Research Statement

Nick Feamster

I am interested in designing, building, analyzing, and measuring networked systems that are composed of multiple autonomous, potentially untrusted entities. There are numerous examples of such systems: the Internet routing infrastructure, public wireless networks, peer-to-peer systems, large-scale distributed computing infrastructures (*e.g.*, grid computing), Web services, content delivery networks, and enterprise networks.

These systems present three challenges to system designers:

1. Developing communication protocols that ensure correct, predictable, and robust operation;
2. Designing practically deployable techniques for secure operation;
3. Supporting fault diagnosis, troubleshooting, and monitoring.

Some fields of engineering have reached a level of maturity where there are concrete design principles that ensure a level of correctness and robustness: for example, it is reasonably well understood how to build bridges and buildings that withstand earthquakes and high winds. Engineers can apply such principles to help them design robust, reliable systems. Unfortunately, network system designers and engineers lack such a rubric.

My research uses four techniques towards the ultimate goal of developing sound methods for designing and implementing networked systems: measurement; modeling; design and implementation; and deployment. Measurement provides evidence and intuition for the severity of a problem. Modeling—the process of deriving a simpler representation that abstracts irrelevant details and concisely describes the aspects that affect the properties under study—facilitates a more thorough understanding of a problem's fundamental causes. Design and implementation provide the opportunity to apply the intuition gained from measurement and modeling to make tangible improvements to real-world systems. Deployment demonstrates the feasibility and practicality of an implementation and provides the excitement of seeing research ideas applied in practice. In my future work, I intend to apply these techniques to specific problems in routing, network security, anonymous communication, anomaly detection, and monitoring in resource-limited environments.

Dissertation Work: Robust, Predictable Internet Routing

My dissertation solves some of the challenges raised above in the context of Internet routing, which requires that competing, autonomous networks ("autonomous systems", or ASes) cooperate to establish global connectivity. I have designed and implemented models and tools that make today's routing infrastructure more robust and easier to manage. I have examined fundamental tradeoffs between routing stability and expressiveness in generic policy-based routing protocols. Finally, I have proposed a new Internet routing architecture that solves many of the problems we discovered (*i.e.*, through measurement and analysis) with today's routing infrastructure.

Routing configuration is essentially a complex distributed program. Each AS independently configures local *policies* that control how routers select and re-advertise routes. These policies implicitly codify bilateral business relationships between ASes. Each AS may contain tens to

hundreds of routers, each of which is individually configured with hundreds to thousands of lines of code. The collection of configurations within an AS determines whether the routing protocol operates correctly. Faults in configuration can induce routing failures, such as forwarding loops, partitions, and instability, that can prevent packets from reaching their destinations.

My research has applied measurement, modeling, design and implementation, and deployment to help make today's Internet routing infrastructure less prone to failure, as well as more predictable. Network operators need assurances that today's routing protocols will operate correctly, and they need to know which route each router will select, given a set of configurations. My work on a *routing logic* defines a correctness specification for policy-based routing. Based on this specification, I developed *rcc* ("router configuration checker"), a tool that analyzes the set of router configurations from a single AS and detects configuration faults that could induce routing failures. *rcc* has been used by over sixty network operators and has successfully identified faults in the configurations of several Internet service providers with nationwide backbone networks. Experience with *rcc*'s deployment in real-world networks has provided a better understanding of the nature and extent of configuration faults that occur in practice. Additionally, my work on modeling route selection led to a tool that makes routing more predictable by helping network operators predict the effects of a configuration change before deploying it.

