Large-Scale Collection and Sanitization of Network Security Data: Risks and Challenges (position paper)

Phillip Porras
SRI International

Vitaly Shmatikov
The University of Texas at Austin

ABSTRACT

Over the last several years, there has been an emerging interest in the development of wide-area data collection and analysis centers to help identify, track, and formulate responses to the ever-growing number of coordinated attacks and malware infections that plague computer networks worldwide. As large-scale network threats continue to evolve in sophistication and extend to widely deployed applications, we expect that interest in collaborative security monitoring infrastructures will continue to grow, because such attacks may not be easily diagnosed from a single point in the network. The intent of this position paper is not to argue the necessity of Internet-scale security data sharing infrastructures, as there is ample research [13, 48, 51, 54, 41, 47, 42] and operational examples [43, 17, 32, 53] that already make this case. Instead, we observe that these well-intended activities raise a unique set of risks and challenges.

We outline some of the most salient issues faced by global network security centers, survey proposed defense mechanisms, and pose several research challenges to the computer security community. We hope that this position paper will serve as a stimulus to spur groundbreaking new research in protection and analysis technologies that can facilitate the collaborative sharing of network security data while keeping data contributors safe and secure.

1. INTRODUCTION

Computer (in)security has become a global phenomenon. Distributed denial of service attacks, rapidly propagating viruses, self-replicating worms are a bane of computer networks worldwide, and attacks constantly grow in severity and sophistication. Following the popular success of such initiatives as DShield [17] and DeepSight [43], there has been a growing interest in the creation of large-scale analysis centers that collect network security information from a diverse pool of contributors and provide a real-time warning service for Internet threats, as well as a source of real data to drive new research in large-scale collaborative defense.

New Security Paradigms Workshop, September 19–22, 2006, Schloss Dagstuhl, Germany.

Availability of rich, comprehensive datasets collected from a broad cross-section of intrusion detection systems, firewalls, honeypots, and network sensors has the potential to cause a paradigmatic shift in computer security research. Real-time detection of zero-day attacks; large-scale picture of Internet security trends and inflection points; automatic extraction of signatures for polymorphic malware; blacklisting of attacker-controlled hosts and networks; new understanding of malicious software, its propagation patterns and attack vectors — the possibilities are almost limitless.

It has been recognized, however, that open access to 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, not to mention insider compromise. Moreover, even legitimate access to the data can be abused, and the data contributed by well-intentioned collaborative partners can be turned against them. For example, security alerts contributed by network sensors can be used to fingerprint these sensors and to map out their locations [5]. Security and audit logs may passively leak information about the contributor's vulnerabilities, as well as the data about topology of protected networks, enabled services and applications, egress filtering policies, and so on. From the attacker's viewpoint, this information is complementary to the directly observable attack effects (*e.g.*, network responses of attacked systems), It is especially useful for tracking detection and disposition of *unsuccessful* attacks.

Protection and sanitization of network security data has received some attention in the past two years [26, 28, 41, 51], but the problem is far from being solved. The objective of this position paper is to formulate several crisp research challenges for the computer security community. We believe that these challenges will stimulate the discussion, spur design and implementation of efficient sanitization technologies that balance the utility of network security data for collaborative analysis against the need to protect contributors' privacy and security, and even lead to new paradigms for large-scale sharing of network data, including security alerts, packet traces, and so on. (In the rest of this paper, we will overload the term "privacy" to refer to confidentiality of corporate and organizational data, as well as sensitive information about individuals.)

Our work on formulating these challenges is motivated in part by our involvement in actual design and implementation of Internet-scale centers for privacy-preserving collaborative threat analysis. Our objective, however, is not to promote or champion specific solutions, but to raise awareness of the risks and challenges in this area, and to bring

a well-informed perspective of both theoretical and pragmatic issues involved in architecting strong privacy guarantees into collaborative sharing infrastructures for network security data.

We group risks and challenges into three areas of concern, associated with, respectively, network sensors that generate the data, repositories that collect the data and make them available for analysis, and the network infrastructure which delivers the data from sensors to repositories. We pay special attention to a class of threats we refer to as *fingerprinting* attacks on network data, which have proved devastatingly effective in many contexts [5, 23]. In a fingerprinting attack, an attacker may search for natural patterns in the data that uniquely identify a particular host (*e.g.*, clock skew [23]). Alternatively, the attacker may actively influence data patterns by triggering rare rules in signature-based intrusion detection systems, employing rare port combinations, or generating certain event sequences or timing patterns that can later be recovered from the repository. (This is known in the literature as the *probe response attack* [26, 5].)

