International Journal for Rapid Research in Engineering Technology & Applied Science
Vol IX Issue I January 2023
ISSN Online:- (2455-4723)

Machine Learning based Intrusion Detection System for Web-Based Attacks

Nitika jain , Dr Jitendra Singh Chouhan
Department of computer science and engineering, with specialization in software engineering Aravali Aravali institute of technical studies, udaipur, rajasthan rajasthan technical university, kota, rajasthan
nitjain06@gmail.com , Jitendrasingchauhan1984@gmail.com

Abstract: The number of people who use and rely on computers and computer networks grows at the same rate as the need to keep a computer network safe. Industries have to deal with a lot of new threats every day. Artificial intelligence can be used in a lot of different ways to make an intrusion detection system. This study will look at how an Intrusion Detection System (IDS) neural network works. There is a term for this method called "Self-Organizing Maps" (SOM). A lot of people are surprised by how neural networks can help people figure out how to categorise things. In order to use the neural technique, which claims that each person is unique and leaves a unique computer system footprint, a neural network component will be added to the system. This could mean that there is a security breach in progress if a user leaves behind a trail that isn't what the system administrator or security officer would expect from normal system use. Self-Organizing Maps can be used to make an Intrusion Detection System at the end of this article. We'll also talk about the pros and cons of doing so

Keywords: Intrusion Detection, Security System, Deep learning, Attack

I INTRODUCTION

Security vulnerability may be found in as little as a day or two, depending on how fast you are. A hacker is familiar with current security measures and is always seeking for methods to exploit them. Additionally, cyber criminals are well-versed in a variety of technologies that enable them to circumvent standard security measures and steal personal information. Root Kits, zero-day vulnerabilities, and Browser Exploit Packs (BEP) may all be obtained for free on the black market. Zero-day vulnerabilities are also known as zero-day threats. Personal data and hacked domains may be purchased by attackers for use in future assaults [1]. It's unavoidable to have a security breach. Your greatest line of defense against an assault is early identification and mitigation.

If your personal information is compromised, the ramifications might be severe. Leaking private information might cause major problems for governments, organizations, and people. Hackers may get access to your computer whether or not you are connected to the internet, Bluetooth, text messaging, or other online services. Allowing even minor issues to go unnoticed may result in a big data breach. People, for the most part, don't pay enough attention to current security dangers because they don't understand how they function. Private, sensitive, or confidential information is made accessible to a third party without authorization in the case of a data breach. Without sufficient authorization, data breach files may be read and/or disseminated.

II RELATED WORK

Tommaso Zoppi (2020) et.al Anomaly detection aims at identifying patterns in data that do not conform to the expected behavior, relying on machine-learning algorithms that are suited for binary classification. It has been arising as one of the most promising techniques to suspect intrusions, zero-day attacks and, under certain conditions, failures. This tutorial aims to instruct the attendees to the principles, application and evaluation of anomaly-based techniques for intrusion detection, with a focus on unsupervised algorithms, which are able to classify normal and anomalous behaviors without relying on input data with labeled attacks.

Wei Zhong (2020) et.al With vast amounts of data being generated daily and the ever increasing interconnectivity of the world's internet infrastructures, a machine learning based Intrusion Detection Systems (IDS) has become a vital component to protect our economic and national security. Previous shallow learning and deep learning strategies adopt the single learning model approach for intrusion detection. The single learning

model approach may experience problems to understand increasingly complicated data distribution of intrusion patterns. Particularly, the single deep learning model may not be effective to capture unique patterns from intrusive attacks having a small number of samples. In order to further enhance the performance of machine learning based IDS, we propose the Big Data based Hierarchical Deep Learning System (BDHDLS). BDHDLS utilizes behavioral features and content features to understand both network traffic characteristics and information stored in the payload. Each deep learning model in the BDHDLS concentrates its efforts to learn the unique data distribution in one cluster. This strategy can increase the detection rate of intrusive attacks as compared to the previous single learning model approaches. Based on parallel training strategy and big data techniques, the model construction time of BDHDLS is reduced substantially when multiple machines are deployed.

