International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 03 | March - 2023 Impact Factor: 7.185 ISSN: 2582-3930

DETECTION OF RANSOMWARE ATTACKS IN NETWORK USING MACHINE LEARNING

[1] B Ramakrishna [2] A Ajith Reddy [3] B Srinivas [4] Dr V Shanmukha Rao
[123] Department of Information Technology, Andhra Loyola Institute of Engineering andTechnology, Vijayawada, Andhra Pradesh.
[4] Associate Professor, Department of Information Technology,
Andhra Loyola Institute of Engineering and Technology, Vijayawada, Andhra Pradesh.

1.ABSTRACT
The development of computer and communication technology has resulted in considerable changes from the past. Despite the fact that utilising new inventions benefits people, organisations, and governments greatly, some people are prejudiced against them. For instance, information security in file systems, information accessibility, and so on.Due to several groups, including the criminal underworld, professionals, and digital activists, including dread of the digital world, which has caused many issues for individuals and organisations, has reached the point where it may jeopardise national and open security. Intrusion Detection Systems (IDS) were created as a result to maintain a safe distance from internet threats. The new CICIDS2017 dataset was utilised to train the Support vector machine (SVM) computations, which are currently being used to identify port sweep efforts. Instead of SVM, we might use other algorithms like CNN, ANN, and random forest.

Keywords: data security, information accessibility, digital fear, intrusion detection systems.

2.INTRODUCTION
Increasingly, political and economic entities interrupt, obliterate, or conceal information content in computer networks utilising sophisticated cyberwarfare. Network protocol resilience must be ensured against intrusions by strong attackers who can even control a portion of the network's parties. The controlled parties are capable of launching both passive (eavesdropping, non-participation) and active (jamming, message dropping, corruption, and forging) attacks. The system that continually monitors activity on a computer system or network, analyses it for indicators of possible issues, and, in many circumstances, stops unauthorized access. This is often done by automatically compiling data from a variety of networks and systems to check for potential security flaws. When it comes to effectively protecting networks and systems against more sophisticated attacks like denial of service, traditional intrusion detection and solutions like firewalls, access restricting mechanisms, and encryptions have serious shortcomings. Moreover, the majority of systems built using these approaches frequently identify false positives and false negatives and are unable to respond to evolving dangerous behaviour. Yet during the past ten years, a number of Machine Learning (ML) techniques have been applied to the problem of intrusion detection in the hopes of increasing detection rates and adaptability. These techniques are routinely used to keep attack information bases complete and up to date. The problem of cyber-security and defence against various cyber-attacks has recently gained a lot of attention. The exponential growth of computer technology is the main cause of this. a lot of useful apps used by individuals or groups for private or professional purposes, especially once the Internet of Things was recognised (IoT).

On a massive scale, the cyber-threats disrupt networks and cause considerable financial losses. Already implemented hardware and software

solutions Security measures include things like firewalls, user authentication, and data encryption technologies. Sadly, there was not enough protection for all of the machines connected to the computer network to deal with the problem of projected demand. Cyber-threats.

These conventional security measures are ineffective. Because intrusion detection systems have evolved more quickly and rigorously, they provide enough protection. Just access coming via the firewall is under control. The phrase "network to network" describes when two networks are unable to interact with one another.

Networks. In the case of an emergency, it does not, however, issue any notifications. So, it goes without saying that precise defence must be established. Machine learning-based methods to intrusion detection system (IDS) for system security Typically, an invasion A software or system that detects anything is called an identification system (IDS).

3.LITERATURE SURVEY

1.Neethu B. (2014) The Naive Bayes collection of features was given PCAA in order to create a network intrusion detection system. The KDD 1999 dataset was used in the study's experiments. The outcome shows the efficiency of the methodology when compared to neural network and tree algorithm techniques, obtaining a greater discovery rate, low time consumption, and low cost factor, with an accuracy of 94%.

2. Naseer et al. (2018) conceived, built, and trained their models utilising a range of deep neural network frameworks such as RNNs, Autoencoders, and CNNs; their models were also trained and evaluated using NSLKDD datasets. The accuracy of the DCNN and LSTM models was 85 percent and 89 percent, respectively.

