Computer Communications 35 (2012) 875–884

Contents lists available at SciVerse ScienceDirect

Computer Communications

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

Improved anomaly detection using block-matching denoising

S.P. Kozaitis *, W. Petsuwan

150 W. University Blvd., Department of Electrical and Computer Engineering, Florida Institute of Technology, Melbourne, FL 32901, USA

ARTICLE INFO

Article history:
Received 30 June 2010
Received in revised form 30 September 2011
Accepted 20 January 2012
Available online 28 January 2012

Keywords:
Anomaly
Denoising
Noise
Traffic
Wavelet

ABSTRACT

We present a new approach for network traffic anomaly detection based on a denoising algorithm that uses wavelet transforms. Using a block-matching technique and considering network traffic as noise, we suppress the traffic in order to detect anomalies. This approach is data-driven in the sense that samples of network traffic determine the amount of background traffic suppression. Therefore, the output of the algorithm is an anomaly that can be easily detected. To improve the performance, the block-matching technique is combined with a method that can detect very short attacks. Results show that attacks can be detected under a variety of conditions.

1. Introduction

Some of the main challenges of network traffic anomaly detection are due to the large amount of data involved. Even a small false alarm rate can render a detection system ineffective. Other problems are that real-time analysis is desirable and often required and that detectors must keep up with constantly changing scenarios. A computationally efficient and adaptive process is needed.

Denial of Service (DoS) attacks are prevalent and may result in disruption of services. They generate a stream of traffic to overwhelm the resources of a target in order to deny or delay services. DoS attacks are typically the easiest to produce because they require little knowledge of the target system. If an attack signature is known, then it may be detected. However, an effective approach for anomaly detection should do so without knowing the signature of an attack or without requiring a detailed model of network traffic.

There has been some success in detecting DoS attacks because they usually result in a change of a measured statistic of network traffic flow and can be considered anomalous behavior [1–3]. Advanced signal processing methods involving wavelets have been used with much success to analyze a variety of signals and have also been applied to the detection of DoS attacks. Analysis methods using wavelets have a firm theoretical basis and are particularly effective at multiresolution analysis, representing transient behavior, modeling statistical dependence, and denoising [4,5]. Applying methods using wavelets to detect anomalous behavior could provide a fruitful approach toward detecting DoS attacks.

Wavelets are simply basis functions of the wavelet transform that meet certain conditions [6]. These transforms represent data much like the Fourier transform, by using waves. Unlike the basis functions of the Fourier transform, which are sines and cosines that extend to infinity, wavelets have local support, meaning each wavelet extends over a limited range. This makes wavelets useful for representing transient behavior. For example, by separating network traffic into regions of low and high transient behavior using wavelets, network anomaly detection can be improved [7].

The Fourier transform represents a signal in terms of different frequencies of sine and cosine functions. A wavelet transform represents a signal in terms of scaled versions of a wavelet. The wavelet representation has been very useful for examining data at different resolutions, for example to extract features from network traffic. In one approach, mulitresolution analysis of network traffic was used to examine the decay of wavelet coefficients as a function of scale to detect anomalies [8]. Another approach found that DoS attacks could be detected by using several metrics that were decomposed into a multiresolution framework [9]. For use in self-similar traffic, multiresolution analysis was applied to determine both the presence and nature of an anomaly [10]. By exploiting the favorable characteristics of wavelet transforms, progress in anomaly detection has been achieved.

There have been other advances using wavelets for anomaly detection in DoS attacks. For example, a method based on the correlation of destination IP addresses in outgoing traffic at an egress router was developed [11]. In this method, the address correlation data are wavelet transformed for detection of anomalies through statistical analysis. Another method considers the traffic as a time-series and uses an exponential moving average for smoothing of network traffic [12]. This is followed by a wavelet transform to

* Corresponding author. Tel.: +1 321 6747312.
E-mail address: kozaitis@fit.edu (S.P. Kozaitis).

0140-3664/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.comcom.2012.01.019

determine the distribution of energy between two consecutive time windows. Here, anomaly detection is achieved when the energy distribution reaches a threshold level. In yet another approach, a new signal modeling approach using wavelet approximation and system identification theory is used for detecting network anomalies. [13]. These methods illustrate that wavelets provide a rich set of tools to analyze network traffic for anomaly detection.

Despite wavelets having success in removing noise from signals, little work has been done using denoising for anomaly detection. Wavelet methods have been shown to be useful for the detection of small transient signals. In terms of network traffic, these signals are represented by a small percentage of the traffic as compared to the total traffic. Detecting small anomalous packets is significant when dealing with low-rate DoS attacks, because the attack rate is smaller than for a flood-based attack and may not result in significant traffic volume change. By identifying small statistical changes in network traffic, it is possible to identify attack patterns and low-rate attacks that are not detectable with volume-based detection systems.

