Research Statement

Nick Feamster

Summary

My research focuses on developing tools, algorithms, and protocols that make the network easier to manage, more secure, and more available.

Nobody notices when the network works well, but everyone suffers when it doesn't. Thus, communications networks should be both secure and available. Network *security* has many facets, ranging from the ability to stop "unwanted traffic" (e.g., spam and denial-of-service attacks) to the ability to trace back attacks to their perpetrators ("accountability"). *Availability* means that the network must provide good performance for users whenever they want to use it—unfortunately, the increasing complexity of the network, coupled with hardware faults, software bugs, misconfigurations, and malice, make it difficult to achieve this goal. Unfortunately, these two important goals have also been among the most evasive. Breakthroughs require not only extensive domain knowledge, but also the ability to techniques from a wide range of areas, ranging from economics to machine learning. My work combines domain knowledge, extensive interactions with network operators, techniques from a wide range of disciplines, and—perhaps most importantly—the competence and tenacity to implement and deploy these systems in practice. This unique combination has allowed me to build one of the few networking research groups in the world that interacts directly with network operators to deploy fundamentally new systems and technologies in real-world networks.

I discover interesting and challenging practical problems through frequent discussions and meetings with network operators and people in industry. I then tackle these problems from first principles, develop new methods, and transfer these solutions back to practice in the form of working systems. I have tackled a variety of problems in network operations, ranging from real-time network diagnosis to stemming unwanted traffic like spam to architectures for fast failure recovery. Many people—most notably, operators "in the trenches"—are also working on these problems. Unfortunately, many of the people who have the domain knowledge that best equip them to solve these problems are busy with day-to-day operations, putting out fires as they arise but rarely taking time to think about fundamental changes to the network that might eradicate these problems. My research fills this niche. I first devise methods to understand the nature of the problem in practice. I tackle domain-specific problems with tools and techniques from other disciplines—ranging from machine learning to economics to program analysis—whose principles might provide insights into a new, previously undiscovered solution. I then devise a new approach or solution, and I transfer it to practice through implementation and deployment of real-world systems.

My work in this broad area follows four themes: (1) making the network more secure; (2) improving network availability and performance by making the network easier to operate and manage; (3) designing platforms for virtual networks that facilitate technical innovation in both network security and operations; (4) making performance and reachability problems more transparent to users. The first two themes involve developing solutions that make the network more robust and resilient in the face of faults, misconfiguration, and malice; the third theme provides an avenue to evaluate and deploy these solutions in practice. The fourth theme concerns my ongoing and planned work concerning emerging problems and threats in the battle for information and influence on the Internet.

Theme 1: Network Security

My research explores the role that communications networks—in particular the network layer—can play in improving computer and communications security. This line of research began with my arrival at Georgia Tech in 2006. A cornerstone of this research is a system that was published in August 2009 called "SNARE"

(Spatio-temporal Network-level Automated Reputation Engine). This work appeared at the *USENIX Security Symposium*, a top-tier security conference. The main idea behind SNARE—and the key insight behind my research in spam filtering—is that spammers have different sending behavior than legitimate senders. Filters can distinguish spammers from legitimate senders by examining their *sending behavior* (i.e., how they send traffic), rather than what is in the messages themselves.

Prior to my research, conventional spam filters attempted to distinguish spam from legitimate email by looking at message contents: that is, they would look at the words or language used in the messages themselves and try to detect spam based on what the message said. This approach has become increasingly untenable, since spammers have begun to embed their messages in all sorts of media, ranging from images to PDFs to audio files to spreadsheets—by the time developers perfected their content filters for one type of medium, spammers moved onto the next. My line of work has taken an entirely different, but complementary approach: I look at features of the senders' *behavior* (e.g., the time of day they are sending, whether there are other "nearby" senders on the network, whether and how the sizes of the messages of the senders vary over time) to distinguish spamming behavior from legitimate email use. The method is harder for spammers to evade, it is more flexible because it can be deployed anywhere in the network, and it can work at much higher traffic rates than conventional approaches. This idea was first laid out in the initial award paper at *SIGCOMM* and finally realized in the SNARE paper from August 2009 at *USENIX Security*.