Probe response and fingerprinting attacks turn the usual intrusion detection game on its head. In contrast to the standard situation, where the attacker's goal is to evade detection, here the attacker *wants* to be detected so that he can analyze the resulting report for evidence of vulnerabilities and gain better understanding of the defender's security posture. Rigorous formalization of fingerprinting attacks and development of provably secure defense mechanisms against fingerprinting are among the most important challenges identified in this paper.

We believe that techniques and methods developed for sanitizing network security data will find applications well beyond the immediate area of collaborative threat detection and analysis. For example, privacy-preserving transformation and anonymization of Internet packet traces [45, 37, 36] and routing configuration data [29] have received a lot of attention in the network research community, and could potentially benefit from techniques developed for security data anonymization and anonymity-preserving data publishing protocols. Similarly, analysis techniques for flow-level network traffic [20] and locality-based anomaly detection [30] can be used to extract security-related features from sanitized traffic data.

In this paper, we use the term *repository* somewhat loosely to denote both open- and restricted-access analysis centers, which collect network security data from contributors and make it available either in raw, or in sanitized form.

Acknowledgements. This material is based upon work supported by the Department of Defense under Contract No. H98230-05-C-1650. The authors are grateful to anonymous reviewers for their helpful comments, and to the participants of the 2006 New Security Paradigms Workshop for their comments during public discussion of this paper, many of which have been incorporated into the final version.

2. SANITIZATION TECHNOLOGIES FOR DATA CONTRIBUTORS

The first line of protection when sharing sensitive Internet security information is the contributor's selection of the elements of local security log data that may be shared with other collaborators. Traditionally, we have observed two main approaches to addressing the contributor's security and privacy concerns. The first approach is to release a bare minimum of content to the data repository (an extreme example may be to include hashed source and destination IP addresses, source and target ports, and a rounded timestamp of the event). Unfortunately, not only does this approach significantly limit the utility of the collected data for downstream collaborative analysis, but it also fails to provide protection against fingerprinting attacks [5].

The second approach is to collect security data under a strict non-disclosure agreement, with significant liability accepted by the repository operator should the data be released and used to harm the contributor. We argue that the assumption of blind trust between contributors and the data collector neither addresses the underlying privacy concerns, nor is workable in the context of truly collaborative international grids of security sensors. Contributors' security in this approach relies solely on the goodwill of the repository operator and, most importantly, on his ability to protect the repository. A repository containing unsanitized network security data from multiple sources is likely to be of great interest to a variety of attackers. Basing contributors' safety on the assumption that the operator will be able to defend the repository against even the most determined attack is not a good security practice and may discourage potential contributors.

Thus far, we have used the term "network security data" rather loosely to refer to the data produced locally by the contributor to capture security-relevant operations within the contributor's network perimeter. These data can represent a diverse range of information, depending on the type of security device that produced it. In our context, this includes, but is not limited to, security logs produced by services such as firewalls, intrusion detection systems, network flow logs, and so on. The raw data produced by these sensors tend to contain fine-grained information about observed communication patterns, as well as policy decisions regarding connectivity and content analysis conclusions. In addition, each network security alert may divulge IP address information, protocol and port usage, event timing, sensor identity, and potentially even information related to payload or header contents.

Traditional objectives of large-scale network defense require high precision and accuracy from collected security data. For example, high-trust security alert repositories such as DShield [17] and DeepSight [43] rely on precise data from a diverse pool of contributors to identify Internet-scale security trends and provide an early warning service.

Paradoxically, if the collected data are to be made publicly available for large-scale sharing and collaborative analysis, then precision and accuracy come into conflict with security. The traditional assumption of intrusion detection — that all collected data are supplied to a trusted system administrator or an automated software program that performs analysis and assessment — is not valid anymore, because in the open-access environment, the attackers may easily gain access to the data and (mis)use it to identify their own attacks and analyze their propagation and effects. At the very least, defeating these attacks requires local sanitization of security data before they leave the organization that collected them.

A variety of techniques have been proposed for sanitizing local data before releasing them to Internet-scale analysis centers. These vary from hashing IP addresses [26] to compressing them into Bloom filters [28] to "generalizing"

elements of local datasets so that each element of the sanitized dataset corresponds to multiple elements of the original dataset [51].

All of these techniques are non-cryptographic in nature, and do not require any keys to be shared between contributors. Unfortunately, none of them provide *provable* security, especially in the face of an attacker who has access to auxiliary information. For example, IP address hashing is trivially defeated by pre-computation and dictionary attacks, especially when the set of candidate addresses is small. Similarly, if sensitive attributes are rounded up and generalized, an attacker with specific knowledge of attribute values that might have been present in the original dataset can easily infer valuable information from the presence of a generalized attribute in the sanitized dataset. Techniques based on extracting metadata such as Bloom filters [28], k-ary sketch data structures [24], and data cubes [47], do not provide cryptographically strong security guarantees, either.

