UNIVERSITI PUTRA MALAYSIA

"A comparison of different technique in flow based anomaly detection"

Mr. Mohammad Salah (GS41892)

FSKTM 2017 10

"A comparison of different technique in flow based anomaly detection"

By: Mr. Mohammad Salah (GS41892)

Supervisor: Dr.Fahrul Hakim Ayob

.

Thesis Submitted to the School of Graduate Student, University Putra Malaysia, in Fulfillment of the Requirement for the Degree of Master of Computer science

January 2017

COPYRIGHT

DEDICATION

This thesis is dedicated to:

The sake of Allah, my Creator and my Master,

My great teacher and messenger, Mohammed (May Allah bless and grant him), who taught us the purpose of life,

My beloved Parents,

My Wife,

My Son

My Brother and Sister,

And all my friends,

For

Their Endless Patience and Support

ABSTRACT

Abstract of the thesis presented to the Senate of University Putra Malaysia in fulfilment of the requirement for the degree of Master of Science

Stability Improved of Improved Low Energy Adaptive Clustering Hierarchy Protocol for Wireless Sensor Networks

By

Mohammed Salah Taha (GS41892)

January 2017

Supervisor: : Dr.Fahrul Hakim Ayob

Faculty: Computer Science and Information Technology

By performing network traffic analyzing in different datasets, Intrusion Detection Systems (IDS) that works based on anomaly techniques learn the pattern of anomalous and normal behavior. The huge data size in IDSs dataset to process is known as the trend challenge. It causes high false alarms rates and low rates of detection. In this proposal, a new method which functions based on the Online Sequential Extreme Learning Machine (OS-ELM) is introduced for detecting intrusions in the network. Our proposed method detect anomaly by using alpha profiling technique and by utilizing a group of filtered, feature selection techniques based on

Consistency and Correlation has eliminated the inappropriate features. Beta profiling technique has been used in order to decrease the training dataset's size, as an alternative for sampling technique. In order to evaluate the efficiency of the proposed method we used the standard version of Network Security Laboratory-Knowledge Discovery and Data Mining (NSL-KDD 2009) dataset. According to the primary achieved results from our experiments, it is assumed that our proposed IDS method can achieve lower rate of false positive and higher accuracy when using NSL-KDD dataset. It can also be seen that our proposed method is more efficient than conventional methods in intrusion detection.

ACKNOWLEDGMENTS

To my Lord Allah Almighty, I am thankful for the blessings and virtues, and for reconcile, strength, patience, courage, and determination he gave me to complete this work to the fullest, Alhamdulillah.

I would like to extend my gratitude to : Dr.Fahrul Hakim Ayob, for his supervision, advice, and guidance from the very early stage of this project as well as giving me extraordinary experiences throughout the work. Above all and the most needed, he provided me unflinching encouragement and support in various ways.

My warmest gratitude goes to all of my family members, especially my father, my mother who always believed in me, gave me all the possible support, and being patient with me for years, providing me with everything, just to make me focus on my goals.

I would like to thank my wife for her endless support in so many aspects, by giving me advice and guidance throughout my research and, of course, sharing my happiness and sorrow. I am also thankful for my brothers and sisters for them support and concern about my study, and them willing to provide me with any support I needed.

Finally, I must extend my sincere thanks to the Ministry of defense in Iraq, especially University of Baghdad for their support by sponsoring me in my study. None the less, my gratitude to the Malaysian people in general for their perfect hospitability in their green land during my study period.

APPROVAL SHEET

This thesis submitted to the faculty of Computer Science and Information Technology of University Putra Malaysia and has been accepted as partial fulfillment of the requirement for the degree of Master of Computer Science.

The member of the Supervisory Committee were as follows:

Supervisor: Dr.Fahrul Hakim Ayob
Department of Communication Technology and Network
Faculty of Computer Science and Information Technology
University Putra Malaysia
Date and Signature: _____

Assessor: Mr. Mohd Noor Derahman
Department of Communication Technology and Network
Faculty of Computer Science and Information Technology
University Putra Malaysia
Date and Sinagutre: _____

DECLARATION

I declare that the thesis is my original work except for quotation and citations which have been duly acknowledge. I also declare that it has not been previously, and is not concurrently, submitted for any other degree at University Putra Malaysia or other institution.

