[1] P. Lincoln, P. A. Porras, and V. Shmatikov. Privacy-preserving sharing and correlation of security alerts. In *Proceedings of the 13th USENIX Security Symposium*, pages 239-254, 2004. [bib | .html]

We present a practical scheme for Internet-scale collaborative analysis of information security threats which provides strong privacy guarantees to contributors of alerts. Wide-area analysis centers are proving a valuable early warning service against worms, viruses, and other malicious activities. At the same time, protecting individual and organizational privacy is no longer optional in today's business climate. We propose a set of data sanitization techniques that enable community alert aggregation and correlation, while maintaining privacy for alert contributors. Our approach is practical, scalable, does not rely on trusted third parties or secure multiparty computation schemes, and does not require sophisticated key management.

[2] U. Lindqvist and P. A. Porras. Detecting computer and network misuse through the production-based expert system toolset (P-BEST). In *Proceedings of the IEEE Symposium on Security and Privacy*, pages 146-161, 1999. [bib | DOI]

This paper describes an expert system development tool set called the Production-Based Expert System Toolset (P-BEST) and how it is employed in the development of a modern generic signature-analysis engine for computer and network misuse detection. For more than a decade, earlier versions of P-BEST have been used in intrusion detection research and in the development of some of the most well-known intrusion detection systems, but this is the first time the principles and language of P-BEST are described to a wide audience. We present rule sets for detecting subversion methods against which there are few defenses- specifically, SYN flooding and buffer overruns-and provide performance measurements. Together, these examples and performance measurements indicate that P-BEST-based expert systems are well suited for real-time misuse detection in contemporary computing environments. In addition, the simplicity of the P-BEST language and its close integration with the C programming language makes it easy to use while still being very powerful and flexible.

[3] P. A. Porras, M. W. Fong, and A. Valdes. A mission-impact-based approach to INFOSEC alarm correlation. In *Lecture Notes in Computer Science, Proceedings of Recent Advances in Intrusion Detection (RAID)*, volume 2516, pages 95-114, October 2002. [bib | http]

We describe a mission-impact-based approach to the analysis of security alerts produced by spatially distributed heterogeneous information security (INFOSEC) devices, such as firewalls, intrusion detection systems, authentication services, and antivirus software. The intent of this work is to deliver an automated capability to reduce the time and cost of managing multiple INFOSEC devices through a strategy of topology analysis, alert prioritization, and common at-tribute-based alert aggregation. Our efforts to date have led to the development of a prototype system called the Mission Impact Intrusion Report Correlation Sys-tem, or M-Correlator. M-Correlator is intended to provide analysts (at all experience levels) a powerful capability to automatically fuse together and isolate those INFOSEC alerts that represent the greatest threat to the health and security of their networks.

Keywords: Network security, intrusion report correlation, alert management, alert prioritization

[4] P. A. Porras and P. G. Neumann. EMERALD: Event monitoring enabling responses to anomalous live disturbances. In *Proceedings of the 1997 National Information Systems Security Conference (NISSC)*, pages 353-365, October 1997. [bib]

The EMERALD (Event Monitoring Enabling Responses to Anomalous Live Disturbances) environment is a distributed scalable tool suite for tracking malicious activity through and across large networks. EMERALD introduces a highly distributed, building-block approach to network surveillance, attack isolation, and automated response. It combines models from research in distributed high-volume event-correlation methodologies with over a decade of intrusion detection research and engineering experience. The approach is novel in its use of highly distributed, independently tunable, surveillance and response monitors that are deployable polymorphically at various abstract layers in a large network. These monitors contribute to a streamlined event-analysis system that combines signature analysis with statistical profiling to provide localized real-time protection of the most widely used network services on the Internet. Equally important, EMERALD introduces a recursive framework for

coordinating the dissemination of analyses from the distributed monitors to provide a global detection and response capability that can counter attacks occurring across an entire network enterprise. Further, EMERALD introduces a versatile application programmers' interface that enhances its ability to integrate with heterogeneous target hosts and provides a high degree of interoperability with third-party tool suites.

[5] A. Valdes and M. Fong. Scalable visualization of propagating internet phenomena. In *Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security (VizSEC/DMSEC)*, pages 124-127, 2004. [bib | DOI]

The Internet has recently been impacted by a number of large distributed attacks that achieve exponential growth through self-propagation. Some of these attacks have exploited vulnerabilities for which advisories had been issued and for which patches and detection signatures were available. It is increasingly apparent, however, that such prevention and detection mechanisms are inadequate, and that the attacker's time to exploit is shrinking relative to the defender's ability to learn of a new attack and patch systems or update intrusion detection signatures. We introduce visual, scalable techniques to detect phenomena such as distributed denial-of-service attacks and worms. It is hoped that these new approaches will enable detection of such events at an early stage and enable local response actions even before the publication of advisories about a new vulnerability and the availability of patches.

[6] A. Valdes and M. Fong. Data cube indexing of large-scale infosec repositories. In *Proceedings of the AusCERT Asia Pacific Information Technology Security Conference*, May 2006. (Best Paper). [bib]

Analysts examining large-scale infosec repositories for propagating network events are interested in quickly identifying temporal and spatial (IP address and/or port) regions containing interesting phenomena, or correlating events from different time periods. The size of these datasets strains current query capabilities provided by, for example, relational databases. We introduce a scalable, animated data cube representation and viewer, suitable for a broad range of observables, to permit coarse-grain detection and correlation in such data sets. We scale from the LAN to the Internet through flexible, locality-preserving hash algorithms mapping traffic source and destination (IP addresses or IP and port considered simultaneously). Data streams considered include inherently suspicious traffic such as packets rejected at a firewall, IDS alerts, or traffic to unused address space, as well as Netflow data. We display observables as intensity plots, where X and Y coordinates are the hashed source and target address and the intensity is proportional to traffic volume. Source and target address space may or may not be the same and may or may not be mapped the same way. Propagating events have distinct visual signatures that can be enhanced through matched filtering techniques. Future work will correlate cubes efficiently through cell-by-cell multiplication. An analyst will be able to, for example, examine whether plots representing two time periods (hours or days) exhibit similar patterns. Multiplication of a cube with its transpose permits identification of nodes that respond to potentially malicious probes. These data cubes permit coarse-grained detection and correlation without expensive data base queries.
