data is stored, and (2) they provide no guarantee that routing paths remain local, within an organizational domain, whenever possible. Over a summer internship, I was a member of the Herald project at Microsoft Research, which developed SkipNet [8], an overlay network that supports these two properties that are important in practice. SkipNet allows for both fine-grained and coarse-grained control over data placement: content can be placed either on a pre-determined node or distributed uniformly across the nodes of a hierarchical naming subtree.

During another summer internship, I was a member of the Consumer Real-Time project at Microsoft Research. 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 and developed a predictive model [9] of practical performance characteristics and requirements of a popular soft modem. My model predicted and my findings confirmed the benefits that a real-time scheduler would bring to the performance and reliability of a commodity operating system.

Future Interests

In my area of work, research can be described in terms of the following pipeline:

Gather insight into the problem: When a new, popular problem space is identified, a flood of new research invariably follows. However, the initial surge of work often lacks insight into the real tradeoffs and design issues involved with the problem. This insight serves to focus the community's research efforts on the most relevant issues. Past examples of "breakthrough insight" abound, including Patterson's work on characterizing CISC vs. RISC, and Wolman's work demonstrating the limits of cooperative web caching.

Build a model of the problem: Once enough insight has been gathered to understand the true nature of the problem, the second step in the research pipeline is to develop models of the problem that expose the underlying fundamental processes that are at work, and that facilitate rigorous and controlled experimentation.

Devise and evaluate solutions: In the third step of the pipeline, the community proposes, analyzes, and compares alternative architectures and systems that solve the problem.

All my work fits this pipeline profile. Most of my dissertation work has focused on providing insight into peer-to-peer systems research. My most recent work on measuring the extent of the spyware problem provides initial insight into a new research problem. My early work on modeling the behavior of "soft devices" fits the modeling pipeline step, as does my recent collaborative work on modeling peer-to-peer file-sharing workloads. My work on developing designs and building practical overlay networks fits the implementation and evaluation step of the pipeline approach.

In the future, I would like to continue to apply the same pipeline style to new research challenges. In this context, I present two problems I intend to pursue. First, I plan to continue attacking the spyware problem and Internet security vulnerabilities in general. Second, I plan to focus on a new research area that has received little attention so far: the emergence of the broadband infrastructure as the largest growing Internet segment and its implications to content delivery systems.

Addressing the Security Vulnerabilities of the Internet. I believe that an important problem the Internet is facing today is its vulnerability to a variety of attacks. Although computing in general has seen tumultuous change over the past decade, we are still relying on decade-old security mechanisms to protect Internet hosts. In the past, a few million early adopters occasionally dialed into the Internet, software was shrink-wrapped and purchased from stores, few applications used the network, and viruses were the dominant threat to end users. Today, hundreds of millions of technologically unsophisticated users have permanent broadband Internet connections, they acquire and upgrade their software over the Internet, their applications depend on the network to function correctly, and as a result of all this, they must contend with worms, viruses, spyware, denial-of-service (DoS) attacks, and hackers that exploit remote vulnerabilities to install Trojan horses.

Current mechanisms for providing and enforcing security on computer desktops have clearly failed. Commodity operating systems are bloated with complicated and poorly understood mechanisms. Decisions are being delegated to end-users who have a poor understanding of their consequences. This is coupled with a lack of any rigorous system administration for these millions of highly available and resourceful home machines. The Internet needs an automated system administrator service, an *Internet SysAdmin*. This system should combine general functionality, such as updating and patching software or distributing signatures for intrusion detection systems with specific functionality, such as tailoring security policies to users or monitoring individual network usage. This system will benefit from a distributed design because an Internet-wide view of the network will help to detect newly arrived, unknown threats or to make local decisions for the good of the global network.

Current Internet attacks, such as DoS attacks and worms, have received a great deal of interest from the research community. Several projects have mapped the spread and frequency of these attacks along with characterizing their impact upon victims. An aspect of these attacks that is currently poorly understood is their impact to "bystanders": users that are not directly targeted but happen to share the same infrastructure as the attacks. I believe that the frequency of DoS attacks and worms in the Internet today is so high that a large fraction of legitimate traffic is being affected by it, but masked by the network and applications congestion control mechanisms. I am interested in quantifying the extent of this problem and exploring the suitability of solutions from the prism of network bystanders, rather than the victims.

