

Distributed Intrusion Detection and Attack Containment for Organizational Cyber Security

Stephen G. Batsell¹, Nageswara S. Rao², Mallikarjun Shankar¹

¹ Computational Sciences and Engineering Division

² Computer Science & Mathematics Division

Abstract

The sheer number and sophistication of cyber attacks are making our nation's critical computer networks increasingly vulnerable. At the same time, these networks are being called upon to play a key role in processing, data storage, monitoring and control of critical infrastructures such as energy, transportation, and finance. Disruption of these networks can have highly damaging affects on our Nation. Current cyber security systems are not capable of protecting from all attacks nor providing near real-time response. Host-based intrusion detection systems are not sufficient to protect these networks due to the sheer volume, distributed nature of data, and real-time response requirements. Further they only detect known attacks. We developed an integrated cyber security framework for identifying and containing attacks within an organizational network domain. This framework is distributed, autonomous, and capable of detecting new attacks. It integrates existing cyber security systems and provides a single picture of the entire network, which allows real-time situational awareness of large scale network systems. It consists of individual components for host-level anomaly detection, attack source localization, and attack containment.

Introduction

Throughout the 1990's the rise of commercial interest in the Internet has lead to the integration of the information infrastructure as a core component of the United States economy. However, an increasing number of cyber attacks and threats of cyber attacks on our national networks have shown that our energy, transportation and finance infrastructures are vulnerable with potentially dire consequences. While a significant fraction of these attacks have been ineffective, the cyberspace has become an arena for warfare and acts of terrorism, since it controls various critical infrastructures. Protecting these infrastructures has become a critical and key area of interest for homeland defense.

Current cyber security capabilities have evolved largely as patches and add-ons to the Internet, which was designed on the principles of open communication and implicit mutual trust. It is now recognized that it is no longer sufficient to follow such evolutionary paths and that security must be an integral part of the information infrastructure. Existing intrusion detection systems have evolved as separate ad hoc capabilities and are not sufficient for responding to sophisticated and disguised cyber attacks expected from well-funded terrorist organizations. This created an opportunity to develop a new direction in large-scale and integrated intrusion detection and response systems, which is the main motivation for this project.

Technical Approach

We developed an integrated cyber security framework for identifying and containing cyber attacks at the level of an organizational network domain. This framework consists of three components: *intrusion detection*, *attack source localization*, and *attack containment*. For the first and third components, we utilized the existing methods as well as developed several new components. In particular, we developed novel information fusion methods for host-level anomaly detection and also for network-level diagnosis and attack source identification.

Our integrated framework for intrusion detection and containment provides for scalability by allowing multiple sensor engines to operate in parallel. Single network sensors can only handle small traffic loads and are often limited in their functionality. Only by operating in a distributed fashion can a large-scale approach be successful. This new approach operates in an autonomous manner that allows near real-time response to events and to ensure that the latest signature updates are available to the sensors as soon as they exist. Current intrusion detection systems are not autonomous and rely on human intervention as a key part of their operation, which makes them orders of magnitude slower than needed. Direct human-in-the-loop operation cannot always effectively counter the newer cyber attacks, particularly, at high network speeds. By using an autonomous and distributed framework shown in Figure 1 the sensor outputs from multiple parts of the administrative network domain can be rapidly correlated.

Our framework is capable of addressing the routers to set packet filters and the firewalls to block specific ports. Together they form an active response that is activated by the source isolation component. For any suspected attack, its signature is obtained by the detection module together with the physical paths leading to the regions of the attack source. This component activates the filters along the physical paths from the attack source to deny passage rights to the attack packets. Thus the extent of the attack's reach is contained. We investigated two classes of attack containment methods. The first method is suited for attacks that generate low levels of traffic such as unauthorized logins. Here the fusers can readily exchange data with sensors and activate the firewalls closest to the source to filter out the packets from the attack machine. This method, however, does not work in the case of attacks that generate high traffic such as denial of service attacks. To handle these cases, in this framework the fuser expands the rate controls gradually from the nearby filters to farther ones.

