International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(2) 2026 pp. 601-610
https://doi.org/10.70917/ijcisim-2026-2962
Article

Development of Intrusion Detection for Secure Social Internet of Things Based on Collaborative Edge Computing by Generative Adversarial Network using Hybrid Algorithms

Ashwini P[1], Ramesh Sengodan[2]

[1]Presidency University,iamashwinip@gmail.com
[2] PSAIAC, ramesh.sengodan@presidencyuniversity.in

Abstract: Abstract: As cyber-attacks and cybercrime against cyber-physical systems (CPSs) increase, it is become harder to spot these intrusions. In the modern cyber environment, intrusion detection is one of the major security issues. A sizable number of methods that are based on machine learning strategies have been created. Thus, we have developed machine learning techniques for spotting the infiltration. When compared to conventional machine learning (ML) solutions, deep learning (DL) offers higher performance. When there is enough information, DL models nearly always produce great results. In contrast to other domains like NLP, image processing, software vulnerability, and many more, DL models have just recently been used to address the CPS cybersecurity issue.. Furthermore, it has been noted that a large number of DL models have been put out in recent papers to identify CPS cyber-attacks. The degree of complexity when superimposing cybersecurity on CPSs was attributed as a widely recognised explanation for why it is difficult to detect cyberattacks on CPSs. The UNSW-NB15 dataset was used in this system by referencing a repository of data. Finally, we must put the various categorization algorithms—including CNN, GAN, and logistic regression (LR)—into practise. The experimental findings demonstrate the correctness of the aforementioned algorithms.

Keywords: CNN, GAN, cyber-attacks, CPS etc.

1. Introduction

The Social Internet of Things (SIoT) now penetrates our daily lives. As a strategy to alleviate the escalation of resource congestion, collaborative edge computing (CEC) has become a new paradigm for solving the needs of the Internet of Things (IoT). CEC can provide computing, storage, and network connection resources for remote devices. Because the edge network is closer to the connected devices, it involves a large amount of users' privacy. This also makes edge networks face more and more security issues, such as Denial-of-Service (DoS) attacks, unauthorized access, packet sniffing, and man-inthe-middle attacks.

To combat these issues and enhance the security of edge networks, we propose a deep learning-based intrusion detection algorithm. Based on the generative adversarial network (GAN), we designed a powerful intrusion detection method. Our intrusion detection method includes three phases. First, we use the feature selection module to process the collaborative edge network traffic. Second, a deep learning architecture based on GAN is designed for intrusion detection aiming at a single attack. Finally, we propose a new intrusion detection model by combining several intrusion detection models that aim at a single attack. Intrusion detection aiming at multiple attacks is realized through the designed GAN-based deep learning architecture.

Cyber-physical systems (CPSs) suffer from cyber-attacks when they are increasingly connected to the cyber space. Cyber-attacks have become increasingly sophisticated and prevalent as automated attacking tools, and professional hacking groups have started to get involved. A successful cyber attack against a CPS may be disastrous, catastrophic, or even fatal.

Many CPS systems lack cyber security mechanisms like message authentication, resulting in challenges to detect false data injection attacks. A lack of universal encryption, especially on the systems employing dated technologies, makes it challenging to defend against eavesdropping attacks.

Hacking incidents are increasing day by day as technology rolls out. A large number of hacking incidents are reported by companies each year. Distributed Denial of Service (DDoS) attack was launched against Estonian websites in 2007.On June 17, 2008, Amazon started receiving some authenticated request from multiple users in one of its location.

The requests began to increase significantly causing the servers slow down. On Jan 2013, European Network and Information Security Agency (ENISA) reported that Dropbox was attacked by DDoS and suffered a substantial loss of service for more than 15 hours affecting all users across the globe. Facebook was hit by suspected distributed denial of service attack on Sept 28, 2014. Reported that some form of network scanning activity precedes 50% of the attacks against cyber systems.

Attackers are not only launching flooding and probing attacks but also spreading malware files in the form of virus, worm, spams to exploit the vulnerabilities present in existing software, causing a threat to the sensitive information of users stored on machines. Cisco Annual Security report mentioned that spam related to the Boston Marathon bombing comprised 40% of all spam messages delivered worldwide on April 17, 2013. On a recent survey done by Cisco in 2017, Trojan was classified as one of the top five malware which is used to gain initial access to the user's computers and organizational networks.

