Information Sciences 420 (2017) 313–328

Contents lists available at ScienceDirect

Information Sciences

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

Anomaly detection using the correlational paraconsistent machine with digital signatures of network segment

Eduardo H.M. Pena[a,*], Luiz F. Carvalho[b], Sylvio Barbon Jr.[b], Joel J.P.C. Rodrigues[c,d,e,f], Mario Lemes Proença Jr.[b]

[a] *Federal University of Technology Paraná, Toledo, Brazil*
[b] *Computer Science Department, State University of Londrina, Londrina, Brazil*
[c] *National Institute of Telecommunications (Inatel), Brazil*
[d] *Instituto de Telecomunicações, Universidade da Beira Interior, Portugal*
[e] *ITMO University, Russia*
[f] *University of Fortaleza (UNIFOR), Brazil*

ARTICLE INFO

Article history:
Received 13 January 2015
Revised 18 August 2017
Accepted 23 August 2017
Available online 24 August 2017

Keywords:
Traffic characterization
Anomaly detection
Network management
Paraconsistent logic
ARIMA
ACO

ABSTRACT

This study presents the correlational paraconsistent machine (CPM), a tool for anomaly detection that incorporates unsupervised models for traffic characterization and principles of paraconsistency, to inspect irregularities at the network traffic flow level. The CPM is applied for the mathematical foundation of uncertainties that may arise when establishing normal network traffic behavior profiles, providing means to support the consistency of the information sources chosen for anomaly detection. The experimental results from a real traffic trace evaluation suggest that CPM responses could improve anomaly detection rates.

1. Introduction

Considered to be some of the most important resources in a modern environment, computer networks must provide means to satisfy demanding requirements. Moreover, along with their rapid growth comes a need for automating management functions to prevent network abuse and reduce the cost of ordinary operations. In this context, a network anomaly detection system is an important component of a security management infrastructure for computers and networks.

Network traffic anomalies have become a troubling issue for both network administrators and end users because they typically change the normal behavior of a network traffic in a malicious or unintentional manner, resulting in the congestion and depletion of available resources. Apart from reducing performance, abnormal activities may also interrupt the operation of services on a network, incurring substantial losses for universities, government agencies, and companies in general [25].

Many challenges restrict the widespread setting of anomaly detection techniques. For example, defining the normal behavior is very ambitious because it can evolve over time and domains [21]. Moreover, filtering techniques may not be able to remove only the actual noise from training data [40]. In addition, the complexity in the tuning, configuration, and de-

[*] Corresponding author.
E-mail addresses: eduardohmpena@gmail.com (E.H.M. Pena), luizfcarvalhoo@gmail.com (L.F. Carvalho), barbon@uel.br (S. Barbon Jr.), joeljr@ieee.org (J.J.P.C. Rodrigues), proenca@uel.br (M.L. Proença Jr.).

http://dx.doi.org/10.1016/j.ins.2017.08.074
0020-0255/© 2017 Elsevier Inc. All rights reserved.

ployment of available solutions may lead to rough requirements and constraints [39]. Another major difficulty is the current low detection efficiency of available systems, for example, high false-positive rates [7]. However, despite all these challenges, anomaly detection techniques have still been widely investigated because they consider several interesting research problems.

As practical measures that mitigate the trespassing of networks, anomaly detection techniques are typically divided into two main categories: signature- and anomaly-based [7]. Signature-based techniques rely on templates of well-known attacks to match and identify intrusions, require a regular update of their signature rules, and are not generally efficient against novel intrusions. In contrast, anomaly-based techniques can detect unknown attacks as a result of a strategy that analyzes deviations in the real traffic behavior from normal patterns. Anomaly-based techniques are generally related to a pair (M, λ), where M is the model of a normal network operation, and λ is a defined rule to estimate the deviation from M used to detect anomalous activities [7].