3. Zhang et al. (2017) offered two methods for detecting network intrusions: direct and combination. Direct employs a single technique, whereas combination applies a number of procedures. Their suggestion is for a new belief-based directed acyclic graph (DAG) detection paradigm (BRB). The findings show that the DAG-BRB combinational model outperforms standard detection methods in terms of rate detection.

4.Wang et al. (2017) suggested a hierarchically spatial system that autonomously learns network traffic features.The system is an intrusive detection system that uses deep CNNs to learn spatial components and LSTM network features.

5. Shen et al. (2018) introduced Tiresias as a system for forecasting hazardous acts using deep roots learning. To predict what will happen on a computer, the system employs RNNs based on past observation. The testing was done with a commercial IPS dataset. Even in a complex environment, high precision and consistent findings were maintained, demonstrating that the approach was useful in forecasting future activities that may occur on a system with a 93% correctness rate.

6. Zhao et al. (2017) introduced a network attack identification paradigm that incorporates deep learning, flow computation, and an instantaneous detection and classification approach.Numerous tests were run and comparisons were made using the CICIDS2017 dataset. When compared to previous methodologies, the results showed a higher level of instant detection efficacy.

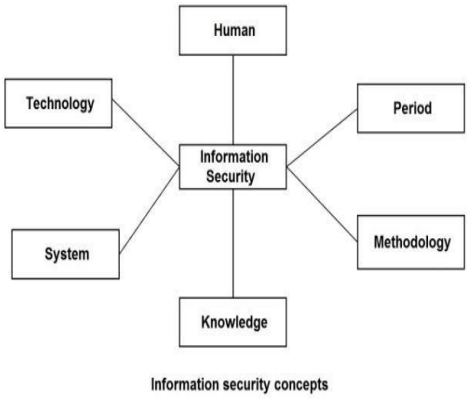

Fig 1:Information Security Concepts

4.IMPLEMENTATION STUDY

CICIDS2017 Dataset

Our investigation makes use of the CICIDS2017 dataset.The Canadian Institute for Cyber Security generated the dataset, which includes many fundamental attack kinds.This is currently focused on port output initiatives.There are 692703 records that have been transformed into 691406 such as source IP, source port, goal port, stream term, all out fwd parcels, all out in reverse bundles, and so on. Table I contains a portion of the records.

Sharafaldin H. et al demarked and performed two systems, Attack-Network and VictimNetwork, while creating the dataset. They collected data for the dataset from July 3, 2017 to July 7, 2017.

5. ALGORITHMS

A.Support Vector Machine :

The foundation of Support Vector Machine (SVM) computations is built on factual learning and arched improvement in light of the concept of fundamental risk reduction. SVM was developed by Vapnik et al as a solution to a variety of problems [16]. For example, it might be used in a variety of fields, including learning, design recognition, relapse, grouping, and research.

Table 1: A Sample Set of Records from Dataset

Destination Port	Flow Duration	Total Fwd Packets	Total Bwd Packets
49666	3	2	0
49413	4	3	0
3268	4955405	33	24
49441	130	1	2
44459	28460	14	9
13792	29	1	1
13791	54	1	3
35215	35831	15	6

B. Random Forest

Random Forest is a classifier that contains a number of decision trees for the various data sets supplied and takes steps to enhance the data's prediction accuracy.

Step 1: Begin by selecting samples from the specified dataset.

Step 2: This will then generate a decision tree for each sample. The prediction for each decision tree will then be available.

Step 3: During this phase, each anticipated outcome will be voted on.

Step 4: Finally, as the final forecast result, choose the predicted conclusion with the most votes.

C. Artificial Neural Network (ANN)

MLP is an abbreviation for multilayer perceptron, which is a sort of feedforward artificial neural network. Artificial neural networks are a machine learning approach inspired by how the human brain learns and extracts new knowledge. An MLP is made up of three layers. MLP leverages the unsupervised learning technique backpropagation for training.

D. Convolutional Neural Network (CNN)

Convolutional Neural Networks are one of the major types of neural networks used for image categorization and recognition. Convolutional

neural networks are commonly utilised in domains such as scene labelling, object identification, and facial recognition, among others.

