REVIEW RESEARCH

Improve the Efficiency of Large RFID Network Using Enhanced Security Data Delivery Model for Machine Learning Based Network Intrusion Detection System – A Survey

Nagarathna C[1] . B Muthu Kumar[2] . Bhavana N[2] . Manjushree T L[1] . Deepa Pattan[1]

[1]Department of Information Science and Engineering
Sai Vidya Institute of Technology, Bengaluru, Karnataka, India
[2]School of Computing and Information Technology
REVA University, Bengaluru, Karnataka, India

Received: 15 October 2022 / Revised: 19 October 2022 / Accepted: 07 December 2022

Abstract – The main issue in both computer and computer networks is security. Intrusion Detection System (IDS) has faced many problems constantly growing methods and techniques by attackers and also the increase of connected devices from interpretation to operation. A major research problem in network security is IDS. Machine Learning algorithms were adapted as a result of this for Network IDS. To analyze network traffic datasets are used and a framework developed which enables the use of network traffic that is frequently updated and concerned to entitle the complete solutions for model deployment. The framework consists of (i)generation of attacked Dataset, (ii) the bonafide dataset, (iii) training models using machine learning techniques, (iv) training the model, and (v) deployment and evaluation of the model. This framework has the following characteristics: frequently updated network traffic, reproducible attacks, and addresses that examine the model's realization and deployment.

Index terms – Machine Learning, Cyber Security, Intrusion Detection System

I. INTRODUCTION

In this era, malicious jobs such as stolen credentials, illegal access to data, data alteration, probing, intrusion and impersonation, and many more are growing all over the world. Existing prevention and security methods are not adequate to provide protection completely from malware and complex attacks where detection of these types of attacks are more challenging and complex. To design security is the nature of interconnected devices, which brings advantages and drawbacks to building security protection. In order to prevent malicious attacks network administrators adopts an intrusion detection system. Any misuse of data or any type of intrusion attack is detected and reported by IDS. If the Computer Networks in the appropriate place do not have a security plan, they are vulnerable to attack. Nowadays data transformed through electronic media is experiencing many problems such as security, confidentiality and privacy of information.

A network attack effectuated by an insider where the attack is started by an inside entity where the attacker has complete authorization, and tries to access some resources or by an outsider they are unauthorized users may be criminals or international terrorists or pranksters. This paper gives a complete review and taxonomy of the research work carried out on cyber security. The signature-based IDS (SIDS) and anomaly-based IDS (AIDS) methods are reviewed. The main concern of the research area is IDs dataset problems and machine learning techniques to provide security to the data. An IDS might be a Hardware or software system that recognizes malicious actions on computer systems, IDS main aim is to recognize all kinds of malicious activities on network traffic, and computer usage. There are two main groups of IDS systems: the first one is SIDS and second one is AIDS.

II. DIFFERENT TYPES OF IDS

SIDS- Signature-based Intrusion Detection System (SIDS)

SIDS adopts pattern matching methods to identify an unknown attack, this type of models is also known as Misuse Detection systems or Knowledge-based Detection. when an intrusion signature is matched with an existing intrusion in the dataset the alarm will be triggered. The major drawbacks of the SIDS are zero-day attacks are not detected where no corresponding signature stored in the repository.

AIDS- Anomaly-based Intrusion Detection System (AIDS)

To solve the issues raised by SIDS Anomaly-based IDS was introduced. In AIDS, using statistical-based, machine learning, or knowledge-based methods behavior of the computer system is modeled. Any abnormal behavior which is dissimilar to normal behavior is grouped as intrusions. AIDS contains two stages first stage dataset is trained and in the second stage dataset undergoes testing phase. Machine learning techniques are adopted to train the system misbehave of the system. While testing the data new dataset is created to generalize the unknown intrusion activities

TABLE 1: DEFINITION OF VARIOUS TYPES OF ALERT

Name	Definition
True Positive (TP)	Correctly classification of anomalous attack
False Positive(FP)	Incorrectly detects the attacks
True Negative(TN)	Correct classification of ordinary instances of non-attack sample
False Negative(FN)	Incorrect classification of attack as an ordinary case

