

# Research Statement

Micah Sherr

## 1 Summary

**My research focuses on the study of privacy-enhancing technologies to defeat eavesdropping in communications networks.** Well-established confidentiality protocols protect the *contents* of network messages from eavesdroppers. However, such protocols do not mask the *identities* of communicating parties and the times and durations of their exchanges, permitting eavesdroppers to assemble profiles of users' network access. Since David Chaum first proposed anonymizing mix networks nearly 30 years ago [4], a myriad of *anonymity systems* have been proposed that further aggravate traffic analysis by obfuscating the identities of the communicating parties. Such systems provide network privacy beyond that achieved using standard end-to-end cryptography, enabling constituents of restrictive governments to gain access to censored information, bloggers to publish without fear of punishment, Internet users to browse the web without their information being captured in data warehouses, the military to hide network command-and-control hierarchies, and law enforcement agencies to conduct undercover online sting operations, etc.

Unfortunately, existing anonymity systems often impose rigid “one-size-fits-all” routing policies without regard to the communication requirements of networked applications. For example, an anonymity network that produces high bandwidth anonymous paths does not necessarily benefit a user who requires a low latency channel. The inability to provide anonymous channels that are compatible with network applications hinders the adoption of these anonymity systems, leaving a significant fraction of network communication susceptible to traffic analysis.

My research focuses on the problem of constructing anonymous channels that adhere to the communication requirements of modern networked applications. My dissertation proposed an anonymity framework called *Application-Aware Anonymity* ( $A^3$ ) [1] in which applications specify their communication requirements using a declarative policy language. Unlike other anonymity systems,  $A^3$  is sufficiently flexible to support diverse communication requirements (for example, constraints may be specified in terms of bandwidth, latency, loss, jitter, or some combination of the above), making the system of practical use to applications.

My approach to defeating traffic analysis is three-fold: (i) Analyze the capabilities and limitations of eavesdropping systems; (ii) Examine the communication properties of networked applications; and (iii) Develop network architectures that both resist traffic analysis and provide tunable performance. Defeating traffic analysis while retaining efficient communication characteristics requires cross-domain knowledge of distributed systems, applied cryptography, systems security, and databases. I take an interdisciplinary approach, and actively involve students with different expertise into my research.

## 2 Research Contributions

My research focuses on eavesdropping countermeasures and anonymous communication. To conduct my research, I use a principled and systematic approach. I first identify the assumptions made by traffic analysis systems and evaluate the security implications of violating those assumptions. I then examine eavesdropping systems to enumerate their capabilities and identify their weaknesses. Finally, based on the underlying assumptions and capabilities of eavesdropping technologies, I propose protocols that achieve flexible and high performance anonymous communication.

## 2.1 Dissertation Work: Tunable and High Performance Anonymity

My dissertation introduces the A<sup>3</sup> anonymity architecture that enables applications to select anonymous routes that meet their specific communication requirements while providing an acceptable level of anonymity [11]. For example, A<sup>3</sup> enables a voice-over-IP application to produce anonymous paths with low latency and jitter while providing high bandwidth paths for anonymized file transfers. Unlike existing anonymity systems that implement rigid relay selection strategies, A<sup>3</sup> incorporates the use of *declarative networking* [9] to enable applications to concisely specify their communication requirements (e.g., bandwidth and latency). Additionally, A<sup>3</sup>'s declarative policy language is sufficiently flexible to encode both existing and new anonymity protocols (for example, Tor's relay selection algorithm can be specified in just five lines of the declarative language [18]), making A<sup>3</sup> an ideal platform for protocol designers and anonymity researchers.

To permit flexible routing using a variety of performance metrics, A<sup>3</sup> leverages virtual coordinate embedding systems (e.g., Vivaldi [8]). Virtual coordinate systems map nodes to multidimensional coordinates such that the Euclidean distance between any two nodes' coordinates corresponds to the actual network distance between them. A<sup>3</sup> estimates the performance of potential anonymous paths by aggregating the Euclidean distances between adjacent nodes in the paths, and instantiates only the paths that meet the application's performance constraints.

