NADA – Network Anomaly Detection Algorithm

Sílvia Farraposo[1], Philippe Owezarski[2], and Edmundo Monteiro[3]

[1] School of Technology and Management of Leiria
Alto-Vieiro, Morro do Lena, 2411-901 Leiria, Apartado 4163, Portugal
[2] LAAS – CNRS, 7 Avenue du Colonel Roche
31077 Toulouse, CEDEX 4, France
[3] Laboratory of Communications and Telematics, Computer Science Department
Pólo II – Pinhal de Marrocos, 3030-290 Coimbra, Portugal
silvia@estg.ipleiria.pt, owe@laas.fr, edmundo@dei.uc.pt

Abstract. This paper deals with a new iterative Network Anomaly Detection Algorithm – NADA, which accomplishes the detection, classification and identification of traffic anomalies. NADA fully provides all information required limiting the extent of anomalies by locating them in time, by classifying them, and identifying their features as, for instance, the source and destination addresses and ports involved. To reach its goal, NADA uses a generic multi-featured algorithm executed at different time scales and at different levels of IP aggregation. Besides that, the NADA approach contributes to the definition of a set of traffic anomaly behavior-based signatures. The use of these signatures makes NADA suitable and efficient to use in a monitoring environment.

Keywords: Traffic Anomaly Identification, Anomaly Signature.

1 Introduction

The lack of security in networks is an issue that network administrators would like to solve on the fly, independently of the network size. Being anomalies a structural part of traffic, it is important to completely detect, classify (i.e., determining the type of anomaly) and identify (i.e., determining all the packets and flows involved in the anomaly) them in order to act adequately.

NADA aims being completely generic and work on any kind of time series issued from incoming traffic (online) or packet traces (offline). To illustrate NADA, in this paper we will consider three different data time series: Number of packets per unit of time; Number of bytes per unit of time; and Number of new flows per unit of time.

Other approaches for detecting traffic anomalies exist. However, as far as we know, none permits simultaneously the detection, classification and identification of traffic anomalies. At most, some recent works introduced some level of classification in the algorithms being proposed, using information provided by IP features [1] [2]. Nevertheless, anomaly classification and identification remains an important, unmet challenge, since none of the proposals exploited exhaustively the richness of IP

A. Clemm, L.Z. Granville, and R. Stadler (Eds.): DSOM 2007, LNCS 4785, pp. 191 – 194, 2007.

feature information to provide accurate information by the involved parties. NADA's classification and identification stages are developed in an easy way for both configuring the tool and analyzing its outputs. This aspect is particularly important when one of the main goals is to limit the negative effects of an anomaly occurrence in real networks.

The rest of this paper is organized as follows: Section 2 gives an overview of the NADA algorithm presenting its main features. Section 3 presents anomaly signatures, and how these signatures can be used for anomaly classification, and section 4 concludes the paper, summarizing our ongoing research.

2 Network Anomaly Detection Algorithm – NADA

NADA has been defined as a multi-scale, multi-criteria, and multi-level of IP aggregation approach [3]. NADA's algorithm has two phases. The first one is devoted to the detection and classification of traffic anomalies, while the second phase targets the anomalous flows by fully identifying them.

The core idea used in NADA's detection stage is that any anomaly will be responsible for some level of variation at least on one of the criterions considered, at some time-scale and at some level of IP aggregation. Variations are pointed by using the formula below (1), in which *X* is a data time series directly obtained from traffic traces, and *P* is a data series that is obtained from *X*, and in which each value is the difference between two consecutive values of *X*. Each value p_i of *P* corresponds then to a variation. Significant variations might be associated to an anomaly. Significant variations were named deltoids by Cormode et al. [4] who used them to detect significant traffic changes.

$$\begin{array}{l} X=\{x_1,x_2,\dots,x_n\}, x_i=\{\#packets | \#bytes | \#flows\}/\Delta \\ P=\{p_1,p_2,\dots,p_{n-1}\}, p_i=x_{i+1}-x_i \\ \begin{cases} pi \geq E(p)+k\sigma, select \\ pi < E(p)+k\sigma, reject \end{cases} \end{array} \quad (1)$$

The mean and the standard deviation, *E(p)* and *σ* respectively, of each time series are calculated and used to define a threshold. Each value of the time series that exceeds the threshold might point a traffic anomaly. This sort of filtering can be more or less coarse grained depending on the value of the adjustment parameter *k* of the formula, where smaller values of *k* fine-grain the search. Currently, the value of *k* is assigned manually, ranging from 0.5 to 2.5, being the value 2.0 the most used. These values were obtained empirically, after successive executions of NADA from where it was seen that for values of *k* greater than 3.0 no significant variations are detected, while for values of *k* smaller than 0.5 the formula is not effective because $E(p)+k\sigma \approx E(p)$.

