Journal of Applied Logic 24 (2017) 109–118

Contents lists available at ScienceDirect

Journal of Applied Logic

www.elsevier.com/locate/jal

Comparison of classification techniques applied for network intrusion detection and classification

Amira Sayed A. Aziz [a,c,*], Sanaa EL-Ola Hanafi [b], Aboul Ella Hassanien [b,c]

[a] *Future University in Egypt (FUE), Cairo, Egypt*
[b] *Faculty of Computers and Information, Cairo University, Egypt*
[c] *Scientific Research Group in Egypt (SRGE), Egypt*

ARTICLE INFO

Article history:
Available online 14 November 2016

Keywords:
Intrusion Detection
Artificial Immune Systems
Intrusion Classification
Machine Learning

ABSTRACT

In a previous research, a multi-agent artificial immune system for network intrusion detection and classification was proposed and tested, where a multi-layer detection and classification process was executed on each agent, for each host in the network. In this paper, we show the experiments that were held to chose the appropriate classifiers by testing different classifiers and comparing them to increase the detection accuracy and obtain more information on the detected anomalies. It will be shown that no single classifier should be used for all types of attacks, due to different classification rates obtained. This is due to attacks representations in the train set and dependency between features used to detect them. It will also be shown that a basic and simple classifier such as Naive Bayes has better classification results in the case of low-represented attacks, and the basic decision trees such as Naive-Bayes Tree and Best-First Tree give very good results compared to well-known J48 (Weka implementation of C4.5) and Random Forest decision trees. Based on these experiments and their results, Naive Bayes and Best-First tree classifiers were selected to classify the anomaly-detected traffic. It was shown that in the detection phase, 90% of anomalies were detected, and in the classification phase, 88% of false positives were successfully labeled as normal traffic connections, and 79% of DoS and Probe attacks were labeled correctly, mostly by NB, NBTree, and BFTree classifiers.

1. Introduction

Computer systems today are usually interconnected, where they are networked into large distributed systems which are essential in industrial computing world. Today's software systems require more trends such as interconnection, intelligence, and ubiquity. This all led to the arise of multi-agent systems. The multi-agent systems mimic human intelligent behavior, and the way humans interact with each other and towards their environment to achieve certain goals. One of the systems that can be implemented successfully and efficiently as a multi-agent system is Artificial Immune Systems (AIS). Artificial Immune System [1,11]

[*] Corresponding author.
E-mail address: amira.aly.fci@gmail.com (A.S. A. Aziz).

http://dx.doi.org/10.1016/j.jal.2016.11.018
1570-8683/© 2016 Elsevier B.V. All rights reserved.

is a research area that involves immunology, computer science, and engineering. Inspired by natural immune systems, three main fields of research lie under AIS: immune modeling, theoretical AISs, and applied AISs. Immune modeling is concerned about immunity models and immune systems simulations. Theoretical AISs research is about explaining and digging into the theoretical aspects of AIS algorithms, their mathematical models, and their performance and complexity analysis. Finally, applied AISs research is about developing and implementing algorithms and computers systems inspired by immune systems, applying them to a variety of real world applications.

Looking into computer systems, we can find that the self/non-self concept applies in the form of normal/anomalous activities and elements. Intrusion Detection Systems (IDS) are powerful security systems that have a variety of types for all protection purposes. They do not replace a certain security tool, but instead they add a defense line against intrusions and threats from inside the system (a computer or a network) as well as from outside – just like and immune system. IDSs can be categorized in different ways, but basically they can be classified into Misuse-based and Anomaly-based IDSs. Misuse IDS, which is also known as signature-based or knowledge-based, depends on detecting intrusions using patterns representing known attacks. These patterns or signatures are compared to captured events to find possible intrusions. Anomaly IDS, also known as behavior-based, depends on building a profile that represents normal behavior of a system by monitoring its activities over time. Then any deviation from that profile is considered an anomaly. Profiles can be static or dynamic, and are developed using many attributes of the system [13,17].

With the diversity and complexity of attacking techniques, many issues are common with IDSs that need to be considered while building an IDS [17]. The most common issue is deriving an expert rule set, which in most cases is updated through a human expert, especially with misuse IDS. With anomaly detection, such an issue is solved using proper representation and definition of attacks such that different types of attacks can be detected using a limited set of rules with variations of these rules. Another issue is the training of behavioral models, where usually normal data only is used for training or two sets of normal and anomalous data. Machine learning techniques are mostly used for training and learning.

In an attempt to overcome the mentioned issues above, we suggested a multi-agent, two-layer classification algorithm, that detects and classifies anomalies in a network. The suggested system IDS combines Genetic Algorithm with Negative Selection Approach as a first layer of anomalies detection. Then selected classifiers are trained and applied to label the detected anomalies in both the normal and anomalous traffic. The immune system is a distributed system composed of different specialized cells with high interactivity between its components to give a coordinated response. Taking this into consideration, many approaches adopted the implementation of an AIS as a multi-agent system.

In previous researches [14,16,19,22] that are similar some way to the suggested system, either specific classifiers are used for each attack type, or classifiers – which were trained using the labeled data – are applied directly to the data set for intrusion detection and labeling unlabeled attacks. The contribution of this paper is to do some comparative analysis to answer questions about how the system will act to different parameters, and which techniques to use for best results and why. It will also investigate whether it is rewarding to feed normal traffic into the classifiers or not.

In a previously published work [8], a multi-agent artificial immune system was implemented, for network intrusion detection and classification. The algorithm applied as an artificial immune system technique is Negative Selection Approach, using Genetic Algorithm. As an intelligent system, data mining is applied throughout the process for best results. Two classifiers were used for anomalies classification, Naive Bayes and Best-First Tree classifiers. Naive Bayes classifier was used for attacks that have low representation in the training data set as it has proven to give better results than other classifiers in a previous experiment. The BFTree classifier was used for the remaining attacks classification, as it also proved to give better results than other more complex classifiers in the experiments shown in this paper.

The paper is organized as follows. Section 2 presents a background about the basics of the classification techniques. In section 3, the proposed approach and system model is explained with its different components