Monika D. Rokade (2021) et.al Computer network and virtual machine security is very essential in today's era. Various architectures have been proposed for network security or prevent malicious access of internal or external users. Various existing systems have already developed to detect malicious activity on victim machines; sometimes any external user creates some malicious behavior and gets unauthorized access of victim machines to such a behavior system considered as malicious activities or Intruder. Numerous machine learning and soft computing techniques design to detect the activities in real-time network log audit data. KKDDCUP99 and NLSKDD most utilized data set to detect the Intruder on benchmark data set. In this paper, we proposed the identification of intruders using machine learning algorithms. Two different techniques have been proposed like a signature with detection and anomaly-based detection. In the experimental analysis, demonstrates SVM, Naïve Bayes and ANN algorithm with various data sets and demonstrate system performance on the real-time network environment.

III PROPOSED APPROACH

For self-organizing maps, training and mapping are the two ways that they work. This is true for most artificial neural networks, as well. First, training uses an input data set (the "input space") to make the input data more compact (the "map space"). The map is then used to classify more data that comes in. As a rule of thumb, the goal of training is to make an input space with p dimensions look like a two-dimensional map space, most of the time. There must be at least one variable for an input space to have at least one dimension with at least one other variable. Nodes, also called "neurons," are the building blocks of a two-dimensional map space. They are placed in a hexagonal or rectangular grid, and they make up the map. The bigger goals of data analysis and exploration set the number of nodes and how they are laid out in advance.

Node weights are based on where each node is in the input space. The weight of each node in the map space is based on its location in the input space. Learning how to use a map is all about pushing weight vectors toward the input data (lowering a distance metric like Euclidean distance) while still keeping the topology of the map space. The map can be used to find the node that is closest in weight to an input space vector after it has been trained.

Results Discussion: We send the packets through the SOM once they've been collected, vectored, and trained. Figure 1 depicts the end product. Fig. 3 shows the categorization of input vectors, which represents user behaviour, and its mapping to specific neurons, which comprise single potential user behaviour states. Intrusion is indicated by the form, but is it really an intrusion, or is it only a possibility? Typical - Typical. Based on the results of the tests, the SOM network seems to be an appropriate core for IDS systems.

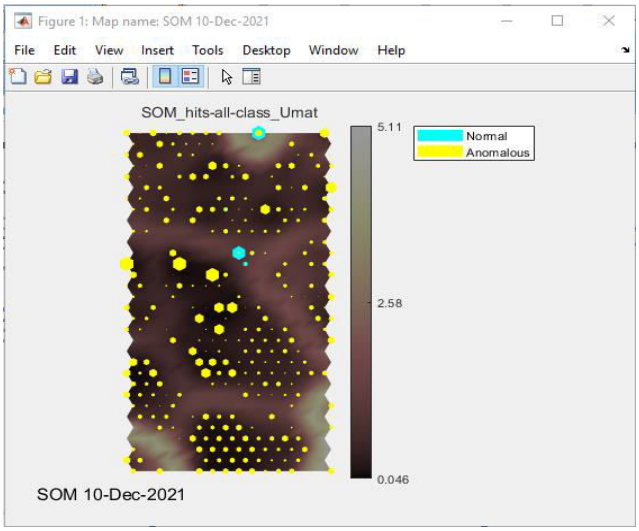


Fig 1 Normal and abnormal activity in the network

As part of our research, we sent an attack flow into the network that had been trained. As a result, the attacks spread to two classes and to both normal and weird activities.

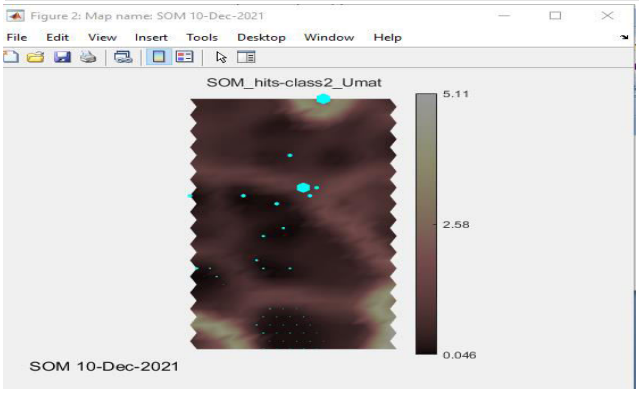


Fig 2 Normal activity in the network

During this experiment, we will use attack flow to train the network. We will not use the regular flow. For this reason, we thought it would be easier to tell attack patterns from normal flow if we trained with attack flow. This suggests that we are not looking for unusual things, but rather looking for signs of abuse or signatures.