The formula above is applied recursively. Each level of iteration uses a different level of traffic aggregation. At the first iteration the all IP space is considered, and time slots of duration Δ, with possible anomalies, are spotted. At each new iteration, flows in the time slots previously spotted are analyzed, from more generic ones (mask /1) to more specific ones (mask /32).

The classification stage is based on behavior-based signatures. These signatures were obtained through the execution of NADA over several traces. The anomalies signaled by our algorithm presented always a set of characteristics that could be used to identify them in a univocal way. Finally, the purpose of anomaly identification is to allow its complete mitigation. This third stage then includes an exhaustive description of the anomaly, using all the information previously collected.

3 Classification Based on Anomaly Signatures

To assess the accuracy and performance of NADA, a set of traces created in the framework of the MetroSec Project, between 2004 and 2006 was used. This repository spans different types of anomalies, legitimate and illegitimate ones, with different levels of intensity. Also, different types of anomaly generators (as Hping, Iperf, Trinoo, TFN2k) were used in order to improve the quality of the database.

The successive utilization of NADA showed that anomalies of a specific type have a consistent signature. Such signatures are obtained by looking how the distribution of source and destination IP addresses and ports relate to each other in candidate-anomalous flows. Running NADA on the traces collected permitted the isolation of several types of anomalies. In this paper we will focus on DDoS attacks.

When analyzing the traces we have obtained two different types of DDoS signatures, depending on the number of destination ports that are flooded. It was also observed that these signatures are independent of the tool being used to perpetrate the attacks, and because of that may be considered as behavior-based signatures, instead of regular signatures, which are dependent of specific parameters of a specific anomaly.

Each behavior-based signature obtained can be represented as a sequence of four plots that constitute what we have called the graphical signature. Figure 1 is a representation of the DDoS signature behavior-based of type *n*IP sources using *n*Ports attacking *1*IP destination using *n*Ports. The four plots show how the different source and destination addresses/port relate to each other in flows associated to a DDoS. This sequence of shapes is detected at all levels of IP aggregation at the destination, ranging from /8 to /32. This signature is unique for a given type of attack when analyzing the correct time series, packets and bytes.

All plots in Figure 1 relate the distribution of source with destination information. So, the leftmost plot shows the IP sources that are flooding the destination IP address. The next plot, inserts information about the ports that were used by the sources. It can be seen that each packet sent was using a different port number. In the plot this is denoted by a full straight line. The third plot of the signature adds the port number information to each destination. In our DDoS case it is possible to see that different ports of the target are being flooded (diagonal line). Finally, the rightmost plot shows how destination ports are affected by the anomaly. This plot is important to differentiate network scan from port scan attacks.

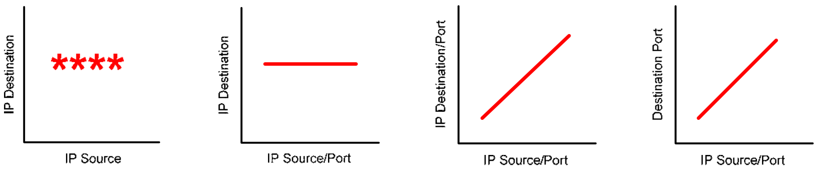

Fig. 1. DDoS behavior-based signature without noise. Type of DDoS: *n* IP Sources, *n* Source Ports : *1* IP Destination, *n* Destination Ports.

The Receiver Operating Characteristic (ROC) curves were used to obtain some information about the sensitivity of NADA (they are not showed due to the lack of space). The analysis of the curves permitted us to verify that NADA is efficient and that whatever the value of *k* is, the detection probability is always higher than the rate of false alarms.

4 Conclusion

In this paper, we have presented NADA an algorithm for detecting, classifying and identifying anomalies of any type in network traffic, and that provides information about the parties responsible of the anomaly, in a way easily understandable by technicians who are operating and managing networks.

Moreover, the information provided by NADA is delivered in graphical and textual format. If the first format could be interesting for administrator to discover, at a glance, what is happening in the network, the latter one could be easily used to trigger other types of signals or actions, suited to the anomaly that is occurring.

To conclude this work, we intend to run NADA over traces for which we do not know about the presence of anomalies, to test the efficiency and robustness of NADA. Future work also includes the design of a election method for the *k* factor, as it is for the moment hand made.

References

1. Kim, S., Reddy, A., Vannucci, M.: Detecting Traffic Anomalies through Aggregate Analysis of Packet Header Data. In: Networking 2004, Athens (2004)
2. Lakhina, A., Crovella, M., Diot, C.: Mining Anomalies Using Traffic Feature Distributions. In: ACM SIGCOMM, Philadelphia (2005)
3. Farraposo, S., Owezarski, P., Monteiro, E.: A Multi-Scale Tomographic Algorithm for Detecting and Classifying Traffic Anomalies. In: IEEE ICC 2007, Glasgow (2007)
4. Cormode, G., Muthukrishnan, S.: What's New: Finding Significant Differences in Network Data Streams. In: IEEE/ACM Transactions on Networking, vol. 13 (2005)