Nick Feamster

Research Statement

Summary

My research focuses on developing techniques, algorithms, and protocols that make the network easier to manage, more secure, and more available. I am an experimental networked systems and security researcher. A hallmark of my work is building, designing, and monitoring real networked systems.

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 difficult to achieve. Breakthroughs require not only extensive domain knowledge, but also the ability to apply 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 stopping 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 research in this broad area is currently focusing on several themes: (1) Internet censorship and open access; (2) home and access networks; and (3) software defined networking. These themes, which I have been developing in the past several years since receiving tenure, build on the broader research themes I have developed on network security and operations. I first survey the new leadership roles that I have assumed in research, teaching, and service. Then, I discuss each of the new research themes I have developed since tenure and the impact that they have had on both other researchers and on industry.

Ongoing and Future Research

In the past three years, I have focused on three research themes: Internet censorship and information control, home and access networks, and software defined networking.

- *Access Networks.* We began our research studying the performance of home networks and mobile networks in June 2010, when we began studying the performance of DSL networks in France. Upon realizing that accurate measurements would require deploying infrastructure in the home router itself, we began developing BISmark (Broadband Internet Service Benchmark), custom router firmware which has now been deployed in more than 300 home networks in nearly 30 countries around the world. We also developed a version of this software that

runs on Android phones and has been installed by more than 4,000 users in 130 countries. The testbed that we have developed is the first of its kind to study access and cellular networks.

- *Censorship and Information Manipulation.* Although my first work on censorship circumvention dates back to my work on the Infranet system in 2002, we revived this work in a system called Collage Recently, I founded the *USENIX Workshop on Free and Open Communications on the Internet*. This work has broadened to include *profile pollution attacks*, whereby an attacker can affect the content that a user sees (*e.g.*, search results, recommended products) by polluting the users profile with cross-site scripting attacks.
- *Shared Programmable Networking.* Our work on shared programmable networking over the past three years includes research in support of network virtualization, fast packet forwarding in programmable hardware, and control systems for Software Defined Networks (SDN) that make it easier for network operators to write programs that control network behavior under changing network conditions.

In the following subsections, I describe my contributions in these areas in more detail and discuss both the intellectual depth and impact of the contributions of the work in each of these areas.

Access Networks

Access networks (*i.e.*, cellular networks and home broadband networks) are proliferating: Over 90% of US households now have Internet access, and networks have become an essential part of every home. Video streaming already accounts for over 60% of the peak download bandwidth for the Internet; remote learning is flourishing, with Khan Academy alone delivering over 86 million videos; and within five years, Forrester Research expects 63 million Americans to telecommute from home. Bandwidth to the home is also growing rapidly: Huge investments by industry and government mean that over 60% of US homes have broadband access. Inside the home, 55% of traffic is delivered to game consoles, set-top boxes, smart TVs, and mobile devices. Further, cellular networks have become the predominant mode of Internet access for many people: For example, in Brazil, Russia, India, China, and Indonesia, there are 610 million Internet users, but 1.8 billion mobile-phone connections.

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.

Intellectual depth. Despite the great extent to which billions of users around the world depend on broadband access networks every day, very little is known about how well they perform, and even less is known about the causes of poor performance when it does arise. Yet, understanding the performance of broadband access networks has important implications for people who make decisions about how to invest in technology and about how policies should be set to improve network reliability and performance. To design networks to be more reliable and secure, and to provision and connect them so that they perform better first requires a deep understanding of where the reliability problems and performance bottlenecks exist. Our work in this area has thus developed a variety of new measurement methods and statistical techniques to both reliably measure access network performance and locate the source of performance bottlenecks. For example, our work in diagnosing home network performance problems required the design of a maximum likelihood detector to identify the location of performance faults, using only observations of indirect features (*e.g.*, queueing delay). Such detectors have been used in other fields of science where the phenomenon trying to be measured can only be observed indirectly (*e.g.*, radar, astronomy).

The work we have done is important, but at the same time it is also difficult to carry out because it requires deploying measurement observation points in the home networks of "real users" (in contrast to previous measurement infrastructure, which has primarily been deployed on academic networks, which face are subject to much different constraints than commercial networks). We have spent several years developing both reliable metrics and software platforms that are reliable enough to collect accurate measurements without disrupting the connectivity of the users who are hosting our measurement devices. As a result of our efforts, more than ten other research groups are now using the BISmark platform for their own research projects.

