ISSN: 2278-4632
Vol-10 Issue-6 No. 8 June 2020

APPLICATION OF MACHINE LEARNING TECHNIQUES IN INTRUSION DETECTION SYSTEM- A LITERATURE REVIEW

D. RADHIKA & M. SUGHASINY

***D. RADHIKA**, Ph.D Research Scholar, PG & Research Department Of Computer Science, Srimad Andavan Arts and Science College (Autonomous), Tiruchirappalli- 620005. Tamil Nadu, India. Mobile: 9942141110 Email:kvradhika2014@gmail.com*

***M.Sughasiny**, Assistant Professor & Head, PG & Research Department Of Computer Science, Srimad Andavan Arts and Science College (Autonomous), Tiruchirappalli- 620005.Tamil Nadu, India. Mobile:9944547931 Email: drsugha@andavancollege.ac.in*

ABSTRACT

Securing a network from the attackers is a challenging task at present as many users involve in variety of computer networks. To protect any individual host in a network or the entire network, some security system must be implemented. In this case, the Intrusion Detection System (IDS) is essential to protect the network from the intruders. The IDS have to deal with a lot of network packets with different characteristics. A signature-based IDS is a potential tool to understand former attacks and to define suitable method to conquest it in variety of applications. In this paper, a detailed literature review on the Network based Intrusion Detection system and Host based Intrusion Detection system using Data Mining techniques like Feature Selection, Classification, Association Rule Mining.

Keywords: Data Mining, Intrusion Detection System, Feature Selection, Classification, Association Rule Mining, Network based IDS, Signature based IDS

1. Introduction

Today, political and commercial entities are increasingly engaging in sophisticated cyber-warfare to damage, disrupt, or censor information content in computer networks [6]. In designing network protocols, there is a need to ensure reliability against intrusions of

ISSN: 2278-4632
Vol-10 Issue-6 No. 8 June 2020

powerful attackers that can even control a fraction of parties in the network. The controlled parties can launch both passive (e.g., eavesdropping, nonparticipation) and active attacks (e.g., jamming, message dropping, corruption, and forging).

Intrusion detection is the process of dynamically monitoring events occurring in a computer system or network, analysing them for signs of possible incidents and often interdicting the unauthorized access [4]. This is typically accomplished by automatically collecting information from a variety of systems and network sources, and then analyzing the information for possible security problems.

Traditional intrusion detection and prevention techniques, like firewalls, access control mechanisms, and encryptions, have several limitations in fully protecting networks and systems from increasingly sophisticated attacks like denial of service. Moreover, most systems built based on such techniques suffer from high false positive and false negative detection rates and the lack of continuously adapting to changing malicious behaviors. In the past decade, however, several Machine Learning (ML) techniques have been applied to the problem of intrusion detection with the hope of improving detection rates and adaptability. These techniques are often used to keep the attack knowledge bases up-to-date and comprehensive.

2. Intrusion Detection System

The approaches for deflecting and recognizing the attacks are either a host-based or network-based IDS, which are the most customary IDS. The suspicious intent or malicious are specified by specific patterns and attack signatures are appeared in the product. When these patterns are appeared in the network traffic, then it is network-based IDS. If it is appeared in the log files, then it is log-based IDS. The IDS will be accurately active when it is comprised of both host-based IDS and Network-based IDS.

2.1 Network based Intrusion Detection System

The data source for Network-based IDS utilizes the raw packets. A network-based IDS typically utilizes a network adapter running in promiscuous mode to monitor and analyze all traffic in real-time as it travels across the network. Its attack recognition module uses four common techniques to recognize an attack signature: (i) Pattern, expression or byte code matching, (ii) Frequency or threshold crossing, (iii) Correlation of lesser events, (iv) Statistical anomaly detection.

ISSN: 2278-4632
Vol-10 Issue-6 No. 8 June 2020

Once an attack has been detected, the IDS' response module provides a variety of options to notify, alert and take action in response to the attack. These responses vary by product, but usually involve administrator notification, connection termination and/or session recording for forensic analysis and evidence collection.

2.2 Host based Intrusion Detection System

Host-based intrusion detection started in the early 1980s before networks were as prevalent, complex and interconnected as they are today. In this simpler environment, it was common practice to review audit logs for suspicious activity. Intrusions were sufficiently rare that after the fact analysis proved adequate to prevent future attacks.

Today's host-based intrusion detection systems remain a powerful tool for understanding previous attacks and determining proper methods to defeat their future application. Host-based IDS still use audit logs, but they are much more automated, having evolved sophisticated and responsive detection techniques. Host based IDS typically monitor system, event, and security logs on Windows NT and syslog in UNIX environments. When any of these files change, the IDS compare the new log entry with attack signatures to see if there is a match. If so, the system responds with administrator alerts and other calls to action.

Host-based IDS have grown to include other technologies. One popular method for detecting intrusions checks key system files and executable via checksums at regular intervals for unexpected changes. The timeliness of the response is in direct relation to the frequency of the polling interval. Finally, some products listen to port activity and alert administrators when specific ports are accessed. This type of detection brings an elementary level of network-based intrusion detection into the host-based environment.