Mohammed Salah Taha (GS41892)

Date: _____

TABLE OF CONTENTS

PAGE

LIST OF FIGURE

CHAPTER 1

INTRODUCTION

1.1 Background

Nowadays that the technology is emerging with a fast pace, threats like viruses, Trojan horses, worms, adware, spyware, root kits are facing networks. Before any kind of data loss happens to the organizations, it is essential that all of these intrusions be identified. All parts of network infrastructures even the internal Local Area Network (LAN) are critically under risk of intrusions. These threats are reducing the network bandwidth efficiency and other resources of computer networks. Advance characteristics such as IP address spoofing, dynamic ports, encrypted payload have been deployed by hackers to avoid detection. In order to detect this kind of intrusions the first step is to discover some patterns among network traffic dataset. Processing the entire data could be very problematic, since the dataset in Intrusion Detection System (IDS) which are machine learning based is enormous and not balanced. Therefore, intrusions need to be recognized through the behavior of network traffic. The main duty of IDS is to detect the malicious activities in the network. To detect the malicious activities and attacks, the normal behavior should be learnt from network traffic dataset patterns by anomaly based IDS. The IDSs which work based on soft computing, holds some artificial intelligence methodologies. These methodologies including but not limited to artificial neural networks (ANN), fuzzy logic, evolutionary computation, artificial immune systems, probabilistic computing, and etc. Here we present a new intrusion detection method which functions based on several factors such as feature selection process, massiveness of dataset for network traffic, high false alarms and low accuracy rate. In order to detect intrusions through performing network traffic dataset process,

we use Online Sequential Extreme Learning Machine (OS-ELM) introduced by (Liang, Huang, Saratchandran, & Sundararajan, 2006). This method is more accurate and faster with only one hidden feed forward layer works with neural network and known as (SHLFN). It is capable to perform instances network process in a group or one by one. The usability of this method in classification has been verified by single iteration performing. By utilizing NSL-KDD 2009 as benchmarked dataset, the evaluation performance of the proposed method has been carried out.

1.2 Related work

1.2.1 Online sequential extreme learning machine (OS-ELM) Method

In order to tackle the restrictions of feed forward neural network caused by slow learning the OS-ELM method has been designed. It has make faster learning pace with better performance for generalization available (Liang et al., 2006). The classification and estimation issue has been solved with fuzzy OS-ELM. In order to enhance the system's learning pace, the extreme learning machine has been used (Avci, 2012). The online sequential extreme learning machine with kernels (OS-ELMK) has been used to predict the non-stationary time series. During the elapsed time for learning process, a memory prediction with limited accessibility will enhance the accuracy when there is at least a function for reduction of an order-of-magnitude is present. Based on the node classification method, the functionality of selected features and links in the social networks is defined by OS-ELM method. For considering the nodes interactions among each other, the node features are used (Sun, Yuan, & Wang, 2015). In order to perform

intrusion detection in IDS the Extreme Learning Machine (ELM) is used. The efficiency of ELM and SVM has been study in many researches. It has been observed that the accuracy of ELM and SVM is similar while ELM might have quicker response comparing with SVM method. However, when performing intrusion detection, the ELM requires less time in comparison with SVM. In this method, when we have a large dataset of network traffic, a smaller set of samples is required to accomplish the evaluation performance of OS-ELM (Cheng, Tay, &Huang,2012). Based on the reviewed literatures, it has been observed that OS-ELM is known as an emerging technique for classification. In the process of overcoming many issues related to classification procedure this technique has presented a solution. The advantage of this method is that in dealing with large size dataset process it requires less time so it becomes the most suitable method for IDS. Therefore, according to the literature researchers are encouraged to utilize the OS-EM method as the most suitable method to be used in network intrusion detection systems.

1.3 Background of the problem

In intrusion detection systems (IDS) several improvements in detection and accuracy taken place which enable these systems to detect all kind of network attacks and more types of anomalous traffics in the existing environments. This system is located in the target network which is meant to be protected. It functions by continuously collecting packets from network in the similar way as a packet sniffer performs in the network. It can detect anomalous activities in the network by collecting and analyzing the network

packets, then warn the network administrator, and then find the attack type and its connections so it can help the administrator to stop the attack from making more damage to the system. Also it can work with the firewall which is known as the essential tool in the category of network security. Commonly, the algorithms of intrusion detection systems are characterized by their two different approaches: misuse and anomaly detection (Depren, Topallar, Anarim, & Ciliz, 2005). Each of these detection systems have its own advantages and disadvantages that need to be study and considered. Nevertheless, each of this detection systems are useful and efficient in different situation.

