Quest Journals
Journal of Software Engineering and Simulation
Volume 7 ~ Issue 5 (2021) pp: 03-07
ISSN(Online) :2321-3795 ISSN (Print):2321-3809
www.questjournals.org

Research Paper

An Enhanced Network Intrusion Detection System Using Data Mining

Barida Baah[1], Chioma Lizzy Nwagbo[2]
[1]*Department of Computer Science, Ebonyi State University, Abakaliki- Nigeria*
[2]*Department of Computer Science and Robotics Education, College of Education, Nsugbe, Anambra State, Nigeria*

ABSTRACT
This research work treats the enhanced network intrusion detection system via the utilization of data mining techniques to detect attacks cause by intruder in a network for attackers. Many research works have done in area of intrusion detection in a network we will be reviewing some of the techniques used in data mining for intrusion detection system: In our research we will improve on intrusion detection problems in terms of the data overload, false positive and false negative by introducing an effective algorithm that combine the artificial neural network with the re-enforcement learning in the training of data in the network intrusion detection system. Our system is expected achieved the following results: (1) Provides the evaluation results obtained from the proposed system in the form of a confusion matrix showing how well the system is capable of performing its job, {2) Find a clue that will assist in predicting or detecting any kinds of attacks by an intruder to the network and (3) Show the metrics based approach that evaluates the system from different perspectives; using performance metrics.

KEYWORD: *Network Intrusion, false positive, false negative and data mining*

Received 14 May, 2021; Revised: 28 May, 2021; Accepted 30 May, 2021

I. INTRODUCTION

In today world many cases of network intrusion attack involves flooding or overloading the network, gathering data about the network to attack it from a weak point later, or inserting information into the network to spread and gain access from inside. It's important to keep hacker detection tools active, so you can prevent these vulnerabilities from getting into your system in the first place.

Recently, many institutions have been experiencing a heavy usage of networks within their systems. However, the broad technological expansion that accompanied these networks has brought along various threats to them. These threats included many kinds of malicious programmes that affect the efficiency of networks, such as the transmission of data through the network or data that can be accessible via the network. This issue has urged researchers to improve and develop new techniques to explore and contain such threats [1].

This gives rise to cyber security. Cyber security is a branch of computer technology known as information security. It can be applicable to computer systems and networks within the sectors of communication (email, cell phones), entertainment (digital cable, mp3s), transportation (car engine systems, airplane navigation), shopping (online stores, credit cards) and medicine (equipment, medical records). Cyber security involves protecting the system information by effectively preventing, detecting and responding to attacks [2].

One of the main challenges face in the security management of large-scale high-speed networks is the detection of suspicious anomalies in network traffic patterns is due to Distributed Denial of Service (DDoS) attacks or worm propagation [3][4]. In determination of a secure network it must be able to provide the following merits as stated and explained below:

Data confidentiality: It ensures that all data that are being transferred through the network should be accessible only to those that have been properly authorized.

Data integrity: It is the process of ensuring that all data should maintain their integrity from the moment they are transmitted to the moment they are actually received. No corruption or data loss is accepted either from random events or malicious activity.

Data availability: The network should be resilient to Denial of Service attacks.

The first threat for a computer network system was realised in 1988 when 23-year old Robert Morris launched the first worm, which override over 6000 PCs of the ARPANET network. On February 7th, 2000 the first DoS attacks of great volume where launched, targeting the computer systems of large companies like Yahoo!, eBay, Amazon, CNN, ZDnet and Dadet [5].

These threats and others that is likely to appear in the future which may lead to the design and development of Intrusion Detection Systems. According to webopedia an intrusion detection system (IDS) inspects all inbound and outbound network activity and identifies suspicious patterns that may indicate a network or system attack from someone attempting to break into or compromise a system

II. REVIEW OF RELATED WORKS

Ashraf et al. [6] proposed random forest which is data mining technique after comparing with the two other techniques like naïve bayes and J48 algorithms the result shows that random forest has accuracy level of 99.71% compare to naïve bayes and J48 which produces accuracy level of 96.27% and 99.17% respectively. The work only compares three algorithms but did to extend to other data mining techniques which may be better in terms of performance.

