Research Statement

Nick Feamster

Summary

My research focuses on building available, secure communications networks in the face of failures, misconfiguration, and malice. My approach to this problem is three-fold: (1) Design and implement approaches to improve the network's inherent robustness, to help prevent downtime in the case of network faults; (2) Design and implement techniques to help network operators reduce unwanted and malicious traffic (*e.g.*, spam and phishing attacks); (3) Design and implement tools and techniques that help network operators restore network connectivity when failures do occur.

Two of the Internet's most pressing problems, availability and security, have been around since its inception. They are the most difficult to solve because they require making the network easier to manage and operate. Breakthroughs require domain knowledge, techniques from a wide range of areas, and implementation of the resulting solutions in working systems that are ultimately deployed in practice. My unique approach and expertise will allow me to continue to make important contributions to these areas.

I draw inspiration for problems from practice, apply a principled approach to develop solutions, and transfer these solutions back to practice. I discover interesting and challenging practical problems through frequent discussions and meetings with network operators and people in industry; in particular, I search for real-world network operations problems that present an opportunity for making breakthroughs by applying a principled approach. I then tackle these problems by applying first principles, developing new methods, and transferring these solutions back to practice in the form of working systems.

Contributions and Impact

Today's network operators face a major challenge: spam, phishing, denial of service attacks, and even the increasing size and complexity of the network itself have made the network increasingly difficult to manage. At the same time, everyday users are becoming more dependent on the Internet: even minutes of downtime can result in inconvenience or loss for users and companies. Nobody notices when the network works well, but everyone suffers when it doesn't. My research focuses on developing tools and algorithms that make the network easier to manage, more secure, and more available.

Approach: Principled approach to network operations problems. 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 gap. I first devise methods to understand the nature of the problem in practice. I then 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.

Theme 1: Designing inherently robust and resilient networks. 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 (“multipath 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.

When networks do not perform as expected, network operators need ways to determine the underlying cause for an attack or performance degradation. In other words, the network should afford some level of *accountability*. The current Internet architecture provides little to no accountability whatsoever: 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: Both Path Splicing and AIP appeared in *ACM SIGCOMM* in 2008, the premier conference on communications networks; early versions of both papers also appeared in *ACM Workshop on Hot Topics in Networking (HotNets)*. I am in the process of incorporating both of these technologies into working systems and transferring them to practice. I am working with BBN on a DARPA proposal 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). I am working with several people at Cisco on incorporating path splicing into Cisco router functionality, and we are also working on a prototype implementation in existing software routers.

Theme 2: Defending against unwanted traffic. I have applied my research approach to defend the network against unwanted traffic, such as spam. Prior to my work, nearly all previous approaches to filtering spam relied in some way on analyzing an email’s contents to determine the legitimacy of the message. This approach, content filtering, continues to improve dramatically, and many people are working on tuning content filters. Nevertheless, the approach has a fundamental shortcoming: email content is malleable, meaning that it is very easy for a spammer to alter the content of an email to evade a filter without changing the actual meaning of the message. To keep pace with continually changing content, both operators and implementors of spam filters must continually tune their filters in a game of catch-up. For example, last year saw a rise in spam whose main content was transported in images; when content filters incorporated optical character recognition for these image-based messages, spammers switched to portable document format (PDF) messages. My research has taken a complementary approach: rather than classifying an email message based on *what* is in the message, classify the message based on *how it is sent* (e.g., what country it coming from, when it was sent). In other words, examine the *network-level behavior* of the email sender and classify the spam based on whether the observed sending behavior likely corresponds to a legitimate sender or a spammer.

Impact: Based on my study of network-level behavior of spammers, which received the Best Student Paper Award at *ACM SIGCOMM* in 2006, I have developed a new class of blacklisting techniques, called *behavioral blacklisting*, which leverages machine learning algorithms to classify email senders as spammers based on previously observed spamming patterns. I am working with both Cisco/Ironport and Secure Computing, vendors of several large spam filtering appliances, to deploy and evaluate these algorithms in practice. I sit on the board of Anchor Intelligence, who develop solutions for detecting click fraud; this company is also applying some of our ideas in studying network-level behavior for automatically detecting click fraud.

Theme 3: Improving fault detection and diagnosis. I have also developed several new techniques that improve Internet availability by helping operators run their networks better. Much of my work has focused on fault detection and troubleshooting. Prior to my work, operators relied on detecting problems with networks “at runtime” on a live network. My 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 need to 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 a sea of heterogeneous data, none of which intuitively points them to the true source of the problem. 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. More recently, I have begun developing a system that collects and 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.

Impact: One cornerstone of this work is a system called “rcc” (router configuration checker), which 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. We are applying these detection algorithms to routing data in several large ISPs and enterprise networks. Finally, we are working with Google to design, test, and deploy a system for detecting violations of network neutrality.