III. LITERATURE SURVEY

R. Vinaykumar and Mamoun Alazab [1] in their work propose Deep Neural Network (DNN) model which detects and classifies abnormal and unexpected cyberattacks. It collects the network and host level activities in distribute manner using DNNs. Natural language processing (NLP) of Advance text representation methods was used at host-level events as a system call to capture contextual and semantic patterns matching and to preserve system calls sequence of information. It failed to give detailed characteristics of the malware and information on the structure. A technique is suggested by *Lu Lv, Wenhai Wang, Zeyin Zhang* [2] to distinguish between malicious activities and normal behavior of system by using hybrid kernel function (HKELM). Differential evolution (DE) and Gravitational search algorithm (GSA) were used to optimized the parameters of HKELM which helps to predict the attacks. Reduction and feature extraction of data Kernel principal component analysis(KPCA) algorithm were used.

Bedi, P., Gupta, N. & Jinda let.al [3] adopted a IDS model called Improved-Siama-IDS which uses Binary Neural Network (BNN), DNN classifiers and eXtreme Gradient Boosting algorithm to handle the problem class in first layer which will filter the network data. The network data is filtered multiple times by different machine learning classifiers to reduce malicious traffic. Attacked data is the input to I-SiamIDS second layer which consists of multiclass eXtreme Gradient Boosting for classification. *Vishwa Teja Alaparthy et al. [*4] developed a technique using danger and immune theory to safeguard wireless sensor network which will envision the different energy depletion attacks. When an attack is recognized in multi-level IDS an alarm signals are generated using immune cell which reacts to stress and abnormal cells.

Kaja, N., Shaout, A. et al. [5] have proposed a two-phase IDS using machine learning methods which detects and protects malicious attacks. K-Means algorithm is used in the first phase which detects attacks and in second phase J47, Adaptive Boosting, random forest and Navie Bayes algorithm is used to classify the attacks. This model eliminated the false positives. *Z. K. Maseer, R. Yusof, N. Bahaman et al.* [6] gives reviews on AIDS which is applied on different datasets. Machine Learning based AIDS was designed to detect the attacks on a binary dataset. Performance of the system was tested using Machine Learning methods like SVM, K-Nearest Neighbors, Random Forest.

Preethi Devan, Neelu Khare [7] proposed feature selection of network intrusion and classification of intrusion by XGBoost-DNN model. The model was proposed in three different steps: normalization of data, feature selection of dataset and classification of dataset. Dataset used is NSL_KDD for the experiment and testing the model. For learning rate optimization Adam optimizer were used, for training and classification of network intrusion softmax classifier is used in this model. *M. H. Ali, B. A. D. Al Mohammed et al.* [8] had proposed a fast-learning network based on swarm optimization using an artificial neural network (ANN) which monitors malicious user activity or network traffic by triggering the alarm, ANN capacity of learning from real example reduces the number of wrong negative or false positives.

*Salama et al. [*9] suggested framework to detect SQL injection attack with a combined misuse and anomaly detection algorithms. In the training phase to legitimate database behavior a profile is created which was extracted for XML files using association rules containing SQL queries, where application will submit these queries to database. *Lu Lv, Wenhai Wang, et al*. [10] have developed IDS model based on machine learning algorithms to find the malicious attack using gravitational search and differential evolution (DE) algorithms which will boost the characteristics of the hybrid kernel function. *Mahmoud elsisi, minh-quang tran* [11] proposes a IoT framework for online monitor and detect gas-insulated switchgear (GIS) defects and to detect the fake data which provides authentic and secured GIS data. The model can present GIS defects and network status with different alarms on power system.