Shanmugavadivu R. [7] proposed system a designed for fuzzy logic-based system to effectively identify the intrusion activities within a network. The proposed fuzzy logic-based system was able to detect an intrusion behavior of the networks since the rule base contains a better set of rules. The system used automated strategy for generation of fuzzy rules, which are obtained from the definite rules using frequent items. The experiments and evaluations of the proposed intrusion detection system are performed with the KDD Cup 99 intrusion detection dataset. The experimental results clearly show that the proposed system achieved higher precision in identifying whether the records are normal or attack one.

Markey [8] enumerated the advantages of decision tree over other classification techniques, one of the main advantages is that it generates a set of rules which are transparent, easy to understand, and easily deployed into real-time technologies as Intrusion Detection systems. However, Rokach and Mimon [9] pointed out that this technique works only with target attributes having discrete values. They also said the greedy characteristic of decision trees leads to another disadvantage that should be pointed out.

h Aslam Khan

Arif Jamal Mali et al.[10] proposed an intrusion detection mechanism based on binary particle swarm optimization (PSO) and random forests (RF) algorithms called PSO-RF that investigate the performance of various dimension reduction techniques along with a set of different classifiers. Binary PSO is used to find more appropriate set of attributes for classifying network intrusions, and RF is used as a classifier. In the preprocessing stage it's applied a reduction in the dimension of the dataset by using different state-of-the-art dimension reduction techniques, and then its reduces the dataset that is presented to the PSO-RF approach that further optimizes the dimensions of the data and then finds an optimal set of features. PSO is an optimization method that has a strong global search capability and is used for dimension optimization.

From the above literature review so far shows in most cases an Intrusion Detection Systems (IDS) have become a standard component in the security of infrastructures as they tend to give room for network administrators to detect policy violations. These policy violations in-turns provide room for range of external attackers trying to gain unauthorized access to insiders abusing their access.

Currently, an Intrusion detection system have major drawback as stated and explained below:

- ***Data overload***: Another aspect which does not relate directly to misuse detection but is extremely important is how much data an analyst can efficiently analyze. That amount of data he/she needs to look at seems to be growing rapidly. Depending on the intrusion detection tools employed by a company and its size there is the possibility for logs to reach millions of records per day.
- ***False positives:*** A common complaint is the amount of false positives that an intrusion detection system will generate. A false positive occurs when normal attack is mistakenly classified as malicious and treated accordingly.
- ***False negatives:*** This is the case where an intrusion detection system (IDS) does not generate an alert when an intrusion is actually taking place. (Classification of malicious traffic as normal)

Data mining can help improve intrusion detection by addressing each and every one of the above mentioned problems effectively. Figure 1 below shows an existing intrusion detection system with firewall for security of the system from intruders while Figure 2 shows the process of data movement from data stage to final knowledge based that does the interpretation and validation of data.