3. Challenges in Intrusion Detection System

An IDS generally has to deal with problems such as large network traffic volumes, highly uneven data distribution, the difficulty to realize decision boundaries between normal and abnormal behavior, and a requirement for continuous adaptation to a constantly changing environment [14]. In general, the challenge is to efficiently capture and classify various behaviors in a computer network. Strategies for classification of network behaviors are typically divided into two categories: misuse detection and anomaly detection [4]. Misuse detection techniques examine both network and system activity for known instances of misuse using signature matching algorithms. This technique is effective at detecting attacks that are already known. However, novel attacks are often missed giving rise to false

ISSN: 2278-4632
Vol-10 Issue-6 No. 8 June 2020

negatives. Alerts may be generated by the IDS, but reaction to every alert wastes time and resources leading to instability of the system. To overcome this problem, IDS should not start elimination procedure as soon as the first symptom has been detected but rather it should be patient enough to collect alerts and decide based on the correlation of them.

Anomaly detection systems rely on constructing a model of user behavior that is considered normal. This is achieved by using a combination of statistical or machine learning methods to examine network traffic or system calls and processes. The detection of novel attacks is more successful using the anomaly detection approach as any deviant behavior is classified as an intrusion. However, normal behavior in a large and dynamic system is not well defined and it changes over the time. This often results in a substantial number of false alarms known as false positives. A network-based IDS looks at the incoming network traffic for patterns that can signify whether a person is probing the network for vulnerable computers. Since responding to each alert consumes relatively large amounts of time and resources, IDS should not respond to every alert it generates. Disregarding this fact may result in a self-inflicted denial-ofservice. To overcome this problem, alerts should be aggregated and correlated in order to produce fewer but more expressive and remarkable alerts.

4. Machine Learning Approaches

In this section, ML-based approaches to intrusion detection into two categories: approaches based on Artificial Intelligence (AI) techniques and approaches based on Computational Intelligence (CI) methods. AI techniques refer to the methods from the domain of classical AI like statistical modeling and while CI techniques refer to nature-inspired methods that are used to deal with complex problems that classical methods are unable to solve. Important CI methodologies are evolutionary computation, fuzzy logic, artificial neural networks, and artificial immune systems. CI is different from the well-known field of AI. AI handles symbolic knowledge representation, while CI handles numeric representation of information. Although the boundary between these two categories is not always clear and many hybrid methods have been proposed in the literature, most previous work are mainly designed based on either of the categories. Moreover, it would be quite useful to understand how well nature-based techniques perform in contrast to classical methods.

ISSN: 2278-4632
Vol-10 Issue-6 No. 8 June 2020

4.1 Artificial Intelligence based Techniques

The supervised methods evaluated in this work include decision trees, k-Nearest Neighbor (kNN), Multi-Layer Perceptron (MLP), and Support Vector Machines (SVM). The unsupervised algorithms include γ-algorithm, k-means clustering, and single linkage clustering.

4.2 Computational Intelligence based Techniques

Based on the four core techniques of computational intelligence: genetic algorithms, artificial neural networks, fuzzy logic, and artificial immune systems.

5. Literature Survey of IDS using Supervised Machine Learning Approaches

Alom, Zahangir, Venkata Ramesh Bontupalli, and Tarek M. Taha [1] the authors explored the capabilities of Deep Belief Networks (DBN) – one of the most influential deep learning approaches – in performing intrusion detection after training with the NSL-KDD dataset. Additionally, they examined the impact of using Extreme Learning Machine (ELM) and Regularized ELM on the same dataset to evaluate the performance against DBN and Support Vector Machine (SVM) approaches.

Anumol, E. T [2] proposed an intrusion prediction system (IPS) with the extension of a commercial SIEM framework, namely open source security information management (OSSIM), to perform the event analysis and to predict future probable multistep attacks before they pose a serious security risk. Security information and event management (SIEM) framework affirms network protection by the correlation and management of network log files.

Lin, Wei-Chao, Shih-Wen Ke, and Chih-Fong Tsai [3] proposes a novel feature representation approach, namely the cluster center and nearest neighbor (CANN) approach. In this approach, two distances are measured and summed, the first one based on the distance between each data sample and its cluster center, and the second distance is between the data and its nearest neighbour in the same cluster. Then, this new and one-dimensional distance-based feature is used to represent each data sample for intrusion detection by a k-Nearest Neighbor (k-NN) classifier.

Wagh, Sharmila Kishor, and Satish R. Kolhe [4] introduced a new semi-supervised mechanism for intrusion detection, which efficiently reduces false alarms, still maintaining a high detection rate. In the proposed semi-supervised learning approach, only a small quantity of labelled data and a large amount of unlabelled data has been used.

ISSN: 2278-4632
Vol-10 Issue-6 No. 8 June 2020

Selvi, R., S. Saravan Kumar, and A. Suresh [5] proposed new classification algorithm for effective decision making in the network data set. Moreover, an attribute selection algorithm called modified heuristic greedy algorithm is used to select itemsets from redundant data.