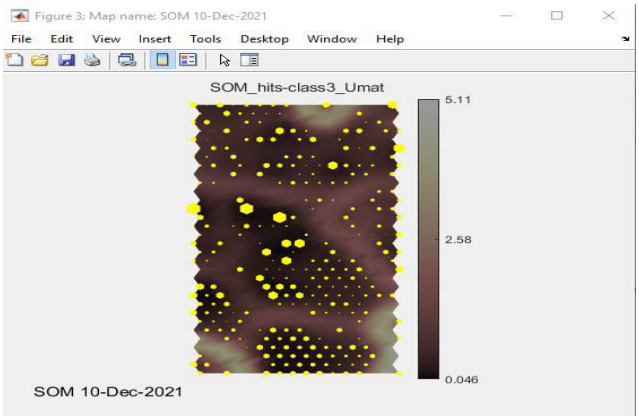


Fig 3 self organization map distribution network with normal activity

Cost-wise, the NIDS is better than any other kind of detector. This makes it more useful than any other kind. NIDS The NIDS collects the packets that go over the network and sends them to a central server, where they are looked at. This is done by having a lot of single-purpose sensors spread out across the network.

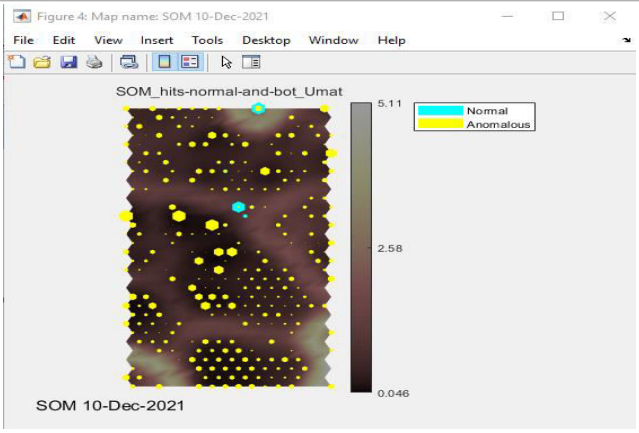


Fig 4 self organization map distribution network with normal activity and abnormal activity

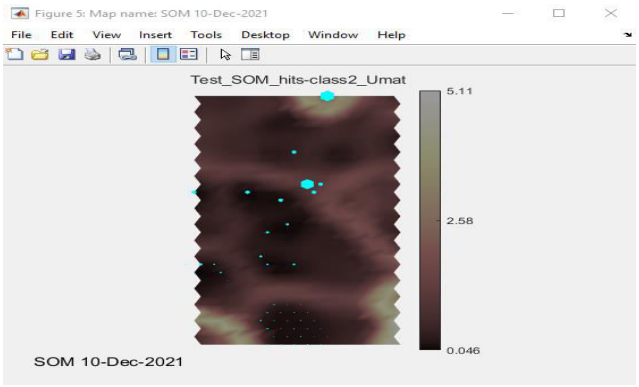


Fig 6 self organizing map with abnormal activity

As you train on the same set of data, the UMAT shows how far apart the neurons are from each other. Training data colour codes show how many neurons are close together and far apart. Brighter colours in the training data show how close together and far apart neurons are.

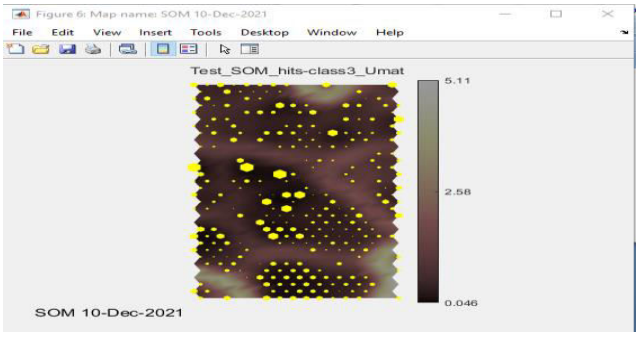


Fig 5 releasing a number of packets

An attacker might look for a weak spot in the system before he or she starts a fight. In order to find vulnerability, many packets are sent to a lot of different hosts until they find one that is.