1.4 Problem Statement

The algorithms of misuse detection are able to identify and detect the attacks by analyzing it based on a technique known as signature of attack. This method is effective when detection of known attacks with truncated error rate is critical. Nevertheless, they are not able to detect zero-day attacks with different properties and feature which are not consider as known attacks. In the other hand, algorithms of anomaly detection are able to analyze and evaluate the normal network traffic and make a pattern from normal network traffic. The functionality of the mentioned techniques is based on the hypothesis that the behavior and activities of attacker is unlike of normal user behavior. The primary function of this method is on the classification of network traffic for an anomalous activity as if its traffic features are totally dissimilar to normal network traffic patterns. The algorithms of anomaly detection are useful for detecting patterns of zero-day attack, though their effectiveness is not in the same level with misuse detection algorithms in the aspect of detection rate when analyzing false positive rates and known attacks, which is

considered as a ratio of normal traffic when they are misclassified. For the currently in used IDS systems, the common challenge is the huge size of dataset which needs to be processed. It causes high false alarms rates and low rates for detection.

For the currently in used IDS systems, the common challenge is the huge size of dataset which needs to be processed.

It causes:

- high false alarms rates
- low rates for detection.
- reduce the number of comparisons and features.

1.5 Project Aim

Whit respect to the fact that the anomaly based IDSs application is to detect new attacks based on the anomalous traffic. And the current challenge in existing systems is to analyze and find anomalous pattern of activity from a big size dataset of network traffic. In order to eliminate this problem several methods have been proposed and tested. But the existing methods are still not efficient enough. Therefore, we propose a system that will be able to tackle this issue and enhance the efficiency of the anomaly based IDS. In order to enhance the network security by utilizing our proposed method, the Online Sequential Extreme Learning Machine (OS-ELM) technique has been used to improve the intrusion detection efficiency.

1.6 Project Objectives

The objective of this proposal it to increase detection accuracy rate and decrease the false alarm rate with utilizing Online Sequential Extreme Learning Machine (OS-ELM) method. In this method the dataset will be analyzed using some preprocessing technique in order to select suitable features using continuous and categorical values in dataset. Then two techniques known as Alpha and Beta profiling as for its final feature selection technique. There is assumed that by utilizing proposed method the result will be close to our desirable result.

The final result of the proposed method is estimated to meet the following outcomes:

- The IDS detection accuracy rate will increase
- The IDS false positive rate will decrease
- By utilizing Alpha and Beta profiling the number of comparisons and features are reduced

1.7 Project Scope

This proposal, describes and clarifies the item used to develop this project. In that direction, as for preparing the classifier, a preprocessing will be performed to mix continues and categorical features in the dataset. Then, a k-fold cross validation method will validation the performance of the proposed method. In the next step, a process called Alpha profiling will make profiles base on the protocol and service features. To accomplish intrusion detection process in the dataset, the OS-ELM has been used. Then

Beta profiling also known as sample reduction to minimize the overall computational time and memory has been performed. This procedure has been performed for three experiments, respectively named Alpha-Full Features, Alpha-FST and Alpha-FST-Beta. The dataset used for these experiments is a version of the well-known NSL-KDD dataset.

6.References

[1]Kim, Gisung, Seungmin Lee, and Sehun Kim. "A novel hybrid intrusion detection method integrating anomaly detection with misuse detection." Expert Systems with Applications 41.4 (2014): 1690-1700.

[2]Depren, O., Topallar, M., Anarim, E., & Ciliz, M. K. (2005). An intelligent intrusion detection system for anomaly and misuse detection in computer networks. Expert Systems with Applications, 29(4), 713–722.

[3]Zhang, J., & Zulkernine, M. (2006). A hybrid network intrusion detection technique using random forests. In Proceedings of the first international conference on availability, reliability and security (pp. 262–269).

[4]Farid, D. M., & Rahman, M. Z. (2008, December). Learning intrusion detection based on adaptive bayesian algorithm. In *Computer and Information Technology, 2008. ICCIT 2008. 11th International Conference on* (pp. 652-656). IEEE.

[5]Al-mamory, S. O., & Jassim, F. S. (2015). On the designing of two grains levels network intrusion detection system. *Karbala International Journal of Modern Science*, *1*(1), 15-25.

[6]Fossaceca, J. M., Mazzuchi, T. A., & Sarkani, S. (2015). MARK-ELM: Application of a novel multiple kernel learning framework for improving the robustness of network intrusion detection. *Expert Systems with Applications*, *42*(8), 4062-4080.

[7]Lin, W. C., Ke, S. W., & Tsai, C. F. (2015). CANN: An intrusion detection system based on combining cluster centers and nearest neighbors. *Knowledge-based systems*, *78*, 13-21.

[8]Eesa, A. S., Orman, Z., & Brifcani, A. M. A. (2015). A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems. *Expert Systems with Applications*, *42*(5), 2670-2679.