Content Delivery in Tomorrow's Internet. One of the most important trends that the Internet is currently experiencing is the emergence of broadband network infrastructure. Fast home Internet connections have brought millions of new users to the Internet and it certainly is one of the catalysts of the surge in popularity of peer-to-peer systems. Broadband is a peculiar network environment, where communication bandwidths and latencies are assymetric, and traffic is being multiplexed across thousands of users. Surprisingly little is known about the conditions in which broadband networks operate, yet there is a flurry of new system designs that are addressed to them, such as peer-to-peer file-sharing systems. Therefore, I believe it is important to understand how network characteristics differ in broadband networks from better-understood, well-connected Internet network environment, such as a University medium.

The emergence of good home Internet connectivities and efficient and cheap ways of creating audio and video content has led to the recent surge in popularity of peer-to-peer systems. I believe that peer-to-peer is the first application in a continuous flux of new content delivery applications that the Internet will experience in the near-term, such as computer telephony, desktop video conferencing, distance learning, interactive video games, telecommuting, telemedicine, webcasting, and video on demand. As with peer-to-peer, these applications will rapidly change the nature of Internet traffic. I am interested in acquiring an understanding of the characteristics of this new traffic and in developing the right, practical content delivery mechanisms for these future applications.

References

[1] Citeseer. Most cited articles in computer science published in 2002, May 2003. `http://citeseer.nj.nec.com/articles2002.html`.

[2] Stefan Saroiu, P. Krishna Gummadi, and Steven D. Gribble. A measurement study of peer-to-peer file sharing systems. In *Proceedings of Multimedia Computing and Networking 2002*, San Jose, CA, January 2002. Received *Best Paper Award*.

[3] Stefan Saroiu, Krishna P. Gummadi, Richard J. Dunn, Steven D. Gribble, and Henry M. Levy. An analysis of Internet content delivery systems. In *Proceedings of the Fifth Symposium on Operating Systems Design and Implementation (OSDI 2002)*, Boston, MA, December 2002. Received *Best Student Paper Award*.

[4] Krishna P. Gummadi, Richard J. Dunn, Stefan Saroiu, Steven D. Gribble, Henry M. Levy, and John Zahorjan. Measurement, modeling, and analysis of a peer-to-peer file-sharing workload. In *Proceedings of the 19th Symposium on Operating Systems Principles (SOSP 2003)*, Bolton Landing, NY, October 2003.

[5] Stefan Saroiu. Sprobe: A fast tool for measuring bottleneck bandwidth in uncooperative environments, January 2001. `http://sprobe.cs.washington.edu`.

[6] Krishna P. Gummadi, Stefan Saroiu, and Steven D. Gribble. King: Estimating latency between arbitrary internet end hosts. In *Proceedings of the Second SIGCOMM Internet Measurement Workshop (IMW 2002)*, Marseille, France, November 2002. Received *Best Student Paper Award*.

[7] Stefan Saroiu, Steven D. Gribble, and Henry M. Levy. Measurement and analysis of spyware infections in a university environment. In *Proceedings of the 1st USENIX/ACM Symposium on Networked Systems Design and Implementation (NSDI 2004)*, 2004.

[8] Nicholas J. A. Harvey, Michael B. Jones, Stefan Saroiu, Marvin Theimer, and Alec Wolman. Skipnet: A scalable overlay network with practical locality properties. In *Proceedings of the 4th USENIX Symposium on Internet Technologies and Systems (USITS 2003)*, Seattle, WA, March 2003. Received *Best Paper Award*.

[9] Michael B. Jones and Stefan Saroiu. Predictability requirements of a soft modem. In *Proceedings of the ACM SIGMETRICS Conference on Measurement and Modeling of Computer Systems*, Cambridge, MA, June 2001.