I have also worked on sweeping changes to the Internet architecture that could improve *accountability*, thus making it more difficult for malicious parties to operate unfettered in the first place. The current Internet architecture provides little to no accountability. Malicious end systems can conceal the source of their traffic ("spoofing"), and edge networks can provide false information about their reachability to various Internet destinations ("route hijacking"); both of these attacks make it difficult to track down perpetrators of attacks. Current approaches to solving these problems require manual configuration and operator vigilance, which make them weak and error-prone. Towards building networks that are inherently accountable, I have developed the Accountable Internet Protocol (AIP). One of my contributions to the design was to make the addresses in this protocol self-certifying, which forms the cornerstone of the basic design. I also demonstrated how to apply AIP to secure BGP, the Internet's interdomain routing protocol.

Impact. My research in network security has had impact in research, in industry, and on the national level. My research on this topic has earned the Presidential Early Career Award for Scientists and Engineers (PECASE), a Sloan fellowship, and the Best Paper Award at *ACM SIGCOMM* (the premier computer networking conference). Aspects of my work have also been incorporated into commercial spam filtering products and Web mail clients at companies including Yahoo, Cisco/Ironport, and McAfee, as well as a project for the Department of Defense on high-speed network monitoring. My paper on understanding the network-level behavior of spammers—which won the Best Student Paper award at *SIGCOMM* in 2006—has been cited over 300 times since its initial publication in August 2006—it spawned a variety of high-impact follow-on work, including looking at network-level behavior not only to develop better spam filters, but also to detect botnets more effectively and defend against phishing attacks, click fraud, and other serious threats to the Internet infrastructure. I have also been working on similar approaches to help detect and dismantle the Internet's scam hosting infrastructure (e.g., Web sites that attempt to steal user passwords, money, and so forth). My initial paper on this topic ("Dynamics of Online Scam-Hosting Infrastructure") won the Best Paper award at the *Passive and Active Measurement* conference in April 2009.

My work on SNARE has also garnered significant attention in industry. This work was featured in *Technology Review* and on Slashdot (a popular, high-traffic site for news in information technology). Several companies including Yahoo have incorporated the network-level features that SNARE identifies into its spam filters, and companies that develop spam filtering appliances, such as McAfee, are also using these features to improve the accuracy and performance of their spam filtering appliances.

AIP appeared in *ACM SIGCOMM* in 2008; an early version of the design also appeared in *ACM Workshop on Hot Topics in Networking (HotNets)*. I am incorporating a version of this technology into a working system and transferring them to practice. I am working with BBN on a DARPA project that will ultimately result in

incorporating AIP's mechanisms into a military network protocol that allows attribution of traffic to sources (the details may ultimately be classified).

My impact on the broader field of cybersecurity goes beyond my own research. I am also having impact in the national arena in several ways. Last year, I was involved in setting the nation's agenda for cyber security, through multiple additional activities. First, I led a community-wide effort to develop a "wish list" document that describes the security community's needs for access to better data—ranging from network traffic data, to data about our country's infrastructure. This report was ultimately delivered to Tom Kalil, the deputy director for policy in the Office of Science and Technology Policy. Second, with program managers Karl Levitt and Lenore Zuck at NSF, I organized a community-wide, multi-agency workshop on "Security-Driven Architectures". The workshop included participants from computer science, with an eye towards setting a research agenda for developing more holistic approaches to computer security that consider *all* aspects of computer and communications systems, rather than just a single piece (like the network). Finally, my work on developing next-generation Internet protocols to improve accountability (which could eradicate spam in the first place), based on work that appeared at *ACM SIGCOMM* in 2008, was included in reports for the National Cyber Leap Year.

Representative Publication

> S. Hao, N. Syed, N. Feamster, A. Gray and S. Krasser. "Detecting Spammers with SNARE: Spatio-temporal Network-level Automatic Reputation Engine". *USENIX Security Symposium*, August 2009. Montreal, Quebec, Canada.

Most Cited Publication (334 Citations)

> A. Ramachandran and N. Feamster. "Understanding the Network-Level Behavior of Spammers". *ACM SIGCOMM*, August 2006. Pisa, Italy. **Best Student Paper Award.**

Theme 2: Network Operations

A second major theme of my work is *network operations*, which is what I call the field of designing networks so that they are easier to run and manage. Much of my work in this area has focused on fault detection and troubleshooting. Prior to my dissertation work, operators relied on detecting problems with networks "at runtime" on a live network. My dissertation work demonstrated that, in fact, many routing problems could be detected simply by examining the configuration of the routing protocols, before the configuration is even deployed. I applied techniques from static program analysis to routing configuration to help network operators catch mistakes and predict dynamic network behavior before the configurations are deployed on a live network, preventing costly and catastrophic network downtime.

