New Cyber Security Techniques in Network Using Logistic Regression and Naïve Bayes Techniques

DR. GURU KESAVA DASU GOPISETTY[1], P. JAGADEESH BABU[2], M. BHARATHI RANI[3], SUDHAVALI ADUSUMALLI[4]

[1]*Professor & HOD, Dept. of CSE, KITS Akshar Institute of Technology, Yanamadala, Guntur India*
[2]*Assistant professor, Dept. of CSE, Eluru College of Engineering and Technology, ELURU*
[3]*Assistant professor, Dept. of CSE, Malineni Lakshmaiah Woman's Engineering College, Guntur*
[4]*Assistant professor, Dept. of IT, KKR & KSR Institute of Technology & Sciences, Guntur*

Abstract— The network traffic should be monitored and analyzed to detect malicious activities and attacks to ensure reliable functionality of the networks and security of users' information. Machine learning techniques can be applied to detect the network attacks. Network security is one of the major concerns of the modern. With the rapid development and massive usage of internet over the past decade the vulnerabilities of network security have become an important issue. Our approach is to use three learning techniques in parallel gated recurrent unit (GRU), convolution neural network as deep techniques and Random Forest as an ensemble technique. Our main goal is that the task of finding attacks is fundamentally different from these other applications, making it significantly harder for the intrusion detection community to employ machine learning effectively. The performance of the proposed system is compared with conventional machine learning algorithms namely, Logistic Regression (LR), Naïve Bayes (NB), K-Nearest Neighbor (KNN), Decision Tree (DT) and Random Forest (RF) methods. Machine learning methods can automatically discover the essential differences between normal data and abnormal data with high accuracy. In addition, machine learning methods have strong generalizability, so they are also able to detect unknown attacks. The much more popular kinds of cyber security risks are evaluated using machine learning algorithms which describe how machine learning is used for computer defense such as the identification and avoidance of attacks, vulnerability scanning and recognition and public internet risk assessment.

Index Terms- Cyber Security, Malware Detection, Machine Learning, Cyber Threat Intelligence. Cyber-Attacks

I. INTRODUCTION

Concerns over security and privacy regarding computer networks are increasing in the world, and computer security has become a requirement as a result of the spread of information technology in daily life [1]. The first intrusion detection system was proposed in 1980 [2]. In fact detection of the attack types is not constantly mandatory, because the imperative idea is detecting a survival of the attack and then removing it for making sure the protection [2]. Therefore, systematic techniques have been developed to discover attacks and estimate the suitable security policies for different attack situations [3]. As novel attacks can be found using anomaly detection techniques it is highly advantageous than signature based intrusion detection techniques. Intrusion Detection algorithms can be applied for both n [4]. Cyber security involves techniques and technologies to protect the device's software network and data from unauthorized and unauthenticated access, malware attacks and network attacks [5]. A recurrent neural network (RNN) with gated recurrent unit (GRU) base, a convolution neural network (CNN) and a non-parametric method named Random Forest, are used for detecting the type of connection which classifies them as normal or attack [6]. We present some datasets with descriptions and compare the performance to represent the approaches to show current state of working of attack detection methods with deep learning. Finally, we summarize this work and discuss some ways to improve the performance of attack detection under thoughts of utilizing deep learning structures [7].

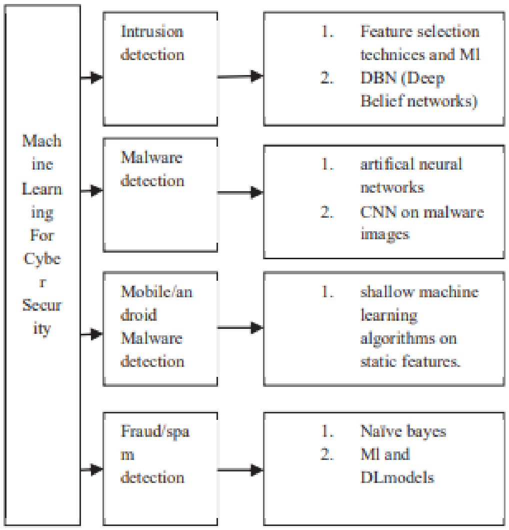
Machine Learning For Cyber Security
Intrusion detection
1. Feature selection technices and Ml
2. DBN (Deep Belief networks)
Malware detection
1. artifical neural networks
2. CNN on malware images
Mobile/android Malware detection
1. shallow machine learning algorithms on static features.
Fraud/spam detection
1. Naïve bayes
2. Ml and DLmodels

