Analysis of Network Intrusion Detection Technique Based on Machine Learning

Hande Bhagyashree M.[1] Prof.Rokade M.D.[2]
[1,2]Department of Computer Engineering
[1,2]Sharadchandra Pawer College of Engineering, Dumbarwadi, Otur, Maharashtra, India

Abstract— In today's world internet usage is on the rise. Every human task is accomplished with the help of the internet. Network security is an important factor in such conditions. Maintaining network availability, integrity and confidentiality is important. This requires a network administrator to accept various types of intrusion detection systems (IDS) that help monitor network origins of unauthorized and dangerous activities. An intrusion Detection System (IDS) is a system that monitors traffic on the network for suspicious activity and it alerts the issues when that activity is available. It is a software application that scans the network or the program for malicious activity or policy violations. Any malicious or illegal engagement is usually reported to the regulator or collected through an event security information and event management system (SIEM). The SIEM system integrates output from multiple sources and uses alarm filtering methods to distinguish hazardous activity from false alarms. The presented paper gives a clear idea about what is the need for IDS and also what are the network intrusion detection techniques using machine learning. The techniques are mainly divided into two parts. first is classical artificial intelligence (AI) and the second is computational intelligence (CI). The proposed paper will explore the brief scenario about these two methods for developing effective IDS.

Keywords: Network Intrusion Detection, Artificial Intelligence, Machine Learning, Neural Networks, Genetic Algorithm, Fuzzy Logic, K-Nearest Neighbors, Naive Bayes, Decision Tree, Support Vector Machine, Swarm Intelligence

I. INTRODUCTION

The Internet is a global social network. Increasing growth of the Internet and its usage, lots of changes are made in the business model of organizations around the world. More and more people are connected to the Internet every day with the benefits of a new business model known as e-Business.

It results in the increasing importance of internet Communication activity. There are harmless as well as harmful users on the Internet. various organizations around the world uses firewalls to protect t their private network from the Public network. But it is difficult to protect private network using firewall against the network threats and malicious activity. Intrusion detection system completes firewall security. . The firewall protects the organization from malicious attacks from the Internet. In such cases intrusion detection system detects if someone tries to access the firewall or is able to break firewall security and attempt to access any system in on the reliable side and warns the system administrator in the event of a security defence. Therefore, the Intrusion Detection System (IDS) is a secure security system that observes computer systems and network traffic and analyses that dangerous network attacks from outside the organization and system abuse or attacks from within the organization.

II. BACKGROUND

Classification of Intrusion Detection System: IDS are classified into 5 types:

A. Network Intrusion Detection Systems

(NIDS) are set in a fixed location within the network to check traffic on all devices in the network. It makes street-view passing through the entire subnet and resembles a subnet street with a collection of well-known attacks. When an attack has been detected or an unusual character appears, a warning may be sent to the controller. An example of NIDS is putting it in a subnet where firewalls are available to see if someone is trying to crack a firewall.

B. Host intrusion Detection Systems

(HIDS) work on private sites or on network devices. HIDS monitors incoming and outgoing packets from the device only and will notify the administrator if suspicious or dangerous activity is detected. It takes a snapshot of existing system files and compares them to a previous image. If the analysis program files are edited or deleted, a warning is sent to the administrator to investigate. An example of the use of HIDS can be seen in the complex mechanical machinery, which is not expected to change its structure.

C. A Protocol-based Intrusion Detection System

(PIDS) contains a program or agent that stays on the server end, manages and interprets the protocol between the user / device and the server. It tries to protect the web server by constantly monitoring the HTTPS protocol stream and accepting HTTP protocol. Since HTTPS is not encrypted and before installing its web presentation layer then this program will need to stay in this interface, between using HTTPS.

D. Application Protocol-based Intrusion Detection System

(APIDS) is a program or agent that resides within a group of servers. Identifies interventions by monitoring and interpreting communications in specific application. For example, this will monitor SQL protocol explicitly displayed in middleware as it works with a web server database.