Research challenge 1: *Develop local sanitization methods for network security data that provide cryptographically strong security guarantees for data owners, while preserving the ability to perform at least some analyses on the aggregated data.*

Sanitization is inherently in conflict with usability, and tends to destroy usefulness of the data for subsequent analysis. Design of sanitization methods must be informed by the types of analyses that network security researchers may want to perform on the sanitized data.

A complete survey of data analysis techniques that are relevant in the context of network security is beyond the scope of this paper. Typically, data analysis involves search for commonalities (*e.g.*, a particular IP address identified as a source of the attack by multiple observers, or similar topologies of attack propagation in multiple networks), extraction of common behavioral patterns associated with the attacker's actions in the network, as well as common structural patterns associated with a particular strain of malicious code (the objective is to create an attack signature that can be used to detect and filter a particular network threat), and detection of "inflection points," *i.e.*, rapid changes in some parameter (*e.g.*, number of packets destined to a particular port) that may provide an early warning of a large-scale attack.

To illustrate the conflict between data security and usability, consider IP addresses that appear in security alerts. If IP addresses are protected by hashing, Bloom filters, or generalization, then testing equality is the only operation that can be performed on the sanitized addresses. Unfortunately, capturing *topological relationships* between addresses is necessary for many types of security analyses such as detecting scanning probes, tracking attack vectors, and identifying propagation trends. For instance, in order to characterize a new infection, the analyst may need to determine whether the attack follows the physical network topology or (as in the case of an email virus) the social network. Topological information, however, is usually lost when addresses are sanitized.

Prefix-preserving hashing [52] is one of the few sanitization techniques that preserves some topological information. To be secure against dictionary attacks, however, hashing algorithms must be keyed. *Key management* is a notoriously difficult problem in a massively distributed setting, such as an Internet-scale analysis center with thousands of contributors. To enable cross-contributor comparison, the keys must be shared across all contributors, which means that security of the scheme is as weak as security of the weakest machine on which keys are stored. It is not clear whether there exist scalable solutions to this problem, although introduction of small, tamper-proof, special-purpose hardware devices to which cryptographic operations could be outsourced locally may solve the problem in some scenarios. (The authors are grateful to Tadayoshi Kohno for pointing this out.)

An alternative is to have the repository consistently process all IP addresses before releasing the data, *e.g.*, by applying HMAC with a key which is known only to the repository itself. This would make the repository the single point of failure, and increase its liability. It would also require a high level of trust between the contributors and the repository. We wish to avoid this in order to increase the pool of potential contributors.

To enable Internet-scale analysis centers to track propagation and topology of Internet threats, we believe it will be necessary to effectively *virtualize* the contributors' IP address space, so that reported data capture all topological relationships without revealing the actual IP addresses contained in the original data.

An important challenge is how to determine *which* topological relationships are essential for analysis and must be preserved, and which can or even should be removed from the virtual address space. It may appear at first glance that retaining as much topological structure as possible in the virtual address space is desirable, but this presents serious security risks. If the attacker succeeds in inverting even a small number of mappings between real and virtual addresses, in a virtual address space with rich topological structure this may open the door to complete de-anonymization. Understanding and quantifying this risk requires development of a rigorous, formal security model for address space virtualization.

The choice of topological features to be preserved by the sanitization algorithms should be heavily influenced by our knowledge of computer worms and self-propagating malware. In particular, sanitization should preserve the analysts' ability to recognize distinct propagation patterns (*e.g.*, distinguish random scanning and sequential propagation).

Research challenge 2: *Develop a formal security model and practical techniques for IP address virtualization that preserves some topological relationships between IP addresses without revealing the contributors' true addresses.*

Sanitization techniques can and should exploit the difference between the objectives and incentives of attackers and honest contributors. The goal of legitimate Internet-scale collaborative analysis is to discover *global* trends and inflection points, while the goal of the adversary is to pinpoint vulnerabilities within a specific *local* system or network. Therefore, we would like each contributor to release locally collected data *if and only if* other collaborative partners have observed the same or similar events. Ideally, this should be achieved with minimal involvement of a global coordinating authority.

Research challenge 3: *Design efficient distributed protocols and similarity metrics for network security data to ensure that each contributor only reveals the data if a threshold number of other participants are ready to reveal similar data.*

3. SANITIZATION TECHNOLOGIES FOR DATA REPOSITORIES

Network security data stored inside global repositories are arguably *more* vulnerable than those stored at the contributors' local sites. In open-access repositories, the attacker may browse and analyze the data at will, looking for evidence of his own and other attacks and actively discovering vulnerabilities such as obsolete intrusion detection systems, old versions of network services, and so on. Even restricted-access repositories are vulnerable, because they are bound to attract malicious attention and become the single point of failure of the system. Finally, it is difficult to prevent malicious insiders with legitimate access to the data from abusing their access privileges.

