TAXONOMYOFNETWORK INTRUSION DETECTION SYSTEM BASED ON ANOMALIES

Jaime Daniel Mejía Castro, Jorge Maestre Vidal, Ana Lucila Sandoval Orozco, Luis Javier García Villalba

Group of Analysis, Security and Systems (GASS)
Department of Software Engineering and Artificial Intelligence (DISIA)
School of Computer Science, Office 431, Universidad Complutense de Madrid (UCM)
Calle Profesor José García Santesmases s/n, Ciudad Universitaria, 28040 Madrid, Spain
{j.mejia, jmaestre, asandoval, javiergv}@fdi.ucm.es

Abstract

On a daily basis, we see new forms of malware which are completely different from those known, so there are no signatures to allow their detection. Hence intrusion detection techniques have arisen in networks that do not rely on malware structure, but on identifying ways of using the system that are not within the usual and legitimate form. When a Network Intrusion Detection System adopts this type of strategy it is said to be based on anomalies. This paper aims to introduce main fundamentals of these systems and presents a classification of them. For each of them, it identifies their main features besides giving a number of considerations that should be taken at the time of this installation.

Keywords: Network Intrusion Detection System, NIDS, Anomalies, Malware

1 INTRODUCTION

Intrusion Detection System or IDS is a defense mechanism whose behavior is based on the analysis of the different events that occur in protected system looking for signs of malicious activity. IDS classify events as legitimate and illicit, the latter being considered as intrusions. If an IDS to detect also has the ability to take action to prevent or mitigate intrusion effects is called Intrusion Prevention System or IPS. When the IDS operate in a network environment, it says it is a Network-based Intrusion Detection Systemor NIDS and if it does at host is a Host-based Intrusion Detection System or HIDS. Any of them can adopt preventive behavior, besides the detection.

In the last 20 years different techniques have been proposed to classify events NIDS. Similarly to IDS,earlier approaches are based on signature detection. This involves having prior knowledge of the specific features of threats, the situation that ceases to be manageable to popularize its use of new technologies. Rapid proliferation of intrusion strategies and the constant appearance of new malware takes raise new analysis mechanisms, able to identify unknown attacks, the so-called zero-day attacks. These mechanisms include statistical methods, machine learning and data mining strategies, to complete aspects not covered by the signature-based approach. Given the satisfactory results obtained we increased its use, making it an indispensable element in any current security perimeter.As the network protocols evolved have also had to NIDS, a NIDS may detect attacks from different sources,as, for example, attacks inserted in the packet header or malware content payload.They are also capable of operating on any media conventional wired, wireless or virtual and face the most sophisticated evasion techniques.

This article is structured in three sections, with this introduction being the first. Section 2 presents a classification of anomaly-based NIDS. Section 3 contains the conclusions of this work.

2 CLASSIFICATION OF ANOMALIES BASED NIDS

Figure 1 illustrates a classification of anomalies based NIDS. The above classification is based on the behavior of the model data processing [1].