Hence, security in such a complex technological environment is a big challenge and needs to be tackled intelligently. Researchers have considered a different category of attacks for intrusion detection. For example, Denial of Service (DoS) attacks (Bandwidth and Resource Depletion), Scanning attacks (Probe) and Remote to Local (R2L) attacks and User to Root (U2R) attacks which are based on KDD'99 dataset. A recent attack dataset (UNSW-NB), classifies attacks into nine categories: Fuzzer, Analysis, Reconnaissance, ShellCode, Worm, Generic, DoS, Exploit and Generic.

All these attacks have been discussed in detail in Section III. Current security solutions include the use of middle-boxes such as Firewall, Antivirus and Intrusion Detection Systems (IDS). A firewall controls traffic that enters or leaves a network based on source or destination address. It alters the traffic according to the firewall rules. Firewalls are also limited to the amount of state available and their knowledge of the hosts receiving the content. An IDS is a type of security tool that monitors network traffic and scans the system for suspicious activities and alerts the system or network administrator.

A Network based Intrusion Detection System (NIDS) is usually placed at network points such as a gateway and routers to check for intrusions in the network traffic. At high-level, the detection mechanism used by these IDSes are of three types: misuse detection, anomaly detection, and hybrid detection. In misuse detection approach, IDS maintains a set of the knowledge base (rules) for detecting the known attack types.

Misuse detection techniques can be broadly classified into Knowledge based and machine learning based techniques. In the knowledge based technique, network traffic or host audit data (such as system call traces) are compared against predefined rules or attack patterns. Knowledge based techniques can be categorized into three types: (i) Signature matching (ii) State transition analysis and (iii) Rule based expert systems.

Some papers investigated data-driven methods for detecting cyber-attacks against CPS systems. However, there is no detailed discussion on applying DL methods to detect CPS cyber-attacks. A short survey was provided with a four-step framework to apply DL methods on CPS issues, including cybersecurity, adaptability, recoverability, and many more, without a specific focus on cybersecurity.

A comprehensive survey on the cyber-attacks against CPSs was presented in without investigating the DL models. Various methods of detecting cyber-attacks in the CPSs were summarized in without using DL methods. A comprehensive list of CPS attacks and challenges were provided in but overlooking ML, or DL approaches.

1.1 Objective

The main objective of our project is,

- To classify or predict the cyber-attacks in network effectively.
- To implement the feature selection for select the best features.
- To implement the feature extraction for extract the features by using PCA.

- To implement the different classification algorithms for better performances.
- To enhance the overall performance for classification algorithms.

1.2 Existing System

In existing system, a holistic view of recently proposed DL solutions is provided to cyber-attack detection in the CPS context. A six-step DL driven methodology is provided to summarize and analyze the surveyed literature for applying DL methods to detect cyber-attacks against CPS systems. The methodology includes CPS scenario analysis, cyber-attack identification, ML problem formulation, DL model customization, data acquisition for training, and performance evaluation. The reviewed works indicate great potential to detect cyber-attacks against CPS through DL modules. Moreover, excellent performance is achieved partly because of several high-quality datasets that are readily available for public use. Furthermore, challenges, opportunities, and research trends are pointed out for future research.

1.3 Proposed System

In this system, the UNSW-NB15 dataset was taken as input. The input data was taken from the dataset repository. Then, we have to implement the data pre-processing step. In this step, we have to handle the missing values for avoid wrong prediction, and to encode the label for input data. Then, we have to split the dataset into test and train. The data is splitting is based on ratio. In train, most of the data's will be there. In test, smaller portion of the data's will be there. Training portion is used to evaluate the model and testing portion is used to predicting the model. Then, we have to implement the classification algorithm (i.e.) machine and deep learning. The machine learning algorithms such as Logistic regression and deep learning algorithm such as CNN and GAN. Finally, the experimental results shows that the performance metrics such as accuracy and comparison results.

Advantages:

- It is efficient for large number of datasets.
- Time consumption is low.
- The process is implemented with removing unwanted data.