E. The Hybrid Intrusion Detection System

(HIDS) is performed by combining two or more intrusion detection methods. In a hybrid access detection system, the host agent or system data is combined with network information to develop a complete overview of the network system. The Hybrid intrusion detection system works better compared to other intrusion detection systems.

III. OVERVIEW OF MACHINE LEARNING

Machine learning (ML) is a study of computer algorithms that develop automatically through experience and data usage. It appears to be part of artificial intelligence. Machine learning algorithms create a model based on sample data, known as "training data", to make predictions or decisions without being programmed to do so. Machine learning algorithms are used in a variety of applications, such as email filtering and computer vision, where it is difficult or impossible to develop general skills to perform the required tasks. The subset of machine learning is closely related to computational statistics, which focus on computer-generated forecasts; but not all machine learning is mathematical learning. The study of the application of mathematics brings methods, theoretical and practical contexts into the field of machine learning. Data mining is a coherent field of study, focusing on the analysis of experimental data by unchecked learning. In its application to business problems, machine learning is also called forensic analytics.

IV. MACHINE LEARNING APPROACHES FOR INTRUSION DETECTION SYSTEM

We separate the Machine learning techniques for IDS methods having two-phase acquisition: Artificial Intelligence (AI) technology-based approaches and methods based on Computational Intelligence (CI) methods. AI techniques refer to approaches from the ancient AI domain such as mathematical modelling and while CI strategies refer to the inspired methods used to deal with complex problems that are old methods that cannot be solved. Key CI- method gives evolution calculations, abstract concepts, artificial neural networks, and artificial immune system. CI event comes from the long-known best of AI. AI handles figurative information representation, while CI manages the numerical representation of information. Al- however the boundary between these two categories is not always clear and has many hybrids methods are suggested in the books, many previous works are specially designed based on one of the categories. Besides, it can be helpful to understand how well-based strategies work in contrast to traditional methods.

A. Artificial Intelligence (AI):

Artificial Intelligence is the construction of smart machines by the branch of computer science, where the machine is a system that recognizes its nature and takes over actions that increase the chances of success. It uses computer models to achieve objectives and psychological study wisdom. In short, AI is a study of how computers can perform things, meanwhile, people do better. These AI learning strategies can be used for better interventions Use of the detection system. Different artificial intelligence Strategies can be used for the Intrusion Detection System as well as Re-entry Protection Program. The current trend in file to improve the implementation of the Expert System. Once via Neural Network, Genetic Algorithm, Fuzzy Logic, and other, Professional Access System, they are designed to see and learn through patterns. An expert plan contains the knowledge base where it ends the rules also make for a purposeful consideration.

1) K-Nearest Neighbors:

A new process known as K's closest neighbor could be used to distinguish system performance from normal and non-standard. A separate system call details to detect undesirable functionality of various programs and curriculum profiles you invest a lot of time and money. Neighbors close to K use the number of systems calls to detect malicious activity in the network. Text segregation is used by neighbors closest to K used to separate documents to convert systems into vectors again use it to compare and contrast between the two systems' jobs. Text separation is a grouping process the call logs of the program have been transcripts of previously defined categories. Several machine learning strategies can be used in the text divided into categories in the same way. Uses vector space model, first translated into words. Neighbor's notes are weighed using k-adjacent to the neighbor dividing algorithm. Classes are measured using the parallel parameter of the vector for each neighbor. Therefore, the analogy is compared among them to find. Since program implementation profiles are deleted, computer computing is reduced. Also, test results show that the use of neighbors close to K produces a much lower number of false positives compared to alternatives.

2) Naive Bayes:

Most of the time we know the reliance on statistics or an unusual relationship between system variability. Basic the idea that there is a chance for some of the things described earlier can affect these variables of the system. Also, there may be great difficulty in accurately expressing a relationship between these changing elements. Statistical, constructive, and free the relationship between these variables is exploited to form a possible graph model called the Naïve Bayes network. This model provides answers to questions that can be raised such as, what are the chances that a particular type of attack is known they also occur due to the recording of past events give evidence in the same way. The same answer is given using a conditional probability formula. IDAG (Exactly Acyclic Graph) is used to represent the formation of the Naïve Bayes Network. Here, Here, each node in the graph represents one of the variations of the system, and links are provided details about the influence a node has on another.