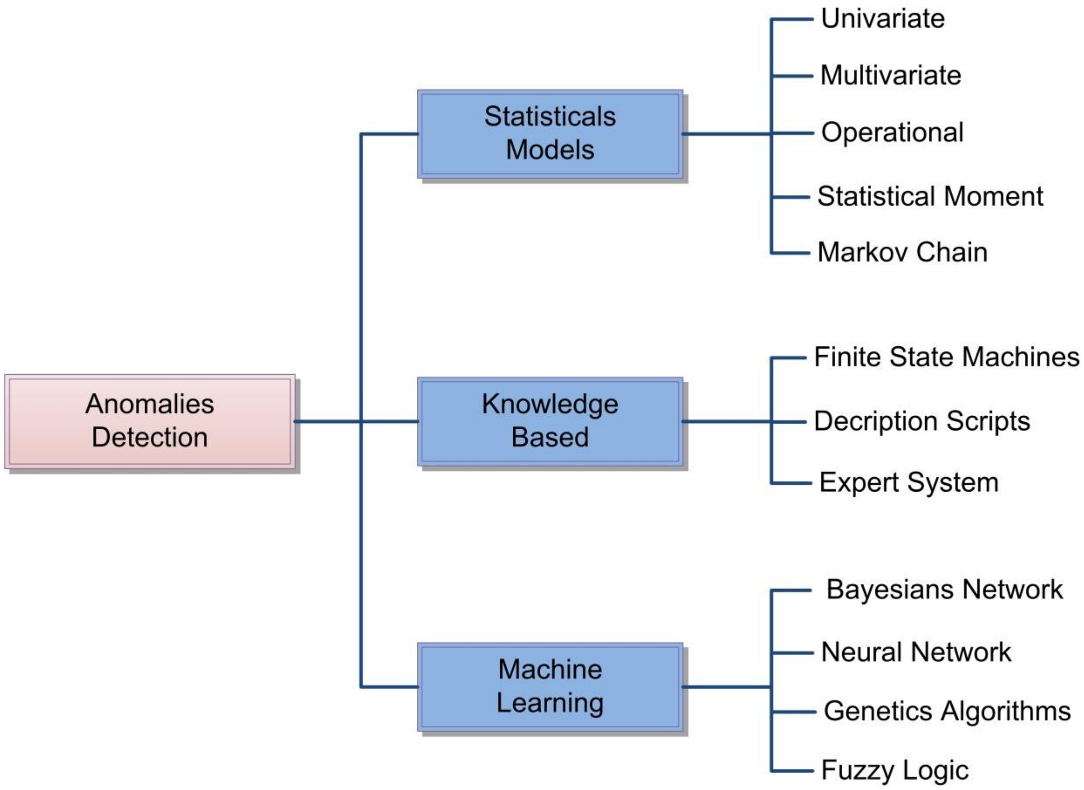

Figure1. Clasification Anomalíes based NIDS

Statisticals Models: Great amount of current NIDS have opted for the use of statistical models. Such systems are intended to build statistical-predictive models can identify anomaly use of the system against legitimate use.This technique captures network traffic and creates a profile that represents the stochastic behavior. This profile uses metrics like traffic rate, number of packets for each protocol, connection rate, or number of different IP addresses, among others, that can represent different system use modes.In detection process are considered two sets of network traffic. One of them represents the characteristics observed at the time of analysis, while the other represents the previously known features. Classification is produced based on degree of similarity between the two sets, anomaly labelling traffic is significantly different from legitimate traffic.

Neumann & Porras [2]introduce EMERALD into your systemand use distributed methodologies correlation of large amounts of events. For this, signature analysis combines with statistical profiles that allow you to perform traffic classification in real time to any of the services available to the networks, marking a milestone in the field of anomaly-based NIDS and statistical models.Years later, Sang& Won[3]propose first anomalies detection method using data clustering algorithms.

From these early works, we are beginning to apply different statistical tools. So, Yu&Zhou [4]present an anomaly detection approach based on an adaptive non-parametric modeling in symmetric network traffic, which has the ability to adjust its parameters to the position detection in which the network is operating. Previously, Ye *et al.*[5]had taken the robustness of the Markov model for classification of events, but only achieved good results with bit distorted data.

One advantage of the design based on the statistical model is that it requires training on a set of known attacks on model generation process.Moreover, most of these proposals have been quite good in real traffic conditions[6], not just causing network overhead. However, it should be noted that in applying these tools, we assume that the behavior of the network traffic is quasi-stationary, something not always guaranteed.

Knowledge-Based Models: A knowledge base is a database type, adapted to the management and representation of knowledge.NIDS incorporating these mechanisms require a training phaseable to identify the most representative parameter sets legitimate and malicious traffic that is intended train. Once extracted, it generates a rule base able to classify the nature of analyzed traffic.

Lee*et al*. [7]present a NIDS with a knowledge base that analyzes traffic based on the contents of the packet payload considering the characteristics of the connection. Jiang *et al.*[8]incorporate a distributed intrusion detection based on finite state machines, with a detection scheme based on cluster, which periodically selects a node as one monitor in the cluster.Tran et al.[9]propose a multi-frame expert classification for detecting different types of anomalies network through detection techniques are selected in which different attributes and learning algorithms.

In general, the most significant advantages of using knowledge bases in NIDS design are the high degree of robustness and flexibility that give them. However, using rule-based analysis can overload the operation of the network, if rule aggregation methods are not used. In addition, certain designs may require too much prior knowledge of the threats it facesand maybe very close tosignature detection.

Machine Learning Techniques: Use of these techniques allow the NIDS to learn of events known to carry out classifications on unknown events, generalizing knowledge gained. Consequently, an anomaly-based NIDSwith machine learning hasthe ability to change its classification strategy by acquiring new information. Precisely a unique feature of these schemes isthe need of labeled data to train model behavior, a condition that can sometimes be a problem, because it can lead to incorrect labels and an unwanted behavior. To avoid this, wetypically employ learning mechanisms tolerant to a noise margin.

Machine learning have a high degree of similarity with the aforementioned statistics strategies, but their approach is directly based on computational cost optimization of those algorithms that can overload the network. Despite its high performance, no works have followed other branches, such as Song &Lockwood[10], to develop a hardware solution for an efficient packet classification operation on a system of network intrusion detection based on FPGA (*Field Programmable Gate Arrays*), a widely used technology in real time.

One of the major efforts in the application of these techniques is due to Mahoney [11]who proposes three intrusion detection systems: packet header anomaly detector(PHAD), application layer anomalydetector(ALAD)and network traffic anomaly detector (NETAD). Each extracts certain information from the traffic analyzed and generates a classification according to the previously received training.