For the purposes of protecting network security datasets stored within the repository, we might as well assume that the repository is completely controlled by the attacker. Defense methods that are robust even under this assumption will also defeat more restricted attacks.

3.1 Understanding and defeating fingerprinting attacks

Fingerprinting attacks effectively enable the attacker to recognize the "signature" of a particular site within the data related to that site. This attack can be passive (*i.e.*, the attacker simply observes the site's unique natural characteristics, such as clock skew [23]), or active (*i.e.*, the attacker probes the system and actively induces a particular attack signature with the goal of eventually recognizing the victim's response in the dataset reported by multiple sites). More generally, the objective of the fingerprinting attack is to uncover the identity of an object within a sanitized dataset by associating the object's attributes to actions that the adversary has knowledge of or control over. More speculatively, probe-response methods have been considered as a method for automatically generating IDS-evading attacks [46].

Effectiveness of fingerprinting and probe response attacks has been shown for TCP traces [23] as well as the data reported by network security sensors [5]. We expect that fingerprinting attacks based on unique event sequences, patterns of intrusion alert production, triggering of rare intrusion detection rules, and so on will prove devastating for naive data collection and sanitization schemes. To the best of our knowledge, there have been no attempts to rigorously define fingerprinting attacks as a class, nor to design sanitization schemes that are provably secure against fingerprinting.

Research challenge 4: *Rigorously formalize fingerprinting attacks and design sanitization schemes for network security data that are provably secure against fingerprinting.*

3.2 Privacy-preserving data mining

Restricting access to repositories containing network security data may provide some protection for data contributors *if and only if* (a) the repository itself is trusted (which may or may not be realistic, depending on the deployment scenario), and (b) the repository manager takes active measures to protect the data contained within, while making them available in some form for collaborative analysis.

Privacy-preserving data mining has been a subject of very intensive research, so we limit our attention to a few common approaches, focusing in particular on the data mining and learning tasks that are most relevant in the context of collaborative analysis of Internet threats.

3.2.1 Non-interactive data mining

In non-interactive data mining, the dataset is sanitized and then released to the users, who access it locally in any way they want. Sanitization may involve statistical randomization of the data [1, 18, 10], which enables users to compute certain statistical properties of the original dataset while preserving privacy of individual data entries. An alternative is provably secure database obfuscation [33], which restricts the types of *queries* that users can feasibly evaluate on the sanitized dataset. The latter can be viewed as a form of uncircumventable access control that does *not* rely on tamper-proof enforcement software or hardware. The class of access control policies that can be enforced in this way, however, is relatively small.

There has been very little research to date on rigorously defining which functions must be efficiently computable (and to what degree of precision) on the sanitized datasets in order to enable common forms of collaborative security analysis. For example, it is clear that classifying network traffic, extracting malware signatures and tracking propagation of attacks through the Internet are among the most critical tasks of collaborative analysis. Yet, to the best of our knowledge, there has been no research on adapting privacy-preserving data mining algorithms to support evaluation of functions that are relevant for these tasks.

Research challenge 5: *Design and implement efficient privacy-preserving data mining algorithms that enable traffic classification, signature extraction, and propagation analysis on sanitized data without revealing the values of individual dataset entries.*

3.2.2 Interactive data mining

Privacy-preserving data mining can also take place as an *interactive* protocol. Instead of publicly releasing a sanitized dataset, the repository accepts queries from users and, to ensure that no sensitive information is revealed, *audits* each query [22, 21]. Auditing queries, however, is unusually difficult in the context of large-scale network security repositories. It is not always feasible to tell the difference between a legitimate security researcher and a malicious user masquerading as a researcher, especially if the only difference between their queries is *intent*. Simplistic auditing algorithms such as geographic discrimination (*e.g.*, only IP addresses located in the U.S. are assumed to be trustworthy) are unlikely to provide meaningful security.

Another approach is to randomize the response to each query in order to hide individual data entries, but enable computation of some global properties. This approach is subject to fundamental limitations [16], but there have been proposed practical techniques that satisfy a rigorous definition of privacy while supporting non-trivial learning and data mining algorithms [6]. To apply these techniques, it is necessary to investigate which functions other than simple statistical calculations and machine learning algorithms such ID3 and Perceptron must be supported in order to enable

collaborative analysis, detection, and tracking of Internet threats.

Another form of randomization was suggested by Michael Collins in order to provide protection against probe-response and fingerprinting attacks. When the user requests a subset of the dataset, the repository would randomly release *some* subset of the requested size. If the database is very large relative to the subset, it is unlikely that the data contributed by the probed system will be in the response. This approach requires that the repository be able to recognize repeated queries by the same user or a group of colluding users, which is a difficult challenge.