Figure 1.Machine Learning in Cyber Security

II. RELATED WORKS

The rate of attacks against networked systems has increased melodramatically, and the strategies used by the attackers are continuing to evolve. For example, the privacy of important information, security of stored data platforms, availability of knowledge[9].The selection is performed by a DAE, where a key processes to add weights to its loss function, which helps improve the selection results by placing more emphasis on the attack samples [10]. Misuse detection is also called signature-based detection. The detection process matches the signatures of samples using a signature database. The main problem in constructing misuse detection systems is to design efficient signatures [11]. Artificial intelligence is a tool that identity thieves use to mislead and manipulate users to supply sensitive data or function, for example to perform the wire transfer and to press on something like a harmful link. ML takes advantage of the activities of the crooks by making [12].The paper presently enhances SVM categorization accurateness and faster training and testing times. Moreover it reveals upright calculations in two-category and five-category classification [13]. Unlike previously mentioned works, [8] analyses several ML-based approaches for intrusion detection for identifying various issues. Issues related to the detection of low-frequency attacks are discussed with a possible solution to improve the performance further [14].Some strategies are followed for the proposed unified method, including a fuzzy set concept for asset criticality, a device study classifier for random prediction, and a composite evaluation version for comparing the effectiveness of controls [15].

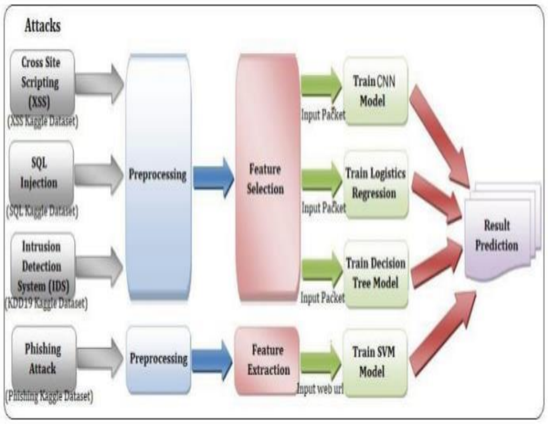

Figure 2: System Architecture

III. SYSTEM ARCHITECTURE

Network data is mostly encapsulated in network packets, which provide the load in the network. Network traffic is the main component for network traffic measurement, network traffic control and simulation. [16]. It is important to note that this entire process requires no supervised information. Many famous auto encoder variants exist such as de noising auto encoders and sparse auto encoders [17]. RBMs distinguish between the forward and backward directions the weights in both directions are the same. RBMs are unsupervised learning models trained by the contrastive divergence algorithm and they are usually applied for feature extraction [18]. We have analyzed 3widelyused tools for learning strategies, namely: Selection Tree, Deep Belief Network and Support Vector Machine. Most review articles target only one specific risk[19].

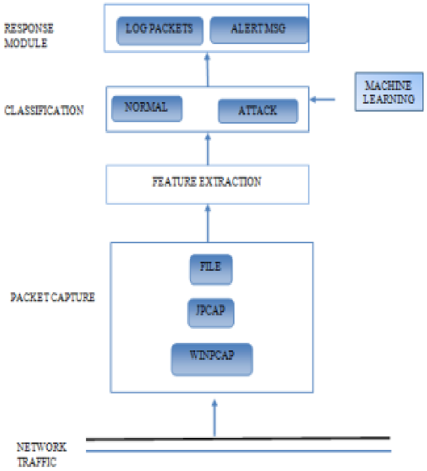

Figure 3. The structure of an auto encoder

IV. PROPOSED SYSTEM

System is an orderly group of interdependent components linked together according to a plan to achieve a specific objective. Its main characteristics are organization, interaction, interdependence, integration and a central objectivAs it is mentioned earlier RNN networks have the ability to remember previous entries [20]. This means that we can use RNN as a time analysis tool. In many modern and sophisticated attacks malicious codes is injected in distributed patterns using bonnets or embedding the codes among many legal packets [21]. Data pre-processing is required to transform the data into a format usable by machine learning algorithms. After these operations, the properties to be used by the algorithms are decided in the feature selection. The randomness of these two deep methods obliges us to use a more robust classifier.

