Available online at www.sciencedirect.com

SciVerse ScienceDirect

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

Computers & Security

WAVE-CUSUM: Improving CUSUM performance in network anomaly detection by means of wavelet analysis

*C. Callegari**, *S. Giordano*, *M. Pagano*, *T. Pepe*

Dept. of Information Engineering, University of Pisa, Via Caruso 16, 56122 Pisa, Italy

ARTICLE INFO

Article history:
Received 23 December 2011
Received in revised form
2 April 2012
Accepted 6 May 2012

Keywords:
Network security
Intrusion detection system
Network anomaly detection
CUSUM
Wavelet analysis

ABSTRACT

The increasing number of network attacks causes growing problems for network operators and users. Thus, detecting anomalous traffic is of primary interest in IP networks management and many detection techniques, able to promptly reveal and identify network attacks, mainly detecting Heavy Changes in the network traffic, have been proposed. Among these, one of the most promising approach is based on the use of the CUSUM (CUmulative SUM). Nonetheless, CUSUM performance is strongly affected by its sensitivity to the presence of seasonal trends in the considered data.

For this reason, in this paper we propose a novel detection method based on the idea of performing a pre-processing stage of the data by means of wavelets, aimed at filtering out such trends, before applying the CUSUM algorithm.

The performance analysis, presented in the paper, demonstrates the efficiency of the proposed method, focusing on the performance improvements due to the pre-processing stage.

1. Introduction

In the last few years the Internet has experienced an explosive growth. Along with the wide proliferation of new services, the quantity and impact of attacks have been continuously increasing. The number of computer systems and their vulnerabilities have been rising, while the level of sophistication and knowledge required to carry out an attack have been decreasing, as much technical attack know-how is readily available on Web sites all over the world.

As a consequence, many research groups have focused their attention on developing novel detection techniques, able to promptly reveal and identify network attacks, mainly detecting Heavy Changes (HCs) in the traffic volume (Brutlag, 2000; Lakhina et al., 2004; Zhang et al., 2005; Thottan and Ji, 2003).

Nevertheless the seasonal nature of the Internet traffic, characterized by cyclic variation (e.g., daily and weekly trends), makes somehow difficult to distinguish a network anomaly from a "normal" variation of the distribution of the traffic, taking to systems that are strongly affected by high percentages of false positives. For such a reason particular attention has to be devoted to the development of methods able to correctly filter out the seasonality of the data so as to reduce the number of false alarms.

To this aim, in this work, we propose to use one of the most promising techniques for detecting changes in the traffic volume, namely the CUSUM (CUmulative SUM) algorithm (Salem et al., 2010), combined with a pre-filtering stage, realized by means of the wavelet transform.

The main idea of the CUSUM (Basseville and Nikiforov, 1993) is to detect changes in the distribution of a given time

* *Corresponding author.*
E-mail addresses: christian.callegari@iet.unipi.it (C. Callegari), stefano.giordano@iet.unipi.it (S. Giordano), michele.pagano@iet.unipi.it (M. Pagano), teresa.pepe@iet.unipi.it (T. Pepe).
0167-4048/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cose.2012.05.001

series, and is applied in the anomaly detection field, considering that the distribution of the traffic descriptors should change between before and during the occurrence of a network anomaly. Hence, it is clear that the method performance is strongly affected by the cyclic variation of the Internet traffic (as an example, let us consider that the traffic distribution change between day and night could deceive the method).

To solve such an issue, in this paper, we propose to combine a "classical" CUSUM based approach together with wavelet analysis. In more detail the latter is used to filter out the seasonal trends in the network traffic before applying the "real" anomaly detection algorithm, based on the CUSUM method.

It is worth noticing that the aim of our work is to demonstrate how the proposed pre-filtering stage can lead to some improvements in the anomaly detection performance. In this paper we have chosen the CUSUM method because it has been shown in the literature to be promising, but in general we can think of also applying wavelet pre-filtering to other detection method.