1.4 Design Methodology

The methodology shown in figure 1 is explained in the following modules. there are 7 modules are explained in detail, they are

1. Data selection
2. Pre-processing
3. Data splitting
4. Feature Selection
5. Feature Extraction
6. Classification
7. Result generation

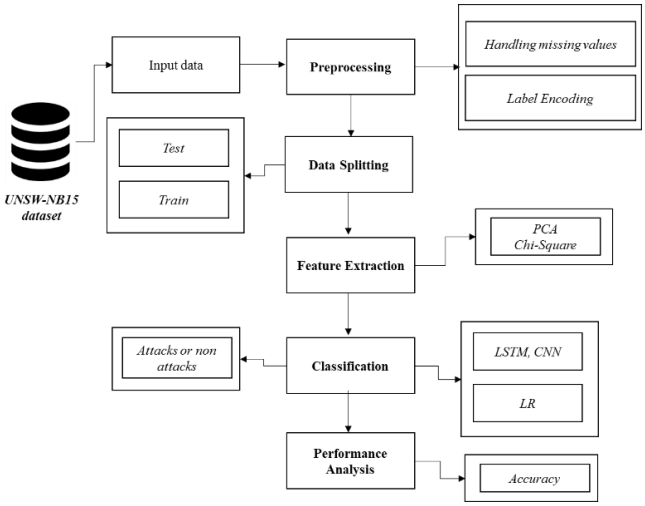


Fig.1. The design methodology of the proposed technique

Data Selection:

The input data was collected from dataset repository. In our process, UNSW-NB15 dataset is used. The data selection is the process of detecting the cyber-attacks. The input dataset was taken from dataset repository such as UCI repository. The dataset contains the information such protocol, duration, host error rate, attack category, label and so on. In python, with the help of panda's package, we can read or load our input dataset. Our dataset is in the format is '.csv'

Data Pre-processing:

Data pre-processing is the process of removing the unwanted data from the dataset. Pre-processing data transformation operations are used to transform the dataset into a structure suitable for machine learning. This step also includes cleaning the dataset by removing irrelevant or corrupted data that can affect the accuracy of the dataset, which makes it more efficient. Missing data removal Encoding Categorical data Missing data removal: In this process, the null values such as missing values and Nan values are replaced by 0. Missing and duplicate values were removed and data was cleaned of any abnormalities. Encoding Categorical data: That categorical data is defined as variables with a finite set of label values. That most machine learning algorithms require numerical input and output variables.

Data Splitting:

During the machine learning process, data are needed so that learning can take place. In addition to the data required for training, test data are needed to evaluate the performance of the algorithm in order to see how well it works. In our process, we considered 80% of the input dataset to be the training data and the remaining 20% to be the testing data. Data splitting is the act of partitioning available data into two portions, usually for cross-validator purposes. One Portion of the data is used to develop a predictive model and the other to evaluate the model's performance. Separating data into training and testing sets is an important part of evaluating data mining models. Typically, when you separate a data set into a training set and testing set, most of the data is used for training, and a smaller portion of the data is used for testing.

Feature Selection:

The chi-square test is a hypothesis test designed to test for a statistically significant relationship between nominal and ordinal variables organized in a bivariate table. In other words, it tells us whether two variables are independent of one another.

Feature Extraction:

In our system, we have to implement the PCA (principle Component Analysis) for extracting the features from splitted data. Principal component analysis, or PCA, is a statistical procedure that allows you to summarize the information content in large data tables by means of a smaller set of "summary indices" that can be more easily visualized and analysed. It is used to reduce the number of dimensions in healthcare data. PCA can help resize an

image. It can be used in finance to analyze stock data and forecast returns. PCA helps to find patterns in the high-dimensional datasets.

Classification:

In our process, we have to implement the different classification algorithm such as LR and CNN, GAN.

Logistic Regression is a Machine Learning algorithm which is used for the classification problems, it is a predictive analysis algorithm and based on the concept of probability. The hypothesis of logistic regression tends it to limit the cost function between 0 and 1.

Generative adversarial networks (GANs) are an exciting recent innovation in machine learning. GANs are generative models: they create new data instances that resemble your training data. For example, GANs can create images that look like photographs of human faces, even though the faces don't belong to any real person.

