ACADEMICIA: An International Multidisciplinary Research Journal
ISSN: 2249-7137 Vol. 11, Issue 12, December 2021 SJIF 2021 = 7.492
A peer reviewed journal

AN OVERVIEW OF THE USAGE OF ARTIFICIAL INTELLIGENCE-BASED INTRUSION DETECTION METHODS

Dr. Vinay Kumar Mishra*

*Associate Professor,
Department of Computer Science, Faculty of Engineering,
Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, INDIA
Email id: vinaym.computers@tmu.ac.in

DOI: 10.5958/2249-7137.2021.02666.5

ABSTRACT

Huge numbers of computer across the globe, operating on a variety of hardware or software platforms, link to provide communication and commercial services over the Internet. However, since computers are interconnected, malevolent individuals may abuse resources and launch Internet assaults. The ever-increasing number of Internet assaults makes it difficult to create flexible, adaptable security solutions. One of the most significant components used to identify Internet assaults is the intrusion detection system (IDS). Various methods from various disciplines have been used in the literature to create effective IDS. Artificial intelligence (AI)-based methods have an important part in the development of IDS as well as provide many advantages over other approaches. There is, however, no thorough evaluation of AI-based methods to evaluate and comprehend their present position in solving intrusion detection issues. Various AI-based methods have been discussed in this article, with an emphasis on IDS creation. The sources of data sets, processing criteria, method utilized, dataset, classifier design, dimension reduction technique used, and other experimental environment setup have all been compared. The advantages and disadvantages of AI-based methods were addressed. The article will aid in a better understanding of the many avenues in which IDS research has been conducted. The results of this article are helpful for individuals interested in applying AI-based methods to IDS and related areas, since they offer important insights into the literature. The study also discusses the research's future directions in this field.

KEYWORDS: *Artificial Intelligence, Ensemble System, Hybrid System, Intrusion Detection System, Intrusion, Network Security.*

1. INTRODUCTION

The Internet has become an integral aspect of everyday life, with applications in sectors such as commerce, communication, entertainment, education, and personnel day-to-day operations, among others. The Internet, in particular, has become an essential component of commercial transactions for accessing information. Access to information through the Internet allows for a variety of attacks on a computer system. More and more businesses are becoming susceptible to cyber-attacks and invasions. “Any collection of activities that try to undermine the security objectives” is what an intrusion or attack is described as. Availability, Integrity, Confidentiality,

Accountability, as well as Assurance are all key security goals. Probing, Denial of Services (DoS), User to root (U2R), and Remote to user (R2L) assaults are the four types of intrusions that may be categorized[1], [2]. A variety of anti-intrusion solutions have been developed to prevent a huge percentage of Internet assaults. Prevention, detection, preemption, deterrent, deflection, as well as countermeasures are the six anti-intrusion systems identified. The precise detection of an incursion is the most essential of these components. Many IDS have been created in the literature using AI-based methods as stated in Sect. 2. Some IDSs were created using a single classification method, while others (known as hybrid/ensemble IDS) used several categorization techniques. However, there is no thorough evaluation of these intrusion detection methods (ID).

The paper's three goals are as follows: The first goal is to provide a short overview of IDS, including its architecture and categorization. Second, the article examines the source of audit data, process criteria, or classification method utilized, classifier design, dataset, feature reduction, and other experimental environment setup to offer a review of different research of AI based approaches for ID. The focus of this article is on AI's basic methods, which include Decision Trees, Rule-Based Decision Making, Support Vector Machines (SVM), Neural Networks (NN), Fuzzy Logic, Data Mining, Genetic Algorithms, Bayesian Networks, Markov Models, and Clustering Techniques. The third goal is to identify current research problems and prospective research initiatives[3].

The remainder of the paper is laid out as follows. The intrusion detection architecture and categorization of IDS are briefly described in Section 2 based on the source of audit data, processing criteria, and response generating components. The existing IDSs are also highlighted in this section. Section 3 discusses different methods for identifying people. Section 4 delves into the specifics of the AI-based methods used for ID. The hybrid/ensemble method employed in ID is highlighted in Section 5. Section 6 compares several AI-based methods studies based on the audit data source, processing criteria, methodology utilized, classifier design, dataset, feature reduction approach used, and classification classes. Finally, Section 7 wraps up the paper and looks forward to the future[4].