Enache, Adriana-Cristina, and Valentin Sgârciu [6] proposed a network anomaly IDS which merges the Support Vector Machines classifier with an improved version of the Bat Algorithm (BA). The authors used the Binary version of the Swarm Intelligence algorithm to construct a wrapper feature selection method and the standard version to elect the input parameters for SVM.

Pawar, Sunil Nilkanth, and Rajankumar Sadashivrao Bichkar [7] proposed to use variable length chromosomes (VLCs) in a GA-based network intrusion detection system. Fewer chromosomes with relevant features are used for rule generation. An effective fitness function is used to define the fitness of each rule. Each chromosome will have one or more rules in it. As each chromosome is a complete solution to the problem, fewer chromosomes are sufficient for effective intrusion detection. This reduces the computational time.

Divya, T., and Kandasamy Muniasamy [8] propose a framework for predicting future attacks by combining two machine learning methods: genetic algorithm (GA) and hidden Markov model (HMM). It has two major components in which the first component makes use of GA to derive efficient intrusion detection rules and thereafter a precise detection of attacks. The second component uses HMM to predict the next attack class of the attacker.

Singh, Raman, Harish Kumar, and R. K. Singla [9] a technique based on the Online Sequential Extreme Learning Machine (OS-ELM) is presented for intrusion detection. The proposed technique uses alpha profiling to reduce the time complexity while irrelevant features are discarded using an ensemble of Filtered, Correlation and Consistency based feature selection techniques.

Kim, Jihyun, and Howon Kim [10] applied recurrent neural network with hessian-free optimization which is one of the deep learning algorithm for intrusion detection. We use DARPA dataset in order to train and test the intrusion detection model. It was used for the 1999 KDD Cup contest dataset. It composed of 41 features and 22 different attacks. We chose salient features for training the model and analysed a result of experiment with various metrics.

Sharma, Rupam Kr, Hemanta Kumar Kalita, and Parashjyoti Borah [11] This paper discussed some commonly used machine learning techniques in Intrusion Detection System

ISSN: 2278-4632
Vol-10 Issue-6 No. 8 June 2020

and also reviews some of the existing machine learning IDS proposed by authors at different times.

Mehmod, Tahir, and Helmi B. Md Rais [12] In this paper ant colony optimization has been applied for feature selection on KDD99 dataset. The reduced dataset is validated using support vector machine. Results show that accuracy of the SVM is significantly improved with reduced feature set.

Narudin, Fairuz Amalina, et al [13] proposed an alternative solution to evaluating malware detection using the anomaly-based approach with machine learning classifiers. Among the various network traffic features, the four categories selected are basic information, content based, time based and connection based. The evaluation utilizes two datasets: public (i.e. MalGenome) and private (i.e. self-collected).

Yu, Yang, et al [14] In this paper, the authors reviewed some important work related to machine learning and visualization techniques for intrusion detection. The authors presented a collaborative analysis architecture for intrusion detection tasks which integrate both machine learning and visualization techniques into intrusion detection.

Reddy, R. Ravinder, Y. Ramadevi, and KV N. Sunitha [15] described Discriminant function is very critical in separating the normal and anomaly behavior accurately. The support vector machine-based classification algorithm is used to classify the intrusions accurately by using the discriminant function. The effective discriminant function will be accurately identifying the data into intrusion and anomaly.

Kolosnjaji, Bojan, et al [16] attempted to transfer these performance improvements to model the malware system call sequences for the purpose of malware classification. The authors constructed a neural network based on convolutional and recurrent network layers in order to obtain the best features for classification. This way we get a hierarchical feature extraction architecture that combines convolution of n-grams with full sequential modelling.

Jabbar, M. A., K. Srinivas, and S. Sai Satyanarayana Reddy [17] dealt with a novel ensemble classifier based on naïve Bayes and ADTree for intrusion detection. ADTree is a well-known supervised boosting decision tree algorithm. Naive bayes is a linear classifier and assumes that all features are independent. Naive Bayes will not perform well, where complex attribute dependencies are present. The proposed ensemble combines ADTree and Naïve Bayes to improve classification accuracy of the detection system.

Kang, Min-Joo, and Je-Won Kang [18] A novel intrusion detection system (IDS) using a deep neural network (DNN) is proposed to enhance the security of in-vehicular network. The parameters building the DNN structure are trained with probability-based

ISSN: 2278-4632
Vol-10 Issue-6 No. 8 June 2020

feature vectors that are extracted from the in-vehicular network packets. For a given packet, the DNN provides the probability of each class discriminating normal and attack packets, and, thus the sensor can identify any malicious attack to the vehicle.

Nishani, Lediona, and Marenglen Biba [19] presented the most prominent models for building intrusion detection systems by incorporating machine learning in the MANET scenario. The authors have structured our survey into four directions of machine learning methods: classification approaches, association rule mining techniques, neural networks and instance-based learning approaches. The authors analyze the most well-known approaches and present notable achievements but also drawbacks or flaws that these methods have.