Wang &Stolfo presentPAYL[12]. This system classifies traffic based on three characteristics: the port, the packet size and flow direction (input or output). Through these three parameters classified payload creating a series of patterns to define what would be normal behavior within each class. Following this work, Bolzoni*et al*.propose POSEIDON [13]in order to solve certain deficiencies of PAYL when performing clustering techniques. Another important contribution of POSEIDON is the use of self-organizing maps (SOM) that besides reducing network overload inNIDS reduces the number of generated classes in the training process allowing it to operate with greater precision.These two works are especially important since most of the current proposals anomaly detection in the payload of network traffic is based on these.

CONCLUSIONS

This paper presents taxonomy of networks intrusion detection systems based on anomalies. Technique used depends on the type of anomaly which has to detect, type and behavior of data, environment in which the system operates, limitations of cost and calculation and, finally, level of security required. To make an IDS implementation should be noted that good training is vital in system effectiveness. Furthermore, model used should reflect the performance,as faithfully as possible, of the system in absence of attacks, for which there must be traffic as clean as possible. Training may be defined as large enough that builds up a complete model of the application environment, but a system update before repeating this phase may have a higher cost overrun. If the duration of this phase is very short,it may be an inadequate classification, seeing an increase in the detection phase legitimate traffic alerts marked as anomaly (false positives).

ACKNOWLEDGMENTS

This work was supported by the Agencia Española de Cooperación Internacional para el Desarrollo (AECID, Spain) through Acción Integrada MAEC-AECID MEDITERRÁNEO A1/037528/11.

REFERENCES

[1] V. Jyothsna, V. V. Rama Prasad, K. Munivara Prasad.A Review of Anomaly based Intrusion Detection Systems.*International Journal of Computer Applications*, New York, USA, Vol. *28* (7), pp. 26-35. August 2011.

[2] P.G. Neumann, P.A. Porras. EMERALD: Event Monitoring Enabling Responses to Anomalous Live Disturbances. *In Proceedings of the 20th NIST National Information Systems Security Conference*, Baltimore, USA, pp. 353–365, October 1997.

[3] H.O. Sang, S.L. Won. An Anomaly Intrusion Detection Method by Clustering Normal User Behaviour.*Computers and Security*, Elsevier B.V. Amsterdam, The Netherlands, Vol. 22 (7), pp. 596–612, October 2003.

[4] W. Yu, X.Y. Zhou.An Adaptive Method for Anomaly Detection in Symmetric Network Traffic.*Computers and Security*, Elsevier B.V. Amsterdam, The Netherlands, Vol. 26 (6), pp. 427–433, September 2007.

[5] N. Ye, Y. Zhang, C.M. Borror. Robustness of the Markov-Chain Model for Cyberattack Detection.*IEEE Transactions on Reliability*, Vol. 53 (1), pp. 116–123, March2004.

[6] Ch. Chen, Y. Chen, H. Lin. An Efficient Network Intrusion Detection.*Computer Communications*, Elsevier B.V. Amsterdam, The Netherlands, Vol. 33(4), pp. 477-484, March 2010.

[7] W. Lee, S.J. Stolfo, K.W. Mok. A Data Mining Framework for Building Intrusion Detection Models.*InProceeding of the IEEE Symposium on Security and Privacy*, Oakland, California, USA, pp. 120-132, May 1999.

[8] Y. Jiang, Y. Zhong, S. Zhang. Distributed Intrusion Detection for Mobile Ad Hoc Networks.*InProceeding of the Symposium on Applications and the Internet*, Trento, Italy, pp. 94-97, January-February2005.

[9] T. Tran, P. Tsai, T. Jan.A Multi-expert Classification Framework with Transferable Voting for Intrusion Detection.*InProceeding of the Seventh International Conference on Machine Learning and Applications*, San Diego, California, USA, pp. 877-882, December2008.

[10] H. Song, J.W. Lockwood. Efficient Packet Classification for Network Intrusion Detection using FPGA.*In Proceedings of the 2005 ACM/SIGDA 13th International Symposium on Field-programmable Gate Arrays*, Monterey, California, USA, pp. 238–245,February 2005.

[11] M.V. Mahoney, P.K. Chan. PHAD.Packet Header Anomaly Detection for Identifying Hostile Network Traffic, *Florida Institute of Technology*, Melbourne, USA, Technical Report, 2001.

[12] K. Wang, S. J. Stolfo. Anomalous Payload-based Network Intrusion Detection. *In Proceedings of the 7th International Symposium on Recent Advances in Intrusion Detection (RAID)*, Sophia Antipolis, France. *Lecture Notes in Computer Science*, Vol. 3224, pp. 203-222, September 2004.

[13] D. Bolzoni, S. Etalle, P. Hartel, E. Zambon. POSEIDON.A 2-tier Anomaly-Based Network Intrusion Detection System. *In Proceedings of the Fourth IEEE International Workshop on Information Assurance*, London, United Kingdom, pp. 144-156, April 2006.