Propagative attacks constitute a growing subclass of cyber intrusions, which rely on steadily compromising hosts and using them as launch pads to attack other hosts. Certain types of worms (e.g., Code Red II) that perpetuate by spreading from host to host belong to this subclass. Coordinated denial-of-service attacks that accumulate zombies into an arsenal of compromised hosts to activate them at a later point, and spam generators that utilize a suite of compromised hosts to send email floods, belong to this subclass. It is important to isolate the origin of attacks, which can potentially distinguish between insider and external attacks. The ability and speed with which such diagnosis can be performed depends on the precise nature of the attack and sensors that detect various attack symptoms. In this paper, we present an abstract framework that utilizes the propagative nature of these attacks to obtain efficient fault source isolation algorithms by utilizing the knowledge (when available) of the sensor activation times and attack propagation times.

The common characteristic of this subclass of cyber attacks of interest to us is that the attack propagates across the network by "infecting" one host or node after another. The other attack characteristics could vary significantly in the type of host compromise, strategy for choosing and attacking hosts, and the generated time and traffic scales, could vary widely. In some worm attacks, the goal is to propagate rapidly, often randomly, in order to infect as many hosts as possible [Weaver et al 2003, Shankar et al 2003]. This behavior typically results in the classical S-curve of the number of infected hosts: the rate of infection starts slowly during the initial phase, quickly becomes very high as the worm grows in strength, and then tapers off when most of the vulnerable targets are compromised. Zombies that are created for denial-of-service or spam attacks utilize a more deliberate approach of compromising hosts without generating high traffic levels, and typically spread more slowly. More intelligent worms like Nimda and Code Red II scan the local networks more frequently than they scan remote networks. The lack of knowledge of the enterprise's internal network addresses suggests that such worms would select a strategy of scanning and spreading systematically and not randomly within the enterprise intranet.

We investigated analytical and algorithmic aspects of diagnosing a generic class of propagative attacks that spread across enterprise networks by steadily compromising hosts and then using them to attack other hosts. Certain types of worms, and preparatory phases of coordinated denial-of-service and spam attacks belong to this class. Symptoms of such attacks are detected at the network sensors by packet signatures and traffic characteristics, and at the hosts by performance degradations and anomalous system behavior. We showed that information about worm propagation times and dynamic sensor activation times can be fused with the network structural information to: (a) isolate the regions of network that contain the original attack origin, and (b) predict the next set of target hosts. We developed the attack propagation graphs that capture the above three types of information, and solved the source isolation and forewarning problems using graph algorithms.

As the attack propagates, its symptoms are detected by the sensors located at the nodes, which could themselves vary in their capabilities and performance. Based on the locations and activation times of the sensors that detect an attack, we showed that source can be isolated within certain regions of the network. We considered two types of sensors deployed to detect the symptoms of cyber attacks, namely host and network sensors. Host sensors typically detect attacks by utilizing packet signatures, system misbehavior and performance degradations, and anomalous traffic levels to and from the host. Network sensors operate on the traffic streams within the vicinity of routers, switches and firewalls; they detect attacks by inspecting packet signatures as well as by observing anomaly patterns of individual and aggregate traffic streams. These two types of sensors could provide qualitatively different information, which is typically localized in either case. An enterprise network deploys a combination of host sensors and strategically located network sensors. We developed algorithms to combine the information from various sensors together with the structural connectivity information to isolate the regions that contain the attack origin. In particular, these methods decide if the attack originated outside or within the enterprise; in the former case, firewalls at gateway routers can be activated to drop the attack packets, and in the latter case, appropriate local firewalls can be activated to quarantine the sources. We also developed algorithms to predict the next set of potential target nodes based on the current sensor information so that local firewalls can be activated ahead of time to prevent the further propagation of attack.