Pozi, Muhammad Syafiq Mohd, et al [20] proposed a new classifier to improve the anomalous attacks detection rate based on support vector machine (SVM) and genetic programming (GP). Based on the experimental results, our classifier, GPSVM, managed to get higher detection rate on the anomalous rare attacks, without significant reduction on the overall accuracy. This is because, GPSVM optimization task is to ensure the accuracy is balanced between classes without reducing the generalization property of SVM.

Roy, Sanjiban Sekhar, et al. [21] This paper checked the potential capability of Deep Neural Network as a classifier for the different types of intrusion attacks. A comparative study has also been carried out with Support Vector Machine (SVM). The experimental results show that the accuracy of intrusion detection using Deep Neural Network is satisfactory.

Li, Zhipeng, et al [22] proposed an image conversion method of NSL-KDD data. Convolutional neural networks automatically learn the features of graphic NSL-KDD transformation via the proposed graphic conversion technique.

Bu, Seok-Jun, and Sung-Bae Cho [23] proposed a hybrid system of convolutional neural network (CNN) and learning classifier system (LCS) for IDS, called Convolutional Neural-Learning Classifier System (CN-LCS). CNN, one of the deep learning methods for image and pattern classification, classifies the queries by modeling normal behaviors of database. LCS, one of the adapted heuristic search algorithms based on genetic algorithm, discovers new rules to detect abnormal behaviors to supplement the CNN.\

Ashfaq, Rana Aamir Raza, Yu-lin He, and De-gang Chen [24] proposed a fuzziness based instance selection technique for the large data sets to increase the efficiency of supervised learning algorithms by improving the shortcomings of designing an effective intrusion detection system (IDS). The proposed methodology is dependent on a new kind of

ISSN: 2278-4632
Vol-10 Issue-6 No. 8 June 2020

single layer feedforward neural network (SLFN), called random weight neural network (RWNN).

Pajouh, Hamed Haddad [25] proposes a novel two-tier classification models based on machine learning approaches Naive Bayes, certainty factor voting version of KNN classifiers and also Linear Discriminant Analysis for dimension reduction.

Yu, Yang, Jun Long, and Zhiping Cai [26] The primary goal of this research is utilizing unsupervised deep learning techniques to automatically learn essential features from raw network traffics and achieve quite high detection accuracy. In this paper, the authors proposed a session-based network intrusion detection model using a deep learning architecture.

Mukesh, Sharma Divya, Jigar A. Raval, and Hardik Upadhyay [27] This paper influenced the precision of Semi-supervised learning in identifying new malware classes. The author show the adequacy of the framework utilizing genuine network traces. Amid this research, we will execute and design the proactive network security mechanism which will gather the malware traces. Assist those gathered malware traces can be utilized to fortify the signature-based discovery mechanism.

Al-Yaseen, Wathiq Laftah, Zulaiha Ali Othman, and Mohd Zakree Ahmad Nazri [28] proposed a multi-level hybrid intrusion detection model that uses support vector machine and extreme learning machine to improve the efficiency of detecting known and unknown attacks. A modified Kmeans algorithm is also proposed to build a high-quality training dataset that contributes significantly to improving the performance of classifiers. The modified K-means is used to build new small training datasets representing the entire original training dataset, significantly reduce the training time of classifiers, and improve the performance of intrusion detection system.

Wang, Yunpeng, et al [29] proposed a machine learning model, advanced Naive Bayesian Classification (NBC-A) which is based on NBC and ReliefF algorithm, to be used in the novel IDS. The authors used ReliefF algorithm to give every attribute of network behavior in KDD'99 dataset a weight that reflects the relationship between attributes and final class for better classification results.

Jones, Andrew, and Jeremy Straub [30] A two-stage intrusion detection system is proposed which consists of a signature detection component and an anomaly detection component. The anomaly detection component utilizes a deep neural network that is trained to detect commands that deviate from expected behavior.

ISSN: 2278-4632
Vol-10 Issue-6 No. 8 June 2020

Almomani, Ammar, et al [31] proposed online intrusion detection cloud system (OIDCS) adopts the principles of the new spiking neural network architecture called NeuCube algorithm. It is proposed that this system is the first filtering system approach that utilizes the NeuCube algorithm. The OIDCS inherits the hybrid (supervised/ unsupervised) learning feature of the NeuCube algorithm and uses this algorithm in an online system with lifelong learning to classify input while learning the system.

Aksu, Doğukan, et al [32] intrusion detection systems have been developed to avoid financial and emotional loses. In this paper, we used CICIDS2017 dataset which consist of benign and the most cutting-edge common attacks. Best features are selected by using Fisher Score algorithm. Real world data extracted from the dataset are classified as DDoS or benign with using Support Vector Machine (SVM), K Nearest Neighbour (KNN) and Decision Tree (DT) algorithms.

Deshpande, Prachi, et al [33] reported a host-based intrusion detection model for Cloud computing environment along with its implementation and analysis. This model alerts the Cloud user against the malicious activities within the system by analyzing the system call traces. The method analyses only selective system call traces, the failed system call trace, rather than all. An early detection of intrusions with reduced computational burden can be possible with this feature. The reported model provides security as a service (SecaaS) in the infrastructure layer of the Cloud environment.