Elie Alhajjar, Paul Maxwell [12] conducted many experiments on UNSW-NB15 and NSL_KDD datasets to study the nature of adverbial problem in Network Intrusion Detection Systems (NIDS), highlighted the vulnerability of machine learning based NIDS. The results of this experiment shows high misclassification rates in different machine learning algorithm. *Raisa Abedin Disha, Sajjad Waheed* [13] proposes a IDS using Random Forest feature selection technique which was based on source. "Gini Impurity-based weighted Random Forest (GIWRF)" model used to select features based on score which is calculated by adjusting weights for imbalanced class distribution. Data preprocessing and training the dataset and testing the dataset is done to adopt this model.

Ilhan Firat Kilincer et.al [14] proposes a binary class classification and multiple class classification IDS model which has high detection rate, low false alarm rate with high accuracy. KDD99 and NSLKDD datasets were used, reduction of dataset was done using feature selection method and classified using AdaBoost and Bagging techniques. By applying resampling methods this model can be improved further. *Iqbal H. Sarke et.al* [15] work shows how to identify cyber anomalies and multiple attacks using machine learning based model. DoS, Worms Backdoor and many attacks were detected by using binary and multi-class classification model. Machine learning algorithms were used for pattern security and detection of anomalies attacks.

Mayank Gupta, Manu Gupta [16] proposes mechanism for sharing a secret images between two parties. Using neural cryptography single secret image sharing method was developed Common key between two parties were identical weight vectors. Encryption was done using neural cryptography and shared to public channel and secure key exchange protocol was used for safety communication. The weights are used as key between two parties. The proposed framework works for single secret image. In the future, multiple secret images identification can be extended. *A.R Gad, A. A. Nashat et al.* [16] To improve traffic safety and collision prevention over the communication of vehicles over wireless communication infrastructure on network ToN-IoT dataset a intelligent transportation system(ITS) was proposed.Chi and SMOTE technique was adopted to choose featuresa and class of dataset.

Li Yang, Abdallah Moubayed [17] developed a IDS which is build on machine learning models based tree strcutures to identify cyber-attacks in autonomous vehicles. This model uses Decision tree, extra trees, random forest algorithm and extreme gradient boosting algorithms to identity the threats, SMOTE and feature selection methods which is tree-based were used to reduce class imbalance and computational cost. *S.Singh et al.* [18] have developed IDS framework collaborative for the cloud, to protect the services and cloud resources from different kinds of threats and attacks. The proposed model is combined with a cloud cluster and collaboration between the cluster is done by co-relation unit which is placed in any one of the cluster. For selection of cluster for correlation unit is done by bully election algorithm based on workload. Snort tool were used identify the attacks based on signature matching using decision tree algorithm and support vector machine anomaly detection system is built.

N *Modi et al*.[19] developed a IDS of hybrid techniques which consist of seven units: first unit is capturing the packet , in second unit Anomaly activities are identified, in third unit signature based identification of malicious activities done, in fourth unit score function is used to calculate error or misbehave of the system, in fifth unit alerting system is designed to give the signals when malicious activities is detected, in sixth unit to detect external and internal network attack central log is maintained in cloud and in last unit Snort and Apriori algorithm used by signature module where generates the detection of known and derivative attacks.

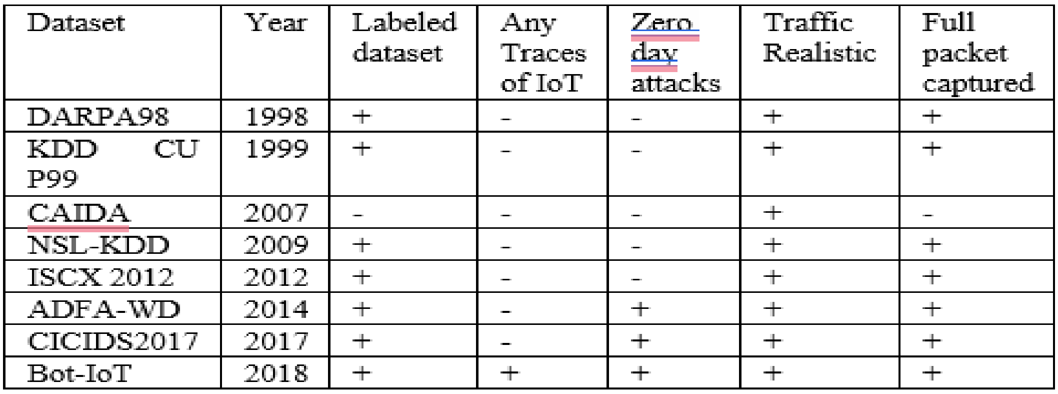
Gisung Kim et al. [20] have developed a SVM Model which uses a ChiSqSelector feature selection and using SVM Apache Spark Big Data platform is developed. They have used KDD99 dataset which can be trained and tested. Here dataset was loaded first and exported to resilient datasets and categorical data is converted into numerical data by preprocessing the data in Apache Spark. ChiSqselector and SVM combined which is applied to dataset numTop Features method to feature selection of data. Data can have processed and analyzed with high speed using this model.