Convolutional neural network (CNN) is a deep learning neural network sketched for processing structured arrays of data such as portrayals. CNN are very satisfactory at picking up on design in the input image, such as lines, gradients, circles, or even eyes and faces.

Flow Diagram: The flow diagram of the working of this model is as shown in figure 2.

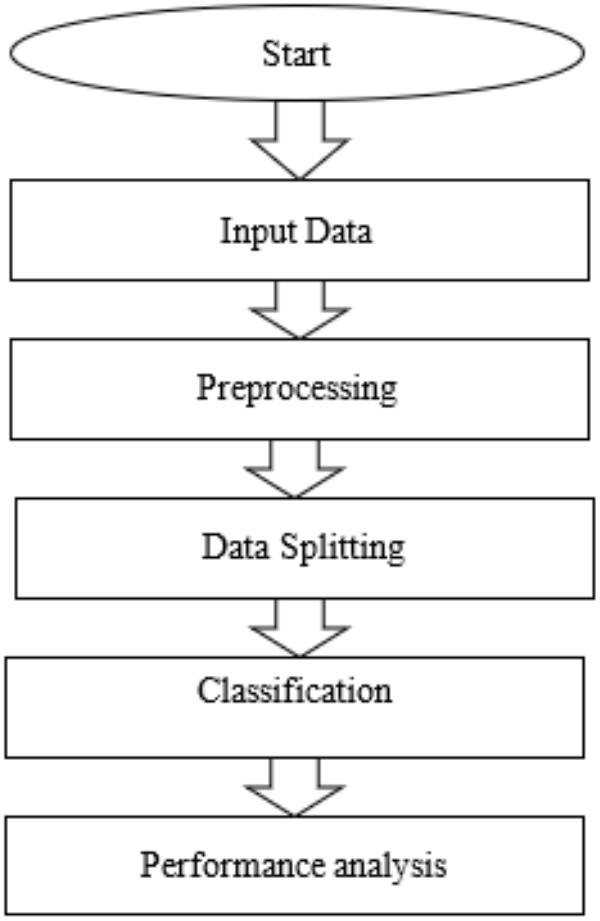
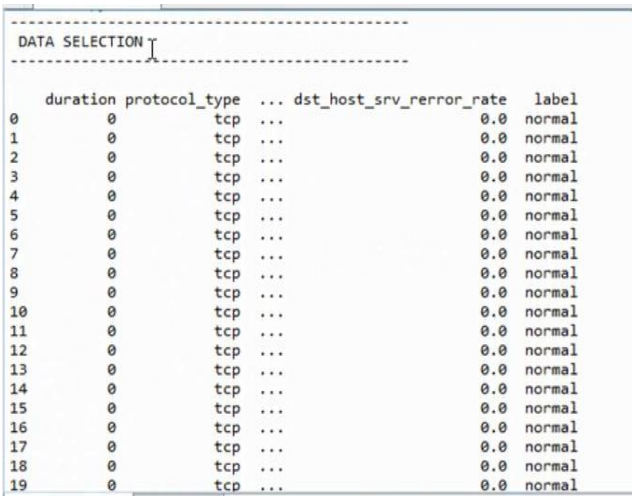


Figure 2: Flow diagram of the proposed techniques

1.5 Experiments and Discussion

The Final Result will get generated based on the overall classification and prediction. The performance of this proposed approach is evaluated using some measures like,