Other methods for sanitizing network security data include random alert sampling and/or suppression of rare data attributes. For example, the repository may only disclose records that have some information in common if the total number of such records exceeds a given threshold. (The threshold may be randomized to prevent flushing attacks — see section 3.3). Finally, the repository may add synthetic or artificial records to the datasets in order to introduce uncertainty into the attacker's analysis of the data.

None of these method provide cryptographically strong security guarantees. The objective, however, is worthwhile: to introduce uncertainty to the attacker's observations of the data repository and make it difficult to determine whether the absence of a particular fingerprint is due to the lack of collaborative detection, sampling percentage, selective or threshold-based event filters, or the repository's distribution policy, all of which can be controlled and dynamically adjusted by the repository manager and/or data contributors.

The absence of cryptographically strong security should not discourage attempts to quantitatively measure the degree of protection accorded by various sanitization technologies. The metrics should focus on the attacker workfactor needed to stage a successful attack, *e.g.*, the number of probes that must be launched before the response can be recognized in the reported data, number of addresses scanned, number of packets generated, and so on.

The resulting workfactor estimates should be explicitly conditioned on the attacker's knowledge. They must clearly specify trust assumptions underlying security of sanitization. For example, prefix-preserving hashing is useful only if all participants use the same key (to ensure consistent hashing of data originating from different participants) and secure only under the assumption that this common key is *not* known to the attacker.

Research challenge 6: *Develop quantitative metrics for estimating attacker workfactor for different data sanitization and protection technologies. Explicitly condition workfactor estimates on trust assumptions about components of the distributed data collection system.*

Taking into account the expected *value* of protected information may help in developing sanitization metrics and technologies that provide an adequate level of protection at a reasonable cost (this is similar to the "pay-for-privacy" approaches). For example, if the value of a single IP address is relatively low, then even weak sanitization that can be inverted at a small computational cost may be adequate because the attacker will not be willing to invest significant computational resources into de-anonymizing millions of addresses. Charging a small amount for each access to the repository is another possibility, although a feasible economic model for this is yet to be developed.

Time can be considered as another form of payment. Arguably, network data become less sensitive with time, as old data are less likely to reflect the current security posture of the contributing systems. Therefore, delaying data release by several months or even years may solve many privacy problems considered in this paper. The resulting data may still be useful for security research, but global analysis centers could no longer be used for real-time detection of emerging Internet threats.

3.3 Preventing data poisoning and enforcing accountability

Protecting sources of data and identities of data contributors (discussed further in section 4) is inherently in conflict with preserving utility of the collected data. A completely anonymous system can be abused in many ways. For example, the attacker may stage a blending attack [39] by submitting a large number of fake records that contain some information in common with some record of interest that may or may not be contained in the dataset. The attacker's hope is that these records will be released together (*e.g.*, because the number of records sharing this information exceeds the threshold) and he will then be able to recognize the target record in the released set.

Attackers may also stage a denial of service attack by flooding the repository with spam and fake records, poisoning the dataset and rendering it unusable. Even more seriously, attackers may deliberately inject "plausible-looking" data intended to mislead intrusion detection and worm fingerprinting algorithms [38, 11, 34].

Combining anonymity with accountability is a difficult task. In the context of network security data collection, one possible solution involves a registration phase, during which each contributor is issued an *anonymous credential* or a cryptographic key that enables him to compute a digital group signature on his messages.

Anonymous credentials (*e.g.*, [8]), group signatures and group authorization mechanisms have been a subject of very intensive research — see bibliographies in [27, 50]. The particular flavor of group credentials that is relevant in our context should enable the repository to verify that a given contributor is an authorized member of the group (*i.e.*, he successfully passed through the registration protocol at some point in the past), yet his identity remains anonymous (up to the entire set of group members). Obviously, availability of revocation mechanisms is extremely important, in case one of the contributors is compromised.

Research challenge 7: *Investigate how anonymous credential and blind authorization schemes may be applied to prevent spam and enforce accountability in large-scale network security data collection.*

4. ANONYMOUS DATA DELIVERY

We expect that many of the voluntary contributors to global data analysis centers will be interested in protecting their identities even from the centers themselves. There are several reasons for this: (i) probe response and fingerprinting attacks become more difficult if the source of the data is hidden; (ii) anonymous data are more secure against insider attacks; (iii) even direct compromise of data repositories will not necessarily enable attackers to link data records with

their creators.

To support broad participation in data collection efforts, the data delivery infrastructure must provide (perhaps optionally) *anonymous message delivery* mechanisms for transmitting the data from contributors to data repositories. Obviously, standard Internet protocols reveal source IP addresses and thus do not provide anonymity even against passive eavesdroppers. Therefore, we envision anonymous data delivery technologies that will "piggyback" on existing mix-based anonymity networks.

