# Privacy-Enabled Global Threat Monitoring

Phillip A. Porras
*SRI International*

The history of intrusion detection research gives a nice example of a community in a perpetual race to stay relevant. While we once focused on detecting user account misuse in mainframes, we then moved on to local area network abuse, and then to address the scalability problems in enterprise-wide detection. With the rise of e-commerce in the late 1990s, we intrusion detection developers have had to react to the emergence of script kiddies and Web defacements. Distributed denial-of-service attacks and wide-scale virus propagation soon followed, as did a new term, *malware research*, to address the growing concern about viruses and self-replicating worms spreading across the Internet at alarming speeds. More recently, we've had to consider the problem of botnets, which can organize and maintain illicit control of thousands of machines for months at a time to spread spam, conduct phishing attacks, or steal data or computing resources. Over the past decade, intrusion detection research has rarely been boring.

Our ambition as defenders has also grown substantially, to a goal that we might one day solicit any and all willing networks to fight back against global network attacks. Following the popular success of such initiatives as DShield (www.dshield.org) and DeepSight (http://tms.symantec.com), we've seen an increased interest in large-scale analysis centers that collect network security information from a diverse pool of contributors and provide a rapid warning service for Internet threats and a resource for new defense strategies. The availability of rich, comprehensive network security data sets that are collected and analyzed in real time and culled from a broad cross-section of intrusion detection systems (IDSs), firewalls, honeypots, and network sensors could shift how we identify and formulate responses to malware.

However, IDS research is an area that also must constantly heed the Hippocratic Oath. For those who contribute data to global threat reconnaissance, the open sharing of raw network security data is fraught with peril. A repository of such data becomes a single point of failure and a natural target for attackers. Moreover, outsiders can abuse legitimate access to a contributor's security logs and use that data against the contributor. Attackers can use security alerts from network sensors, for example, to fingerprint those sensors and map out their locations. Security and audit logs can passively leak information about a contributor's vulnerabilities or divulge its topological details, enabled services and applications, egress filtering policies, and so on.

Successful deployment of global analysis centers requires us to resolve several fundamental trade-offs among global network security, user privacy, data abuse, data repository liability, the utility of data for large-scale attack diagnosis, and practical efficiency. We can't circumvent these issues by placing faith in well-intentioned analysis center operators. Defining these challenges requires an interdisciplinary perspective that spans from information privacy, cryptography, and network traffic anonymity to the needs of those people designing large-scale malware analysis services.

## What's the harm in sharing?

It doesn't take a giant leap of intellect to imagine that sharing the details of your local security and network operations activity might negatively affect your own security posture. Most organizations that propose collecting this information usually attempt to alleviate concerns with two general strategies:

- establish some degree of trust between the repository and the contributors (standard tactic of commercial organizations who advertise agreements not to disclose the contents of collected logs) or
- extract a bare minimum of data from contributors, under the argument that the minimal data collected doesn't significantly disclose vital details about the contributor's site (IP address anonymization often helps disassociate logs from their site of origin).

Unfortunately, there's no reason for members of a large and diverse contributor pool to place their trust in the first approach without the repository's owners accepting a significant

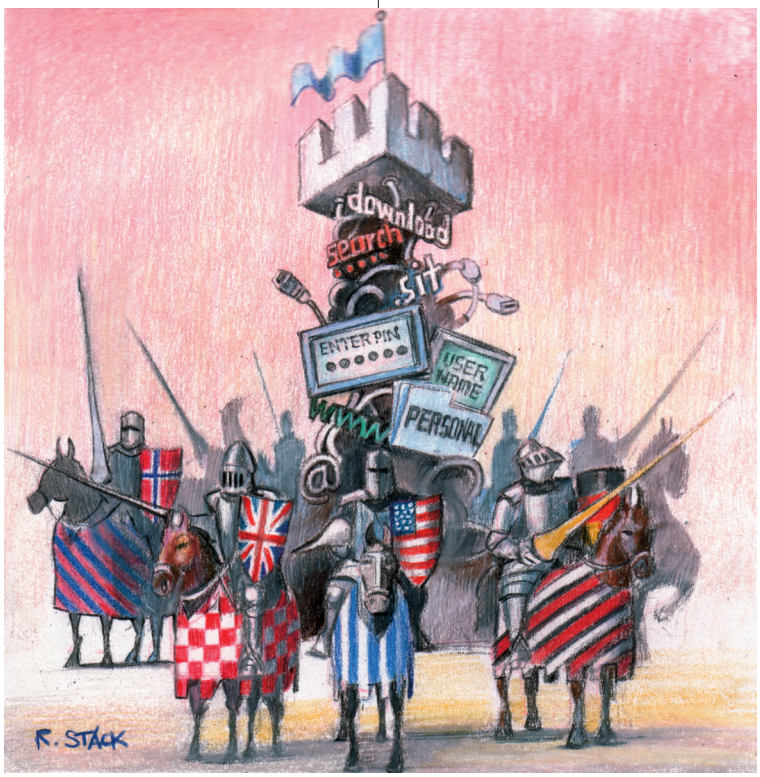
liability should it fail. Malicious repository insiders, accidental data disclosure, traffic interception, or inference attacks against any results the repository produces are potentially serious problems. Moreover, researchers have recently shown that the minimalist approach to data extraction, even in the presence of extensive address anonymization or blacklisting, offers virtually no defense against a determined adversary who wants to map someone's security posture.