In the paper we also face the problem of scalability. Indeed "classical" approaches based on the analysis of the single traffic flows result to be definitely infeasible in real backbone networks, even using dedicated hardware. Moreover the use of standard aggregation techniques (e.g., ingress link aggregation) has been demonstrated to be ineffective in such a context. For such reason in our method we make use of the sketches, probabilistic data structures that allow us to randomly aggregate the traffic flows. In more detail, in this paper we have chosen to use the reversible sketches that are able to also identify the traffic flows (inside an aggregate) responsible for the detected anomalies, overcoming the limitations of "classical" sketch.

The experimental results, obtained by testing our system over real traffic traces, collected in the Internet2/Abilene backbone network, demonstrate the efficiency of our method, which outperforms the "classical" method in terms of both correct detection rate and false positive rate.

The remainder of this paper is organized as follows: next section discusses some related works, while Section 3 gives a detailed description of the proposed architecture. The in Section 4 we discuss the experimental results and, finally, Section 5 concludes the paper with some final remarks.

2. Related work

Due to their properties, both Wavelet transform and CUSUM algorithm are quite "classical" approaches to detect irregular patterns in traffic traces.

Given the vast amount of literature on anomaly detection, in the following we only review some of the most relevant works, referring to (Thottan et al., 2010) for a more complete overview on the topic.

The first detailed work on the study of network anomalies by means of wavelet analysis is represented by Barford et al. (2002), where the authors apply general wavelet filters to the data strings of Internet traffic measurements, seen as a generic signal. The authors propose a platform for network measurement, called IMAPIT (Integrated Measurement Analysis Platform for Internet Traffic). Even if this work has inspired us the idea of performing a pre-filtering stage, before applying the CUSUM algorithm, the detection stage is based on a completely different approach.

An important work (Donoho et al., 2002) discusses the application of the wavelet transform to detect anomalies, by monitoring the access link connecting a given site to the Internet. The idea is to characterize the packets generated by the keystrokes in a SSH connection. The authors also discuss some evasion techniques. Even if quite interesting the work just focuses on a very particular kind of Internet traffic, while our work is much more "general".

A framework for real time wavelet-based analysis of network traffic anomalies, called Waveman, is then proposed in Huang et al. (2008). The system is based on the application of signal processing techniques in IDSs (Intrusion Detection Systems) and works on two distinct metrics, namely percentage deviation and entropy. The aim of the work is to evaluate the performance of various wavelet functions in detecting different types of anomalies like Denial of Service (DoS) attacks and port-scans. The tests are only performed on a few minutes of traffic leading up to the attack, the attack itself, and a few minutes of normal traffic after the attack. This experimental methodology strongly affects the general applicability of the proposed work.

Recent work (Lu and Ghorbani, 2009) proposes a new network signal modeling technique for detecting network anomalies, combining wavelet approximation and system identification theory. The system is based on the selection of 15 network flow-based features modeling the normal daily network traffic; wavelet approximation and the ARX (Auto Regressive eXogenous) system prediction technique are used together to detect network attacks. The authors also perform the comparison of four different wavelet basis functions and a completed analysis for the full 1999 DARPA network traffic dataset is carried out. The results of this work have been used in our method for selecting the best mother wavelet function to be used.

Finally it is worth noticing that the combined use of wavelet transform and the sketches to detect network anomalies has only been applied in a couple of works. In more detail, in (Pukkawanna and Fukuda, 2010) the authors apply the method already described in (Barford et al., 2002) to the time series given by the temporal evolution of the single buckets of a sketch.

Regarding the CUSUM algorithm, it has been widely used in the field of network anomaly detection. This is strongly justified by the fact that being a sequential analysis technique, typically used for monitoring change detection, the application to network anomaly detection is quite straightforward. Nevertheless, the original version of the algorithm imposes several constraints that are usually not respected in the field of anomaly detection (see next Section for more details). Hence the non parametric version of the CUSUM is usually applied (Tartakovsky et al., 2006).

To cite some of the most interesting works, in Wang et al. (2002) the authors propose to aggregate the whole traffic in one flow, and to use a non parametric version of CUSUM for detecting TCP SYN flooding, while in Siris and Papagalou