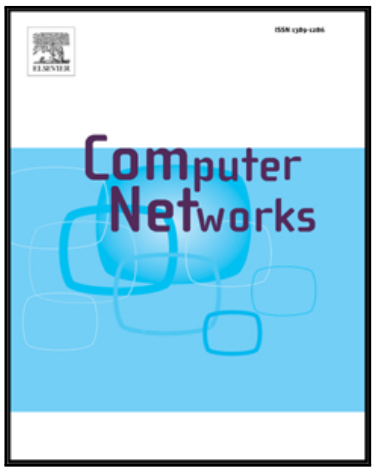

Toward a Reliable Anomaly-Based Intrusion Detection in Real-World Environments

Eduardo K. Viegas , Altair O. Santin , Luiz S. Oliveira

PII: S1389-1286(17)30322-5
DOI: 10.1016/j.comnet.2017.08.013
Reference: COMPNW 6286

To appear in: *Computer Networks*

Received date: 23 December 2016
Revised date: 11 July 2017
Accepted date: 14 August 2017

Please cite this article as: Eduardo K. Viegas , Altair O. Santin , Luiz S. Oliveira , Toward a Reliable Anomaly-Based Intrusion Detection in Real-World Environments, *Computer Networks* (2017), doi: 10.1016/j.comnet.2017.08.013

Toward a Reliable Anomaly-Based Intrusion Detection in Real-World Environments

Eduardo K. Viegas, Altair O. Santin
Pontifical Catholic University of Parana
{*eduardo.viegas, santin*}*@ppgia.pucpr.br*

Luiz S. Oliveira
Federal University of Parana
luiz.oliveira@ufpr.br

Abstract—*A popular approach for detecting network intrusion attempts is to monitor the network traffic for anomalies. Extensive research effort has been invested in anomaly-based network intrusion detection using machine learning techniques; however, in general these techniques remain a research topic, rarely being used in real-world environments. In general, the approaches proposed in the literature lack representative datasets and reliable evaluation methods that consider real-world network properties during the system evaluation. In general, the approaches adopt a set of assumptions about the training data, as well as about the validation methods, rendering the created system unreliable for open-world usage. This paper presents a new method for creating intrusion databases. The objective is that the databases should be easy to update and reproduce with real and valid traffic, representative, and publicly available. Using our proposed method, we propose a new evaluation scheme specific to the machine learning intrusion detection field. Sixteen intrusion databases were created, and each of the assumptions frequently adopted in studies in the intrusion detection literature regarding network traffic behavior was validated. To make machine learning detection schemes feasible, we propose a new multi-objective feature selection method that considers real-world network properties. The results show that most of the assumptions frequently applied in studies in the literature do not hold when using a machine learning detection scheme for network-based intrusion detection. However, the proposed multi-objective feature selection method allows the system accuracy to be improved by considering real-world network properties during the model creation process.*

Index Terms— *Multi-objective feature selection; Anomaly-based Classifier; Machine Learning-based Intrusion Detection; Intrusion Databases.*

1 Introduction

Machine learning techniques for performing anomaly-based network intrusion detection have been extensively studied over the years. Despite the extensive and promising results reported in the literature [1], pattern recognition in intrusion detection remains mostly a research topic, rarely being deployed in real-world (production) environments [2].

Typically, the main reason for using machine learning is the assumption that it can detect new attacks. This is achieved by specifying only the expected (normal, background traffic) while considering the remaining activity as an intrusion attempt. However, when using a machine learning technique, during the system training examples from all classes are normally required. In other words, an anomaly-based system using a machine learning technique must have a significant number of examples from all the considered classes, and thus requires representative training examples [3].

The lack of public and updated training datasets, as well as of specific evaluation methods that take the intrusion detection properties into account, makes it difficult to adopt anomaly-based intrusion detection in production environments. A typical machine learning evaluation scheme relies on a test dataset. It assumes that the classifier's accuracy rate obtained in the test dataset is the same as the accuracy rate obtained during its usage in production environments.

Unlike other fields, where machine learning is extensively used, the field of intrusion detection involves significantly different characteristics. Because of the highly variable and constant changes in open-world network traffic behavior [4], to create a representative training dataset is a difficult task. Thus, normally researchers incorrectly assume an immutable network traffic behavior and thus evaluate the conceived system on a single dataset, discarding the open-world network traffic behavior characteristics [5], such as non-site-specific, new (different) attack behaviors, highly variable content (service), and continuously occurring changes in the traffic behavior.

Several techniques have been proposed in the literature for the creation of an intrusion dataset [3] [6] [7]. However, the most frequently used dataset [8] remains the well-known DARPA1998 dataset [9], which is now almost 20 years-old. Most approaches used to create a public intrusion dataset attempted to statistically model the user behavior [10]. Normally, a typical and real user is monitored during a certain period and its traffic characteristics are reproduced in a statistically similar manner. Thus, a static user behavior is imposed during the monitored period. Thus, these approaches generate a site-specific traffic behavior that are difficult to reproduce.

The great majority of research studies on the subject were aimed to improve the classifier accuracy on a specific dataset [3] [11] [12], usually on KDD'99 [13], which was created in 1999, through the DARPA1998 database, with several limitations [14] [15], making the obtained results unrealistic [1]. Moreover, little or no effort has been invested in using the obtained models in