Since both the performance and anonymity offered by A<sup>3</sup> depends upon the accuracy of the underlying coordinate embedding system, my research also investigates mechanisms to protect such systems from manipulation. Prior to my research, existing techniques for protecting coordinate systems relied on *a priori* trusted nodes or centralized authorities, both of which present central points of failure (and eavesdropping) for anonymity networks. I developed *Veracity* [14], a fully distributed vote-based protocol that protects the truthfulness of network coordinates by introducing a verification step in which the accuracy of a coordinate must first be assessed by both deterministic and random sets of network peers before being used. *Veracity* protects the underlying coordinate system, enabling A<sup>3</sup> to provide tunable and high performance anonymous paths even when a large percentage of A<sup>3</sup> participants are malicious. Additionally, since virtual coordinate embedding systems are also used in proximity-based routing, neighbor selection in overlays, and replica placement in content-distribution networks, *Veracity* may also be applied to protect the performance of these systems as well.

**Impact:** A<sup>3</sup>'s relay selection algorithm appears in the 2009 Privacy Enhancing Technologies Symposium (PETS) [13], the premier conference for privacy and anonymity research; an earlier version is published in the Workshop on Hot Topics in Security (HotSec) [16] in 2007. The complete A<sup>3</sup> design will appear in the 2010 Network and Distributed System Security Symposium (NDSS) [18]. We plan on deploying A<sup>3</sup> as a service on PlanetLab, providing protocol designers with a testbed for rapidly specifying and implementing anonymity protocols. The *Veracity* coordinate protection protocol first appears in the 2008 International Workshop on Peer-to-Peer Systems (IPTPS) [17]; a more mature version of the work appears in the 2009 USENIX Technical Conference [14]. I am working with project leaders of the Tor anonymity system to incorporate the use of virtual coordinate embedding systems into Tor to estimate link latencies between potential relays.

## 2.2 Analyzing Internet Eavesdropping and Telephone Wiretapping Systems

A second theme of my research is the analysis of interception architectures, particularly fielded Internet and telephone eavesdropping systems. To investigate the accuracy of eavesdropping systems, I have developed unilateral eavesdropping countermeasures called *confusion* and *evasion* that impair an eavesdropping system's ability to reliably reconstruct communication [5]. *Confusion* and *evasion* exploit ambiguities in protocol implementations and asymmetries in network topologies to either hide actual content in superfluous cover traffic (*confusion*) or cause the eavesdropping system to erroneously discard the legitimate communication (*evasion*). My work shows that a large number of free and commercial Internet eavesdropping systems are susceptible to such traffic analysis countermeasures [6], and at the extreme, may be tricked into interpreting

false messages of the sender's choosing rather than the actual communication [7].

My research also demonstrates weaknesses in older and more tightly controlled eavesdropping infrastructures. Legally authorized telephone wiretapping systems – that is, eavesdropping systems used by the FBI and other law enforcement agencies to conduct telephone wiretaps – are often perceived as supplying infallible interception transcripts for investigative intelligence and legal evidence. My research demonstrates that telephone wiretap systems, despite their long history and importance to the investigative and judicial processes, are vulnerable to the same unilateral traffic analysis countermeasures that impair Internet eavesdropping systems [15, 19]. In particular, both older *loop extender* telephone wiretap systems as well as the newer *CALEA* wiretaps that replace them are susceptible to attacks that enable the subject of the wiretap to insert false call records, mask dialed telephone numbers, stop the recording of call audio at will, and launch denial-of-service attacks against the wiretap.

**Impact:** The confusion and evasion eavesdropping countermeasures were introduced in the Security Protocols Workshop (SPW) in 2005 [5]. The study of current generation computer eavesdropping systems appears in the 2006 International Conference on Digital Forensics [6] and in extended form in the International Journal of Security and Networks (IJSN) [7]. Vulnerabilities in loop-extender wiretap systems were disclosed in the IEEE Security and Privacy Magazine in 2005 [15], and were the subject of articles in the New York Times and several online news and technology magazines. The security analysis of the newer CALEA wiretapping architecture appears in the 2009 ACM Conference on Computer and Communications Security (CCS). The study of CALEA systems has been publicized in Wired and the IDG News Service. I am currently engaged in conversations with the FBI concerning mitigation strategies and best practices.