Chawla, Ashima, et al [34] described a computational efficient anomaly based intrusion detection system based on Recurrent Neural Networks. Using Gated Recurrent Units rather than the normal LSTM networks it is possible to obtain a set of comparable results with reduced training times. The incorporation of stacked CNNs with GRUs leads to improved anomaly IDS. Intrusion Detection is based on determining the probability of a particular call sequence occurring from a language model trained on normal call sequences from the ADFA Data set of system call traces. Sequences with a low probability of occurring are classified as an anomaly.

Othman, Suad Mohammed, et al [35] introduced Spark-Chi-SVM model for intrusion detection. In this model, the authors have used ChiSqSelector for feature selection, and built an intrusion detection model by using support vector machine (SVM) classifier on Apache Spark Big Data platform. We used KDD99 to train and test the model. In the experiment, the authors introduced a comparison between Chi-SVM classifier and Chi-Logistic Regression classifier.

ISSN: 2278-4632
Vol-10 Issue-6 No. 8 June 2020

Bansal, Ashu, and Sanmeet Kaur [36] In this paper, various machine learning algorithms have been used for detecting different types of Denial-of-Service attack. The performance of the models have been measured on the basis of binary and multiclassification. Furthermore, parameter tuning algorithm has been discussed. On the basis of performance parameters, XGBoost performs efficiently and in robust manner to find an intrusion. The proposed method i.e. XGBoost has been compared with other classifiers like AdaBoost, Naïve Bayes, Multi-layer perceptron (MLP) and K-Nearest Neighbour (KNN) on recently captured network traffic by Canadian Institute of Cybersecurity (CIC).

Meng, Weizhi, et al [37] proposed a framework for improving the intelligent false alarm reduction for DIDSs based on edge computing devices (i.e., the data can be processed at the edge for shorter response time and could be more energy efficient). The evaluation shows that the proposed framework can help reduce the workload for the central server and shorten the delay as compared to the similar studies.

Kumar, D. Ashok, and S. R. Venugopalan [38] This paper presents an adaptive algorithm that gets trained according to the network traffic. The presented algorithm is tested with Kyoto University's 2006+ Benchmark dataset. It can be observed that the results of the proposed algorithm outperform all the known/commonly used classifiers and are very much suitable for network anomaly detection.

AL-Maksousy, Hassan Hadi Latheeth, and Michele C. Weigle [39] In this work, we analyze the network behavior of five Internet worms: Sasser, Slammer, Eternal Rocks, WannaCry, and Petya. Through this analysis, we use a deep neural network to successfully classify network traces of these worms along with normal traffic. The proposed hybrid approach includes a visualization that allows for further analysis and tracing of the network behavior of detected worms.

Shenfield, Alex, David Day, and Aladdin Ayesh [40] presents a novel approach to detection of malicious network traffic using artificial neural networks suitable for use in deep packet inspection-based intrusion detection systems. Experimental results using a range of typical benign network traffic data (images, dynamic link library files, and a selection of other miscellaneous files such as logs, music files, and word processing documents) and malicious shell code files sourced from the online exploit and vulnerability repository exploitdb, have shown that the proposed artificial neural network architecture is able to distinguish between benign and malicious network traffic accurately.

Wu, Mingtao, Zhengyi Song, and Young B. Moon [41] intended to learn new vulnerability, the cyber-physical attacks is defined via a taxonomy under the vision of Cyber

ISSN: 2278-4632
Vol-10 Issue-6 No. 8 June 2020

Manufacturing system (CMS). Machine learning on physical data is studied for detecting cyber-physical attacks. Two examples were developed with simulation and experiments: 3D printing malicious attack and CNC milling machine malicious attack.

Alrowaily, Mohammed, Freeh Alenezi, and Zhuo Lu [42] In this work, a range of experiments has been carried out on seven machine learning algorithms by using the CICIDS2017 intrusion detection dataset. It ensued to compute several performance metrics to examine the selected algorithms. The experimental results demonstrated that the K-Nearest Neighbors (KNN) classifier outperformed in terms of precision, recall, accuracy, and F1-score as compared to other machine learning classifiers.

da Costa, Kelton AP, et al [43] this research work focused on the literature on Machine Learning Techniques applied in Internet-of-Things and Intrusion Detection for computer network security. The work aims, therefore, recent and in-depth research of relevant works that deal with several intelligent techniques and their applied intrusion detection architectures in computer networks with emphasis on the Internet of things and machine learning.

Yihunie, Fekadu, Eman Abdelfattah, and Amish Regmi [44] This research aimed to find an efficient classifier that detects anomaly traffic from NSL-KDD dataset with high accuracy level and minimal error rate by experimenting with five machine learning techniques. Five binary classifiers: Stochastic Gradient Decent, Random Forests, Logistic Regression, Support Vector Machine, and Sequential Model are tested and validated to produce the result.