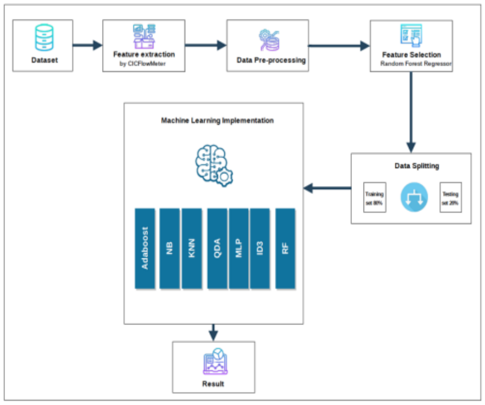

Fig. 4. Proposed Operation Scheme

V. METHODOLOGYS

The network security datasets are available in two ways, First, from packet monitoring software such as Wire shark, Tcp dump, Win Dump etc but these data will not be labeled and a lot of time will go into labeling hence may not be suitable for modeling purposes [22]. One machine learning algorithm or technique for developing an intrusion detection system can be used as a standalone classifier or single classifier.

Decision Tree Creating a classifier for predicting the value of a target class for an unseen test instance, based on several already known instances is the task of Decision tree (DT). Through a sequence of decisions, an unseen test instance is being classified by a Decision tree.

Naive Bayes: On the basis of the class label given Naive Bayes assumes that the attributes are conditionally independent and thus tries to estimate the class-conditional probability [15]. Naive Bayes often produces good results in the classification where there exist simpler relations

K-nearest neighbor: Various distance measure techniques are being used in K-nearest neighbor. K-nearest neighbor finds out k number of samples in training data that are nearest to the test sample and then it assigns the most frequent class label among the considered training samples to the test sample.

Artificial Neural Network: (ANN) is a processing unit for information which was inspired by the functionality of human brains [23]. Typically neural networks are organized in layers which are made up of a number of interconnected nodes which contain a function of activation.

Support Vector Machines: (SVM) was introduced in mid1990's [5]. The concept behind SVM for intrusion detection basically is to use the training data as a description of only the normal class of objects or which is known as non-attack in intrusion detection system and thus assuming the rest as anomalies [11].

Fuzzy Logic: For reasoning purpose, dual logic's truth values can be either absolutely false (0) or absolutely true (1), but in Fuzzy logic these kinds of restrictions are being relaxed [60]. That means in Fuzzy logic the range of the degree of truth of a statement can hold the value between 0 and 1 along with '0' and '1'[13].

VI. GENETIC ALGORITHMS

Genetic algorithms are a family of problem-solving techniques based on evolution and natural selection. This section gives a brief overview of genetic algorithms. The goal of genetic algorithms is to create optimal solutions to problems. Potential solutions to the problem to be solved are encoded as sequences of bits, characters or numbers [23].

An Auto-Encoder (AE) is a type of neural networks with the same number of neurons in both input and output layer [24]. It is mainly used for dimensionality reduction for better representation of data. Auto-Encoder is an unsupervised learning model and applies back propagation. The input and output layer consist of N nodes and hidden layer consist of K nodes. Hidden layer of AE is known as abstract layer. For a given training data set X with m samples, the encoder performs the mapping of input vector to hidden vector using mapping function.

```
Input: Dataset D= {x1, x2, ..…..xm} with m samples, number of hidden layers L
Output: Output of each hidden unit
Step 1: for l ∈[1, L] do
Step 2: initialize Wl= 0, Wl ’= 0, bl= 0, bl ’= 0
Step 3: define the l-th hidden layer representation vector hl= f(Wl hl-1+ bl)
Step 4: define the l-th hidden layer output xl’= f(Wl’hl+bl)
Step 5: while not stopping criterion do
Step 6: calculate hl from hl-1
Step 7: calculate yl
Step 8: calculate the loss function
Step 9: update layer parameters θl= (Wl, bl) and θl’= (Wl ’ , bl ’ )
Step 10: end while
Step 11: end for
Step 12: Initialize (Wl+1, bl+1) at the supervised layer
Step 13: calculate the labels for each sample xi of the training dataset D
Step 14: perform BP in a supervised way to tune parameter of all layers;
```

VII. EXPERIMENT RESULTS

Through the review on cyber-attack detection using deep learning algorithms, the following challenges are addressed. By using multi-layer perceptron (MLP)-based intrusion detection scheme, a mean square error was slightly high.. The performance of Deep auto encoder is compared with classical machine learning algorithms. Support vector machine and artificial neural network are the most popular approaches for single learning algorithm classifiers and number of comparative samples is less but the comparison result implies that Support Vector machine is by far the most common and considered single classification technique. Hybrid classifiers in intrusion detection have established in the mainstream study due to the performance accuracy in recent times Statistics shows hybrid classifiers have the highest number of articles used algorithms in each article and their performance in intrusion detection system.