### 2.3 Additional Work: Evaluating the Security of Electronic Voting Machines

Defeating traffic analysis requires an in-depth understanding of complex security systems. To hone these skills, I participated in two source code evaluations of electronic voting machine systems. These rare opportunities enhanced my ability to locate design and implementation flaws in large and heterogeneous software and hardware systems.

The Top-to-Bottom Review (TTBR) [3], initiated by the California Secretary of State to investigate the security of the state's electronic voting systems, was the first study of voting equipment and software in which academic investigators had access to the complete source-code and developer documentation of major voting systems. By conducting traffic analysis of the communication generated by Sequoia Voting Systems' election management software, I discovered significant implementation vulnerabilities that enable a rogue election worker to bypass all of the software's authenticity and authentication checks, allowing him/her to misconfigure voting equipment and alter election results.

Additionally, I participated in the Ohio Secretary of State's EVEREST study [10] of Ohio's electronic voting systems. In particular, I examined the touchscreen voting equipment produced by ES&S – the largest manufacturer of election equipment in the United States. During the study, I uncovered numerous programming errors and undocumented “quality assurance modes” that enable a malicious voter with a properly configured handheld computing device (e.g., a Palm organizer) to bypass all authentication and authorization checks and load malware onto the touchscreen system. Such software could, for example, erase or alter vote counts and spread virally to other touchscreen terminals located in the same polling place.

**Impact:** As a result of the California review, the California Secretary of State withdrew her approval of the tested touchscreen devices, permitting their use only if a manual tally were subsequently conducted at the close of the election to confirm the election results. Similarly, following the release of the Ohio EVEREST report, the Ohio Secretary of State recommended the elimination of touchscreen voting systems from all polling locations. Both reports received significant media attention and were the subject of articles in the New York Times and the San Francisco Chronicle. A paper that summarized our experiences with the

ES&S voting equipment appears in the 2008 USENIX/ACCURATE Electronic Voting Technology Workshop (EVT) [2], the premier venue for voting machine research. I presented written and oral testimony [12] concerning the results of the California and Ohio studies to the West Virginia legislature's Joint Judicial Subcommittee.

### 3 Future Work

Below, I briefly outline my future research agenda, describing short-term extensions to my graduate work as well as my longer-term research plan.

#### 3.1 Extensions to Dissertation

The effectiveness of an anonymity system depends upon the availability of anonymizing routers. An important aspect of constructing anonymity networks that we have not fully explored is the development of *incentive systems* that motivate participation in the system.

As a first step, I intend to study the economics of *tit-for-tat* schemes in which bandwidth is traded as a commodity: nodes that dedicate a significant portion of their bandwidth resources to forward the traffic of its peers will be given priority access to desirable anonymous channels (i.e., those with low latency and high bandwidth). Incorporating differentiated services into anonymity networks presents interesting research challenges since, by definition, the performance of a differentiated channel will leak some information about that channel's originator. Developing such a system requires knowledge of economics, game-theory, distributed systems, and network security, and presents an exciting opportunity for interdisciplinary collaboration.

Another area of related research is the design of fully decentralized anonymity networks. Most anonymity networks (including A<sup>3</sup>) rely on directory servers to maintain the addresses of anonymizing relays and resolve lookup requests from clients. The directory servers are both central points of failure and security in such anonymity networks, and are consequently attractive targets for conducting eavesdropping and denial-of-service attacks. I will study the use of distributed directory services to maintain membership information in a decentralized fashion. Although there has been significant work in designing distributed hash tables (DHTs) that are resilient to manipulation, existing techniques do not permit the *anonymous lookups* that are required by anonymity systems. The ultimate goal is to develop distributed directory services that are simultaneously robust, scalable, insusceptible to insider manipulation, and resistant to traffic analysis.