A detailed survey of anonymity systems is beyond the scope of this paper (a bibliography can be found at `http://www.freehaven.net/anonbib/`). Mix networks are an especially popular class of systems which provide a practical way of enabling *unlinkable* communications on public networks. A *mix*, first proposed by Chaum [9], can be thought of as a server that accepts incoming connections and forwards them in such a way that an eavesdropper cannot easily determine which outgoing connection corresponds to which incoming connection. To protect message sources even when some of the mixes in the network are compromised, messages are typically routed through a mix chain.

Since real-time detection of Internet threats is one of the envisioned applications of global analysis centers, we are especially interested in *low-latency* anonymity networks such as Tor [15], JAP [4], and mix rings [7]. Unfortunately, low-latency networks tend to be extremely vulnerable to *traffic analysis* based on correlating packet stream characteristics and/or message dispatch and arrival times [19, 25]. An attacker who controls both the data repository and public network links in the vicinity of the sensor generating the data can easily collect traffic observations required for traffic analysis and completely de-anonymize messages received at the repository. Devastating timing attacks have been successfully demonstrated in real-world mix networks [31, 35]. Other traffic analysis attacks on mix networks can be found in [44, 3, 2, 40].

An alternative solution may be provided by peer-to-peer anonymous publishing systems such as Freenet [12]. The primary security objectives of such systems are availability and censorship resistance. It is not clear whether they can be easily adapted to achieve accountability and low distribution latency required for aggregation of information security alerts.

Research challenge 8: *Research new models and implementations for anonymous data delivery networks that provide low- or mid-latency guarantees as well as resistance to traffic analysis attacks even when connection endpoints are directly observable by the attacker.*

Of course, anonymity must go hand in hand with *accountability*. As described in section 3.3, group authorization and anonymous credential mechanisms may need to be deployed to prevent attackers from dumping garbage data into the network. Another promising direction involves reputation systems, which must be combined with identity protection [14, 49].

Research challenge 9: *Investigate applications of reputation systems for ensuring quality of network security data collected anonymously from a broad pool of contributors.*

Finally, we observe that the anonymous communication channels between contributors and repositories should be *bi-directional*. For example, after a certain IP address hash has been identified as suspicious, the repository may inform the contributors, who – with their knowledge of the original, unsanitized address – may take appropriate defense measures. Similarly, many types of global network phenomena can be understood fully only if the raw data are available; therefore, we expect that the results of correlation analysis will be frequently propagated back to contributors.

Another potentially useful feature that is not supported by the existing repositories is the ability of contributors to anonymously verify that their submissions have been received and processed correctly. Techniques borrowed from electronic voting schemes may prove useful in this context.

5. CONCLUSIONS

In recent years, as Internet attacks increased in scale, frequency, and severity, there has been a growing interest in creating global analysis centers that would gather network security data from a wide variety of network sensors, use it for real-time collaborative analysis to detect inflection points and global security trends, identify propagation patterns and attack vectors of malware, and make the data available for network security researchers.

Successful deployment of global analysis centers will require resolving a number of fundamental tradeoffs between increased global network security, privacy of data contributors, potential for malicious abuse of the reported data, liability of data repositories, usefulness of the data for network security research, and practical efficiency. This position paper outlines several specific research challenges in this area. They vary from rigorous formalization of fingerprinting attacks to better understanding of traffic analysis attacks which de-anonymize the data contributed to global analysis centers. We hope that our challenges will become part of the research program for computer scientists working in this area. It is unlikely that global Internet defense will succeed without solving them.

6. REFERENCES

[1] Agrawal, R., and Srikant, R. Privacy-preserving data mining. In *Proc. ACM SIGMOD International Conference on Management of Data* (2000), pp. 439–450.

[2] Back, A., Goldberg, I., and Shostack, A. Freedom Systems 2.1 security issues and analysis. `http://www.freehaven.net/anonbib/cache/freedom21-security.pdf`, May 2001.

[3] Back, A., Möller, U., and Stiglic, A. Traffic analysis attacks and trade-offs in anonymity providing systems. In *Proc. 4th International Workshop on Information Hiding* (2001), vol. 2137 of *LNCS*, pp. 245–257.

[4] Berthold, O., Federrath, H., and Köpsell, S. Web MIXes: a system for anonymous and unobservable Internet access. In *Proc. Workshop on Design Issues in Anonymity and Unobservability* (2000), pp. 115–129.

[5] Bethencourt, J., Franklin, J., and Vernon, M. Mapping Internet sensors with probe response attacks. In *Proc. 14th USENIX Security Symposium* (2005), pp. 193–208.