Nathiya, T., and G. Suseendran [45] In this research work, the effective monitoring of security by a hybrid intrusion detection system (H-IDS) in the virtual network layer of cloud computing technology is discussed and a detailed view of insider and outsider attackers in the virtual network layer is provided. This framework splits into four layers, namely virtual machine layer, node layer, cloud cluster layer, and cloud layer. Signature and anomaly techniques are used to detect known as well as unknown attacks and all virtual machine (VM) host systems which are available in the cloud computing environment are considered. The cloud cluster layer uses a correlation module (CM) to detect distributed attacks, and the Dempster-Shafer theory (DST) is employed in the final decision-making phase of the intrusion detection system (IDS) in order to improve its accuracy.

Park, Kinam, Youngrok Song, and Yun-Gyung Cheong [46] In this paper, the authors analyzed the results of the attacks classified using Intrusion Detection System, and the

ISSN: 2278-4632
Vol-10 Issue-6 No. 8 June 2020

training time of Random Forest algorithm is measured by increasing the size of the KDD dataset in intervals thereby observing the changes in the final evaluation metrics obtained.

Hajimirzaei, Bahram, and Nima Jafari Navimipour [47] This paper proposed a new intrusion detection system (IDS) based on a combination of a multilayer perceptron (MLP) network, and artificial bee colony (ABC) and fuzzy clustering algorithms. Normal and abnormal network traffic packets are identified by the MLP, while the MLP training is done by the ABC algorithm through optimizing the values of linkage weights and biases.

Dey, Saurabh, Qiang Ye, and Srinivas Sampalli [48] proposed a machine learning based intrusion detection scheme for mobile clouds involving heterogeneous client networks. The proposed scheme does not require rule updates and its complexity can be customized to suit the requirements of the client networks. Technically, the proposed scheme includes two steps: multi-layer traffic screening and decision based Virtual Machine (VM) selection.

Lin, Huaqing, Gao Liu, and Zheng Yan [49] explored application-layer tunnel detection and propose a generic detection method by applying both rules and machine learning. Our detection method mainly consists of two parts: rule-based domain name filtering for Domain Generation Algorithm (DGA) based on a trigram model and a machine learning model based on our proposed generic feature extraction framework for tunnel detection.

Saleh, Ahmed I., Fatma M. Talaat, and Labib M. Labib [50] The goal of this paper is to design a Hybrid IDS (HIDS) that can be successfully employed in a real time manner and suitable for resolving the multi-class classification problem. HIDS relies on a Naïve Base feature selection (NBFS) technique, which is used to reduce the dimensionality of sample data. HIDS is a triple edged strategy as it has three main contributions, which are: (i) NBFS, which has been employed for dimensionality reduction, (ii) OSVM,which is applied for outlier rejection, and (iii) PKNN, which is used for detecting input attacks.

6. Research Direction

Intrusion Detection System (IDS) is a vital component of security measures shielding computer systems and networks from potential abuse and misuse. Since then many different efficient approaches for IDS have been proposed and implemented in practice. However, the research on intrusion detection is still an active field and attracts attention of many researchers because of its challenges and necessity of IDS for our computing resources when using Internet. Machine learning is becoming more and more important for solving these challenges as it gives computers the ability to learn without being explicitly programmed.

ISSN: 2278-4632
Vol-10 Issue-6 No. 8 June 2020

However, one of the important research questions before machine learning can be applied for IDSs in practice is about the reliability of detection results provided by automatic learning algorithms. The following are the research direction for IDS using ML techniques.

- To extract the interesting features that improve the effectiveness of intrusion detection systems.
- To learn the normal behaviour of a protected system or its users in an anomaly detection context.
- To automatically generate signatures or rules for misuse or signature based intrusion detection systems.

7. Conclusion

In this paper, we review the research carried out along with tools and solutions available for intrusion detection to lead a secure computer and network systems to the extent possible. Solutions using convergence of various machine learning techniques show a great promise and potential. Still there is a good number of open challenges in the field to explore by researchers. As the future research work, the above-mentioned works to be carried out with Machine Learning Techniques for building the effective IDS.

References

[1] Alom, Zahangir, Venkata Ramesh Bontupalli, and Tarek M. Taha. "Intrusion detection using deep belief network and extreme learning machine." *International Journal of Monitoring and Surveillance Technologies Research (IJMSTR)* 3.2 (2015): 35-56.

[2] Anumol, E. T. "Use of machine learning algorithms with SIEM for attack prediction." *Intelligent Computing, Communication and Devices*. Springer, New Delhi, 2015. 231-235.

[3] Lin, Wei-Chao, Shih-Wen Ke, and Chih-Fong Tsai. "CANN: An intrusion detection system based on combining cluster centers and nearest neighbours." *Knowledge-based systems* 78 (2015): 13-21.

[4] Wagh, Sharmila Kishor, and Satish R. Kolhe. "Effective semi-supervised approach towards intrusion detection system using machine learning techniques." *International Journal of Electronic Security and Digital Forensics* 7.3 (2015): 290-304.

ISSN: 2278-4632
Vol-10 Issue-6 No. 8 June 2020

[5] Selvi, R., S. Saravan Kumar, and A. Suresh. "An intelligent intrusion detection system using average manhattan distance-based decision tree." *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*. Springer, New Delhi, 2015. 205-212.