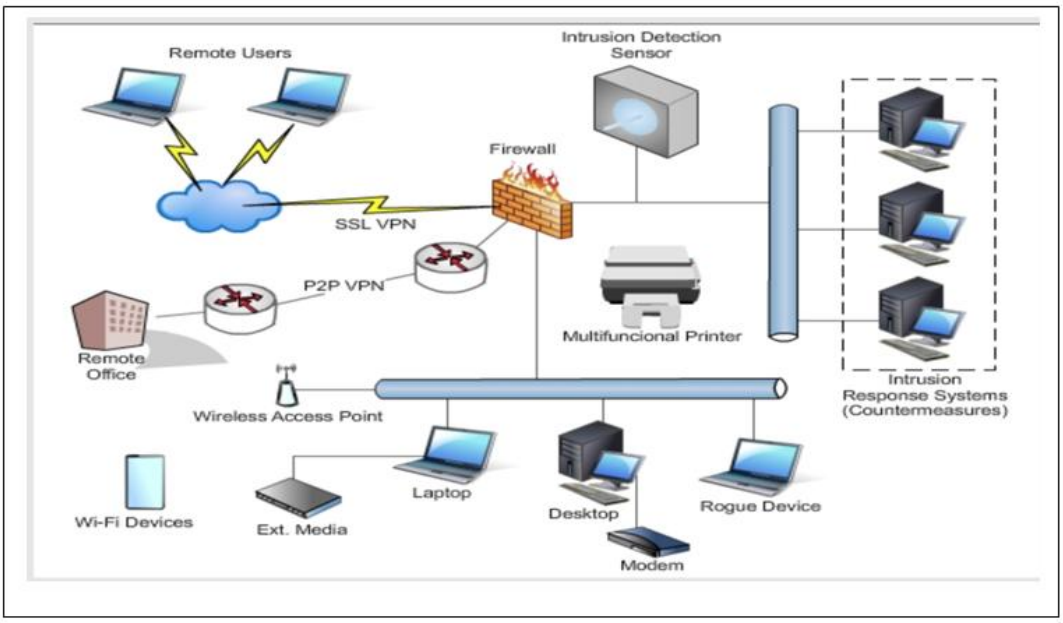

Figure 1: An Intrusion Detection System

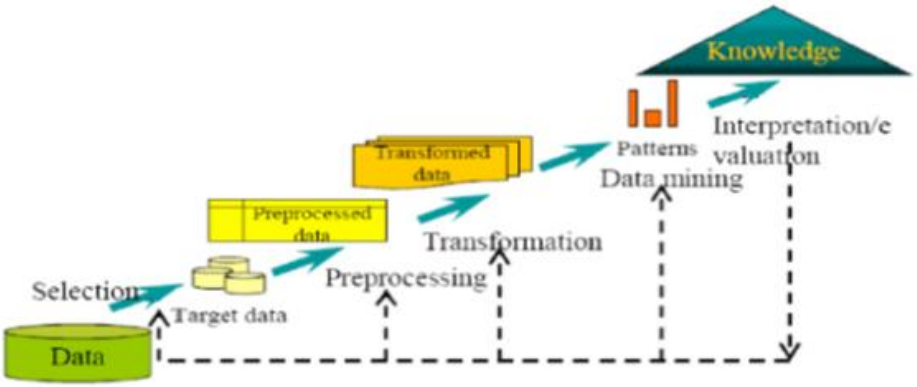

Figure 2: A Transition of raw data to a knowledgeable

In order to demonstrate the originality of our research work our system will be able to monitor all incoming and out coming traffic in our defined network, separate normal activity from false activity, locate false alarm generator and bad signature and locate different ongoing IP address of same activity in a network which will serve as a gap in knowledge that we are aiming to address

III. METHODOLOGY:

The materials that are needed for this research work are laptop with an installed Preprocessor Hypertext Programming (PHP) language and KDD99 dataset for the training in the network intrusion detection system. The research methodology implored here is the Hybrid which combined the qualitative and iterative experimental approach. The qualitative approach will be used to understand some concepts and systems behaviors while the iterative experimental approach will be used when building the systems. In the iterative experimental approach, the system will be initial implemented then tested. Based on the obtained results, the system implementation is modified until reaching to a point where obtained results will be acceptable.

In our proposed system we will be using data mining techniques to detect any malicious packets that SNORT was not able to capture then we will automatically updates all SNORT signatures holder with a new one. It also provides a brief background overview of other data mining techniques used in the proposed system (Decision tree, Neural Network, Fuzzy logic, Reinforcement Learning and Hierarchical clustering). Our proposed system will also goes through the data set used for training and evaluation (KDD99) and which features from which the data will be selected.