Impact. 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 250 home networks around the world; it is also currently deployed on Google's Measurement Lab. To transition some of the technologies we are developing in research to practice, I participated 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. We also received an NSF Innovation Corps grant and Georgia Research Alliance grant to help us commercialize this technology.

More recently, we have been expanding our work on BISmark across developing countries and across a broader range of devices. For example, we recently completed a study with Research ICT Africa (RIA) to characterize fixed and mobile performance across South Africa; we are in the process of expanding this study to other countries in Africa. Second, we have developed a home network performance troubleshooting tool that helps users identify whether performance bottlenecks are within their home network or in the Internet service provider (ISP) network. The Federal Communications Commission (FCC) has recently agreed to back the deployment of our software in 4,000 home networks across the United States, and Comcast has also recently agreed to a trial deployment of this software.

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.

Internet Censorship and Information Manipulation

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.

I have developed several techniques that help users gain access to information that they might not otherwise see as a result of overt censorship. Ten years ago, we 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.

A growing threat to free and open access to information in the coming years is 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, we 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. 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. We are now applying the same statistical techniques that we developed for NANO to detect intentional performance degradation that censors may be perpetrating to discourage users from visiting certain cites.

Intellectual depth. Censorship circumvention is an intellectually difficult problem because of the need to provide not only confidentiality but also *covertness*—in other words, the communication between parties must not only be private, but it must also be inconspicuous (with some degree of measurable confidence that the communication is unobservable). The work thus first involved applying the concept of "deniability", whereby messages can be encoded in traffic streams that look innocuous (or, otherwise, like a user's normal behavior) and subsequently involved developing encoding schemes that produce traffic that is statistically indistinguishable from that of normal user activity.

Measurement of censorship and, more generally, information manipulation also requires a set of statistical tools to attribute *causality* of performance degradation to the ISP (or to some other cause). In many of the research problems we have worked on (*e.g.*, network neutrality violation, censorship detection), we have built on existing theories of causality to improve the confidence concerning the cause of performance degradation or other type of manipulation. The study of information manipulation represents a significant conceptual advance in security, which has conventionally focused on more traditional attacks against infrastructure (*e.g.*, hosts, networks). In contrast, our work focuses on *semantic* attacks, which not only involve more subtle attacks (*e.g.*, on the information that is passed through search engines and other information portals), but also entail attacks on information and ideas, as opposed to simply assets.

Impact. 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*. To measure the effects of personalization, we have developed and released tools such as Bobble (`http://bobble.gtisc.gatech.edu/`) and Appu (`http://appu.gtnoise.net/`), both of which now have large groups of users, to help users track online information manipulation and privacy. Our work on search poisoning, whereby an attacker can affect the search results that a user sees by polluting a user's search history through cross-site request forgery (XSRF) attacks. This work appeared in the 2013 *USENIX Security Symposium*. Our NANO tool appeared in *ACM SIGCOMM CoNext* in 2009, and we have deployed the system on Google's Measurement Lab (`http://www.measurementlab.net/`).

Shared Programmable Networks

I began working on network virtualization during my postdoc at Princeton. 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. With Andy Bavier, Jennifer Rexford, 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. The concepts behind virtual programmable networks, in concert with some of our work on the Routing Control Platform (RCP) ultimately led to the advent of Software Defined Networking (SDN). Since this initial work, I have focused on three aspects of software defined networking: (1) providing Internet connectivity and routing control to software defined networks; (2) designing very fast packet forwarding technologies for software defined networks; (3) designing better languages and control models for software defined 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, a software-defined controller for interdomain routing that provides individual virtual networks the illusion of having a direct, physical upstream connection to multiple Internet service providers. This work appeared in *USENIX Annual Technical Conference* in June 2010. We performed several research projects to follow up on this work, which used the Transit Portal to improve both the reliability and performance of cloud-hosted Internet services. This follow-up work has appeared in *ACM SIGCOMM* in 2012, and *ACM SIGMETRICS* in 2013. The Transit Portal 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, appeared at *ACM SIGCOMM* in August 2010.

Finally, I have been developing new control models and languages to support event-based control for software defined networks. I have focused on how better control models and languages can help solve three problems in network management: (1) enabling frequent changes to network conditions and state; (2) providing support for network configuration in a high-level language (including developing one of the first formal languages for software defined networks, Procera); and (3) providing better visibility and control over tasks for performing network diagnosis and troubleshooting. With my students, I built and deployed software defined networks in campus and home networks to demonstrate how SDN can improve common network management tasks.