6.METHODOLGY

On the basis of the CICIDS2017 dataset, the SVM, ANN, CNN, Random Forest, and deep learning calculations were used to detect port output activities.Figure depicts the flowchart of the invented method. Most notably, 692703 records from that dataset were normalised after they were extracted. After standardisation, tests were divided into two parts: 75% preparing information and 25% testing information. Similarly, the SVM and deep learning IDS models were made reliant on the preparation data. Finally, the models were tested with test data, and the presentation of the models was virtually defined.

Furthermore, the profound realised IDS model is made up of 7 veiled layers, with each layer including a different number of neurons, for example, 100,70,40,150, and 6 separately. We picked optimum quantities based on the model's accuracy based on the amount of neurons and veiled layer model characteristics that were adjusted in this work. But, we didn't have any substantial influence on any component selection computation for SVM and we used all highlights. In the future, we will use various artificial reasoning methods to describe and pick the optimum attributes.

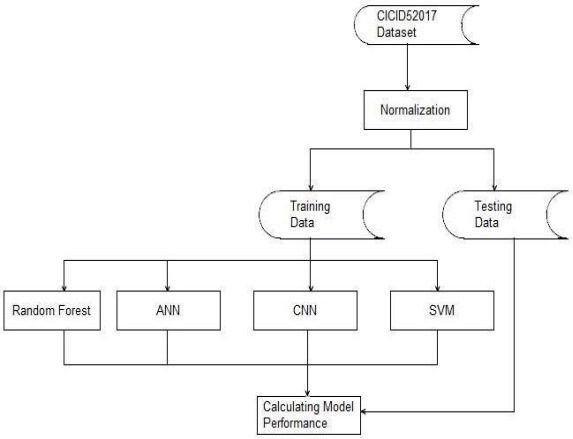

Fig 2:Flowchart of the method

Initially, we have the dataset CICIDS2017, which requires normalisation and data pre-processing. We'll need to separate the data afterwards into training and testing. In order to train the model, we must use techniques such as Random Forest, CNN, ANN, and SVM. Finally, we must assess our data using 25% of the test data.

Important steps of the algorithm are indicated in the "Flowchart of the Method" above.

1) Every dataset should be normalised.

2) Transform that dataset into testing and training datasets.

3) Create IDS models using RF, ANN, CNN, and SVM algorithms.

4) Assess the performance of each model.

7.RESULTS AND EVOLUTION METRICS

For testing, a Computer with an Intel(R) Core(TM) i5 CPU M 460 @2.53 GHz and 4 GB RAM was used. 692703 records from the institutionalised dataset were separated into two groups with 75% planning and 25% testing extents, for example, 518555 models for preparing and 172852 models for testing. Table IV shows execution estimation for the SVM, ANN, CNN, and Random Forest learning models.