[6] Blum, A., Dwork, C., McSherry, F., and Nissim, K. Practical privacy: the SuLQ framework. In *Proc.*

24th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems (PODS) (2005), pp. 128–138.

[7] Burnside, M., and Keromytis, A. Low latency anonymity with mix rings. In *Proc. 9th International Information Security Conference (ISC)* (2006), pp. 32–45.

[8] Camenisch, J., and Lysyanskaya, A. An efficient system for non-transferable anonymous credentials with optional anonymity revocation. In *Proc. Advances in Cryptology – EUROCRYPT* (2001), pp. 93–118.

[9] Chaum, D. Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of the ACM 24*, 2 (1981), 84–88.

[10] Chawla, S., Dwork, C., McSherry, F., Smith, A., and Wee, H. Towards privacy in public databases. In *Proc. 2nd Theory of Cryptography Conference (TCC)* (2005), pp. 363–385.

[11] Chung, S., and Mok, A. Allergy attack against automatic signature generation. In *Proc. Recent Advances in Intrusion Detection: 9th International Symposium (RAID)* (2006), pp. 61–80.

[12] Clarke, I., Sandberg, O., Wiley, B., and Hong, T. Freenet: A distributed anonymous information storage and retrieval system. In *Proc. International Workshop on Design Issues in Anonymity and Unobservability* (2001), vol. 2009 of *LNCS*, Springer-Verlag, pp. 46–66.

[13] Debar, H., and Wespi, A. Aggregation and correlation of intrusion-detection alerts. In *Proc. Recent Advances in Intrusion Detection: 4th International Symposium (RAID)* (2001), pp. 85–103.

[14] Dingledine, R., Mathewson, N., and Syverson, P. Reputation in P2P anonymity systems. In *Proc. Workshop on Economics of Peer-to-Peer Systems* (2003).

[15] Dingledine, R., Mathewson, N., and Syverson, P. Tor: the second-generation onion router. In *Proc. 13th USENIX Security Symposium* (2004), pp. 303–320.

[16] Dinur, I., and Nissim, K. Revealing information while preserving privacy. In *Proc. 22nd ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems (PODS)* (2003), pp. 202–210.

[17] DShield. `http://www.dshield.org`, 2006.

[18] Evfimievski, A., Gehrke, J., and Srikant, R. Limiting privacy breaches in privacy-preserving data mining. In *Proc. 22nd ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems (PODS)* (2003), pp. 211–222.

[19] Fu, X., Graham, B., Bettati, R., and Zhao, W. On effectiveness of link padding for statistical traffic analysis attacks. In *Proc. 23rd IEEE Conference on Distributed Computing Systems* (2003), pp. 340–349.

[20] Gates, C., Collins, M., Duggan, M., Kompanek, A., and Thomas, M. More Netflow tools: For performance and security. In *Proc. 18th Conference on Systems Administration (LISA)* (2004), pp. 121–132.

[21] Kenthapadi, K., Mishra, N., and Nissim, K. Simulatable auditing. In *Proc. 24th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems (PODS)* (2005), pp. 118–127.

[22] Kleinberg, J., Papadimitriou, C., and Raghavan, P. Auditing Boolean attributes. *J. Comput. Syst. Sci. 66*, 1 (2003), 244–253.

[23] Kohno, T., Broido, A., and Claffy, K. Remote physical device fingerprinting. In *Proc. IEEE Symposium on Security and Privacy* (2005), pp. 211–225.

[24] Krishnamurthy, B., Sen, S., Zhang, Y., and Chen, Y. Sketch-based change detection: methods, evaluation, and applications. In *Proc. 3rd ACM SIGCOMM Conference on Internet Measurement* (2003), pp. 235–247.

[25] Levine, B., Reiter, M., Wang, C., and Wright, M. Timing attacks in low-latency mix systems. In *Proc. 8th International Conference on Financial Cryptography* (2004), pp. 251–265.

[26] Lincoln, P., Porras, P., and Shmatikov, V. Privacy-preserving sharing and correlation of security alerts. In *Proc. 13th USENIX Security Symposium* (2004), pp. 239–254.

[27] Lipmaa, H. Group signature schemes. `http://www.cs.ut.ee/~lipmaa/crypto/link/signature/group.php`, 2006.

[28] Locasto, M., Parekh, J., Keromytis, A., and Stolfo, S. Towards collaborative security and P2P intrusion detection. In *Proc. IEEE Information Assurance Workshop* (2005), pp. 333–339.

[29] Maltz, D., Zhan, J., Xie, G., Zhang, H., Hjálmtýsson, G., Greenberg, A., and Rexford, J. Structure preserving anonymization of router configuration data. In *Proc. 4th ACM SIGCOMM Conference on Internet Measurement* (2004), pp. 239–244.