DATA SELECTION

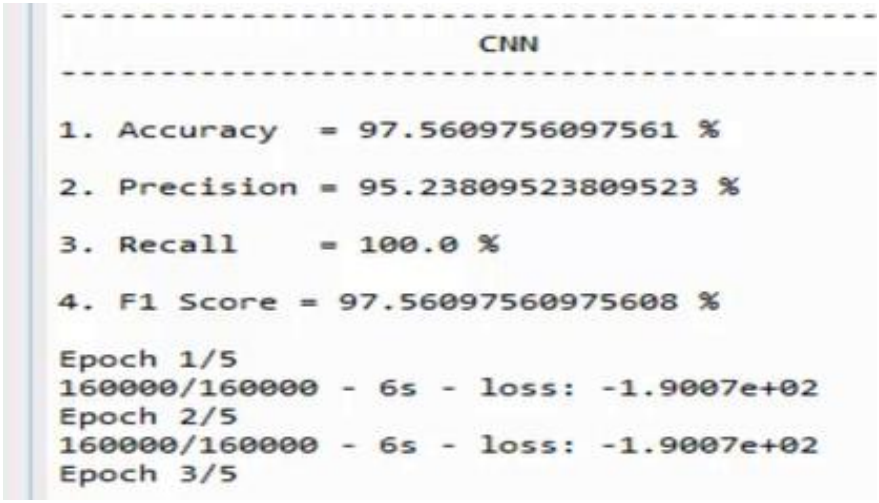
	duration	protocol_type	...	dst_host_srv_rerror_rate	label
0	0	tcp	...	0.0	normal
1	0	tcp	...	0.0	normal
2	0	tcp	...	0.0	normal
3	0	tcp	...	0.0	normal
4	0	tcp	...	0.0	normal
5	0	tcp	...	0.0	normal
6	0	tcp	...	0.0	normal
7	0	tcp	...	0.0	normal
8	0	tcp	...	0.0	normal
9	0	tcp	...	0.0	normal
10	0	tcp	...	0.0	normal
11	0	tcp	...	0.0	normal
12	0	tcp	...	0.0	normal
13	0	tcp	...	0.0	normal
14	0	tcp	...	0.0	normal
15	0	tcp	...	0.0	normal
16	0	tcp	...	0.0	normal
17	0	tcp	...	0.0	normal
18	0	tcp	...	0.0	normal
19	0	tcp	...	0.0	normal

Figure 3. Data Selection with TCP protocol type

Performance measurement by accuracy:

Accuracy of classifier refers to the ability of classifier. It predicts the class label correctly and the accuracy of the predictor refers to how well a given predictor can guess the value of predicted attribute for a new data.

$$AC = \frac{TP + TN}{TP + TN + FP + FN}$$



```
-------------------------------------------
                  CNN
-------------------------------------------

1. Accuracy  = 97.5609756097561 %

2. Precision = 95.23809523809523 %

3. Recall    = 100.0 %

4. F1 Score = 97.56097560975608 %

Epoch 1/5
160000/160000 - 6s - loss: -1.9007e+02
Epoch 2/5
160000/160000 - 6s - loss: -1.9007e+02
Epoch 3/5
```

Figure 4. Description of Accuracy Precision, Recall and F1 score for first data set.

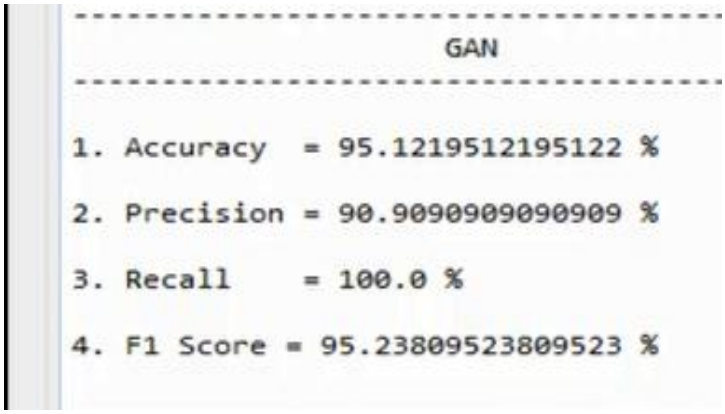
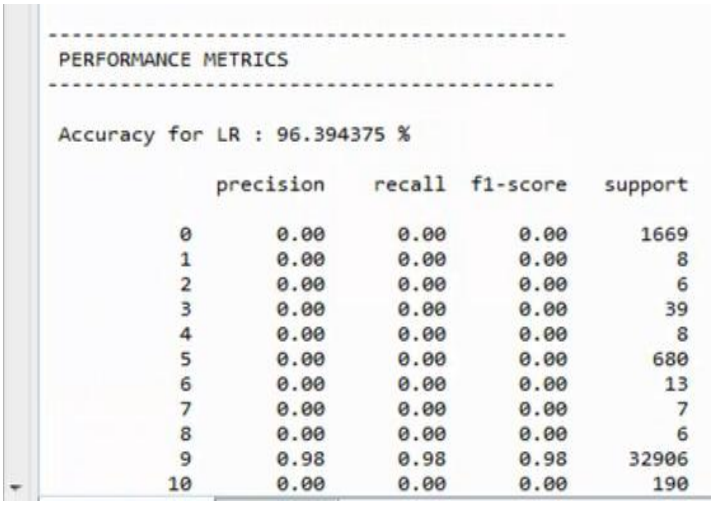