[6] Enache, Adriana-Cristina, and Valentin Sgârciu. "An improved bat algorithm driven by support vector machines for intrusion detection." *Computational Intelligence in Security for Information Systems Conference*. Springer, Cham, 2015.

[7] Pawar, Sunil Nilkanth, and Rajankumar Sadashivrao Bichkar. "Genetic algorithm with variable length chromosomes for network intrusion detection." *International Journal of Automation and Computing* 12.3 (2015): 337-342.

[8] Divya, T., and Kandasamy Muniasamy. "Real-time intrusion prediction using hidden Markov model with genetic algorithm." *Artificial intelligence and evolutionary algorithms in engineering systems*. Springer, New Delhi, 2015. 731-736.

[9] Singh, Raman, Harish Kumar, and R. K. Singla. "An intrusion detection system using network traffic profiling and online sequential extreme learning machine." *Expert Systems with Applications* 42.22 (2015): 8609-8624.

[10] Kim, Jihyun, and Howon Kim. "Applying recurrent neural network to intrusion detection with hessian free optimization." *International Workshop on Information Security Applications*. Springer, Cham, 2015.

[11] Sharma, Rupam Kr, Hemanta Kumar Kalita, and Parashjyoti Borah. "Analysis of machine learning techniques based intrusion detection systems." *Proceedings of 3rd International Conference on Advanced Computing, Networking and Informatics*. Springer, New Delhi, 2016.

[12] Mehmod, Tahir, and Helmi B. Md Rais. "Ant colony optimization and feature selection for intrusion detection." *Advances in machine learning and signal processing*. Springer, Cham, 2016. 305-312.

[13] Narudin, Fairuz Amalina, et al. "Evaluation of machine learning classifiers for mobile malware detection." *Soft Computing* 20.1 (2016): 343-357.

[14] Yu, Yang, et al. "Machine learning combining with visualization for intrusion detection: a survey." *International Conference on Modeling Decisions for Artificial Intelligence*. Springer, Cham, 2016.

[15] Reddy, R. Ravinder, Y. Ramadevi, and KV N. Sunitha. "Effective discriminant function for intrusion detection using SVM." *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, 2016.

ISSN: 2278-4632
Vol-10 Issue-6 No. 8 June 2020

[16] Kolosnjaji, Bojan, et al. "Deep learning for classification of malware system call sequences." *Australasian Joint Conference on Artificial Intelligence*. Springer, Cham, 2016.

[17] Jabbar, M. A., K. Srinivas, and S. Sai Satyanarayana Reddy. "A Novel Intelligent Ensemble Classifier for Network Intrusion Detection System." *International Conference on Soft Computing and Pattern Recognition*. Springer, Cham, 2016.

[18] Kang, Min-Joo, and Je-Won Kang. "Intrusion detection system using deep neural network for in-vehicle network security." *PloS one* 11.6 (2016): e0155781.

[19] Nishani, Lediona, and Marenglen Biba. "Machine learning for intrusion detection in MANET: a state-of-the-art survey." *Journal of Intelligent Information Systems* 46.2 (2016): 391-407.

[20] Pozi, Muhammad Syafiq Mohd, et al. "Improving anomalous rare attack detection rate for intrusion detection system using support vector machine and genetic programming." *Neural Processing Letters* 44.2 (2016): 279-290.

[21] Roy, Sanjiban Sekhar, et al. "A deep learning based artificial neural network approach for intrusion detection." *International Conference on Mathematics and Computing*. Springer, Singapore, 2017.

[22] Li, Zhipeng, et al. "Intrusion detection using convolutional neural networks for representation learning." *International Conference on Neural Information Processing*. Springer, Cham, 2017.

[23] Bu, Seok-Jun, and Sung-Bae Cho. "A hybrid system of deep learning and learning classifier system for database intrusion detection." *International Conference on Hybrid Artificial Intelligence Systems*. Springer, Cham, 2017.

[24] Ashfaq, Rana Aamir Raza, Yu-lin He, and De-gang Chen. "Toward an efficient fuzziness-based instance selection methodology for intrusion detection system." *International Journal of Machine Learning and Cybernetics* 8.6 (2017): 1767- 1776.

[25] Pajouh, Hamed Haddad, GholamHossein Dastghaibyfard, and Sattar Hashemi. "Two-tier network anomaly detection model: a machine learning approach." *Journal of Intelligent Information Systems* 48.1 (2017): 61-74.

[26] Yu, Yang, Jun Long, and Zhiping Cai. "Session-based network intrusion detection using a deep learning architecture." *International Conference on Modeling Decisions for Artificial Intelligence*. Springer, Cham, 2017.

[27] Mukesh, Sharma Divya, Jigar A. Raval, and Hardik Upadhyay. "Real-Time Framework for Malware Detection Using Machine Learning Technique." *International Conference on Information and Communication Technology for Intelligent Systems*. Springer, Cham, 2017.

[28] Al-Yaseen, Wathiq Laftah, Zulaiha Ali Othman, and Mohd Zakree Ahmad Nazri. "Multi-level hybrid support vector machine and extreme learning machine based on modified K-means for intrusion detection system." *Expert Systems with Applications* 67 (2017): 296-303.