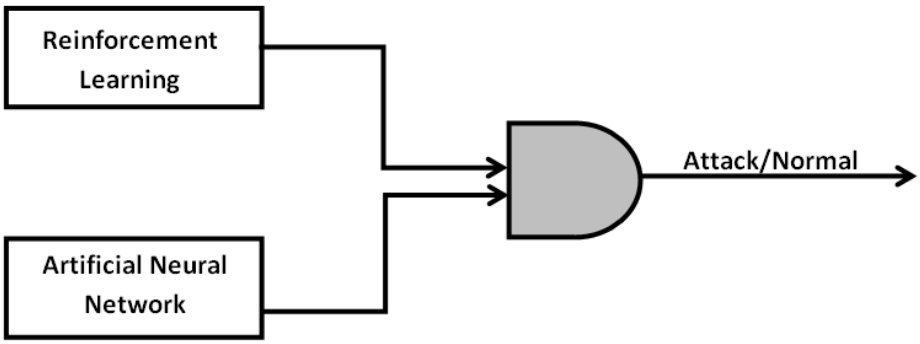

Figure 3: Proposed Hybrid Model Overview

The benefit of using the hybrid approach is that it will increase the intrusion detection rate, some of attacks may not be detected by one of the modules but the other one may be able to detect them. In other words, one module will overcome some of other module shortcomings in detecting malicious traffic. However, there is a chance of increasing the false-positive rate for malicious traffic.