3) Decision Tree:

The decision tree is a supervised learning algorithm there contains a flow-graph similar to each structure and the interior of the node represents the test case of the tree. Separates the sample as if the current decision helps to do something else decisions that will lead us to a conclusion. Such types of decisions are displayed in the form of a decision tree. The passage from the root to the leaf is that each leaf represents some form of segregation. These are the qualities labelled for each node and the decision is made in the background to include all attributes of a particular node. Therefore, the path from the root to the leaf tree represents separation rules. These rules of segregation are defined in terms of inconsistent recordings recorded by the machine or received across history. They are defined by moral and is classified according to characteristics such as upon submission of the sample, the decision tree will continue through the decision making and come to the conclusion that what will be shown is cruel behaviour or not. A decision tree with a wide range of classical (symbolic) labels is called a tree of separation, and the decision tree with a degree of

continuous (numerical) values is called backward. Anyway, the tree can be used for Accessing purposes According to their requests.

4) Support Vector Machine:

Support Vector Machine is used for binary solutions separation problems. It is a supervised learning algorithm that depends on the concept of decision planes defined by the parameters of the decision. Marking specific algorithm to the linear map using a kernel function. Polynomial also radial base functions are used to make such a map, which separates the element space by forming a hyperplane. These kernel functions can be used during training classifiers selecting supporting vectors for this function. Therefore, these hyperplane support carriers are used separate data support system Vector Machine system. A multiclass Support Vector Machine creates sections in Access-to-Access Program training. Therefore, multiclass vector support equipment can be used for the separation of Accessibility System. The one with the biggest line between classes is the leading Support Vector hyperplane the machine is also selected for partition purposes.

B. Computational Intelligence (CI):

Core computational intelligence approaches for IDS are artificial neural networks, fuzzy sets, evolutionary computation, artificial immune systems, swarm intelligence.

1) Artificial Neural Networks:

Artificial neural networks having two approaches for IDS. First is multi-layered feedforward neural networks and second is Kohonen's self-organizing maps. A multilayer feedforward neural network is a connection of perceptron's where data and statistics flow in one direction, from input data to output. The number of layers in the neural network is the number of layers of perceptron's. A very simple neural network has a single input layer and an output layer of perceptron's. Self-Organizing Map (or Kohonen's Map or SOM) is a type of Artificial Neural Network also inspired by the natural models of neural systems that make up the 1970's. It follows an uncontrolled learning approach and trains its network using a competitive learning algorithm. SOM is used to integrate and map (or reduce the size) of multidimensional data entry techniques into low-level environments that allow people to reduce complex problems for easy interpretation. Self-organizing map has two layers, first is the input layer and second is the output layer.

2) Fuzzy Sets:

In recent decades there has been a rapid increase in the number and variety of applications for the Fuzzy concept. Fuzzy understanding, dealing with ambiguity and is appropriate detecting intrusion for two main reasons. First, the login detection problem involves multiple numerical attributes in the audit data collection, as well as the various statistics, obtained steps. Build models directly into numerical data causes high acquisition errors. For example, interventions that only a slight deviation from the model may not be detected either a slight change in normal behaviour can cause a false alarm. Second, security itself includes powerlessness, because the boundary between normal and inconsistent is not well defined.

3) Artificial Immune Systems:

Artificial Immune Systems (AIS) is concerned with deploying the formation and function of antibodies in computer systems and investigating the use of these systems in solving computational problems ranging from mathematics, engineering, and information technology. AIS is a small component of biologically inspired computing, as well as ecological calculations, interested in machine learning and belongs to the broad field of Artificial Intelligence. Artificial Immune Systems (AIS) are flexible systems, which are inspired by theoretical and immune systems, used to solve problems. AIS differs from computational immunology and theoretical theory of imitation immunology using computational and mathematical models in better understanding of the immune system, although those models initiate the AIS sector and continue to provide fertile ground for stimulation.