Figure 4. Description of Accuracy Precision, Recall and F1 score for another data set.



	precision	recall	f1-score	support
0	0.00	0.00	0.00	1669
1	0.00	0.00	0.00	8
2	0.00	0.00	0.00	6
3	0.00	0.00	0.00	39
4	0.00	0.00	0.00	8
5	0.00	0.00	0.00	680
6	0.00	0.00	0.00	13
7	0.00	0.00	0.00	7
8	0.00	0.00	0.00	6
9	0.98	0.98	0.98	32906
10	0.00	0.00	0.00	190

Figure 4. Logical Regression with CNN performance metrics

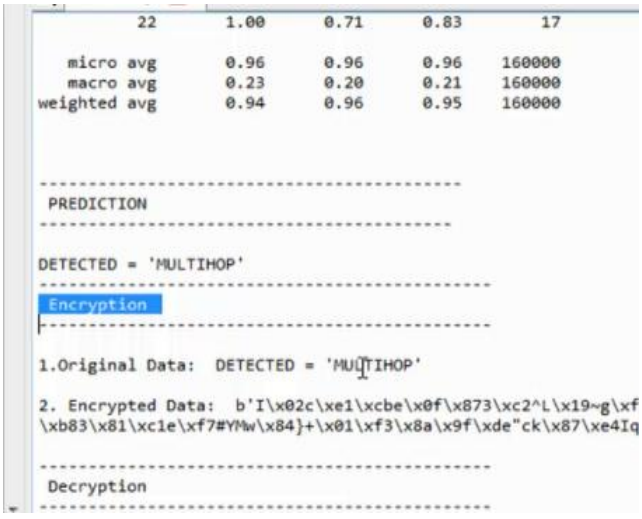


Figure 4. Multihop detection output representation for Hybrid CNN with LR

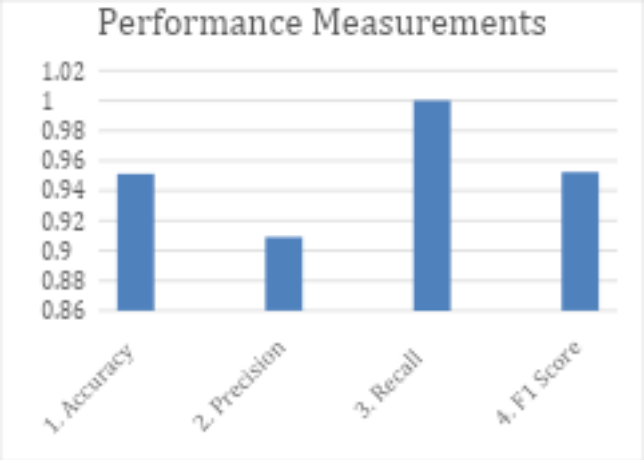


Figure 5. Bar Graph illustrating the Performance Measurements

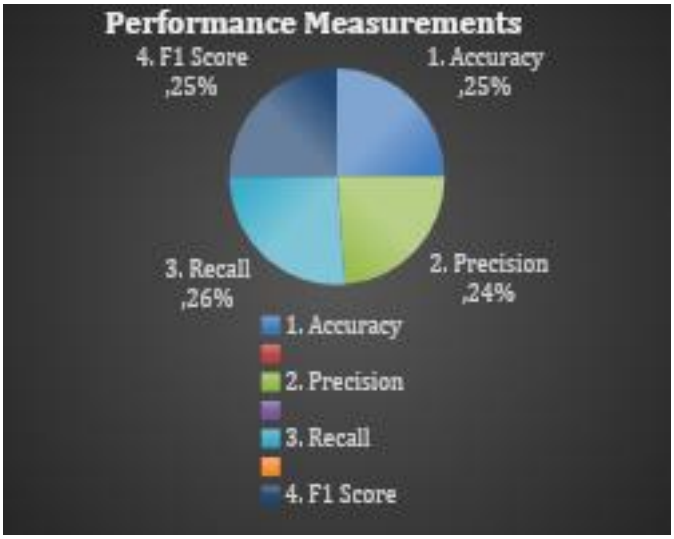


Figure 6. Pie chart illustrating the Performance Measurements

1.6 CONCLUSION AND FUTURE WORK

We conclude that, the UNSW-NB15 dataset was taken as input. The input dataset was mentioned in our research paper. We are implemented the classification algorithms (i.e) machine and deep learning algorithms. Then, machine learning algorithms such as logistic regression and deep learning algorithm such as CNN, GAN. Finally, the result shows that the accuracy for above mentioned algorithm and estimated the performances metrics such as accuracy for both algorithms and comparison graph.

In the future, we should like to hybrid the two different machine learning or to hybrid two deep learning algorithms. In future, it is possible to provide extensions or modifications to the proposed clustering and classification algorithms to achieve further increased performance. Apart from the experimented combination of data mining techniques, further combinations and other clustering algorithms can be used to improve the detection accuracy.

Reference

1. A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, “Survey of intrusion detection systems: techniques, datasets and challenges,” Cybersecurity, vol. 2, no. 1, 2019.
2. H. Hindy, D. Brosset, E. Bayne, A. Seeam, C. Tachtatzis, R. Atkinson,and X. Bellekens, “A Taxonomy and Survey of Intrusion DetectionSystem Design Techniques, Network Threats and Datasets,” vol. 1,no. 1, 2018.