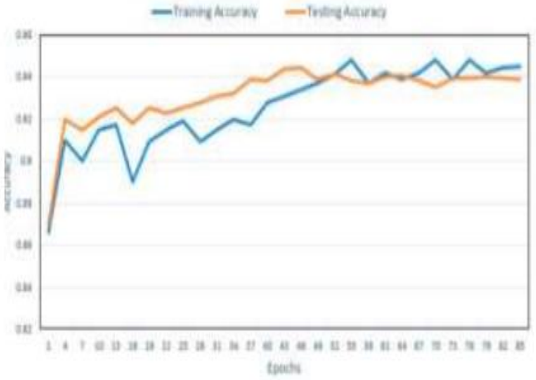

Fig. 5. Learning curve accuracy for the multiclass classification model

VIII. CONCLUSION AND FUTURE WORK

The present time, assessments of help vector machine, ANN, CNN, Random Forest and significant learning estimations reliant upon current CICIDS2017datasetwere presented moderately. Results show that the significantlearning estimation performed generally best results over SVM, ANN, RFand CNN. In this paper a deep auto-encoder based intrusion detection system has been proposed and its performance is compared with classical machine learning algorithms on the NSL-KDD dataset.Our proposed IDS is able to update the dataset and learn to deal with new misclassified records. The response time should be as low as possible for detecting and stopping cyber attacks and the false negative rate should be as low as possible. Since the false negative rate is inversely proportional to recall, the recall score should be as high as possible. There are multiple new intrusions were seen within each broader category. When the model was trained and evaluated on the train-validation split, the model performance was quite high, compared to test set accuracy where new intrusions are seen. Removal of redundant and irrelevant features for the training phase is a key factor for system performance. Consideration of feature selection will play a vital role in the classification techniques in future work we applied various machine learning algorithms independently from each other. In the future, we sould like to combine different machine learning algorithms as a multi-layered model to improve the detection performance.

REFERENCES

[1] Z. N. Zarandi and I. Sharifi, "Detection and Identification of Cyber-Attacks in Cyber-Physical Systems Based on Machine Learning Methods," 2020 11th International Conference on Information and Knowledge Technology (IKT), 2020, pp. 107-112, doi: 10.1109/IKT51791.2020.9345627.

[2] Nurjahan, F. Nizam, S. Chaki, S. Al Mamun and M. S. Kaiser, "Attack detection and prevention in the Cyber-Physical System," 2016 International Conference on Computer Communication and Informatics (ICCCI), 2016, pp. 1-6, doi: 10.1109/ICCCI.2016.7480022.

[3] Ding Chen, Qiseng Yan, Chunwang Wu, and Jun Zhao, “SQL Injection Attack Detection and Prevention Techniques Using Deep Learning,” Journal of Physics: Conference Series, Volume 1757, International Conference on Computer Data and Artificial Intelligence (ICCBDAI 2020) October 2020, Changsha, ChinaGyusoo Kim and Seulgi Lee, “2014 Payment Research”, Bank of Korea, Vol. 2015, No. 1, Jan. 2015.

[4] ErcanNurcanYılmaz, SerkanGönen, “Attack detection/prevention system against cyber-attack in industrial control systems,” Computers & Security Volume 77, August 2018, pp 94-105

[5] Arpitha. B, Sharan. R, Brunda. B. M, Indrakumar. D. M, Ramesh. B. E, “Cyber Attack Detection and notifying system using ML Techniques,” International Journal of Engineering Science and Computing (IJESC), Volume 11, Issue No.06

[6] Yirui Wu, Dabao Wei, and Jun Feng, “Network Attacks Detection Methods Based on Deep Learning Techniques: A Survey,” Security Threats to Artificial Intelligence-Driven Wireless Communication Systems, 2020.

[7] RafałKozik, MichałChoraś, “Machine Learning Techniques for Cyber Attacks Detection,” Image Processing and Communications Challenges 5, pp 391-398, Springer International Publishing Switzerland 2014.