4) Swarm Intelligence:

Swarm intelligence (SI) is an integrated approach to distributed, systematic, environmental, or non-distributed programs. The concept has been applied to the work of artificial technology. This talk was delivered by Gerardo Beni and Jing Wang in 1989, in the context of robotic systems. SI programs contain the frequency of most people or interacting with your local environment and environment. Motivation often comes from nature, especially biological systems. Agencies follow very simple rules, and although there is no key regulatory framework that defines how each agency should behave, locally, and to some extent randomly, communication between those agencies leads to the emergence of "intelligent" global behaviour, unknown to individual agents. Examples of particle ingenuity in natural systems include ant colonies, bee colonies, bird migration, hawk hunting, animal husbandry, bacterial growth, fish studies, and microbial intelligence. The use of swarm terms in robots is called swarm robotics while swarm intelligence refers to the most common set of algorithms. Swarm prediction is used in the context of predictive problems. Similar mechanisms of the proposed robotics type are considered to be genetically modified organisms in a combined intelligence.

V. CONCLUSION

We reviewed the most important algorithms depending on the variety of machine learning strategies. ML strategic features make it possible to design IDS that have a high detection rate and low false positives during the process it quickly adapts to changing behaviour. We separated these algorithms into two types of schemes designed for ML: Artificial Intelligence (AI) and Computational Intelligence (CI). Although these two categories of algorithms share many similarities- bonds, several aspects of CI-based strategies, such as familiarity, error tolerance, high computer speed, and error correction in the face of audio data, synchronize the need to build access systems.

REFERENCES

[1] W. –C. Lin, Shih-Wen K. Chih-Fong T . (2015). Intrusion detection system based on combining cluster centers and nearest neighbors. Knowledge-Based Systems 78 (pp. 13-21). Elsevier.
[2] Bobba Brao and Kailasam Swathi. (2017). Fast KNN Classifiers for Network Intrusion Detection System. Indian Journal of Science and Technology. 10(14). Researchgate. (1-10).
[3] Chie-Hong L. Yann-Yean S. Yu-Chun Lin and Shie-Jue L. (2017). Machine Learning Based Network Intrusion Detect ion. 2nd IEEE International Conference on Computational Intelligence and Applications. (pp. 79-83). IEEE.
[4] Deyban P. Miguel A. A, David P. A, and Eugenio S. (2017). Intrusion detect ion in computer networks using hybrid machine learning techniques. XLIII Latin American Computer Conference (CLEI). (pp.1-10). IEEE
[5] Marzia Z. and Chung-Horng L.(2018). Evaluation of Machine Learning Techniques for Network Intrusion Detection. IEEE. (pp. 1-5)
[6] Kazi A., Billal M. and Mahbubur R. (2019). Network Intrusion Detect ion using Supervised Machine Learning Technique with feature selection. International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST). (pp. 643-646). IEEE.
[7] Yuyang Z., Guang C., Shanqing J. and Mian D. (2019). Building an Efficient Intrusion Detection System Based on Feature Selection and Ensemble Classifier. Computer Networks. Doi:https://doi.org/10.1016/j.comnet .2020.107247
[8] Vinoth Y. and Kamatchi K. (2020). Anomaly Based Network Intrusion Detection using Ensemble Machine Learning Technique. International Journal of Research in Engineering, Science and Management. IJRESM. (290-296)..
[9] Bhavani T . T , Kameswara M. R and Manohar A. R. (2020). Network Intrusion Detection System using Random Forest and Decision Tree Machine Learning Techniques. International Conference on Sustainable Technologies for Computational Intelligence (ICSTCI).(pp. 637-643). Springer.
[10] Maniriho et al. (2020). Detecting Intrusions in Computer Network Traffic with Machine Learning Approaches. International Journal of Intelligent Engineering and Systems. INASS. (433-445)

