Intrusion Detection Systems using Data Mining Techniques: A comparative study

Mohamed Haddadi
Département des sciences commerciales
Faculté des sciences économiques, commerciales, et des sciences de gestion,
Université M'hamed Bougara de Boumerdes,
Avenue de l'indépendence,Boumerdes 35000, Algérie
m.haddadi@univ-boumerdes.dz
haddadimohamed2013@gmail.com

Abdelhamid Khiat
Networks and Distributed Systems Division
Research Center on Scientific and Technical Information, Algeria
a.khiat@dtri.cerist.dz

Nacera Bahnes
1-Industrial Computing and Networks Laboratory (RIIR) Computer science Department, University of Oran 1 Ahmed Ben Bella, PO Box 1524 El M'naouar, Oran, Algeria
2- Computer Science Department, Faculty of Exact Sciences and Computer Science
University of Abdelhamid Ibn Badis, 27000, Mostaganem Algeria.
Nacerabahnes2020@gmail.com

2021 International Conference on Information Systems and Advanced Technologies (ICISAT) | 978-1-6654-7824-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICISAT54145.2021.9678485

***Abstract*— Data mining tools are widely used in computer networks. The well-known and mostly used tools to secure computers and network systems are WEKA and TANAGRA. The purpose of this study is to compare these two tools in terms of detection accuracy and computation time. This comparison was conducted using a well-known NSL-KDD dataset. Experiments show that TANAGRA achieves better results than WEKA in detection accuracy. But, TANAGRA is competitive with WEKA in terms of computation time.**

Keywords*— *tanagra, weka, NSL KDD dataset, data mining, IDS, experimental comparison.

I. INTRODUCTION

Nowadays, cyber-attacks become a real threat for computers and network systems. In practice, the most used attack of cyber-attack is the Denial of Service (DoS) attack. This DoS attack aims to deny the authorized operators to access the services provided by the network system. Therefore, there are plenty of DoS and Distributed Denial of Service (DDoS) attacks that are mentioned in [1], like SYN, UDP, and ICMP flood attacks.

In the last two decades, several data mining tools have been developed to detect cyber-attacks. But, the widely used are WEKA and TANAGRA for classification of several attacks. Each of which contains various well-known algorithms classified in several categories, like decision tree, neural network, and clustering. In the decision tree, there are several types, such as C4.5 [2], which is the well-known algorithm, whereas in the neural networks, the most used algorithm is Multi-Layer Perceptron (MLP) [3]. For clustering, the most utilized algorithm is K-means [4]. All of the machine learning algorithms of the WEKA tool are implemented in the Java language by Waikato University in New Zealand carrying out and analyzing data-sets [5], whereas the TANAGRA tool is implemented in the Delphi language. Both tools are open source softwares which means that the machine learning algorithms can be modified using Java language in WEKA and Delphi language in TANAGRA [6]. For the two data mining tools, there are a few machine learning algorithms with the same appellation, such as Naïve Bayes algorithm [7]. But, others have not the same appellation like C4.5 algorithm in TANAGRA, whereas in the WEKA tool, the C4.5 is called J48 decision tree algorithm [8], which is similar to C4.5 approach.

The rest of the current paper is structured as follows. Section II summaries related work, while Section III presents NSL-KDD dataset. Section IV explains simulation evaluation. Section V provides a brief conclusion.

II. RELATED WORK

Many data mining approaches have been carried out in the last two decades. We start with the scheme of Shrivas et al. [9], who proposed a robust classifier that is a combination of C4.5 and CART classifiers. Experimental results, conducted using NSL KDD dataset in the WEKA environment, demonstrate that the proposed architecture achieves better results than other algorithms. An enhanced J48 algorithm is developed by Aljawarneh et al. [10], who proposed an improved version of the J48 algorithm using the WEKA tool. Results, carried out using NSL-KDD dataset, demonstrate that the enhanced J48 algorithm has high detection accuracy compared to other existing approaches. Another efficient classifier is developed by Hola et al. [11]. This classifier based on various existing feature selection techniques using WEKA and TANAGRA tools to remove insignificant features from NSL-KDD dataset. Experimental results demonstrate that the proposed classifier has the highest accuracy with fewer features. Nalavade et al. [12] conducted a comprehensive study and statistical analysis on KDD99 dataset using two important and popular data mining tools (i.e., WEKA and TANAGRA). Another enhancement of Iterative Dichotomiser 3 (ID3) and Multi-Layer Perceptron (MLP) is developed by Beghdad et al. [13]. This study appropriate attributes of 10% KDD cup 99 dataset and two well-known data mining tools (i.e., WEKA and TANAGRA). Gnanaprasanambikai et al. [14] proposed significant feature extraction, feature selection, and a classification approach for traffic anomaly intrusion detection utilizing NSL-KDD dataset and a well-known data mining tool (i.e., WEKA).

III. NSL-KDD DATASET

The NSL-KDD dataset is a refined variant of the original KDD cup'99 [15]. So, the NSL-KDD dataset is very easy to