TABLE 2: DIFFERENT DATASET USED AND ASSESSED BY DIFFERENT ALGORITHMS

System Used	Dataset Used
SVM (Support Vector Machine), J Decision Tree and Decision Table and Naive Bayes	KDD
C4.5 Decision Tree and Naive Bayes Classifier	KDDCUP99
Bayesian classifier, Artificial Neural Network, Decision Tree algorithm, Random Tree and Forest algorithm	CICIDS, KDD
Apache Spark Big Data Classifier based on Support Vector Machine (SVM)	KDD99
Network based Intrusion Detection system(NIDS) using J48 Decision Tree and Random Forest algorithm	DARP, KDD99

T. Mehmood and H. B. Md Rais [21] gives the uses of supervised learning algorithm for anomaly detection. J Decision Tree, SVM, Decision Table and Naive Bayes methods were used to recognize illegal; activities in network traffic. The data is trained into prediction model and tested. The dataset used is KDD. 492021 instances were used to train the package and 311029 instances were used for checking the dataset. As a result of the model developed DOS attacks were less, high accuracy with J48 and SVM. *Kunal and M. Dua* [22] have developed a model which uses the Naive Bayes Classifier and C4.5 Decision Tree for preprocessing and classification of a dataset. To train the model C4.5 decisions are taken according to the input given. Two parameters are taken out X for the protocol used such as TCP, UDP, IP and Y parameter for the attack name such as U2R, probe, DOS, etc. Binary tree with Naive Bayes classifier which detect the IDS attack like Remote to User(R2L), Denial of Service (DoS), User to Root(U2R), and probing. This model has produced 99% accuracy

TABLE 3. COMPARISON BETWEEN DIFFERENT DATASET



Dataset	Year	Labeled dataset	Any Traces of IoT	Zero day attacks	Traffic Realistic	Full packet captured
DARPA98	1998	+	-	-	+	+
KDD CU P99	1999	+	-	-	+	+
CAIDA	2007	-	-	-	+	-
NSL-KDD	2009	+	-	-	+	+
ISCX 2012	2012	+	-	-	+	+
ADFA-WD	2014	+	-	+	+	+
CICIDS2017	2017	+	-	+	+	+
Bot-IoT	2018	+	+	+	+	+

Alqahtani, H et.al [23] uses different machine learning algorithms such as Artificial Neural Network, Naive Bayes Classifier, Decision Tree, Random Forest, Bayesian Network to identify the Intrusion detection system and to test the effectiveness on cyber-security dataset which has several cyber-attacks. Parameters such as f1-score, recall, precision and accuracy has been evaluated. In this model author has adapted several stages such as data exploration, preprocessing, security modeling. *Othman, S.M., Ba-Alwi, F.M., Alsohybe, N.T. et al* [24] developed a real time IDS model where incoming information through network is recorded online and result is posted instantly, so that administrator can be alerted and can stop ongoing attack using misuse detection technique. To classify the incoming network data several machine learning algorithm were used such as Ripper Rule, Decision Tree, Back-propagation, Radial Basis Function Neural network. False detection rates were reduced, DoS attack and probe attacks was enhanced.