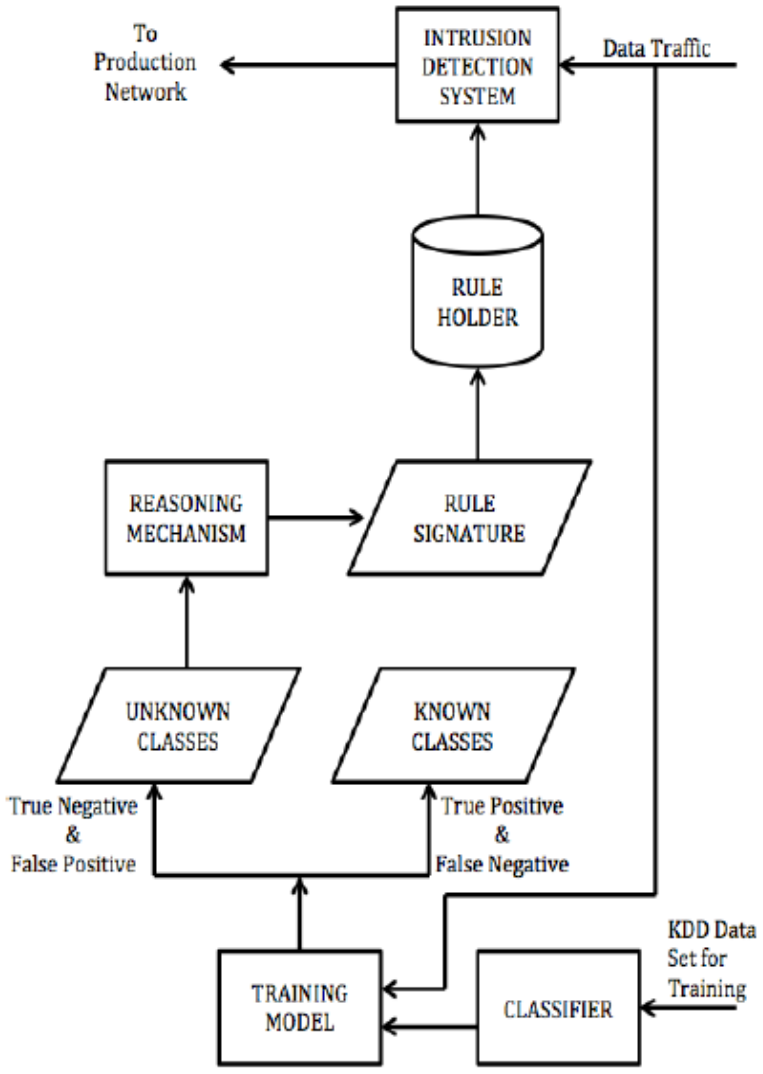

Figure 4: Propose High level Hybrid Process Classifier

The research process consists of the following elements as shown in Figure 4:
• *Intrusion Detection System*: SNORT will be used in this solution as a signature based IDS. In addition to its main functionality as an intrusion detection system it will be used as a network sniffing tool that feeds the training model with the live traffic.
• *Rule Holder*: This contains all signatures used by SNORT to capture attacks matching the stored signatures.
• *Data set and categorisation*: The first step in the research process is to find a reliable high quality network traffic data set, where each packet has been labelled so that the training model is created, and as a result the classification can be used reliably.
• *Feature Selection:* The network packets in the data set and then passed through an attribute evaluator, in order to extract a set of features that can be used effectively to detect intrusions. Non-essential features are known not to be only a bottleneck in terms of cost of computation, but are also factors that contribute towards increased error rates [11].
• *Classifier Module:* This module is responsible for building a classifier using Reinforcement learning and ANN that can compute a model using the most discriminating features in an instance of a data packet, in order to describe a class (concept). This is done by training a classifier, using a pruned set of features, where the objective is that the model created is more generic than the rules (as compared to SNORT) and hence, it outperforms this in accuracy and effectiveness, when compared to general rule-based signature matching systems.
• *Training Model:* This model is the outcome of the classifier module. The results of each classification algorithm will be compared to each other then one of them will be selected to be used as the training model. This model will classify the traffic to either known or unknown classes. The traffic will be passed to the reasoning module in case if it is unknown for further investigation.
• *Reasoning Mechanism:* The purpose of this mechanism is to provide another stage for classifying the network traffic, if the first stage fails to classify it. The reasoning mechanism is based on a hybrid model built using neural network (MLP) and reinforcement learning. The outcome of this module will be in a form of a signature that will be added to the rule base.

IV. CONCLUSION

In conclusion, the system will be of great assist to any business organization or institution as that are on the internet since it will help the checking the of an attack in the network and determine the areas where the attackers is coming from in the network also block such areas also it will identify any false positive or false negative signals through the alarm bell and separate its appropriately.

REFERENCES:

[1]. Al-Saedi, K.H., H. Al-Khafaji, A. Almomani, S. Manickam and S. Ramadass(2011), An approach to assessment of network worm detection using threatening-database mining. Australian Journal of Basic Applied Sci., 5: 2676-2683
[2]. Tjhai, G.C., S.M. Furnell, M. Papadaki and N.L. Clarke (2010) A preliminary two-stage alarm correlation and filtering system using SOM neural network and Kmeans algorithm. Journal of Computer Security, 29: 712-723. DOI: 10.1016/j.cose.2010.02.001
[3]. Christos Douligeris, Aikaterini Mitrokotsa (2004) DDoS attacks and defense mechanisms: classification and state-of-the-art, Computer Networks: The International Journal of Computer and Telecommunications Networking, Vol. 44, Issue 5, pp: 643 - 666, 2004.
[4]. Z. Chen, L. Gao, K. Kwiat(2003), Modeling the spread of active worms, Twenty- Second Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM), Vol. 3, pp. 1890 1900,
[5]. Brian Krebs (2003), A Short History of Computer Viruses and Attacks Washington http://www.securityfocus.com/news/2445
[6]. Nabeela Ashraf1, Waqar Ahmad and Rehan Ashraf(2018), A Comparative Study of Data Mining Algorithms for High Detection Rate in Intrusion Detection System, Annals of Emerging Technologies in Computing (AETiC), Vol. 2, No. 1, pp. 49-5
[7]. LR. Shanmugavadivu (2014), Network Intrusion Detection System Using Fuzzy Logic, Indian Journal of Computer Science and Engineering (IJCSE), Vol. 2 Issue 1 pp. 101-111
[8]. Markey, J. (2011) Using Decision Tree Analysis for Intrusion Detection: A How-To Guide. Sans Institute, Available from http://www.sans.org/reading-room
[9]. Rokach, L. and Mimon, O. (2014) Data Mining with Decision Trees. 2nd Edition, Massachusetts: World Scientific
[10]. Arif Jamal Malik, Waseem Shahzad and (2012), Network intrusion detection using hybrid binary PSO and random forest, Security and Communication Network
[11]. Wei, M., Xia, L., and Su, J. (2011) Research on the Application of Improved K-Means in Intrusion Detection. Communications in Computer and Information Science, 243(2011), pp. 673-678