With collaborators, I have also applied the above techniques to design improvements to today's Internet routing protocols. Ideally, Internet routing should disseminate loop-free routes and converge to a stable routing topology, regardless of how each AS configures its local policies. We have applied an abstract model of today's Internet routing protocol to derive constraints on local policies that must be satisfied to guarantee that a policy-based routing protocol will not oscillate. To guarantee correct dissemination of loop-free routes, we proposed the Routing Control Platform (RCP), a system that computes routes on behalf of routers. By applying two design principles—(1) compute consistent routes with complete routing information and (2) control interactions between different routing protocols (*e.g.*, between the inter-AS routing protocol and an AS's internal routing protocol)—RCP explicitly prevents the forwarding loops and oscillations that plague today's Internet routing infrastructure.

Future Research Directions

I intend to continue working on improving the robustness, security, and diagnosis capabilities of large-scale systems in which potentially untrusted entities must cooperate to provide some service.

Robustness and Predictability

Wireless mesh networks. While Internet routing is perhaps the best studied example of a routing system that requires cooperation among multiple untrusted parties, other domains, such as public wireless "mesh" networks, present interesting issues. These networks are composed of nodes that are typically owned by different parties (*e.g.*, homes, businesses) that must cooperate to provide connectivity. Because each of these entities may have vastly different criteria for ranking preferred paths through the network and for carrying traffic over those paths (*e.g.*, minimizing loss rate, cost, etc.), these wireless networks may benefit from using policy-based routing protocols. I intend to explore routing problems in public wireless networks to see what design principles can tackle wireless-specific challenges (*e.g.*, contention for a shared channel, interference, mobility) to achieve robust and predictable routing.

Routing stability. Policy-based routing protocols provide each participant remarkable flexibility for implementing complex business arrangements in local policy; unfortunately, the interactions between these policies may conflict, resulting in instability. The tradeoff between routing policy flexibility and stability is poorly understood today. I would like to characterize the minimal set of constraints that must be imposed on each participant's policies to guarantee global stability. In the context of Internet routing, I would like to determine whether those constraints are expressive enough to implement important operational tasks (*e.g.*, load balancing traffic). My experience using game theory and mechanism design to study routing protocol oscillation, as well as my knowledge of network operations, should prove useful for solving these problems.

Secure Networked Systems

Data plane security. Previous work has studied routing protocol security, but little attention has been paid to security and policy enforcement in the *data plane* (*i.e.*, the path that data packets actually traverse). Today's Internet architecture provides scant support for a network to thwart unwanted packets (*e.g.*, spam, viruses, denial of service attacks) and essentially no control over the sequence of ASes that outgoing traffic traverses en route to a destination. I plan to design architectural modifications that could facilitate stronger security and policy enforcement capabilities in the data plane.

Anonymous and censorship-resistant communication. Governments of certain countries routinely implement firewalls to restrict communication to various destinations. To enable clients behind these firewalls to access restricted Web content, we designed and implemented Infranet, an anti-censorship system that embeds requests for blocked content in a covert channel that appears to the censor as innocuous traffic to permissible Web sites. I would like to design anti-censorship systems like Infranet that are robust to untrusted or malicious participants. More generally, I intend to examine how incorporating network-layer information can make anonymous communication systems more resistant to eavesdropping attacks.

Fault Diagnosis and Monitoring

Anomaly detection. Supporting both fault diagnosis and secure operation in large-scale networked systems typically requires the ability to collect, analyze, and audit large quantities of data. I intend to explore whether signal processing and clustering techniques can be useful for performing forensic analysis of spam (*e.g.*, determining groups of machines that are being controlled by a single sender). I have also begun investigating whether signal processing-based anomaly detection techniques such as principal component analysis can be useful for detecting routing anomalies.

Monitoring in resource-limited environments. I previously applied signal processing techniques to design and implement a video transcoder that allowed video streams that were originally encoded at very high bitrates to be transmitted over relatively low-bandwidth wireless links in real-time. I plan to investigate how signal processing can reduce communication costs in other bandwidth and resource-constrained environments. For example, sensor networks have strict bandwidth and energy budgets that often require data streams with high data rates to be processed in the network. I would like to design and implement distributed signal processing algorithms that help reduce computation and communication overhead in resource-constrained environments.