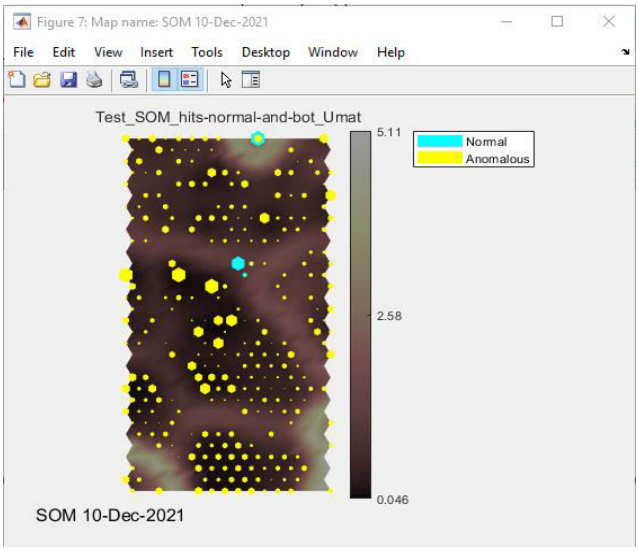


Fig 6 classification of normal and abnormal activity of network

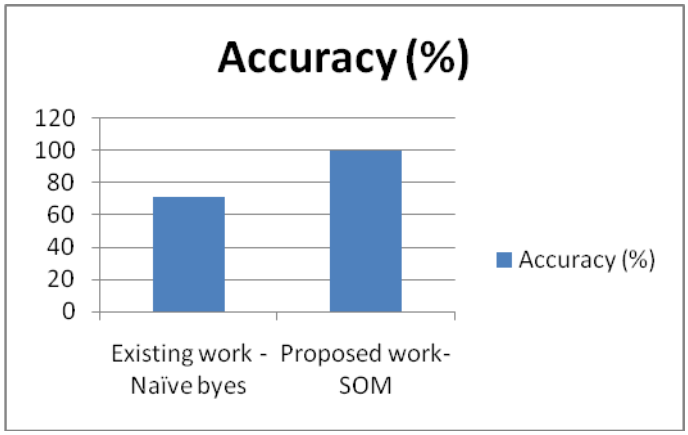
This method was made to look for any problems with the way a machine does its job. It makes a profile for each host or network connection and records everything that happens in the system. If your profile suddenly changes, you'll be seen as suspicious. Regular users who usually log in to their accounts twice a day will make the system think that 20 logins in a single day is unusual. If the user logs on more than twice a day, it will be seen as an attack.

Accuracy is the percentage of time slots that are correctly identified to the total number of time slots. This is called the accuracy rate.

Accuracy =

Table 1: Result comparison

	Techniques	Accuracy (%)
Existing work	**Naïve byes**	71.01
Proposed work	SOM	99.9



IV CONCLUSIONS

The Self-Organizing Map, a strong mechanism, makes it possible to automatically figure out what kind of system activity is allowed. Self Organizing Maps can be used to make an intrusion detection system, as we showed in the study that came before this one, In this lesson, we've talked about the SOMe's architecture and flow diagram as well as its benefits and drawbacks. A basic map that has been trained on normal data can tell the difference between the two types of buffer overflow invasions that we tried it on, according to our real tests. No one needs to tell the self organizing map what invasive behaviour looks like, making this a very strong way to do things, It learns how to describe normal behaviour so that it can recognize when there is something wrong with the network.
Our work has shown a relatively good result in detecting attacks however it is necessary to improve our model further to detect more known and unknown attacks. In addition, a further work that could be an extension of our work to fulfill the need as follows: Using different hyper parameter optimization technique to improve and identify core difference parameters that influence the model performance. Study on additional features and dataset included and selecting relatively high performance models. Hybrid IDS have shown high performance in other studies. So integrating with other signature-based IDS to form a hybrid IDS and measure the performance to what extent is usable the model. Implement with front end applications and using a model for analysis as a back end engine on live network traffic and measure the effectiveness of the whole system.

Acknowledgments

Insert acknowledgment, if any. The preferred spelling of the word “acknowledgment” in American English is without an “e” after the “g.” Use the singular heading even if you have many acknowledgments. Avoid expressions such as “One of us (S.B.A.) would like to thank ....” Instead, write “F. A. Author thanks ... .” Sponsor and financial support acknowledgments are also placed here.