1.1 System for detecting intrusions:

One of the standardized parts of security infrastructures is antivirus software (IDS), which is described as "an effective security technology that can detect, prevent, and potentially respond to computer assaults." It monitors specific sources of activity in computer and network systems, such as audit and network traffic data, and employs different methods to offer security services. The primary goal of an intrusion detection system is to quickly identify all intrusions. Network administrators may identify security goal breaches by using IDS. External attackers attempting to obtain unauthorized access to the shared security architecture or making resources inaccessible to insiders misusing their access to system resources are examples of security goal violations.

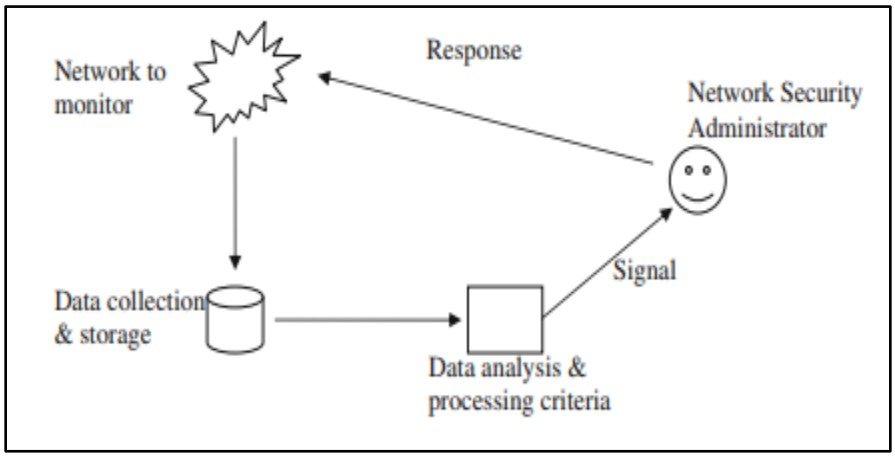

Figure 1:Illustrate the Architecture of IDS[3]**.**

Host-based IDS gathers data from of the host that needs to be safeguarded. They gather information from a variety of sources, including operating system, operating system log files, NT events log files, CPU usage, application log files, and so on. The benefit of host-based IDS is that they've been not operational system dependent and are extremely effective at detecting buffer overflow attacks. In the event of encrypted data and a switched network, these solutions become ineffective. MIDAS is a host-based IDS that is available. Expert System for Intrusion Detection (IDES) Tripwire OSSEC HIDS. Network-based IDS: This kind of IDS gathers data from the network in the form of packets. These IDS are operating system agnostic and simple to install across a variety of platforms. NSM is a network-based intrusion detection system. NADIR (Network Anomaly Detector and Intrusion Reporter) is a network anomaly detector and intrusion reporter[5].

1.2 Techniques for detecting intrusion:

Many methods from various fields have been used to identify incursions in the literature. Statistical methods, knowledge-based techniques, and artificial intelligence (AI)-based techniques are the three main methodologies. The behavior of the system is depicted from a random perspective in statistics-based IDS. Knowledge-based IDS methods, on the other hand, attempt to capture the stated behavior from accessible system data (protocol specifications, or network traffic instances, etc.). Finally, AI-based IDS methods need the creation of an explicit or implicit model that categorizes patterns[6].

1.3 Techniques based on artificial intelligence (AI):

Ponce has outlined a number of benefits of adopting AI-based approaches over more traditional methods. Flexibility (vs. threshold definition of traditional method); Adaptability (vs. precise rules of conventional technique); Pattern matching (and detection of patterns); Fast computing (really quicker than humans); and Learning skills are among the main benefits. Many writers have classified AI-based methods into many categories.

