Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 127 (2018) 521–530

Procedia Computer Science

www.elsevier.com/locate/procedia

The First International Conference On Intelligent Computing in Data Sciences

Building A Fast Intrusion Detection System For High-Speed-Networks: Probe and DoS Attacks Detection

Taha AIT TCHAKOUCHT*, Mostafa EZZIYYANI

University of Abdelmalek Essaadi, Faculty of Sciences and Techniques, Old road of Airport, Km10, PB.416, 90000, Tangier, Morocco

Abstract

Using computers and other intelligent devices associated with internet has become vital in the modern life. Banking transactions, education, trade marketing, social networking, etc. are all examples of those daily and important operations that rely on such technologies, which have generated a large amount of data transiting with high velocity in the last decade. This was accompanied by an extraordinary growth in number and sophistication of cyber threats, going from opportunistic and unstructured to targeted and highly structured. Thus, detecting intrusions in such circumstances requires high levels of accuracy and efficiency, so that heavy losses are prevented. Many intrusion detection models in the literature do not propose real-time solutions to deal with the aforementioned obstacles. This motivates us to propose a lightweight intrusion detection system, for probe and DoS attacks detection. We select the most important set of features using Information Gain (IG), and Correlation-based Feature (CFS) selection filters, applied on a resampled version of KDD'99. Furthermore, we employ four machine learning methods, namely C4.5, Naïve Bayes (NB), Random Forest (RF) and REPTree, as wrappers. Results show good detection and false positive rates, of around 99.6%, and 0.3% for DoS attacks, and 99.8% and 2.7% for Probe attacks. Processing time is also optimized when evaluated using the best selected feature subset.

*Keywords:*Type your keywords here, separated by semicolons ;

* Corresponding author. Tel.: +212-661301799
*E-mail address:*taha.ait@gmail.com

10.1016/j.procs.2018.01.151

1. Introduction

Since 80s and 90s, Internet has revolutionized the whole world. With the launch of WWW, sectors like Education, Politics, Commerce, Press, Tourism, Mail Services, and Banking are now connected, and related services are reachable through Internet. In recent years, the use of such services along with social networking and multimedia streaming has been generating, daily, a large amount of data over the internet, with a very high speed. However, this couldn't be without risks; Valuable data that should be kept out of the procuration and the visibility of unauthorized users are exposed to destruction or/and disclosure, due to the increasing rate of cyber-threats that have evolved as well both in number and complexity. Indeed, exploits are omnipresent and can happen anytime. Internet web servers are exposed to attacks using automated tools and exploit scripts that capitalize on well-known vulnerabilities. There are several sources for these kinds of automated tools maintained by the hacking community. Since 2016, the targeted attacks landscape has changed considerably, as new motives such as subversion and sabotage are emerging [1]. 2016 and 2017 were marked by a myriad of highly structured attacks including the Mirai botnet of infected Internet of Things (IoT) devices, that was responsible for the biggest storm of DDOS attacks ever recorded, the destructive malware Shamoon targeting organizations in Saudi Arabia, A disk-wiping malware causing power outages in Ukraine, as well as the claimed subversive activities attempting to disrupt US elections and targeting other governments [2]. Intrusion Detection systems (IDS) have emerged as one of the most important security solutions to consider. They have several advantages compared to other security tools. Apart from detection, they can archive event data, allow reports, and encounter novel and complex attacks. All these features make IDSs to offer an additional service that helps in protecting the organizations. Since the works of Anderson [3], Denning [4] and Staniford-Chen [5], that inspired researchers, many IDS models have been proposed to extend research works around IDS technology. An intrusion can be defined as any illegal activity that aims at wiping network resources or getting access to the core system data. Based on the mechanisms used to compromise the information system, intrusions can be classified into four categories [6]. Probe attacks are aimed at gathering information about the target network from a source that is often external to the network. Denial-of-Service (DoS) attacks results in an interruption of the service by flooding the target system with illegitimate requests. Remote-to-Local (R2L) is the attempt to gain illegal access to a system's account by exploiting its vulnerabilities, while User-to-Root (U2R) occurs when a user tries to gain super user privileges. Two families of intrusion detection methods are to mention; Signature-based and Anomaly-based. To recognize attacks, Signature-based approaches such as USTAT [7] and IDIOT [8] require a library of signatures of all known attacks and their variants, while Anomaly-based approaches like W&S [9], consist of establishing a user/network 's normal behaviour, and tracking any deviation that results from an intrusive activity. Some IDSs like NIDES [10] and EMERALD [11] combine the two methods. Although anomaly-based methods are not widely commercialized due to the high rate of false alarms generated, they are of a crucial importance as they can detect zero-day attacks, and thus more related research works are conducted. One of the main techniques used in network anomaly-based systems, is to monitor and capture network traffic, and analyze different features of a TCP/IP connection to look for anomalous patterns that indicate the presence of an eventual attack. KDD'99 [12] use 41 features as described in MADAM ID Framework [6], participating in the DARPA Intrusion Detection Evaluation Program [13]. The 41 features are classified into intrinsic features that are used for general analysis, and traffic and content features, each designed to detect a specific type of intrusions when combined with intrinsic features. Using the 41 features in model building is likely to impact both accuracy and efficiency as some features can be redundant or irrelevant. This can be problematic in high speed networks where any delays can make the system to be compromised for some period of time before raising any alarms. Our main contributions in this work are the following:

- Using filters IG and CFS, as well as wrappers NB, C4.5, RF and REPTree for feature subset selection.
- Using classification methods NB, C4.5, RF and REPTree to select the best 19 features for Probe detection, and the most important 9 features for DoS detection.
- Improving detection and False alarm rates. Results show competitive performances w.r.t the literature.
- Optimizing system's processing time as to represent a real-time solution