References

1. Giovanni Apruzzese, Luca Ferretti On the Effectiveness of Machine and Deep Learning for Network security 2018 10th International Conference on Cyber Conflict 2018 © Tallinn Modena, Italy
2. Alexander N. Sokolov, Ilya A. Pyatnitsky, Sergei K. Alabugin Research of Classical Machine Learning Methods and Deep Learning Models Effectiveness in Detecting Anomalies of Industrial Control System 2018 Global Smart Industry Conference (GloSIC) 978-1-5386-7386-7/18/$31.00 ©2018 IEEE
3. ZHENG WANG Deep Learning-Based Intrusion Detection With Adversaries , 2018Digital Object Identifier 10.1109/ACCESS.2018.2854599 2169-3536 VOLUME 6, 2018 Gaithersburg, MD 20899, USA
4. Junyang Qiu Wei Luo Lei Pan Yonghang Tai Jun Zhang Yang Xian Predicting the Impact of Android Malicious Samples via Machine Learning IEEE Access (Volume:7)Page(s): 66304 - 6631 2019 : 2169: 10.1109/ACCESS.2019.2914311 Melbourne, VIC 3122, Australia
5. Khoi Khac Nguyen1, Dinh Thai Hoang2, Dusit Niyato2, Ping Wang2, Diep Nguyen3, and Eryk Dutkiewicz3 Intrusion Detection Based on IDBMOctober 2016 978-1-5386-1734-2/18/$31.00 ©2018 IEEE Sdney, Australia
6. Y. Lin, C. Wang, C. Ma, Z. Dou, X. Ma, "A new combination method for multisensor conflict information", J. Supercomput., vol. 72, no. 7, pp. 2874-2890, 2016.
7. F. Kuang, W. Xu, S. Zhang, "A novel hybrid KPCA and SVM with GA model for intrusion detection", Appl. Soft Comput., vol. 18, pp. 178-184, May 2014.
8. H. Huang, J. Yang, H. Huang, Y. Song, G. Gui, "Deep learning for super-resolution channel estimation and DOA estimation based massive MIMO system", IEEE Trans. Veh. Technol., vol. 67, no. 9, pp. 8549-8560, Sep. 2018.
9. Y. Li, X. Cheng, G. Gui, "Co-robust-ADMM-net: Joint ADMM framework and DNN for robust sparse composite regularization", IEEE Access, vol. 6, pp. 47943-47952, 2018.
10. Y. Lin, C. Wang, C. Ma, Z. Dou, X. Ma, "A new combination method for multisensor conflict information", J. Supercomput., vol. 72, no. 7, pp. 2874-2890, 2016.
11. F. Kuang, W. Xu, S. Zhang, "A novel hybrid KPCA and SVM with GA model for intrusion detection", Appl. Soft Comput., vol. 18, pp. 178-184, May 2014.
12. K. V. Narayana, V. V. R. Manoj, K. Swathi, "Enhanced face recognition based on PCA and SVM", Int. J. Comput. Appl., vol. 117, no. 2, pp. 40-42, 2015.
13. R. Samrin, D. Vasumathi, "Hybrid weighted K-means clustering and artificial neural network for an anomaly-based network intrusion detection system", J. Intell. Syst., vol. 27, no. 2, pp. 135-147, 2016.
14. Y. Tu, Y. Lin, J. Wang, J.-U. Kim, "Semi-supervised learning with generative adversarial networks on digital signal modulation classification", Comput. Mater. Continua, vol. 55, no. 2, pp. 243-254, 2018
15. Y. Lin, C. Wang, J. Wang, Z. Dou, "A novel dynamic spectrum access framework based on reinforcement learning for cognitive radio sensor networks", Sensors, vol. 16, no. 10, pp. 1-22, 2016.
16. T. Liu, Y. Guan, Y. Lin, "Research on modulation recognition with ensemble learning", EURASIP J. Wireless Commun. Netw., vol. 1, pp. 179-187, 2017.
17. S. Chung, K. Kim, "A heuristic approach to enhance the performance of intrusion system using machine learning algoriths", Proc. Korea Inst. Inf. Secur. Cryptol. Conf., 2015.
18. X. Pan, Y. Luo, Y. Xu, K-Nearest Neighbor Based Structural Twin Support Vector Machine, Amsterdam, The Netherlands:Elsevier, 2015.
19. M. Tahir et al., "Hybrid machine learning technique for intrusion detection system", Proc. 5th Int. Conf Comout. Inform. (ICOCI), pp. 11-13, 2015.