A Survey of the High-Speed Self-Learning Intrusion Detection Research Area

Anna Sperotto and Remco van de Meent
University of Twente, The Netherlands
{a.sperotto, r.vandemeent}@utwente.nl

Abstract Intrusion detection for IP networks has been a research theme for a number of years already. One of the challenges is to keep up with the ever increasing Internet usage and network link speeds, as more and more data has to be scanned for intrusions. Another challenge is that it is hardly feasible to adapt the scanning configuration to new threats manually in a timely fashion, because of the possible rapid spread of new threats. This paper is the result of the first three months of a PhD research project in high speed, self-learning network intrusion detection systems. Here, we give an overview of the state of the art in this field, highlighting at the same time the major open issues.

1 Introduction

The continuously increasing number of users, as well as the growing popularity of on-line services, makes the Internet a common place for attacks and misuses. As a consequence, security in 'cyberspace' has become a high priority issue, to protect the end-users from malicious behavior and to provide a safer service. Network Intrusion Detection Systems (NIDSes) have become in the last years a useful way to monitor network traffic to detect signals of attacks. The spread of 1-10Gbps networks technology, the large amount of data and the increasing size of networks present new challenges to NIDS researchers, looking for adaptive and high speed solutions. This paper presents an overview of the state of the art in Intrusion Detection for high speed networks, with reference to the problem of adaptability and self-management. The paper is organized as follows: Section 2 and 3 describe the major trends in High Speed Networks IDSes and the adaptive approaches to the problem, respectively. Section 4, in the end, presents our conclusions and outlines some ideas for future work.

2 State of the Art in High Speed NIDSes

Quickly and thoroughly detecting malicious activity in a network has always been a major aim of NIDSes. This is still true now that gigabit networks are commonly used, with backbone networks of even far more bandwidth capacity. The number of end-users is still increasing, as well as the amount of on-line services. All this is attracting attackers and speeding up worms spread — the consequences result in an always growing damage. Research in this area — as will be described later — shows a great effort in developing high speed (scalable) solutions to the problem of detecting intruders and anomalous activities in backbone networks. Monitoring the network behavior instead of a single host's behavior allows to have a less expensive solution and a more powerful detection: it deals with the state of a set of hosts and not a single machine.

The aim of scalability towards high speed solutions leads to the necessity of having fast systems of detection. According to [8], most NIDSes can currently keep pace only with network traffic of 100-200 Mbps. [5] gives stricter evaluation of Snort (with Bro one of the best-known NIDSes), asserting that it can handle no more than 100Mbps under normal traffic and it has worst performance with heavy traffic (with consequently packet dropping). [1] studies the performance of Snort and Bro in Gbps environments: Snort quickly consumes all the available CPU, while Bro uses all the available memory. The results of these studies indicate that it is no longer possible to have a stateful or even stateless analysis of all the packets that are monitored by a NIDS. Hence, there is the necessity to reduce the amount of data to be processed, for instance by sampling only one out of every n packets, or aggregating packets into flows. Plainly, this drastically changes the type and amount of available data for intrusion detection. This also means that the traditional techniques, i.e. signature based engines, may be less powerful (if usable at all). Furthermore, the attack definitions have to be rewritten according to the new type of data. For example, [9] uses sampled packets to develop a method that statistically estimates the super sources and destinations, i.e., the sources (and destinations) that have an unexpectedly large fan-out (fan-in, respectively; the number of peer hosts) in a small time interval. The identification of super sources and destinations can be useful to detect port scanning and Denial-of-Service (DoS) attacks. The proposed solution aims to memorize the minimally required information to characterize a super source or destination, and combines filtering and sampling techniques to achieve better results. At the same time, [9] provides detection solutions only for a subset of all possible attacks, i.e. attack categories that can be statistically distinguished.

Another approach is the one proposed in [4]. In this case, the data reduction is obtained by, instead of assessing individual packets, looking at trains of packets, i.e., flows (such as TCP connections); the proposed method works directly on flow-level data. The authors propose an Internet backbone monitoring and traffic analysis framework, called UPFrame, that uses NetFlow data exported by routers in a backbone network. The framework is presented as a general purpose platform for Internet monitoring, with possible applications in security. In line with other works by the same authors, e.g., [3], [4] uses UPFrame to detect the propagation of Internet worms in a backbone network, as well as to classify host behavior and to measure a host's activity. Contrary to sampling, flows offer aggregated information about the traffic in a network, moving the analysis towards metadata. These approaches change the nature of data to be analyzed, and, consequently, the analysis methods and the detectable attack types: the amount of analyzed traffic is reduced, but the problem of system accuracy, which may be worse than with systems that take all (raw) data into account, is still open.

3 State of the Art in Self-Learning Systems

As argued earlier, new threats on the Internet, for instance computer viruses, may spread quickly. It is therefore important that network defense systems are able to cope with new threats fast. This motivates the need for adaptive solutions for NIDSes: defense systems that adapt themselves when the environment changes. Adaptive NIDSes have several advantages. First, an adaptive system may recognize attacks that have never been seen

before. Second, the adaptability also entails that less human interaction is needed to update and tune the system.