#### 3.2 Looking Beyond

My future research agenda entails the study of novel applications of privacy-preserving technologies. I plan to extend the themes of my wiretapping and electronic voting system work and further investigate the intersection of systems security, privacy, and public policy. Leveraging my dissertation work, I intend to proceed along the following research directions:

- **Preserving privacy in tightly controlled mobile data networks.** The mobile data networks offered by cellular service providers are tightly controlled and centrally administered, easing the task of monitoring network communication. Given mobile service providers' unobstructed views of their networks, anonymity systems designed for use on the Internet are ill-suited for providing anonymous messaging to mobile devices. I plan on researching methods of achieving private communication under such tightly controlled environments. A potential approach is to circumvent the intended communication model and leverage the multiple interfaces of modern smartphone devices. For example, users may shuffle messages amongst geographically proximate peer smartphones via *ad hoc* 802.11 networks,

Bluetooth, or infrared. I hope to work in partnership with researchers in cellular telephony, ubiquitous computing, and distributed networking in this exploration.

- **Privacy-preserving cloud computing.** Cloud computing services allow the outsourcing of computation and storage to clusters of remote Internet hosts, potentially offering significant cost savings and improved scalability over locally implemented solutions. Often, the nodes in a cluster are administered by a single organization, enabling cloud operators an unobstructed view of the actions of its subscribers. An interesting avenue of future research is the formation of cloud computing models in which no single administrative entity can discern subscribers' computations or data. One potential mechanism for achieving privacy in the cloud is to distribute computation and storage across multiple cloud providers, using anonymous messaging primitives to mask the identities of the cloud nodes used by a given subscriber. A<sup>3</sup> is particularly well-suited for inter-cloud anonymous communication, as it provides the flexibility and performance required for different cloud applications. The design of privacy-preserving clouds requires expertise in distributed computing, systems security, cryptography, and algorithms, and presents an opportunity for collaboration.
- **Secure network provenance.** Recent proposals at *network provenance* provide mechanisms for discerning the path a packet took in the network and the hosts that contributed to its content. Network provenance is applicable to a variety of network administration tasks, including network debugging, forensics, and accounting. Unfortunately, existing network provenance solutions do not adequately address the confidentiality of network provenance information. For example, operators of autonomous systems (ASes) may be reluctant to disclose the structure of its internal network or its peering relationships. I plan to investigate access control and policy mechanisms that permit flexible and fine-grain control over network provenance information. The privacy implications of disclosing message headers and payloads have been well-studied. However, the effects on privacy and confidentiality of disclosing *how* messages came to be has not been adequately addressed. I intend to work with researchers in the database, networking, and visualization communities to develop access control techniques for network provenance as well as to quantify the effects of network provenance disclosure.
- **Consensus anonymity.** Existing Internet anonymity systems attempt to provide permanently untraceable communication. My future research will introduce the notion of *consensus anonymity* in which anonymity can be revoked *a posteriori* if a sufficient fraction of the network's participants elect to do so. Such an architecture enables new classes of self-policing anonymity networks in which communities of users can identify and expel members who overburden the network or use it to transmit unacceptable content (as determined by a majority of its users). Potential mechanisms for constructing consensus anonymity systems incorporate the use of threshold encryption or other cryptographic primitives in which data is revealed only via the cooperation of a predetermined number of participants. I will study consensus anonymity in the context of P2P file sharing systems, anonymous multicast, and anonymous message passing.

## References

- [1] Application-Aware Anonymity. <http://a3.cis.upenn.edu>.
- [2] Adam Aviv, Pavol Cerný, Sandy Clark, Eric Cronin, Gaurav Shah, Micah Sherr, and Matt Blaze. Security Evaluation of the ES&S Voting Machines and Election Management System. In *Third USENIX/ACCURATE Electronic Voting Technology Workshop (EVT)*, August 2008.
- [3] Matt Blaze, Arel Cordero, Sophie Engle, Chris Karlof, Naveen Sastry, Micah Sherr, Till Stegers, and Ka-Ping Yee. Source Code Review of the Sequoia Voting System, July 2007. Part of the California Secretary of State Top-to-Bottom Review of electronic voting machines.
- [4] David L. Chaum. Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms. *Communications of the ACM*, 24(2):84–90, 1981.