[30] McHugh, J., and Gates, C. Locality: a new paradigm for thinking about normal behavior and outsider threat. In *Proc. New Security Paradigms Workshop* (2003), pp. 3–10.

[31] Murdoch, S., and Danezis, G. Low-cost traffic analysis of Tor. In *Proc. IEEE Symposium on Security and Privacy* (2005), pp. 183–195.

[32] myNetWatchman. `http://www.mynetwatchman.com`, 2006.

[33] Narayanan, A., and Shmatikov, V. Obfuscated databases and group privacy. In *Proc. 12th ACM Conference on Computer and Communications Security (CCS)* (2005), pp. 102–111.

[34] Newsome, J., Karp, B., and Song, D. Paragraph: Thwarting signature learning by training maliciously. In *Proc. Recent Advances in Intrusion Detection: 9th International Symposium (RAID)* (2006), pp. 81–105.

[35] Øverlier, L., and Syverson, P. Locating hidden servers. In *Proc. IEEE Symposium on Security and Privacy* (2006), pp. 100–114.

[36] Pang, R., Allman, M., Paxson, V., and Lee, J. The devil and packet trace anonymization. *ACM SIGCOMM Computer Communication Review 36*, 1 (2006), 29–38.

[37] Pang, R., and Paxson, V. A high-level programming environment for packet trace anonymization and transformation. In *Proc. ACM*

SIGCOMM (2003), pp. 339–351.

[38] Perdisci, R., Dagon, D., Lee, W., Fogla, P., and Sharif, M. Misleading worm signature generators using deliberate noise injection. In *Proc. IEEE Symposium on Security and Privacy* (2006), pp. 17–31.

[39] Serjantov, A., Dingledine, R., and Syverson, P. From a trickle to flood: active attacks on several mix types. In *Proc. 5th International Workshop on Information Hiding* (2002), pp. 36–52.

[40] Serjantov, A., and Sewell, P. Passive attack analysis for connection-based anonymity systems. In *Proc. 8th European Symposium on Research in Computer Security* (2003), vol. 2808 of *LNCS*, pp. 116–131.

[41] Slagell, A., and Yurcik, W. Sharing computer network logs for security and privacy: a motivation for new methodologies of anonymization. In *Proc. SECOVAL: The Workshop on the Value of Security through Collaboration* (2005).

[42] Spitzner, L. Know your enemy: Honeynets. `http://project.honeynet.org/papers/honeynet`, 2005.

[43] Symantec. DeepSight threat management system. `http://tms.symantec.com`, 2006.

[44] Syverson, P., Tsudik, G., Reed, M., and Landwehr, C. Towards an analysis of onion routing security. In *Proc. Workshop on Design Issues in Anonymity and Unobservability* (2000), vol. 2009 of *LNCS*, pp. 96–114.

[45] tcpdpriv. Program for eliminating confidential information from traces. `http://ita.ee.lbl.gov/html/contrib/tcpdpriv.html`, 2006.

[46] Templeton, S., and Levitt, K. A requires/provides model for computer attacks. In *Proc. New Security Paradigms Workshop* (2001), pp. 31–38.

[47] Valdes, A., Fong, M., and Skinner, K. Data cube indexing of large-scale Infosec repositories. In *Proc. Australian Computer Emergency Response Team Conference* (2006).

[48] Valdes, A., and Skinner, K. Probabilistic alert correlation. In *Proc. Recent Advances in Intrusion Detection: 4th International Symposium (RAID)* (2001), pp. 54–68.

[49] Walsh, K., and Sirer, E. G. Experience with an object reputation system for peer-to-peer filesharing. In *Proc. 3rd Symposium on Networked Systems Design and Implementation (NSDI)* (2006).

[50] Wang, G. Bibliography on group-oriented signatures. `http://www.i2r.a-star.edu.sg/icsd/staff/guilin/bible/group-oriented.htm`, 2006.

[51] Xu, D., and Ning, P. Privacy-preserving alert correlation: a concept hierarchy based approach. In *Proc. 21st Annual Computer Security Applications Conference (ACSAC)* (2005), pp. 537–546.

[52] Xu, J., Fan, J., Ammar, M., and Moon, S. On the design and performance of prefix-preserving IP traffic trace anonymization. In *Proc. 1st ACM SIGCOMM Workshop on Internet Measurement* (2001), pp. 263–266.

[53] Yegneswaran, V., Barford, P., and Plonka, D. On the design and use of Internet sinks for network abuse monitoring. In *Proc. Recent Advances in Intrusion Detection: 7th International Symposium (RAID)* (2004), pp. 146–165.

[54] Yegneswaran, V., Barford, P., and Ullrich, J. Internet intrusions: global characteristics and prevalence. In *Proc. ACM SIGMETRICS* (2003), pp. 138–147.