Computer Networks 53 (2009) 864–881

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

McPAD: A multiple classifier system for accurate payload-based anomaly detection

Roberto Perdisci [a,b,*], Davide Ariu [c], Prahlad Fogla [d], Giorgio Giacinto [c], Wenke Lee [b]

[a] Damballa, Inc., 817 West Peachtree St. NW, Suite A-110, Atlanta, 30308 GA, USA
[b] College of Computing, Georgia Institute of Technology, Atlanta, 30308 GA, USA
[c] Department of Electrical and Electronic Engineering, University of Cagliari, Piazza D'Armi, 09123 Cagliari, Italy
[d] Google, Inc., Mountain View, CA 94043, USA

ARTICLE INFO

Article history:
Available online 6 December 2008

Keywords:
Network intrusion detection
Anomaly detection
Shell-code attacks
Multiple classifiers
One-class SVM

ABSTRACT

Anomaly-based network intrusion detection systems (IDS) are valuable tools for the defense-in-depth of computer networks. *Unsupervised* or *unlabeled* learning approaches for network anomaly detection have been recently proposed. Such anomaly-based network IDS are able to detect (unknown) *zero-day* attacks, although much care has to be dedicated to controlling the amount of false positives generated by the detection system. As a matter of fact, it is has been shown that the false positive rate is the true limiting factor for the performance of IDS, and that in order to substantially increase the Bayesian detection rate, $P(Intrusion|Alarm)$, the IDS must have a very low false positive rate (e.g., as low as 10^{-5} or even lower).

In this paper we present McPAD (multiple classifier payload-based anomaly detector), a new accurate payload-based anomaly detection system that consists of an ensemble of one-class classifiers. We show that our anomaly detector is very accurate in detecting network attacks that bear some form of shell-code in the malicious payload. This holds true even in the case of polymorphic attacks and for very low false positive rates. Furthermore, we experiment with advanced *polymorphic blending attacks* and we show that in some cases even in the presence of such sophisticated attacks and for a low false positive rate our IDS still has a relatively high detection rate.

1. Introduction

Intrusion detection systems (IDS) are valuable tools for the defense-in-depth of computer networks. Network IDS look for known or potential malicious activities in network traffic and raise an alarm whenever a suspicious activity is detected. Two main approaches to intrusion detection are used, namely *misuse* and *anomaly* detection [25]. Misuse detectors are based on a description of known malicious activities. This description is often modeled as a set of rules referred to as *attack signatures*. Activities that match an attack signature are classified as malicious. On the other hand, anomaly detectors are based on a description of *normal* or *benign* activities. As malicious activities are expected to be different from normal activities, a suitable distance measure allows anomaly-based IDS to detect attack traffic.

The IDS most commonly used in real networks are signature-based, because they are able to efficiently detect known attacks while generating a relatively low number of false positives. Anomaly-based detection systems usually produce a relatively higher number of false positives, compared to the misuse-based or *signature-based* detection systems, because only a fraction of the anomalous traffic actually derives from intrusion attempts. Nevertheless,

* Corresponding author. Address: Damballa, Inc., 817 West Peachtree St. NW, Suite A-110, Atlanta, 30308 GA, USA. Tel.: +1 4048247476.
E-mail address: roberto.perdisci@gmail.com (R. Perdisci).

1389-1286/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.comnet.2008.11.011

anomaly detectors are able to detect *zero-day* (i.e., never-before-seen) attacks, whereas signature-based systems are not.

Because it is very difficult and expensive to obtain a labeled dataset that is representative of real network activities and contains both normal and attack traffic [24], *unsupervised* or *unlabeled* learning approaches for network anomaly detection have been recently proposed [27,12]. These methods aim to learn from datasets of traffic extracted from real networks without the necessity of a labeling process. *Unlabeled* anomaly detection systems are based on the reasonable assumption that the percentage of attack patterns in the extracted traffic traces is usually much lower than the percentage of normal patterns [27]. Furthermore, it is possible to use signature-based IDS in order to filter the extracted traffic by removing the known attacks, thus further reducing the number of attack patterns possibly present in the dataset. Another assumption is that the attack patterns are supposed to be distinguishable from the normal patterns in a suitable feature space. The term "unlabeled anomaly detection" used in the intrusion detection field actually refers to what in machine learning is more often called "novelty detection", "outlier detection" or "one-class classification". One-class classification algorithms pursue concept learning in absence of counter examples [31], and have been shown to be promising for network anomaly detection [12].