For this paper, we use a denoising method for anomaly detection. Our approach considers background traffic as noise that we suppress allowing anomalies to be detected. Our approach is data-driven in the sense that it uses only traffic data to adjust its characteristics. Unlike many other adaptive approaches, it may use non-local data independently at adjacent samples in network traffic. Our approach does not rely on any assumptions or specifications of attack behavior or statistics of network traffic. We do, however, use a single parameter that determines the degree of suppression of traffic. Our particular focus is on setting the degree of suppression in traffic data to detect low-level attacks. However, our approach has the potential to be applied for the detection of other types of attacks in a variety of other environments.

The remainder of this paper is organized as follows. In Section 2, we provide an overview of related work. In Section 3, we describe our approach to anomaly detection. We analyzed the performance of our approach using attacks in terms of synthetic and actual traffic in Section 4, and we finally present some concluding remarks in Section 5.

2. Related work

The following section provides a review of prior work related to ours. Although our method uses a spectral-based method, it is different from others due to the denoising emphasis that includes block matching.

The main assumption for most spectral anomaly detection methods is that data can be transformed into a space where anomalies appear different than normal traffic. Such methods can be used in an unsupervised setting and can be used as a preprocessing step for other methods. Methods based on Principal Component Analysis (PCA) are of interest because they can optimally project data onto an orthogonal subspace. The variances along each principal component will be compacted optimally for that data. Then, anomalies can be easily separated from network traffic and detected [14–16]. Unlike PCA, wavelet transforms do not depend on the input data. However, they do offer similar energy compaction in the transform domain. In addition, PCA is computationally intensive, whereas wavelet transforms can be computed far more efficiently. A significant concern of spectral methods is that they are only useful if anomalies can be separated in the transformed space.

Wavelet transforms allow the quick calculation of features to be used in a statistical approach to anomaly detection. Typically, features are developed to describe normal traffic, and behavior that does not correspond to that is labeled anomalous. In one application, a wavelet filter exposed distinct characteristics of anomalies based on the change of local variance. This approach was useful to detect short- and long-lived traffic anomalies [17]. In another approach, 15 transient and global features from network traffic were found and used as inputs to a system that incorporated a prediction model for normal traffic [13]. The system measured the difference between normal and anomalous activities, and an outlier detection program was then used to detect anomalies. By exploiting redundancy of the continuous wavelet transform rather than a discrete version, volume-based attacks were detected in a cascade approach. [18]. In addition to using the redundancy of scales and coefficients, this approach used the interpretation of the wavelet transform as the cross-correlation function between input traffic and wavelets.

Wavelet transforms simultaneously perform a multiresolution analysis on a signal. Examining a signal at different resolutions allows key features to be extracted. For example, using Bayesian analysis after a wavelet transform allows for smooth or abrupt changes in variance and frequency in a time series [19]. Sampling network traffic at different rates allows self-similarity, or short or long range dependence [20,21], to be characterized. Such an approach is useful because it can account for changes in network activity related to changes in the legitimate demand for services. When traffic does not exhibit self-similarity or changes to its long or short range dependence, then anomalies may exist [22]. Challenges remain for accurately measuring the long-range dependence and the mixture of short and long-range dependent behavior characteristics [23]. Improved options exist to estimate parameters such as using an iterative process [24] or using combinations of approaches [25,26]. The main disadvantage of these approaches is that they depend on an underlying model of the data.

Using histograms of traffic features allows the construction of data-driven models that may describe traffic more accurately than those made with self-similarity methods [27]. Histograms allow finer modeling of traffic features, and as a result they may be applicable to a wider range of networks than possible when using a multiresolution method. Moreover, the histogram approach may be useful for low traffic rate attacks that produce limited change in traffic load and that may be difficult to detect when examining the range-dependent characteristics of the traffic. In this approach, histograms are built to describe characteristics of network features and arranged in a space based on their similarity. Classification techniques can then be used to identify patterns of typical behavior and compared with real-time behavior to identify anomalies. The general idea is that network anomalies may distort normal patterns of features. The use of histograms can be applied to different scenarios based on several parameters including feature selection, similarity measurement, clustering, extracting models, and classification. However, classification methods rely on the availability of accurate labels of traffic and anomalies.

Our approach uses a classification method followed by spectral analysis. The actual anomaly detection could in principle be performed by a variety of methods. Like the histogram approach, our work is also data-driven in that it uses samples of network traffic to characterize traffic. However, we did not construct histograms but instead group similar, small sections of samples using a nearest-neighbor approach. These groups were combined into a two-dimensional (2-D) array and then a 2-D wavelet transform was computed. The 2-D grouping increased the energy compaction of the data when compared to using one dimension (1-D). The wavelet transform of the traffic model was extracted and, when reconstructed in its original space, anomalies could be more easily detected. The extraction of the traffic model is essentially a denoising algorithm, which basically removes some of the variance from the data so the anomaly is more easily detected. Because only small segments of traffic are considered when grouping, many