S Kumar et al. [25] uses NIDS for mobiles and priority was given to detection of malware. They used KDD99 and DARPA1 dataset. Wireshark tool was used for traffic generation. Machine Learning algorithm such as J48, Random Forest, RIDOR, JRIP and PART were used. Cross validation to test the accuracy on dataset was done. The model detected 99.6% of malicious traffic with 97.5% accuracy. *Wesam Almobaideen et al.* [26] to find out the missing data in medical IoT application using genetic algorithm authors developed a hybrid neural network. Jordan network with feedback loop were used to learn the dataset. Genetic algorithm which is high quality for search problem or optimization problems. This model enhances 2% performance on IoT applications and 5% accuracy values but fails to be a dynamic model.

IV. CONCLUSION

Cyber criminals are targeting computer users by complex techniques, where cybercrimes are becoming sophisticated and motivated. we can track the file system by using intrusion detection system (IDS). This paper focuses on survey of intrusion detection system methodologies, technologies. The IDS helps to identify computer misuse and intrusion by gathering data analyzing them. IDS model can be developed for wireless, wired networks. Machine learning methods are discussed which helps to identify the intrusions in the system and in network traffic. This paper also detailed study of IDS models that utilize the machine learning techniques and datasets. Further this paper can analyze traffic in the network which is still challenge. In this paper, we presented the pros and drawbacks, a survey of IDS methodologies using machine learning techniques. Several machine learning algorithms have been suggested to detect the cyber-attacks.

REFERENCES

1. Vinayakumar, A. (2019). Vinayakumar R., Alazab M., Soman K., Poornachandran P., Al-Nemrat A., Venkatraman S. Deep learning approach for intelligent intrusion detection system, IEEE Access, 7, 41525-41550.
2. Lv, L., Wang, W., Zhang, Z., & Liu, X. (2020). A novel intrusion detection system based on an optimal hybrid kernel extreme learning machine. Knowledge-based systems, 195, 105648.
3. Bedi, P., Gupta, N., & Jindal, V. (2021). I-SiamIDS: an improved Siam-IDS for handling class imbalance in network-based intrusion detection systems. Applied Intelligence, 51(2), 1133-1151.