Beyond predicting behavior and proactively detecting configuration faults, operators must understand the network's behavior *as it is running* (e.g., to detect equipment failures, attacks, or unplanned shifts in network traffic). Unfortunately, operators are drowning in heterogeneous data. To help operators better understand network faults "at runtime", I have applied unsupervised learning techniques to Internet routing data to help them efficiently mine the data for network events that require corrective action. This work appeared in *ACM SIGMETRICS* in 2007. My work has also applied statistical inference techniques to help network operators determine the answers to "what if" configuration questions in content distribution networks; we developed a system called "WISE" (What-If Scenario Evaluator) to help network operators determine the effects of configuration changes on network response time. A paper on this system appeared at *ACM SIGCOMM* in 2008 and is now used by operators and network designers at Google. A more mature version of this work that also describes deployment experiences at Google is in submission to *IEEE/ACM Transactions on Networking*.

Users of communications networks also face the potential of intentional performance degradation or manipulation by Internet Service Providers (ISPs); these problems are popularly referred to as "network neutrality violations". This transparency can help users determine whether their network is the cause of performance degradation, or whether performance problems that they are seeing are due to some other cause. With students, I designed, built, and deployed the *Network Neutrality Access Observatory (NANO)*, a system that aggregates measurements from end systems to help users and operators of edge networks infer when transit networks may be discriminating against certain types of traffic. This work appeared in *ACM SIGCOMM CoNext* in 2009, and we have deployed the system on Google's Measurement Lab (`http://www.measurementlab.net/`). More recently, we have been looking at methods for helping users diagnose general problems with access network performance and examining which factors have the most influence on access network performance.

I have developed new network protocols and architectures that improve availability and accountability in communications networks in the face of both faults and malice. Networks face the continual threat of failures and attacks that disrupt end-to-end connectivity. Prior to my work, one promising approach to improving connectivity involved routing traffic along multiple paths between two endpoints ("multi-path routing"); despite the promise of this approach, previous approaches encountered two significant challenges: First, previous approaches for disseminating information about multiple paths through the network did not scale to large networks. Second, end systems had no way to signal to the network that an end-to-end path had failed or was providing inadequate performance. My research applied a new perspective to this problem: rather than simply routing traffic on one of multiple paths to a destination, allow traffic to switch paths at intermediate points en route to the destination, and allow end systems to signal to the network when it should attempt to use a different path to the destination using a small number of bits that can be carried in the traffic itself. This system, called *path splicing*, provides up to an exponential improvement in reliability for only a linear increase in the amount of state that each router in the network must store.

Impact. The foundation of this research theme comes from a system I built called called "rcc" (router configuration checker). This system was the centerpiece of my doctoral dissertation and has had significant impact in both research and industry. The work received the Best Paper Award at *ACM/USENIX Networked Systems Design and Implementation (NSDI)* in 2005 and has been used by hundreds of Internet Service Providers (ISPs) around the world to check their network configurations for errors.

The NANO project is among the most visible of my research projects at Georgia Tech: The project page receives about 2,000 unique visitors every month, and, between January and March 2010, about 300 users have downloaded the code. The system is deployed on Google's Measurement Lab.

The path splicing work resulted in a Sigma Xi undergraduate research award for Megan Elmore. The work was funded by Cisco, and they have considered the possibility of extending their existing multiple routing configuration (MRC) function to support path splicing. A more likely deployment scenario, however, may be the incorporation of path splicing into a network where network elements are more programmable (I discuss the promise of programmable networking in "Future Challenges" below.) We have published an open-source implementation of path splicing on several programmable networking platforms.

Representative Publication

M. Tariq, A. Zeitoun, V. Valancius, N. Feamster, and M. Ammar. "Answering What-if Deployment and Configuration Questions with 'WISE'". *ACM SIGCOMM*, August 2008. Seattle, WA.