The nature of input data remains a common issue, regardless of the technical perspective supported by anomaly detection solutions [7]. In this context, the traffic flow analysis has drawn the interest of the research community because of the increasing support of flow tools from network equipment manufacturers [31]. Abnormal behavior detection can be based on the features extracted from different metrics of network traffic flows. Moreover, a proper correlation between them can be established to provide a better perspective of an event. These procedures have shown great potential in reducing unwanted notifications while also helping establish the underlying problem or condition that produces the anomalous events [47].

One of the foundations of our proposal for anomaly detection is the digital signature of the network segment flow analysis (DSNSF) used as normal traffic behavior profiles for servers or segments of the network [6,35]. In this study, DSNSFs are arranged under the rules of two models, namely the autoregressive integrated moving average (ARIMA) [34] and the ant colony optimization for digital signature (ACODS) [12]. These models have distinctive features. ARIMA is a traditional time series forecasting model, while ACODS is a metaheuristic typically used for optimization. Through the analysis of flow records, each of them can structure different DSNSFs used as a basic standard or level for selected traffic features. Aiming to assimilate DSNSFs from both models and common disturbance of traffic features caused by network-wide anomalies, we take advantage of the paraconsistent criteria to frame a tool for anomaly detection, called the correlational paraconsistent machine (CPM).

The CPM design is based on a nonclassical logic known as the paraconsistent logic (PL) [19], which is applied for the mathematical foundation and interpretation of the uncertainties associated with normal traffic behavior profiles and real measurement evaluation for anomaly classification. The interpretation of uncertainties is inspired by the usual expertise of network administrators that benefits from the historical knowledge of different parameters extracted from network segments to handle events harmful to the network infrastructure. It parallels the behavior of traffic features from DSNSFs and real-time measurements to assess evidence of the following proposition or hypothesis:

P_1: Presence of an anomaly in traffic at a time interval t.

These pieces of evidence determine the levels of certainty and contradiction of the presence of anomalies. If the real-time measurements are behaving properly, the levels of certainty and contradiction are the lowest. If not (i.e., the target of network anomalies), the levels of certainty are the highest, while the levels of contradiction are the lowest.

The proposed CPM accomplishes the following contributions: (a) it combines different computational intelligence approaches in a single cooperative solution; (b) it operates on aggregate traffic at network segments, which is an appealing answer for the current bandwidth transmission technologies; (c) it explores the potential of traffic volume and distribution measures together; and (d) it provides alternative reasoning metrics for the evaluation of the level of certainty for the presence of anomalies.

The rest of this paper is organized as follows: Section 2 presents the related work; Section 3 introduces some fundamentals on the paraconsistent logic for the evaluation of information signals; Section 4 describes the concepts, applications, and generation of DSNSFs; Section 5 discusses the proposed CPM design and application for anomaly detection; Section 6 shows the experimental evaluation results; and finally, Section 7 presents our conclusions and future directions.

2. Related work

Many papers have contributed to the network anomaly detection field by means of several approaches. The use of the subspace method was investigated by [29] for anomaly detection in traffic flow data. The method was applied in the flow time series of the traffic coming from randomly sampled data captured in routers from an academic Internet backbone. The subspace method designated a time interval, in which the traffic was considered anomalous. Through a manual inspection, the method then characterized network-wide anomalies, showing the wealth of information that can be extracted from network traffic flows. More recently, the methodology for applying the subspace method was reviewed in [11]. The study also discussed the limitations of existing solutions based on principal component analysis and investigated the use of statistical process control to overcome their main drawbacks (e.g., to best select the number of principal components and how to incorporate dynamics into the model).

To some extent, many approaches for anomaly detection rely on traffic–feature distributions and correlations [26,42]. Yu et al. [46] found that current DDoS attack flows are usually more similar to each other compared to the flows of flash crowds. Thus, the authors proposed a discrimination algorithm using the flow correlation coefficient as a metric to measure the similarity among suspicious flows and differentiate DDoS attacks from flash crowds. The feasibility of the proposed