In 2004, for example, Patrick Lincoln, Vitaly Shmatikov, and I enumerated several example attack strategies for violating alert contributor privacy, proposed various anonymization strategies, and suggested the use of multiple repositories and countermeasures to basic traffic-monitoring attacks.[1] In subsequent work, John Bethencourt and his colleagues demonstrated active sensor mapping by using alert repository data to inventory sensor locations and map data sinks.[2] More recently, Shmatikov and I enumerated the core problems of ensuring contributor privacy in large-scale threat repositories, surveyed proposed defenses, and posed several central research challenges.[3]

The objectives of large-scale network defense have traditionally assumed the availability of high-precision content from the data contributor pool to track threats, assess security trends, and generally recognize subtle attack patterns. Paradoxically, if the collected data become publicly available for large-scale collaborative analyses, then precision and depth of content in this data come into direct conflict with the contributor's local security posture. The core challenge is in developing a scalable repository and analysis system that strike the balance between the data utility need to drive new large-scale attack forensic algorithms with the need to prevent the linkage of this data back to their associated contributors—and doing so within the practical efficiency constraints necessary to deploy and manage such systems.

Among the more devastating threats to repository contributors is the *fingerprinting* threat, also known as probe-response attacks. Here, the classic intrusion detection paradigm in which the attacker seeks to evade detection is turned on its head—the attacker actually wants to stimulate a contributor's sensors to alarm, intending to later isolate this data from the repository and map such information as the contributor's network defenses, topology, active services, and filtering policies. An adversary might probe a contributor's network in ways that will produce unique or rare alert signatures, for example, or use source and destination port combinations rarely observed together in the wild. In such cases, even complete suppression of addresses and obfuscation of timestamps provide limited anonymity to the contributor. Using static threshold-based filtering on alerts—that is, sharing alerts only when they reach sufficient volume—doesn't work either because the attacker also controls the number of times the probe is performed.

Another challenge lies in preventing an adversary from associating log content with its source during the transport process. A primary method of providing traffic source anonymity is to use an onion-routing system that can provide a circuit-based low-latency anonymous communication channel between the contributor and the alert repository. Because there isn't an assumed trust relationship between the contributor and the repository, the repository must be blind to the contributor's identity, and data transfers must be obfuscated from eavesdroppers located within the untrusted network path. Unfortunately, the most applicable protocols for collaborative alert delivery (long-lived circuits, high-volume alert payloads, and regular posting intervals) impose significant challenges to traffic anonymity. Shmatikov and Ming-Hsiu Wang recently demonstrated attacks and countermeasures that arise when data distribution



through low-latency onion routing networks are subject to correlated timing attacks and other traffic-modification attacks, particularly in the context of long-lived stream collection.[4]

**Will the future of Internet-scale collaborative security frameworks ultimately open a new era of fast-reaction Internet defenses?**

## The Cyber-Threat Analytics project

In June 2006, SRI International began an initiative to help organizations defend against large-scale network threats by creating the underlying technologies that enable next-generation privacy-preserving digital threat analysis centers. These centers must support highly automated threat diagnosis and prioritization, scale to alert volumes and data sources that characterize attack phenomena across millions of IP addresses, and rapidly distribute actionable information back to the broader network community to help mitigate emerging attacks. They must also address fundamental information privacy concerns among the contributor pool, give contributors extensive control over data anonymization policies, and provide traffic delivery anonymization and security.

Accordingly, our multidisciplinary Cyber-Threat Analytics (Cyber-TA) project brings together well-established researchers across the fields of data privacy, cryptography, and malware research, as well as operational experts in Internet-scale sensor management. Our team has four primary research thrusts.

### Data and traffic anonymity

We're currently building anonymization and sanitization operations for all major security log data types, with a special emphasis on understanding how field-level anonymization can provide strong privacy while minimizing its impact on the analytical utility of published logs. We're also using the Tor low-latency onion-routing network[5] to develop countermeasures to traffic-flow-based methods (thus preventing linking contributors to their data submissions). We plan to extend Tor to increase its resistance to application-specific timing and statistical attacks.

### Encrypted computation

We're exploring the application of emerging developments of query, search, and comparison operations on encrypted data for use in the collaborative analysis of high-sensitivity end-node security logs.[6] We're also extending attribute-based encryption methods that provide finer-grained methods of access control than traditional cryptosystems. We envision logging systems that label encrypted data with descriptive attributes (such as IP addresses, ports, and user identities) and then encrypt these attributes in such a way that a mediator can selectively compute private keys that decrypt only on those log entries in which a certain criterion (such as an IDS signature) is met by the associated attributes.[7] We hope to develop IDSs that can analyze fully encrypted security logs for policy and misuse violations without decrypting log content, adjusting and refining these policies well after the data is encrypted and stored. Such systems represent a radical break from current approaches that require full access to sensitive logs to isolate a relative few suspicious records.