The locations of compromised hosts together with the state of sensor activations provide the structural trajectory information about the attack to assist in diagnosis. The sensor activation times together with the estimated attack propagation times provide us the directional information about the attack propagation. We fused the structural and directional information to isolate regions of the network that contain the original attack source. Our methods are effective for attacks that spread deliberately and form the class of topological worms that typically operate in the intranet context as well as for worms that target hosts randomly across the Internet but originate inside the intranet. As is to be expected, the precision of isolation and forewarning depends on: (i) locations of host and network sensors, (ii) network connectivity, and (iii) strategy, propagation times, and sensor activation properties of the attack. In addition, the level of knowledge about each of these items can also have a significant impact both on the algorithms and their precision for isolation and forewarning. We developed propagation graph models that capture the properties (i) and (ii). Using the information about the properties in (iii), we derive a suitable subgraph that will be used both for isolation and forewarning. Such an approach, namely utilizing a propagative graph for diagnosis, has been utilized in process plants [Ira et al 1985], dynamical systems [Rao and Viswanadham 1987] and optical networks [Mas and Thiran 2000]. While these systems are quite different from computer networks, they all share certain foundational properties that make it possible to solve origin isolation and forewarning problems. We extended and adapted the methods developed for graph-based systems [Rao 1993a, 1993b] to propagative cyber attacks. These extensions included identifying and defining the relevant properties of computer networks and cyber attacks in the form of a propagation graph, and then utilizing the appropriate graph algorithms.

A large number of intrusions such as port scans, login attempts, and buffer overflow attacks can be detected at the hosts by matching the headers and contents of network packets with known signatures. These techniques are fairly mature and are available as freeware such as snort, and formed some components of our architecture. While these methods detect known attacks, another key issue in intrusion detection today is the ability to detect new attacks. The principal methodology to accomplish this is the identification of anomalies, namely aberrant deviations from normal behavior, that are hidden within a background of normal activity. Anomaly detection is crucial against new strategies, for which no known signature exists. We developed a method for detecting the programs running on the hosts with anomalous system calls; in particular we use histograms of system calls of a program as a signature. A detector is trained on-line on the host using known programs and a small number of attack programs. Such approach has been used previously based on the Basic System Module (BSM) data that contains the system calls made by a program. The methods based on k-nearest neighbor and support vector machines have been used with good success, but both these methods left residual prediction errors.

We developed an information fusion based approach to train several neural network detectors, wherein these multiple detectors are fused together with a nearest neighbor rule to generate the final answer. Such methods are promising in that they can be shown to perform at least as good as the best among the detectors fused. In fact, a fundamental result in detector theory states that there is no single best detector but each performs well under different conditions. Our fusion approach achieves the best performance among the available detectors. In practice, however, the fuser has to be appropriately chosen to achieve such performance. We previously developed the nearest neighbor projective fusers that have been shown to outperform the individual detector. For the anomaly detection component of our system, we developed a user configuration on BSM data, which performed better than the earlier methods on DARPA benchmark testset. This configuration employed a linear fuser to first combine 10 sigmoid neural networks and a nearest neighbor rule. Then a meta-fuser based on nearest neighbor projective fusion method [Rao 2002] is deployed to combine the original detectors and the linear fuser. The resultant fused detector can be analytically shown to perform at least as well as the best combination of the detectors. This system achieved zero error on the DARPA benchmark dataset, which is the best performance for this dataset.

Results and Discussion

We developed a distributed and autonomous framework capable of quickly detecting existing and new attacks. It consists of individual components for network and host-level intrusion detection, attack source localization, and attack containment. The detection component is a combination of network and host-based sensors that utilize the sensor data together with the network information to identify the attacks. We developed information fusion method for detecting the host programs with anomalous system calls. We are the first to develop attack source isolation methods for propagative network attacks. The source localization component is activated by a suspected attack and locates the attack source(s) by tracing or reconstructing the physical paths of attack packets. The attack containment component utilizes firewalls and packet filters on various host and network locations to regulate or contain the packet flow from the attack source within the organizational network domain. An implementation of these containment modules will be pursued as a follow-on activity to this LDRD project. These modules together with the ones developed under this project will constitute a comprehensive cyber framework for organization-level security.

Benefits

This integrated framework and the specific unique components developed under this LDRD have significant benefits in for any government agency that requires a real-time situational awareness

of their network assets. This would include Department of Energy, Department of Defense, Department of Homeland Security, and Department of State. It allows near real-time detection and response to cyber attack. It further provides an integration platform for cyber defense components to form a single system.