[9]de la Hoz, E., de la Hoz, E., Ortiz, A., Ortega, J., & Martínez-Álvarez, A. (2014). Feature selection by multi-objective optimisation: Application to network anomaly detection by hierarchical self-organising maps. *Knowledge-Based Systems*, *71*, 322-338.

[10]Hosseini, B. M., Amiri, B., Mirzabagheri, M., & Shi, Y. (2015). A New Intrusion Detection Approach using PSO based Multiple Criteria Linear Programming. *Procedia Computer Science*, *55*, 231-237.

.[11] V. Jaiganesh and P. Sumathi, "Kernelized Extreme Learning Machine with Levenberg-Marquardt Learning Approach towards Intrusion Detection," *Int. J. Comput. Appl.*, vol. 54, no. 14, pp. 38–44, 2012.

[12] M. Bahrololum and M. Khaleghi, "Anomaly Intrusion Detection System Using Gaussian Mixture Model," *System*, pp. 1162–1167, 2008.

[13] S. A. Hofmeyr, S. Forrest, and A. Somayaji, "Intrusion Detection using Sequences of System Calls," *J. Comput. Secur.*, vol. 6, no. 3, pp. 151–180, 1998.

[14] S. Devaraju and S. Ramakrishnan, "Performance Analysis of Intrusion Detection

System Using Various Neural," vol. 9, no. 17, pp. 1033–1038, 2011.

[15] L. M. L. de Campos, R. C. L. de Oliveira, and M. Roisenberg, "Network Intrusion Detection System Using Data Mining," vol. 311, pp. 104–113, 2012.

[16] P. M. Mafra, V. Moll, J. Da Silva Fraga, and A. O. Santin, "Octopus-IIDS: An anomaly based intelligent intrusion detection system," *Proc. - IEEE Symp. Comput. Commun.*, pp. 405–410, 2010.

[17] H. T. Elshoush and I. M. Osman, "Alert correlation in collaborative intelligent intrusion detection systems - A survey," *Appl. Soft Comput. J.*, vol. 11, no. 7, pp. 4349–4365, 2011.

[18] R. Singh, H. Kumar, and R. K. Singla, "TOPSIS based multi-criteria decision making of feature selection techniques for network traffic dataset," *Int. J. Eng. Technol.*, vol. 5, no. 6, pp. 4598–4604, 2013.

[19] D. a. M. S. Revathi, "A Detailed Analysis on NSL-KDD Dataset Using Various Machine Learning Techniques for Intrusion Detection," *Int. J. Eng. Res. Technol.*, vol. 2, no. 12, pp. 1848–1853, 2013.

Wang, C. M., & Huang, Y. F. (2009). Evolutionary-based feature selection approaches with new criteria for data mining: A case study of credit approval data. Expert Systems with Applications, 36(3), 5900–5908.

Singh, R., Kumar, H., & Singla, R. K. (2013). Issues related to sampling techniques for network traffic dataset. International Journal of Mobile Network Communications &Telematics, 3(4), 75–85.

Witten, H., Frank, E., & Hall, M. A. (2011). Data Mining: Practical Machine Learning Tools and Techniques (third ed.). Burlington, Massachusetts: Morgan Kaufman Publisher.

Singh, R., Kumar, H., & Singla, R. K. (2014). TOPSIS based multi-criteria decision making of feature selection techniques for network traffic dataset. International Journal of Engineering and Technology, 5(6), 4598–4604.

Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A Density-based algorithm for discovering clusters in large spatial databases with noise. Second International Conference on Knowledge Discovery and Data Mining (pp. 226–231).

Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009). A detailed analysis of the KDD CUP 99 data set. IEEE Symposium on Computational Intelligence for Security and Defense Applications, CISDA 2009, (Cisda) (pp. 1–6)

Liang, N. Y., Huang, G. B., Saratchandran, P., & Sundararajan, N. (2006). A fast and accurate online sequential learning algorithm for feedforward networks. IEEE Transactions on Neural Networks/a Publication of the IEEE Neural Networks Council, 17(6),1411–1423.

Avci, E., & Coteli, R. (2012). A new automatic target recognition system based on wavelet extreme learning machine. Expert Systems with Applications, 39(16), 12340–12348.

Sun, Y., Yuan, Y., & Wang, G. (2015). An on-line sequential learning method in social networks for node classification. Neurocomputing, 149, 207–214

Cheng, C., Tay, W. P., & Huang, G.-B. (2012). Extreme learning machines for intrusion detection. The 2012 International Joint Conference on Neural Networks (IJCNN) (pp. 1–8)