[29] Wang, Yunpeng, et al. "A novel intrusion detection system based on advanced naive Bayesian classification." *International Conference on 5G for Future Wireless Networks*. Springer, Cham, 2017.

[30] Jones, Andrew, and Jeremy Straub. "Using deep learning to detect network intrusions and malware in autonomous robots." *Cyber Sensing 2017*. Vol. 10185. International Society for Optics and Photonics, 2017.

[31] Almomani, Ammar, et al. "An online intrusion detection system to cloud computing based on NeuCube algorithms." *International Journal of Cloud Applications and Computing (IJCAC)* 8.2 (2018): 96-112.

[32] Aksu, Doğukan, et al. "Intrusion detection with comparative analysis of supervised learning techniques and fisher score feature selection algorithm." *International Symposium on Computer and Information Sciences*. Springer, Cham, 2018.

[33] Deshpande, Prachi, et al. "HIDS: A host based intrusion detection system for cloud computing environment." *International Journal of System Assurance Engineering and Management* 9.3 (2018): 567-576.

[34] Chawla, Ashima, et al. "Host based intrusion detection system with combined CNN/RNN model." *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, Cham, 2018.

[35] Othman, Suad Mohammed, et al. "Intrusion detection model using machine learning algorithm on Big Data environment." *Journal of Big Data* 5.1 (2018): 34.

[36] Bansal, Ashu, and Sanmeet Kaur. "Extreme Gradient Boosting Based Tuning for Classification in Intrusion Detection Systems." *International Conference on Advances in Computing and Data Sciences*. Springer, Singapore, 2018.

[37] Meng, Weizhi, et al. "Enhancing intelligent alarm reduction for distributed intrusion detection systems via edge computing." *Australasian Conference on Information Security and Privacy*. Springer, Cham, 2018.

ISSN: 2278-4632
Vol-10 Issue-6 No. 8 June 2020

[38] Kumar, D. Ashok, and S. R. Venugopalan. "A novel algorithm for network anomaly detection using adaptive machine learning." *Progress in Advanced Computing and Intelligent Engineering*. Springer, Singapore, 2018. 59-69.

[39] AL-Maksousy, Hassan Hadi Latheeth, and Michele C. Weigle. "Hybrid intrusion detection system for worm attacks based on their network behavior." *International Conference on Digital Forensics and Cyber Crime*. Springer, Cham, 2018.

[40] Shenfield, Alex, David Day, and Aladdin Ayesh. "Intelligent intrusion detection systems using artificial neural networks." *ICT Express* 4.2 (2018): 95-99.

[41] Wu, Mingtao, Zhengyi Song, and Young B. Moon. "Detecting cyber-physical attacks in Cyber Manufacturing systems with machine learning methods." *Journal of intelligent manufacturing* 30.3 (2019): 1111-1123.

[42] Alrowaily, Mohammed, Freeh Alenezi, and Zhuo Lu. "Effectiveness of machine learning based intrusion detection systems." *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*. Springer, Cham, 2019.

[43] da Costa, Kelton AP, et al. "Internet of Things: A survey on machine learning-based intrusion detection approaches." *Computer Networks* 151 (2019): 147-157.

[44] Yihunie, Fekadu, Eman Abdelfattah, and Amish Regmi. "Applying Machine Learning to Anomaly-Based Intrusion Detection Systems." *2019 IEEE Long Island Systems, Applications and Technology Conference (LISAT)*. IEEE, 2019.

[45] Nathiya, T., and G. Suseendran. "An Effective Hybrid Intrusion Detection System for Use in Security Monitoring in the Virtual Network Layer of Cloud Computing Technology." *Data Management, Analytics and Innovation*. Springer, Singapore, 2019. 483-497.

[46] Park, Kinam, Youngrok Song, and Yun-Gyung Cheong. "Classification of attack types for intrusion detection systems using a machine learning algorithm." *2018 IEEE Fourth International Conference on Big Data Computing Service and Applications (BigDataService)*. IEEE, 2018.

[47] Hajimirzaei, Bahram, and Nima Jafari Navimipour. "Intrusion detection for cloud computing using neural networks and artificial bee colony optimization algorithm." *ICT Express* 5.1 (2019): 56-59.

[48] Dey, Saurabh, Qiang Ye, and Srinivas Sampalli. "A machine learning based intrusion detection scheme for data fusion in mobile clouds involving heterogeneous client networks." *Information Fusion* 49 (2019): 205-215.

ISSN: 2278-4632
Vol-10 Issue-6 No. 8 June 2020

[49] Lin, Huaqing, Gao Liu, and Zheng Yan. "Detection of application-layer tunnels with rules and machine learning." *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*. Springer, Cham, 2019.

[50] Saleh, Ahmed I., Fatma M. Talaat, and Labib M. Labib. "A hybrid intrusion detection system (HIDS) based on prioritized k-nearest neighbors and optimized SVM classifiers." *Artificial Intelligence Review* 51.3 (2019): 403-443.