3. J. Navarro, A. Deruyver, and P. Parrend, “A systematic survey on multi-step attack detection,” Computers & Security, vol. 76, pp. 214–249, 2018.
4. L. Xiao, Y. Li, X. Huang, and X. Du, “Cloud-based malware detection game for mobile devices with offloading,” IEEE Transactions on Mobile Computing, vol. 16, no. 10, pp. 2742–2750, 2017.
5. O. Can, C. Turguner, and O. K. Sahingoz, “A Neural Network Based Intrusion Detection System For Wireless Sensor Networks,” Signal Process. Commun. Appl. Conf. (SIU), 2015 23th, pp. 2302–2305, 2015.
6. R. Zuech, T. M. Khoshgoftaar, and R. Wald, “Intrusion detection and Big Heterogeneous Data: a Survey,” Journal of Big Data, vol. 2, p. 3, 2015
7. S. Yinbiao and K. Lee, “Internet of Things: Wireless Sensor Networks Executive summary,” 2014. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci
8. “A survey on sensor networks,” IEEE Commun. Mag., vol. 40, no. 8, pp. 102–105, 2002.
9. X. Chen, K. Makki, K. Yen, and N. Pissinou, “Sensor network security: a survey,” IEEE Commun. Surv. Tutorials, vol. 11, no. 2, pp. 52–73, 2009.
10. A.-S. K. Pathan, H.-W. Lee, and C. S. Hong, “Security in wireless sensor networks: issues and challenges,” 2006 8th Int. Conf. Adv. Commun. Technol., vol. 2, p. 6 pp.-pp.1048, 2006.
11. P. Yi, Y. Jiang, Y. Zhong, and S. Zhang, “Distributed Intrusion Detection for Mobile Ad Hoc Networks,” 2005 Symp. Appl. Internet Work. (SAINT 2005 Work. pp. 94–97, 2005.
12. H. Sedjelmaci and M. Feham, “Novel Hybrid Intrusion Detection System for Clustered Wireless Sensor Network,” Int. J. Netw. Secur. It’s Appl. (IJNSA), Vol.3, No.4, July 2011, vol. 3, no. 4, pp. 1–14, 2011.
13. L. Khan, M. Awad, and B. Thuraisingham, “A new intrusion detection system using support vector machines and hierarchical clustering,” VLDB J., vol. 16, no. 4, pp. 507–521, 2007.
14. S. K. Sahu, S. Sarangi, and S. K. Jena, “A detail analysis on intrusion detection datasets,” Souvenir 2014 IEEE Int. Adv. Comput. Conf. IACC 2014, pp. 1348–1353, 2014.
15. F. Lu and L. Wang, “Intrusion Detection System Based on Integration of Neural Network for Wireless Sensor Network,” J. Softw. Eng. 2014.
16. Y. Y. Li and L. E. Parker, “Intruder detection using a wireless sensor network with an intelligent mobile robot response,” Southeastcon, 2008. IEEE, pp. 37–42, 2008.
17. A. Kulakov and D. Davcev, “Tracking of unusual events in wireless sensor networks based on artificial neural-networks algorithms,” Inf. Technol. Coding Comput. 2005. ITCC 2005. Int. Conf., pp. 534–539, 2005.
18. M. Panda, “Security Threats at Each Layer of Wireless Sensor Networks,” Int. J. Adv. Res. Comput.Sci. Softw. Eng., vol. 3, no. 11, pp. 61–67, 2013.
19. Karlof and D. Wagner, “Secure routing in wireless sensor networks: attacks and countermeasures,” Proc. First IEEE Int. Work. Sens. Netw. Protoc. Appl. 2003., pp. 113–127, 2003.