References

- Ira, M., K. Aoki, E. Oshima, and H. Matsuyama. 1985. "An algorithm for diagnosis of system failures in the chemical processes", *Comp. Chem. Eng.* **3**, 489-493.
- Mas, C. and P. Thiran. 2001. "An efficient algorithm for locating soft and hard failures in WDM networks", *IEEE Journal on Selected Areas in Communications*. **18**(10), 1900-1911.
- Rao, N. S. V. 1993a. "Computational complexity issues in operative diagnosis of graph-based systems", *IEEE Transactions on Computers*. **42**(4), 447-457.
- Rao, N. S. V. 199b. "Expected-value analysis of two single fault diagnosis algorithms", *IEEE Transactions on Computers*. **42**(3), 272-280.
- Rao, N. S. V. 2002. "Nearest neighbor projective fuser for function estimation", *Proc. Int. Conf. Information Fusion*.
- Rao, N. S. V. and N. Viswanadham. 1987. "Fault diagnosis in dynamical systems: A graph theoretic approach", *Int. J. of Syst. Sci.* **18**(4), 687-695.
- Shankar, M., N. S. V. Rao, and S. Batsell. 2003. "Fusing intrusion data for detection and containment, *Proc. MILCOM*.
- S. Staniford. 2003. "Containment of Scanning Worms in Enterprise Networks", Technical report, Silicon Defense.
- Weaver N., V. Paxson, S. Staniford, and R. Cunningham. 2003. "A taxonomy of computer worms", *WORM*.

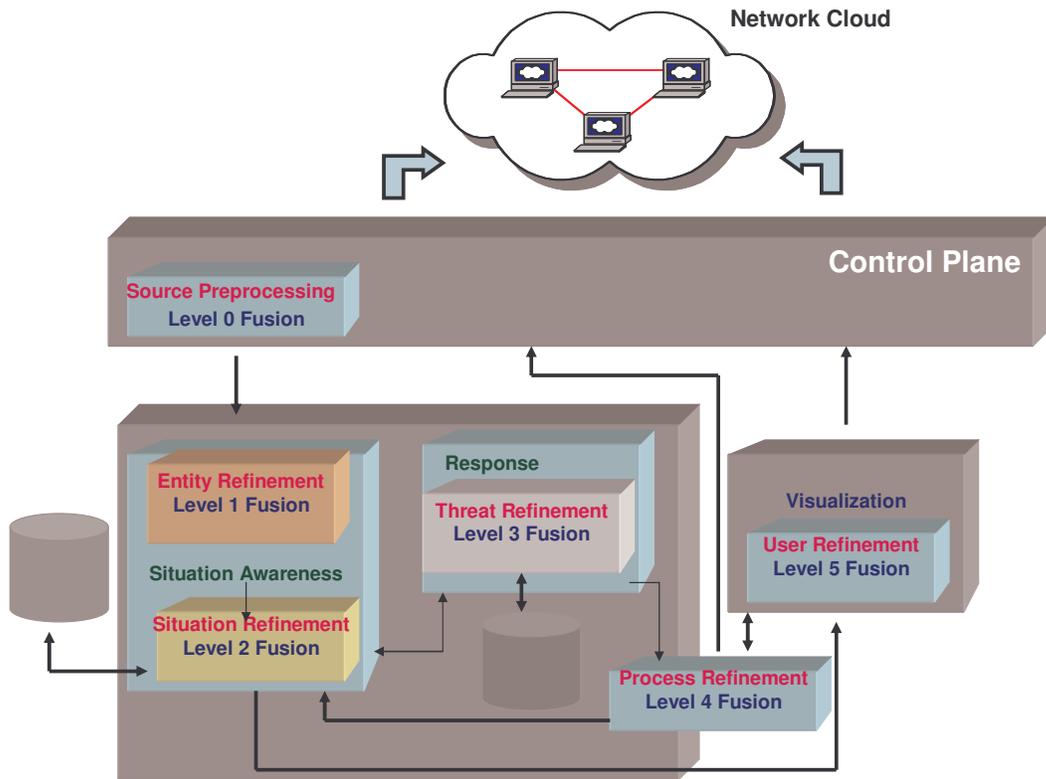


Figure 1. Distributed Framework for Integrated Cyber Security