[8] Nutjahan, FarhanaNizam, ShudarshonChaki, Shamim Al Mamun, M. Shamim, “Attack Detection and Prevention in the Cyber Physical System,” 2016 International Conference on Computer Communication and Informatics (IEEE -2016), Jan. 07 - 09, 2016, Coimbatore, India

[9] Anupong, W., Yi-Chia, L., Jagdish, M., Kumar, R., Selvam, P. D., Saravanakumar, R., &Dhabliya, D. (2022). Hybrid distributed energy sources providing climate security to the agriculture environment and enhancing the yield. Sustainable Energy Technologies and Assessments, 52 doi:10.1016/j.seta.2022.102142

[10] Musleh, D.; Alotaibi, M.; Alhaidari, F.; Rahman, A.; Mohammad, R.M. Intrusion Detection System Using Feature Extraction with Machine Learning Algorithms in IoT. J. Sens. Actuator Netw. 2023, 12, 29. https://doi.org/10.3390/jsan12020029.

[11] Mohammed Maithem and Ghadaa A. Al-sultany, Network intrusion detection system using deep neural networks, 2021 J. Phys.: Conf. Ser. 1804012138.

[12] LirimAshiku, CihanDagli, Network Intrusion Detection System using Deep Learning, Procedia Computer Science, Volume 185, 2021, Pages 239- 247, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2021.05.025.

[13] Plaka, R. (2021). INTRUSION DETECTION USING MACHINE LEARNING FOR

INDUSTRIAL CONTROL SYSTEMS (Dissertation).

[14] M. Pordelkhaki, S. Fouad and M. Josephs, "Intrusion Detection for Industrial Control Systems by Machine Learning using Privileged Information," 2021 IEEE International Conference on Intelligence and Security Informatics (ISI), San Antonio, TX, USA, 2021, pp. 1-6, doi: 10.1109/ISI53945.2021.9624757.

[15] Pallepati, Manvith. (2022). Network Intrusion Detection System Using Machine Learning with Data Preprocessing and Feature Extraction.International Journal for Research in Applied Science and Engineering Technology. 10. 2360-2365. 10.22214/ijraset.2022.44326.

[16] N. Koroniotis, N. Moustafa, E. Sitnikova, and J. Slay, "Towards developing network forensic mechanism for botnet activities in the iot based on machine learning techniques," International Conference on Mobile Networks and Management, pp. 30–44, 2017.

[17] Y. N. Soe, Y. Feng, P. I. Santosa, R. Hartanto, and K. Sakurai, "Rule generation for signature based detection systems of cyber attacks in iot environments," Bulletin of Networking, Computing, Systems, and Software, vol. 8, no. 2, pp. 93–97, 2019.

[18] V. H. Bezerra, V. G. T. da Costa, S. B. Junior, R. S. Miani, and B. B. Zarpelao, "One-class classification to detect botnets in iot devices," Anais do XVIII SimposioBrasileiroemSeguranc¸a da Informac¸ ´ ao e ˜ de SistemasComputacionais, pp. 43–56, 2018.

[19] E. Hodo, X. Bellekens, A. Hamilton, P.-L.Dubouilh, E. Iorkyase, C. Tachtatzis, and R. Atkinson, "Threat analysis of iot networks using artificial neural network intrusion detection system," International Symposium on Networks, Computers and Communications (ISNCC), pp. 1–6, 2016.

[20] D. H. Summerville, K. M. Zach, and Y. Chen, "Ultra-lightweight deep packet anomaly detection for internet of things devices," IEEE 34th international performance computing and communications conference (IPCCC), pp. 1–8, 2015.

[21] F. Y. Yavuz, "Deep learning in cyber security for internet of things," Ph.D. dissertation, 2018.

[22] O. Ibitoye, O. Shafiq, and A. Matrawy, "Analyzing adversarial attacks against deep learning for intrusion detection in iot networks," arXiv preprint arXiv:1905.05137, 2019.

[23] I. Cvitic, D. Perakovi ´ c, M. Peri ´ sa, and M. Botica, "Novel approach for ˇ detection of iot generated ddos traffic," Wireless Networks, pp. 1–14, 2019.

[24] Z. A. Baig, S. Sanguanpong, S. N. Firdous, T. G. Nguyen, C. So-In et al., "Averaged dependence estimators for dos attack detection in iot networks," Future Generation Computer Systems, vol. 102, pp. 198– 209, 2019.

[25] A. H. Lashkari, G. Draper-Gil, M. S. I. Mamun, and A. A. Ghorbani, "Characterization of tor traffic using time based features." ICISSP, pp. 253–262, 2017