1.4 Techniques based on decision trees:

Decision trees are effective and widely used classification and prediction techniques. Nodes, arcs, and leaves are the three major components of a decision tree. Each node has a feature

property that is the most informative among the characteristics not yet evaluated in the route from the root, each arc out of a node has a feature value for the node's feature, and each leaf has a category or class designated on it. Starting at the root of the tree and working through it until you reach a leaf node, a decision tree may be used to categorize a data point. The categorization of the data point would then be provided by the leaf node. Levin developed a collection of locally optimum decision trees from which the best subset of trees for forecasting new instances. For training and testing, 10% of the KDD Cups database is utilized. A random sample of data is taken from the full training data set. Different assault types in the KDD data set are detected using a multi-class detection method[7].

1.5 Techniques based on rules:

To categorize the audit data, rule-based methods usually use a set of association rules including frequent episodes patterns. When a rule says that "if event X happens, then event Y is likely to occur," then events X and Y may be represented as sets of (variable, value)-pairs, with the goal of finding the sets X and Y that "implies" Y. In the classification domain, we fix Y and look for sets of X that are excellent predictors of the correct categorization. The benefit of utilizing rules is that they are usually straightforward and intuitive, as well as unstructured and so less restrictive. However, rules are hard to maintain and, in certain instances, insufficient to describe many kinds of information. Inductive rule generating algorithms have been suggested in the literature in a variety of ways. Some of them start by creating a decision tree, then extracting a set of categorization rules from it. Other algorithms use a divide-and-conquer method to directly infer rules from data[8].

1.6 Network of neurons (NN):

The NN learns to anticipate the behavior of the system's different users and daemons. If developed and executed correctly, NN has the ability to solve many of the issues that rule-based methods face. The primary benefit of NN is their tolerance for erroneous data and unclear information, as well as their capacity to infer solutions from data without previous knowledge of the data's regularities. This, along with their capacity to generalize from learnt data, has made them a viable ID option. To apply this method to ID, we'd have to feed the NN data representing attacks and non-attacks in order for the network's coefficients to change automatically during the training phase. The following are some examples of how to utilize NN.

Unsupervised model: MIT Lincoln Laboratory's Cunningham and Lippmann performed a series of experiments using Neural Networks to identify abuse. In the network traffic, the algorithm was looking for attack-specific keywords. UNIX host assaults and attempts to get root privilege on a server have been detected using a Multi-Layer Perceptron. By categorizing the inputs into two (two) outputs, the system attempted to identify the existence of an assault (normal and attack). The system was able to identify 85% of assaults, with 17 out of 20 being detected. This system's major accomplishment was its ability to identify both old and new threats. The data for the new assaults was not included in the training.

Self-Organizing Maps (SOM) have been shown to be successful in novelty detection, automatic grouping, and visual organization when used in a supervised model. based on network user behavior breaches (Bivens et al. 2002). They used neural networks to evaluate user behavior depending on a time frame. After supervised neural network learning, the network data was categorized and grouped in various time periods using a Self-Organizing Map (SOM) neural

network. Many academics have attempted to combine the Multi-Layer Perceptron (MLP) with the Self-Organizing Map (SOM) for intrusion detection. They tried to build an Intrusion Detection System that used MLP and SOM to identify abuse. They utilized a back-propagation learning feed-forward network with four fully connected layers, nine input nodes, and two output nodes (normal and attack). The network has been programmed to respond to a certain number of assaults. The network has identified the assaults for which it was prepared[9].

1.7 Model of Markov:

In this class, there are two kinds of approaches: Markov chains are a kind of data structure that may be used to A Markov chain is a collection of states that are linked by transition probabilities that define the model's structure and capabilities. The probabilities associated with transitions are calculated during the training phase based on the target system's typical behavior. Anomalies are then detected by comparing the anomaly score (related probability) acquired for the observed sequences to a specified threshold. Hidden Markov models: A hidden Layer is one in which the system of interest is considered to be a Markov process with hidden states and transitions. Only the productions are shown. Several techniques for identifying abnormalities in the use of network protocols by examining packet headers have been described in the literature[10].