1.1. Payload-based anomaly detection

Recent work on unlabeled anomaly detection focused on *high speed* classification based on simple *payload* statistics [19,23,35,36] (the payload is the data portion of a network packet). For example, PAYL [35,36] extracts 256 features from the payload. Each feature represents the occurrence frequency in the payload of one of the 256 possible byte values. A simple model of normal traffic is then constructed by computing the average and standard deviation of each feature. A payload is considered anomalous if a *simplified Mahalanobis distance* between the payload under test and the model of normal traffic exceeds a predetermined threshold. Although PAYL is based on simple statistics extracted from the payload, it has been shown to be quite effective [35]. Nonetheless, we show in Section 5 that PAYL may suffer from a relatively high false positive rate. Wang et al. [35] also proposed a more generic n-gram[1] version of PAYL. A sliding window of length n is used to extract the occurrence frequency in the payload of all the possible n-grams. In this case the payload is described by a pattern vector in a 256^n-dimensional feature space. The n-grams extract byte sequence information, which helps in constructing a more precise model of the normal traffic compared to the simple byte frequency-based model. The extraction of n-gram statistics from the payload can be performed efficiently and the IDS can be used to monitor high speed links in real time. However, given the exponentially growing number of extracted features, the higher n the more difficult it may be to construct an accurate model because of the curse of dimensionality [11] and possible computational complexity problems.

Wang et al. also proposed ANAGRAM [37], an anomaly detector based on n-gram analysis that uses a different approach for modeling the traffic, compared to PAYL. ANAGRAM stores the distinct n-grams extracted from normal packets in a Bloom filter b_1, and the n-grams extracted from known attacks in a second Bloom filter b_2. During the test phase, for each packet all the distinct n-grams are extracted from the payload and compared with the filters b_1 and b_2. Payloads that contain too many n-grams that are not present in b_1 or that are present in b_2 are classified as anomalous [37]. Other anomaly detection systems based on more complex features have been proposed [33,5]. These anomaly detectors involve the extraction of syntax and semantic information from the payload, which is usually a computationally expensive task. Therefore, it may not be possible to use this approach in order to analyze network traffic on high speed links in real time.

1.2. Our contribution: McPAD

In [2], Axelsson has shown that the false positive rate is the true limiting factor for the performance of IDS, and that in order to substantially increase the Bayesian detection rate, $P(Intrusion|Alarm)$ (i.e., the probability of having an intrusion given that an alarm was raised), the false positive rate of the IDS must be very low (e.g., as low as 10^{-5} or even lower). Although fairly effective, payload-based anomaly detectors like PAYL [35] suffer from a relatively high false positive rate, and may therefore have a very low Bayesian detection rate.

Our goal is to devise a new payload-based anomaly detector that uses simple payload statistics to accurately detect network attacks, and in particular shell-code attacks [1] (i.e., attacks that inject executable code), even at a very low false positive rate. We address this challenge using an ensemble of classifiers. Classifier ensembles, often referred to as multiple classifier systems (MCS), have been proved to achieve a better trade-off between false positive and detection rate in many applications, compared to the best single classifier in the ensemble [10,20]. A number of security related applications of MCS have been proposed in the literature. For example, MCS are used in multimodal biometrics for hardening person identification [4], and in misuse-based IDS [16] to improve the detection accuracy.

MCS attain accuracy improvements when the combined classifiers are "diverse", i.e., they make different (ideally independent) errors on new patterns [10]. A way to induce diversity is to combine classifiers that are based on descriptions of the patterns in different feature spaces [20]. The combination of classifiers trained on different feature spaces allows us to effectively exploit the complementarities of the different pattern representations [10]. In [26] we proposed a new approach to construct a *high speed* payload-based anomaly IDS by combining multiple one-class support vector machine (SVM) classifiers using a majority voting rule. As mentioned in Section 1.1, n-gram analysis has been shown to be quite effective for payload-based anomaly detection. However, the exponential increase of the dimensionality of the features space may

[1] Here an n-gram represents n consecutive bytes in the payload.