UNIVERSITI TEKNOLOGI MARA

ENHANCED AI-BASED ANOMALY DETECTION METHOD IN THE INTRUSION DETECTION SYSTEM (IDS)

KAYVAN ATEFI

Thesis submitted in fulfillment
of the requirements for the degree of
Doctor of Philosophy

Faculty of Electrical Engineering

December 2019

ABSTRACT

Intrusion detection systems (IDS) are vital to cybersecurity, particularly with the presence of various networked computer infrastructures. An efficient IDS uses computational methods as techniques of machine learning (ML) to enhance the rates of detection to obtain the lowest false positive rate, although such rates tend to be reduced by the big amount of irrelevant features as an optimization issue. Data clustering, clustering items from information into significant clusters. Based on the above components and circumstances, many studies have been performed on data clustering problems. Despite attempts to solve the data clustering issues, there are also many variants of modified algorithms in traditional information clustering that attempt to solve issues such as clustering algorithms based on condensation. These algorithms are aimed at overcoming them in terms of offering high accuracy and reduced computational time, the quality of their outcomes still does not fulfill researchers. Moreover, shortage of reliable methods on a new dataset for the intrusion detection system and anomaly detection in terms of classification is an issue. Thus, this study is looking for better, new dataset and more reliable optimized method for detecting the intrusion with highest accuracy. One of the major ML problems is classification and it is believed that many previous researchers did not apply their methods on the latest and updated dataset for testing and validation, therefore the results may no longer be applicable and are not reliable within the current attacks. Further, the shortage of efficient feature selection techniques gives rise to low accuracy in anomaly detection. One of the main steps after the data collection stage of any method is selecting a subset of the features to be used for the feature selection process. Some of the previous researchers used a feature set selection which is introduced for IDS but there still shortage in their detection rate and selected amounts of features. To be able to address the challenges that mentioned above in this study, an architecture is proposed in order to select relevant feature subsets and improve clustering accuracy. In this study, researcher is trying to improve the clustering of data using an efficient technique via Enhanced Binary Particle Swarm Optimization (EBPSO) as feature selection. Also, this research will use the most updated dataset wich called CICIDS2017 that it covers the majority of current intrusion and attacks. This approach that is according to the DNN model reduces irrelevant features in the intrusion detection data sets of CICIDS2017 to improve the accuracy and cluster high-scale data sets. This strategy includes a number of components that are a novel approach to clustering generation. In fact a data clustering method is proposed consisting of separate outputs: (i) To select a relevant subset of original features based on our proposed algorithm; which is Enhanced Binary Particle swarm Optimization (EBPSO), (ii) To mine data using various data chunks (windows) and overcome a failure of single clustering. An experimental analysis is conducted by several experiments to assess the efficiency of the suggested methods that have been tested within the benchmark datasets, namely CICIDS2017. In comparison to different metaheuristic algorithms for feature selection, experimental outcomes indicate that the suggested method is capable of reducing dimensionality cost, the number of irrelevant features and produce reasonable accuracy. Experiments demonstrate and prove that the proposed EBPSO method produces better accuracy mining data and selecting subset of relevant features comparing other algorithms. In addition, experiments prove that the enhanced algorithm shows a higher performance through lower false positive, higher accuracy, and better CPU time.

ACKNOWLEDGEMENT

In the name of GOD, who has always blessed me, support and help me during my studying and living. I wish to thank God for giving me the opportunity to embark on my Ph.D. and for completing this long and challenging journey successfully.

Without a doubt, my gratitude and special thanks go to my supervisor Prof. Ir. Dr. Habibah Hashim for all her support, time and considerations. Her patience and encouragement gave me the motivation to work on this research until its successful completion.

This thesis is dedicated to the loving memory of my very dear late father who always has his praying was with me towards my study and life. And also, my special appreciation goes to my mother for the vision and determination to educate me, support and help me during my studying and living. This piece of victory is dedicated to both of you.

I would like to express my gratitude and sincere thanks to my brothers and sister for their blessings, affection, understanding, and support me for completing this long and challenging journey.

In addition, Special thanks to the others who have directly or indirectly helped me in the completion of my work.

TABLE OF CONTENT

Page

CHAPTER ONE
INTRODUCTION

1.1 Overview

This chapter begins with an overview of the chapter and then proceeds with the introduction and background into the research work to provide a clear understanding of the area of study. The chapter goes on to discuss the problem statement, research questions, objectives, scope and significance of the study. Eventually, this chapter concludes with a description of the thesis organization. The purpose of this section would be to offer the introduction and also to provide the generic construct of the thesis.

1.2 Introduction

To understand this area of research, have to study some existing process and definitions from the field of intrusion detection, anomaly detection and digital forensic. The complete details about the aforementioned area will be explained in Chapter Two. Previous works in designing, developing as well as deployment for the Intrusion detection system (IDS) have been referenced prior to embarking on this research. Based on the practical experience of previous researchers, they provided some recommendation about technological choices to support the creation of virtual environments. In this preliminary study, it can be concluded that one of the limitations of previous work is the shortage of an assessment of the accuracy of the data obtained during the reference mapping stage owing to the lack of accurate truth requiring an intrusive strategy.

An intrusion detection system (IDS) usually detect anomalies based on the signature method. Usually, a method for defining intrusions on a network includes storing signature profiles which detect patterns along with network intrusions in a signature database and developing principles according to signature profiles. Dataset passing through the network are classified based on developed classification rules. Classified packets are submitted into a signature engine in order to compare