Most Cited Publication (169 Citations)

> N. Feamster and H. Balakrishnan "Detecting BGP Configuration Faults with Static Analysis" *2nd Symposium on Networked Systems Design and Implementation (NSDI)*, Boston, MA, May 2005. **Best Paper Award.**

Theme 3: Virtual Networking

Network virtualization allows multiple networks to operate in parallel on the same physical infrastructure. Although this concept is not new (commonly used Virtual Private Networks, or "VPNs", come to mind as a prominent real-world example of network virtualization), virtualizing all aspects of the network infrastructure—in particular, both the links *and* the routers themselves—holds great promise for enabling innovation. In 2002, Larry Peterson, Scott Shenker, and Jon Turner argued that networking research had "ossified", because researchers faced a huge deployment hurdle for deploying their research in production environments, and also because the large stakeholders had little incentive to allow disruptive innovation to take place. Their argument was essentially that, by "letting a thousand flowers bloom", multiple networking technologies could be deployed in parallel, thereby providing researchers a path to innovation. The main research challenge was how to design and implement a virtual network infrastructure that supported this philosophy.

Towards solving this challenge, I began working on network virtualization during my postdoc at Princeton. Jennifer Rexford and I wanted to implement a new network protocol we had designed at the end of my graduate career. Our plan was to use PlanetLab—a large testbed with virtualized servers distributed around the world—to do it. Unfortunately, we quickly realized that PlanetLab did not have the necessary functions to instantiate test *networks*; in particular, PlanetLab offered no functions for building virtual routers and links, and also had no support for forwarding traffic at high rates for virtual routers (e.g., every packet needed to be copied several times at each node, significantly slowing the packet forwarding rates). These shortcomings caused us to pursue a larger project to build such a testbed that would support the kinds of experiments that we wanted to run. With Andy Bavier and Larry Peterson, we built a Virtual Network Infrastructure (VINI), a testbed that allows researchers to build virtual networks. This work appeared in *ACM SIGCOMM* in 2006. Although we still strive for more widespread adoption, the testbed is regularly used by several research groups around the country.

Since this initial work, I have focused on two aspects of network virtualization: (1) providing Internet connectivity and routing control to virtual networks; (2) designing very fast packet forwarding technologies for virtual networks. A virtual network—either an experiment or a distributed "cloud" service—typically needs connectivity to the rest of the Internet so that users can actually exchange traffic with it. To provide such connectivity, and to give each virtual network direct control over how user traffic reaches it, I designed, implemented and deployed the Transit Portal. This work will appear in *USENIX Annual Technical Conference* in June 2010; it is also a cornerstone of the larger nationwide GENI effort (featured here, for example: `http://www.geni.net/?p=1682`). Our work on designing faster packet forwarding technologies for virtual networks started with the Trellis project, which moved packet forwarding for virtual networks into the kernel; although this work resulted only in a workshop publication, the software itself was adopted by University of Utah's Emulab, the most prominent emulation-based testbed for networking research. Our current efforts have focused on accelerating packet forwarding further by supporting custom packet forwarding for virtual networks in Field Programmable Gate Arrays (FPGAs); our work on SwitchBlade, a platform for rapidly developing and deploying custom forwarding engines in hardware for virtual networks, will appear at *ACM SIGCOMM* in August 2010.

Impact. The impact of this work thus far has been to support network experimentation for researchers; many other virtual network technologies and platforms have built on this work. Our work on virtual

networks has been over nearly 500 times (the VINI paper has been cited more than 300 times, and our work describing a network architecture based around network virtualization has been cited over 200 times).

The Transit Portal is currently deployed in five locations, and I am using it in my courses to provide students with hands-on experience configuring networks of routers and connecting them to real BGP-speaking routers on the Internet. The course I have developed that uses this technology is likely serves as the first course where students can directly configure networks of routers that are connected to the global Internet.

Representative Publication

> B. Anwer, M. Tariq, M. Motiwala, N. Feamster. "SwitchBlade: A Platform for Rapid Deployment of Network Protocols on Programmable Hardware" *ACM SIGCOMM*, New Delhi, India, August 2010.

Most Cited Publication (311 citations)

> A. Bavier, N. Feamster, M. Huang, L. Peterson, J. Rexford. "In VINI Veritas: Realisitic and Controlled Network Experimentation". *ACM SIGCOMM*, August 2006. Pisa, Italy.