4. Alaparthy, V. T., & Morgera, S. D. (2018). A multi-level intrusion detection system for wireless sensor networks based on immune theory. IEEE Access, 6, 47364-47373.
5. Kaja, N., Shaout, A., & Ma, D. (2019). An intelligent intrusion detection system. Applied Intelligence, 49(9), 3235-3247.
6. Maseer, Z. K., Yusof, R., Bahaman, N., Mostafa, S. A., & Foozy, C. F. M. (2021). Benchmarking of machine learning for anomaly based intrusion detection systems in the CICIDS2017 dataset. IEEE access, 9, 22351-22370.
7. Devan, P., & Khare, N. (2020). An efficient XGBoost–DNN-based classification model for network intrusion detection system. Neural Computing and Applications, 32(16), 12499-12514.
8. Ismail, A. M. A. M. B. (2018). A Zolkipli MF. A new intrusion detection system based on fast learning network and particle swarm optimization IEEE Access, 6, 20255-20261.
9. Salama, S. E., Marie, M. I., El-Fangary, L. M., & Helmy, Y. K. (2012). Web anomaly misuse intrusion detection framework for SQL injection detection. International Journal of Advanced Computer Science and Applications, 3(3).
10. Friedberg, I., Skopik, F., Settanni, G., & Fiedler, R. (2015). Combating advanced persistent threats: From network event correlation to incident detection. Computers & Security, 48, 35-57.
11. Alhajjar, E., Maxwell, P., & Bastian, N. (2021). Adversarial machine learning in network intrusion detection systems. Expert Systems with Applications, 186, 115782.
12. Disha, R. A., & Waheed, S. (2022). Performance analysis of machine learning models for intrusion detection system using Gini Impurity-based Weighted Random Forest (GIWRF) feature selection technique. Cybersecurity, 5(1), 1-22.
13. Kilincer, I. F., Ertam, F., & Sengur, A. (2021). Machine learning methods for cyber security intrusion detection: Datasets and comparative study. Computer Networks, 188, 107840.
14. Sarker, I. H. (2021). CyberLearning: effectiveness analysis of machine learning security modeling to detect cyber-anomalies and multi-attacks. Internet Things 14, 100393 (2021).
15. Gupta, M., Gupta, M., & Deshmukh, M. (2020). Single secret image sharing scheme using neural cryptography. Multimedia Tools and Applications, 79(17), 12183-12204.
16. Gad, A. R., Nashat, A. A., & Barkat, T. M. (2021). Intrusion detection system using machine learning for vehicular ad hoc networks based on ToN-IoT dataset. IEEE Access, 9, 142206-142217.
17. Yang, L., Moubayed, A., Hamieh, I., & Shami, A. (2019, December). Tree-based intelligent intrusion detection system in internet of vehicles. In 2019 IEEE global communications conference (GLOBECOM) (pp. 1-6). IEEE.
18. Singh, D., Patel, D., Borisaniya, B., & Modi, C. (2013). Collaborative ids framework for cloud. International Journal of Network Security, vol. 18(4):699-709.
19. Ahmed, S. S. T., Thanuja, K., Guptha, N. S., & Narasimha, S. (2016, January). Telemedicine approach for remote patient monitoring system using smart phones with an economical hardware kit. In *2016 international conference on computing technologies and intelligent data engineering (ICCTIDE'16)* (pp. 1-4). IEEE.
20. Modi, C. N., & Patel, D. (2013, April). A novel hybrid-network intrusion detection system (H-NIDS) in cloud computing. In 2013 IEEE Symposium on Computational Intelligence in Cyber Security (CICS) (pp. 23-30). IEEE.
21. Kim, G., Lee, S., & Kim, S. (2014). A novel hybrid intrusion detection method integrating anomaly detection with misuse detection. Expert Systems with Applications, 41(4), 1690-1700.
22. Mehmood, T., & Rais, H. B. M. (2016, August). Machine learning algorithms in context of intrusion detection. In 2016 3rd International Conference on Computer and Information Sciences (ICCOINS) (pp. 369-373). IEEE.
23. Dua, M. (2019, June). Machine learning approach to IDS: A comprehensive review. In 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 117-121). IEEE.
24. Alqahtani, H., Sarker, I. H., Kalim, A., Hossain, M., Md, S., Ikhlaq, S., & Hossain, S. (2020, March). Cyber intrusion detection using machine learning classification techniques. In International conference on computing science, communication and security (pp. 121-131). Springer, Singapore.
25. Sreedhar, S., Ahmed, S., Flora, P., Hemanth, L. S., Aishwarya, J., & Naik, R. (2021, January). An Improved Approach of Unstructured Text Document Classification Using Predetermined Text Model and Probability Technique. In *Proceedings of the First International Conference on Advanced Scientific Innovation in Science, Engineering and Technology, ICASISET 2020, 16-17 May 2020, Chennai, India*.
26. Othman, S. M., Ba-Alwi, F. M., Alsohybe, N. T., & Al-Hashida, A. Y. Intrusion detection model using machine learning algorithm on Big Data environment. J. Big Data 5 (1), 1–12 (2018).
27. Sreedhar Kumar, S., Ahmed, S. T., & NishaBhai, V. B. Type of Supervised Text Classification System for Unstructured Text Comments using Probability Theory Technique. *International Journal of Recent Technology and Engineering (IJRTE)*, *8*(10).

28. Kumar, S., Viinikainen, A., & Hamalainen, T. (2016, December). Machine learning classification model for network based intrusion detection system. In 2016 11th international conference for internet technology and secured transactions (ICITST) (pp. 242-249). IEEE.
29. Al-Milli, N., & Almobaideen, W. (2019, April). Hybrid neural network to impute missing data for IoT applications. In 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT) (pp. 121-125). IEEE.