Intellectual depth. Although software defined networking has made it possible for network operators to exert greater control over their networks, the existing protocols for controlling network devices do not make it easy to control network behavior to perform high-level tasks. Moreover, even as new ways of configuring and managing networks take root, they must be designed to co-exist and interconnect with legacy networks. The Transit Portal resulted in developing new abstractions for interconnecting virtual networks; our work on fast, programmable data planes was the first to demonstrate how to provide a hardware-based programmable data plane for virtual networks, and many other works on programmable data planes (*e.g.*, recent work on newer versions of ASICs for SDNs) have proposed using similar abstractions and mechanisms; and our work on developing new control models and languages for SDNs has identified sources of complexity in network management and developed abstractions for reducing the complexity of common tasks.

Impact. The results of our research have not only appeared in top-tier publications, but have also resulted in real, running software systems, some of which have been used by thousands of users. The Transit Portal is currently deployed in six locations—including a recent deployment in the Amsterdam Internet Exchange in May 2013—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 the first course in the world where students can directly configure networks of routers that are connected to the global Internet. The Transit Portal is also being actively used in research and has supported many other research projects, including several projects at the University of Souther California and the University of Washington that have resulted in multiple independent research papers that have appeared at *ACM SIGCOMM*. Our work on better control models and languages for software defined networks is the basis for a controller that has been deployed in tens of home networks around the world and is now in use as part of a trial deployment with Comcast; the contoller we developed was also used in a Coursera course on SDN that I taught in 2013, which was completed by several thousand students.

Older (and Ongoing) Research Themes

Below, I highlight two previous broad research themes—data-driven network security and network operations—which have formed the foundation of much of my work since the start of my faculty career, and which I continue to work in.

Data-Driven Network Security

Spam filtering. 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*. 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.

Lightweight reputation and early-warning systems. I have continued my research in network security by studying how attackers use the underlying Internet infrastructure to achieve *agility*. In particular, we performed a study that explored the initial DNS behavior of spammers that appeared in the *ACM SIGCOMM Internet Measurement Conference* in 2011. We also performed a second study that explored how attackers use the Internet's interdomain routing protocol, Border Gateway Protocol (BGP) to evade detection when sending spam and performing other malicious activities. That study appeared in the *Passive and Active Measurement Conference* in 2011. Finally, we are exploring how to prevent data leaks from cloud-based Web applications, even when the applications themselves may be compromised.

Impact. My network security research 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. SNARE 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. Our work on DNS and BGP reputation has been patented and implemented by Verisign, and is currently in use in several of their security products. Our work on detecting fraudulent voting on webmail messages was implemented and deployed in Yahoo's webmail system.

Network Operations

Much of my work in *network operations* (the practice of designing tools, algorithms, and protocols to help the network operate better) has focused on fault detection, troubleshooting, and optimization.

Proactive fault detection and performance prediction. 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. I have 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* and *IEEE/ACM Transactions on Networking*.

Studying and improving network configuration. In the past several years, I have continued to work on tools and protocols that help operators configure their networks better. To better understand how network operators make changes to network configurations, we performed a study of the evolution of network configuration over five years across two campus networks and have clustered these changes into common tasks, with an eye towards raising the level of abstraction for network configuration. This work appeared in the *ACM SIGCOMM Internet Measurement Conference* in 2011. Second, I have been actively developing systems on top of the Internet routing infrastructure to help network operators optimize the performance of their applications that run in the network. We developed a system called PECAN which jointly optimizes content routing (*i.e.*, the mapping of clients to service replicas) and network routing (*i.e.*, the network-level paths between clients and their respective replicas). We discovered that jointly optimizing network and content routing can significantly improve performance over simply performing each operation independently; our results will appear in *ACM SIGMETRICS* 2013. Finally, we have performed a data-driven econometric analysis that showed how a tiered pricing model can yield both higher profit margins for Internet service providers and greater consumer surplus for users. These results appeared in *ACM SIGCOMM* in 2011.

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. Other work resulted in a Sigma Xi undergraduate research award for Megan Elmore. Our work on tiered pricing was covered extensively in the media, including in *The Economist*.