Theme 4: Internet Transparency and Open Access

Free and open access to information and communications on the Internet is at risk: the Open Net Initiative reports that nearly 60 countries censor some access to information on the Internet. Similarly, ISPs can degrade network performance for certain subsets of users for some or all services. For example, some ISPs have been found to routinely block or throttle certain application traffic (e.g., BitTorrent); additionally, studies of access network performance in the United Kingdom and France have revealed that the level of performance that users achieve in their homes is sometimes as little as half of the rates that ISPs advertise to their users. Although it may not be feasible to always guarantee open, unfettered access to information, users should know when their access to information has been obstructed, restricted, or tampered with.

Towards providing better *transparency* to users concerning their Internet service, I am developing objective, independent third-party services for users that help them both determine whether their Internet service provider or government is restricting access to certain content or services or degrading service for particular applications and gain access to information that they might not otherwise have access to. My research on Internet transparency is focusing on three areas: (1) the *performance* that they receive from their ISP; (2) *connectivity* to various Internet destinations; (3) the *information* that they can discover via search engines and social media.

To provide users better information about the performance that they are receiving, I started Project BISmark (`http://projectbismark.net`) in 2010; BISmark is a software platform for home routers. We have already used BISmark to develop a network measurement suite for access Internet service providers; our first paper on BISmark appeared in *ACM SIGCOMM* in 2011. With collaborators in programming languages and human-computer interaction, I am now exploring ways to use BISmark to simplify the management of home networks by applying some of the same network management principles that we have learned in our studies of transit and enterprise networks.

Second, I am actively developing techniques that help users gain access to information that they might not otherwise see, as a result of overt censorship. Ten years ago, I developed Infranet, a tool to circumvent Internet censorship that was both robust to blocking attempts and deniable—meaning that an adversary could not easily detect that a user was engaged in activities to circumvent censorship; the work won the Best Student Paper Award at the *USENIX Security Symposium* in 2002. Recent developments, such as the rise of user-generated content, have made it easier to deploy censorship circumvention systems, since sites

that host user-generated content can be used as covert "drop sites" for messages; based on this insight, we designed and implemented Collage, a tool that allows users to circumvent censorship firewalls by building covert channels into user-generated content. Collage was presented at the *USENIX Security Symposium* in 2010; it has been downloaded hundreds of times and appeared in various news outlets including *Ars Technica*, *GigaOm*, and *Slashdot*.

Finally, I believe that one of the growing threats to free and open access to information in the coming years will be the emergence of "soft" forms of censorship, such as intentional performance degradation, the spread of propaganda through social media, and selective filtering or placement of search results. To defend against these threats, I have begun developing techniques to identify propagandistic behavior in social media and to allow users to compare their search results with a baseline set of search results assembled through crowdsourced measurements.

Impact. The full impact of our research in this area is not yet clear, since the projects described above are relatively recent, but the initial impact of the work is promising. Our results from the initial BISmark study influenced the design and implementation of the performance measurements used by the Federal Communications Commission's study of broadband connectivity across the United States. The project has been featured in *Ars Technica* and *GigaOm* and has received over 20,000 signups from interested users. We have currently deployed BISmark routers in about 50 home networks around the world and plan to have several hundred routers deployed by the end of 2011 as a measurement and application platform for other researchers. To transition some of the technologies we are developing in research to practice, I am also participating in Georgia Tech's venture program, Flashpoint, to scale our efforts to a larger number of users and learn more about the problems faced by ISPs, content providers, and consumers.

Beyond the impact of the technology itself in industry, I have been developing the BISmark platform as an educational tool. In Summer 2011, I hosted a BISmark "summer camp" at Georgia Tech to help students become familiar with programming network applications on the OpenWrt router platform; the week-long event was attended by about twenty students and faculty members from across the United States, France, and Italy. I have incorporated much of the material into the graduate networking course at Georgia Tech, to give students hands on experience with developing and deploying a variety of network measurement tools. Through these activities, I aim to provide students both concrete exposure to problems and concepts in networking and a platform on which they can innovate.

Representative Publication

> S. Sundaresan, W. de Donato, N. Feamster, R. Teixeira, S. Crawford, A. Pescape "Broadband Internet Performance: A View From the Gateway" *ACM SIGCOMM*, Toronto, Ontario, Canada. August 2011.

Most Cited Publication (76 citations)

> N. Feamster, M. Balazinska, G. Harfst, H. Balakrishnan, D. Karger. "Infranet: Circumventing Web Censorship and Surveillance" *Proceedings of the 11th USENIX Security Symposium*, San Francisco, CA, August 2002. **Best Student Paper Award.**