- [5] Eric Cronin, Micah Sherr, and Matt Blaze. Listen Too Closely and You May Be Confused. In *Security Protocols Workshop (SPW)*, April 2005.
- [6] Eric Cronin, Micah Sherr, and Matt Blaze. On the Reliability of Current Generation Network Eavesdropping Tools. In *Second Annual IFIP WG 11.9 International Conference on Digital Forensics*, January 2006.
- [7] Eric Cronin, Micah Sherr, and Matt Blaze. On the (un)Reliability of Eavesdropping. *International Journal of Security and Networks (IJSN)*, 3(2), February 2008.
- [8] Frank Dabek, Russ Cox, Frans Kaashoek, and Robert Morris. Vivaldi: A Decentralized Network Coordinate System. In *ACM SIGCOMM Conference on Data Communications*, pages 15–26, 2004.
- [9] Boon Thau Loo, Joseph M. Hellerstein, Ion Stoica, and Raghu Ramakrishnan. Declarative Routing: Extensible Routing with Declarative Queries. In *ACM SIGCOMM Conference on Data Communications*, 2005.
- [10] Patrick McDaniel, Kevin Butler, William Enck, Harri Hursti, Stephen McLaughlin, Patrick Traynor, Matt Blaze, Adam Aviv, Pavol Cerny, Sandy Clark, Eric Cronin, Gaurav Shah, Micah Sherr, Giovanni Vigna, Richard Kemmerer, David Balzarotti, Greg Banks, Marco Cova, Viktoria Felmetzger, William Robertson, Fredrik Valeur, Joseph Lorenzo Hall, and Laura Quilter. EVEREST: Evaluation and Validation of Election-Related Equipment, Standards and Testing, December 2007. Commissioned by the Ohio Secretary of State to investigate the security properties of the state’s voting equipment.
- [11] Micah Sherr. *Coordinate-Based Routing for High Performance Anonymity*. PhD thesis, University of Pennsylvania, 2009.
- [12] Micah Sherr. Testimony to the West Virginia Joint Judiciary Subcommittee, August 2009. Charleston, WV. Available at <http://micah.cis.upenn.edu/papers/wv-voting-testimony.pdf>.
- [13] Micah Sherr, Matt Blaze, and Boon Thau Loo. Scalable Link-Based Relay Selection for Anonymous Routing. In *9th Privacy Enhancing Technologies Symposium (PETS)*, August 2009.
- [14] Micah Sherr, Matt Blaze, and Boon Thau Loo. Veracity: Practical Secure Network Coordinates via Vote-based Agreements. In *USENIX Annual Technical Conference (USENIX ATC)*, June 2009.
- [15] Micah Sherr, Eric Cronin, Sandy Clark, and Matt Blaze. Signaling Vulnerabilities in Wiretapping Systems. *IEEE Security & Privacy Magazine*, 3(6):13–25, November 2005.
- [16] Micah Sherr, Boon Thau Loo, and Matt Blaze. Towards application-aware anonymous routing. In *Second USENIX Workshop on Hot Topics in Security (HotSec)*, August 2007.
- [17] Micah Sherr, Boon Thau Loo, and Matt Blaze. Veracity: A fully decentralized service for securing network coordinate systems. In *7th International Workshop on Peer-to-Peer Systems (IPTPS)*, February 2008.
- [18] Micah Sherr, Andrew Mao, William R. Marczak, Wenchao Zhou, Boon Thau Loo, and Matt Blaze. A3: An Extensible Platform for Application-Aware Anonymity. In *17th Annual Network and Distributed System Security Symposium (NDSS)*, February 2010.
- [19] Micah Sherr, Gaurav Shah, Eric Cronin, Sandy Clark, and Matt Blaze. Can They Hear me Now? A Security Analysis of Law Enforcement Wiretaps. In *16th ACM Conference on Computer and Communications Security (CCS)*, November 2009.