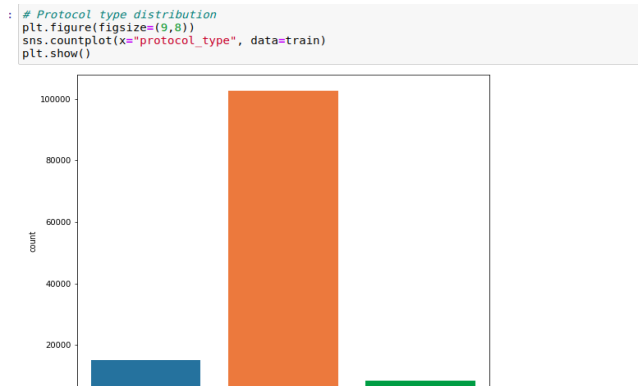

Fig 3:Count of records for training and Testing dataset

8. Conclusion And Future Works

Estimates of assist vector machine, ANN, CNN, Random Forest, and deep learning computations based on the current CICIDS2017 dataset have been introduced very recently. The results reveal that the deep learning calculation outperformed SVM, ANN, RF, and CNN in terms of performance. We will use port sweep initiatives as well as other attack types using AI and deep learning calculations, apache Hadoop and sparkle innovations together later on based on this dataset. All of these calculations assist us in detecting a cyber assault in a network. It occurs in such a manner that when we contemplate far back years, there may have been such many assaults that when these attacks are identified, the characteristics at where these attacks are occurring are recorded in certain datasets. Hence, utilising these statistics, we will be able to forecast whether or not a cyber attack would occur. Four algorithms, including SVM, ANN, RF, and CNN, can make these predictions. This study helps to find which algorithm predicts the greatest accuracy rates, which helps to forecast better outcomes to determine whether or not cyber attacks occurred. Hence, utilising these statistics, we will be able to forecast whether or not a cyber attack would occur. Four algorithms, including SVM, ANN, RF, and CNN, can make these predictions. This study helps to find which algorithm predicts the greatest accuracy rates, which helps to forecast better outcomes to determine whether or not cyber attacks occurred.

9.REFERENCES

[1] K. Graves, Ceh: Official certified ethical hacker review guide: Exam 312-50. John Wiley & Sons, 2007.

[2] R. Christopher, "Port scanning techniques and the defense against them," SANS Institute, 2001.

[3] M. Baykara, R. Das¸, and I. Karado ˘gan, "Bilgi g ¨uvenli ˘gi sistemlerinde kullanilan arac¸larin incelenmesi," in 1st International Symposium on Digital Forensics and Security (ISDFS13), 2013, pp. 231–239.

[4] S. Staniford, J. A. Hoagland, and J. M. McAlerney, "Practical automated detection of stealthy portscans," Journal of Computer Security, vol. 10, no. 1-2, pp. 105–136, 2002.

[5] S. Robertson, E. V. Siegel, M. Miller, and S. J. Stolfo, "Surveillance detection in high bandwidth environments," in DARPA Information Survivability Conference and Exposition, 2003. Proceedings, vol. 1. IEEE, 2003, pp. 130–138.

[6] K. Ibrahimi and M. Ouaddane, "Management of intrusion detection systems based-kdd99: Analysis with lda and pca," in Wireless Networks and Mobile Communications (WINCOM), 2017 International Conference on. IEEE, 2017, pp. 1–6.

[7] N. Moustafa and J. Slay, "The significant features of the unsw-nb15 and the kdd99 data sets for network intrusion detection systems," in Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS), 2015 4th International Workshop on. IEEE, 2015, pp. 25–31.

[8] L. Sun, T. Anthony, H. Z. Xia, J. Chen, X. Huang, and Y. Zhang, "Detection and classification of malicious patterns in network traffic using benford's law," in Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2017. IEEE, 2017, pp. 864–872.

[9] S. M. Almansob and S. S. Lomte, "Addressing challenges for intrusion detection system using naive bayes and pca algorithm," in Convergence in Technology (I2CT), 2017 2nd International Conference for. IEEE, 2017, pp. 565–568.

[10] M. C. Raja and M. M. A. Rabbani, "Combined analysis of support vector machine and principle component analysis for ids," in IEEE International Conference on Communication and Electronics Systems, 2016, pp. 1–5.

[11] S. Aljawarneh, M. Aldwairi, and M. B. Yassein, "Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model," Journal of Computational Science, vol. 25, pp. 152–160, 2018.

[12] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization." in ICISSP, 2018, pp. 108–116.

[13] D. Aksu, S. Ustebay, M. A. Aydin, and T. Atmaca, "Intrusion detection with comparative analysis of supervised learning techniques and fisher score feature selection algorithm," in International Symposium on Computer and Information Sciences. Springer, 2018, pp. 141–149.

[14] N. Marir, H. Wang, G. Feng, B. Li, and M. Jia, "Distributed abnormal behavior detection approach based on deep belief network and ensemble svm using spark," IEEE Access, 2018.

[15] P. A. A. Resende and A. C. Drummond, "Adaptive anomaly-based intrusion detection system using genetic algorithm and profiling," Security and Privacy, vol. 1, no. 4, p. e36, 2018.

[16] C. Cortes and V. Vapnik, "Support-vector networks," Machine learning, vol. 20, no. 3, pp. 273–297, 1995.

[17] R. Shouval, O. Bondi, H. Mishan, A. Shimoni, R. Unger, and A. Nagler, "Application of machine learning algorithms for clinical predictive modeling: a data-mining approach in sct," Bone marrow transplantation, vol. 49, no. 3, p. 332, 2014.