An approach to adaptive NIDSes is self-learning and, as it has emerged from literature, it can play many roles in Intrusion Detection. Self-learning techniques have been applied in anomaly-based detection engines, i.e., systems in which an event is considered malicious if it deviates from the expected behavior. Recently, the work of [6] presented a statistical model to detect flow-level intrusions, which is suitable for high speed networks. The authors have developed an intrusion detection tool called HiFIND (High-speed Flow-level Intrusion Detection system). The anomaly detection engine of HiFIND is based on the error between the expected value for some analyzed metric and the measured value for the same metric. A deviation suggests the presence of an anomaly in the traffic. HiFIND statistically characterizes the traffic according to the measures the system is required to monitor. For example, it is possible to detect TCP SYN flooding DoS attacks by tracing the difference between the number of SYN and SYN/ACK packets for each triplet source IP, destination IP and destination port. The metric, with respect to the set of monitored hosts, gives a clear indication of the distribution of packets over time. A sharp variation points out a DoS attack.

As HiFIND adds adaptability to the detection engine, self-learning methods also are a useful way to improve high level organization between subsystems in distributed environments (i.e., where various NIDSes are working together, exchanging information about threats etc.). An example of this technique is presented in [2], in which the authors describe a distributed architecture based on the concept of autonomous cooperating systems. Each system has the capability of detecting attacks, combining flow-based statistics and packet payload infomation. At the same time, the subsystem can also share its current knowledge with other systems, improving the total detecting ability (self-optimization).

Finally, the scientific community has considered another problem that is common to all kinds of IDSes, in both gigabit-speed and megabit-speed networks: alert management. An IDS can easily produce hundreds of alerts each hour, each of them may be false positive. Hence, there is a clear need to find a way to reduce the amount of alerts to be analyzed by hand — improving the system's accuracy by both achieving false positive reduction and aggregating correlated alerts into attack scenarios [7].

4 Concluding Remarks

The huge spread of high-speed (say 1 Gbps and up) networks and the always increasing number of attacks and network abuses, motivate the interest of both the academic as well as the network operations world in NIDSes. The common goal of NIDS researchers is to improve the system performance, aiming to keep pace with the speed of current networks. At the same time, the NIDS research community seems to show an increasing interest also into adaptive systems, in which less human interaction is required to keep the system running and accurate.

In our research project, we aim to build a high performance NIDS that can cope with speeds of 1Gbps or more. The system, to be competitive, should also achieve the aims of completeness (few failures to detect an intrusion) and accuracy (small number

of false positives and negatives). Moreover, the wish to provide secure services to end-users implies that such a system should work in real-time on actual backbones.

The present literature study is the first step in our research. It outlines the current major trends in high-speed network intrusion detection and has shown that there still are many open issues. As argued in this paper, the first problem is keeping up with speed and massive traffic. Therefore, we are considering metadata (flows) as the most suitable solution for achieve data-reduction: indeed, the computational overhead required by a complete analysis of all packets can not be managed anymore in high speed environments. At the same time, we are looking to integrate flows with the information provided by sampled packets: in our opinion the payload of sampled packets may still be useful to characterize the traffic.

The second goal in this research project is to enhance the system with adaptive mechanisms. The use of metadata itself suggests an anomaly based approach: this would permit the system to perceive new traffic patterns and to react to changes in the environment. Furthermore, adaptability can also lead to system self-optimization and self-reconfiguration, reducing the required human interaction.

Finally, in our project we intend to test and validate the system on real networks.

References

1. H. Dreger, A. Feldmann, V. Paxson, and R. Sommer. Operational experiences with high-volume network intrusion detection. In *SIGSAC: 11th ACM Conference on Computer and Communications Security (CSS'04)*, pages 2–11, 2004.
2. F. Dressler, G. Münz, and G. Carle. CATS - cooperating autonomous detection systems. In *1st IFIP International Workshop on Autonomic Communication (WAC 2004)*, October 2004.
3. T. Dübendorfer and B. Plattner. Host behaviour based early detection of worm outbreaks in internet backbones. In *Enabling Technologies: Infrastructure for Collaborative Enterprise, 2005. 14th IEEE International Workshops on (WETICE'05)*, pages 166–171. IEEE Computer Society, 2005.
4. T. Dübendorfer, A. Wagner, and B. Plattner. A framework for real-time worm attack detection and backbone monitoring. In *Critical Infrastructure Protection, First IEEE International Workshop on (IWCIP'05)*, November 2005.
5. M. Gao, K. Zhang, and J. Lu. Efficient packet matching for gigabit network intrusion detection using TCAMs. In *Advanced Information Networking and Applications, 20th International Conferece on (AINA'06)*, pages 249–254. IEEE Computer Society, 2006.
6. Y. Gao, Z. Li, and Y. Chen. A DoS resilient flow-level intrusion detection approach for high-speed networks. In *Distributed Computing Systems, 2006. ICDCS 2006. 26th IEEE International Conference on*, pages 39–46, 2006.
7. C. Kruegel, F. Valeur, and G. Vigna. *Intrusion Detection and Correlation: Challenges and Solutions*. Springer-Verlag Telos, 2004.
8. H. Lai, S. Cai, H.Huang, J. Xie, and H.Li. A parallel intrusion detection system for high-speed networks. In *Applied Cryptography and Network Security, Second International Conference (ACNS'04)*, volume 3089, pages 439–451. Springer, 2004.
9. Q. Zhao, J. Xu, and A. Kumar. Detection of super sources and destinations in high-speed networks: Algorithms, analysis and evaluation. *Selected Areas in Communications, IEEE Journal on*, 24:1840 – 1852, October 2006.