2. DISCUSSION

Thousands of computers across the world connect to offer communication and commercial services via the Internet, using a range of hardware and software platforms. Due to the linked nature of computers, malicious people may take advantage of resources and conduct Internet attacks. The growing quantity of cyber-attacks makes it harder to develop flexible, adaptive security solutions. The intrusion detection system is one of the most important components used to detect Internet attacks (IDS). In the literature, different techniques from diverse fields have been utilized to develop successful IDS. AI-based techniques have played a major role in the development of IDS and provide many benefits over other approaches. However, there has been no comprehensive assessment of AI-based techniques to assess and understand their current position in resolving intrusion detection problems. This article discusses a variety of AI-based techniques, with a focus on IDS development. In various investigations of artificial intelligence (AI) based methods in ID, several variables such as the source of audit data, processing criteria, method used, classifier architecture, dataset, dimension reduction methodology used, or classification classes are compared. By considering appropriate base classification techniques, training sample size, and combination strategy, the detection accuracy of hybrid or ensemble approaches may be improved. Using a hybrid/ensemble approach, on the other hand, has resulted in a greater computing cost. The following issues with AI-based techniques in ID will be urgently investigated in the future.

3. CONCLUSION

This article has examined different intrusion detection systems (IDS) and their categorization based on various modules, albeit it has done so briefly. A thorough evaluation of different AI-based intrusion detection (ID) methods is provided. The use of a multi-classifier based method (hybrid/ensemble approach) to identify recognized and unknown objects is described. Attacks with great precision Various artificial intelligence (AI)-based methods investigations. Many factors, such as the source of audit data and processing, are taken into account when comparing ID. Criterion, method, classifier design, dataset, and feature reduction technique. There are many

categorization classes. It can be shown that by using the correct base classification, methods, training sample size and combination method, hybrid and/or non-hybrid detection accuracy. The ensemble method may be enhanced. However, using a hybrid/ensemble method has resulted in a higher computational cost. In the future, there will be a pressing need to investigate the following AI-related problems.

REFERENCES

1. M. Alkasassbeh, "An empirical evaluation for the intrusion detection features based on machine learning and feature selection methods," J. Theor. Appl. Inf. Technol., 2017.

2. W. Wang, M. Zhu, X. Zeng, X. Ye, and Y. Sheng, "Malware traffic classification using convolutional neural network for representation learning," 2017, doi: 10.1109/ICOIN.2017.7899588.

3. G. Kumar, K. Kumar, and M. Sachdeva, "The use of artificial intelligence based techniques for intrusion detection: A review," Artif. Intell. Rev., 2010, doi: 10.1007/s10462-010-9179-5.

4. Z. Jinbiao, "Research on intrusion detection system based on clustering fuzzy support vector machine," Int. J. Secur. its Appl., 2014, doi: 10.14257/ijsia.2014.8.3.26.

5. J. Kusyk, M. U. Uyar, and C. S. Sahin, "Survey on evolutionary computation methods for cybersecurity of mobile ad hoc networks," Evol. Intell., 2018, doi: 10.1007/s12065-018-0154-4.

6. V. K. Pachghare, V. A. Patole, and D. P. Kulkarni, "Self Organizing Maps to Build Intrusion Detection System," Int. J. Comput. Appl., 2010, doi: 10.5120/191-328.

7. D. KS and B. Ramakrishna, "An Artificial Neural Network based Intrusion Detection System and Classification of Attacks," Int. J. Eng. Res. Appl., 2013.

8. G. Kumar and K. Kumar, "The Use of Artificial-Intelligence-Based Ensembles for Intrusion Detection: A Review," Appl. Comput. Intell. Soft Comput., 2012, doi: 10.1155/2012/850160.

9. Y. Gao, Y. Liu, Y. Jin, J. Chen, and H. Wu, "A Novel Semi-Supervised Learning Approach for Network Intrusion Detection on Cloud-Based Robotic System," IEEE Access, 2018, doi: 10.1109/ACCESS.2018.2868171.

10. C. Guo, Y. J. Zhou, Y. Ping, S. S. Luo, Y. P. Lai, and Z. K. Zhang, "Efficient intrusion detection using representative instances," Comput. Secur., 2013, doi: 10.1016/j.cose.2013.08.003.