20. J. Song, H. Takakura, Y. Okabe, M. Eto, D. Inoue, and K. Nakao, “Statistical analysis of honeypot data and building of Kyoto 2006+ dataset for NIDS evaluation,” Proceedings of the 1st Workshop on Building Analysis Datasets and Gathering Experience Returns for Security, BADGERS 2011, pp. 29–36, 2011.
21. K. Kendall and A. C. Smith, “A Database of Computer Attacks for the Evaluation of Intrusion Detection Systems by A Database of Computer Attacks for the Evaluation of Intrusion Detection Systems,” 1999.
22. “Unified host and network dataset.” [Online]. Available: https://csr.lanl.gov/data/2017.html
23. [23] R. Sommer and V. Paxson, “Outside the Closed World: On Using Machine Learning For Network Intrusion Detection,” pp. 305–316, 2010.
24. S. Abt and H. Baier, “Are We Missing Labels? A Study of the Availability of Ground-Truth in Network Security Research,” Proceedings - 3rd International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security, BADGERS 2014, pp. 40–55, 2016.
25. M. Baykara and R. Das,, “A Survey on Potential Applications of Honeypot Technology in Intrusion Detection Systems,” International Journal of Computer Networks and Applications (IJCNA, vol. 2, no. 5.
26. I. Yahya, M. Al, P. Chauhan, S. Shukla, and M. B. Potdar, “Review on efficient log analysis to evaluate multiple honeypots using ELK,” no. 6, pp. 492–504, 2016.
27. A. Sperotto, G. Schaffrath, R. Sadre, C. Morariu, A. Pras, and B. Stiller, “An overview of ip flow-based intrusion detection,” IEEE Communications Surveys Tutorials, vol. 12, no. 3, pp. 343–356, 2010.
28. P. Sangkatsanee, N. Wattanapongsakorn, and C. Charnsripinyo, “Practical real-time intrusion detection using machine learning approaches,” Computer Communications, vol. 34, no. 18, pp. 2227–2235, 2011.
29. N. Moustafa and J. Slay, “UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set),” in 2015 Military Communications and Information Systems Conference, MilCIS 2015 - Proceedings. Institute of Electrical and Electronics
30. Y. Li, D. E. Quevedo, S. Dey, and L. Shi, “Sinr-based dos attack on remote state estimation: A game-theoretic approach,” IEEE Transactions on Control of Network Systems, vol. 4, no. 3, pp. 632–642, 2016.

Research Scholar:

Ashwini P receive the BCA Degree from Bangalore University in 2017 and MCA Degree from Bangalore city University in 2021. She is currently working as assistant professor in Vivekananda Institute of management and pursuing the PhD degree in Presidency University. Her current Research area includes Machine Learning, network security, IOT. Author can be contacted at email:iamashwinip@gmail.com

Guide:

Mr. Ramesh Sengodan, Professor, PhD in Computer Science and Engineering at Anna University Chennai, 30 years of experience, expert in Network Security, IoT, AI. Presently Working in Presidency University, Bangalore. In School of Computer Science.

Email:ramesh.sengodan@presidencyuniversity.in