### Malware analysis and mitigation

We're studying the fundamental features of large-scale intrusion phenomena captured in various security logs or observed indirectly through multilog analyses, alternative client-side statistics, or metadata extraction. Our emphasis will be on live high-volume repository correlation that goes beyond standard intensity-based measurements and other single-attribute distribution patterns (such as attacked port statistics or source-address blacklisting). We're developing contributor-side correlation applications that characterize local malicious activity through data structures and statistics, with the repositories providing consensus publishing of malware behavior, content signatures, and other related traffic sequences that help detect internal malware infections. We're also exploring group coordination schemes to publish and distribute consensus threat countermeasure data, schemes for helping sites detect emerging malware and botnet behavior from internal sources, honeynet-driven attack classification, and privacy-preserving self-to-world comparative views of log production patterns relative to the contributor pool.

### Threat operations center

We're releasing our research prototypes via open source software and working on some new core capability demonstrations; we'll deploy an academic release of our core privacy-preserving alert collection infrastructure across our consortium partner sites later this year. We've progressed this study to the point of developing and deploying a reference system implementation that provides IDS and firewall log collection and anonymization, source-anonymity-preserving log distribution through an onion-routing infrastructure, a large-scale data repository implementation, and

a Web-based repository portal that provides remotely controllable data query and analysis services. This initial privacy-preserving threat reconnaissance center is currently undergoing test deployment, with future open source software releases to follow early next year.

Will the future of Internet-scale collaborative security frameworks ultimately open a new era of fast-reaction Internet defenses, or are these systems destined to provide limited deployment and detection power for unclear liability risks? We think the former is unlikely without significant progress in rich-content extraction that addresses the fundamental vulnerabilities inherent in collaborative data sharing. Cyber-TA brings together an established group of researchers across a broad spectrum to search for practical solutions and enable new ways of threat detection. To learn more about our research, software releases, and test deployments, visit our project site at www.cyber-ta.org. □

## Acknowledgments

*The team working on the Cyber-TA project represents a wealth of experience from a variety of sources: Paul Barford (University of Wisconsin), Dan Boneh (Stanford University), Linda Briesemeister (SRI International), Steven Cheung (SRI International), Roger Dingledine (Moria Research Labs), Joan Feigenbaum (Yale University), Ray Granvold (Promia), Wenke Lee (Georgia Tech. Institute of Technology), Karl Levitt (University of California, Davis), Peter Neumann (SRI International), Peng Ning (North Carolina State University), Livio Ricciulli (Force-10 Networks), Marcus Sachs (SRI International), Amit Sahai (University of California, Los Angeles), Vitaly Shmatikov (University of Texas, Austin), Dawn Song (Carnegie Mellon University), Paul Syverson (US Naval Research Laboratories), Johannes Ulrich (SANS Institute), Al Valdes (SRI International), Brent Waters (SRI International), Vinod Yegneswaran (SRI International), and Jian Zhang (SRI International). Cyber-TA is supported through the US Army Research Office grant number W911NF-06-1-0316.*

## References

1. P.D. Lincoln, P.A. Porras, and V. Shmatikov, "Privacy-Preserving Sharing and Correlation of Security Alerts," *Proc. Usenix Security Symp.*, Usenix Assoc., 2004, pp. 239–254.
2. J. Bethencourt, J. Franklin, and M. Vernon, "Mapping Internet Sensors with Probe Response Attacks," *Proc. Usenix Security Symp.*, Usenix Assoc., 2005, pp. 193–208.
3. P.A. Porras and V. Shmatikov, "Large-Scale Collection and Sanitization of Security Data: Risks and Challenges," *Proc. New Security Paradigms Workshop 2006*, ACM Press, 2006; www.cs.utexas.edu/~shmat/privacyframeworks/shmat_nspw06.pdf.
4. V. Shmatikov and M. Wang, "Timing Analysis in Low-Latency Mix Networks: Attacks and Defenses," *Proc 11th European Symp. Research in Computer Security* (ESORICS), Springer-Verlag, 2006, pp. 236–252.
5. R. Dingledine, N. Mathewson, and P. Syverson, "Tor: The Second-Generation Onion Router," *Proc. Usenix Security Symp., Usenix Assoc.*, 2004, pp. 303–320.
6. B. Waters et al., "Building an Encrypted and Searchable Audit Log," *Proc. Network and Distributed System Security Symp.*, Internet Soc., 2004, www.isoc.org/isoc/conferences/ndss/04/proceedings/Papers/Waters.pdf.
7. A. Sahai and B. Waters, "Fuzzy Identity-Based Encryption," *Proc. EUROCRYPT*, LNCS 3494, Springer, May 2005, pp. 457–473.

***Phillip A. Porras** is a program director of network security research and the founding director and chief technical architect of the EMERALD intrusion detection research programs at SRI International. His research interests include intrusion detection, alarm correlation, active networks, and wireless security. Contact him at porras@csl.sri.com.*