PCA-SVM Enabled Intelligent Intrusion Detection System for Detection of DDOS and Botnet Attack in Social Web of Things

Mahyudin Ritonga[1][a], Malik Jawarneh[2][b], Karthikeyan Kaliyaperumal[3,*][c] and Nandula Anuradha[4,†][d]

[1]*Universitas Muhammadiyah Sumatera Barat, Indonesia*
[2]*Oman College of Management and Technology, Muscat, Oman*
[3]*IT@IoT - HH Campus, Ambo University, Ambo, Ethiopia*
[4]*Koneru Lakshmaiah Education Foundation Aziznagar, Hyderabad, Telangana State, India*

Keywords: Social Web of Things, Security, Privacy, Vulnerability, Attacks, Threats, BOTNET, Machine Learning, Accuracy.

Abstract: The Social Internet of Things is growing more pervasive in our everyday lives. Computing at the edge that is shared among several users, also known as collaborative edge computing is a relatively new method that has arisen as a potential solution to the rising resource crisis caused by the Internet of Things. Collaborative edge computing allows for previously inaccessible resources like computers, data storage, and network connections to be made available to devices located in remote locations. Because of the edge network's close proximity to the endpoints, sensitive information about the users of the network might be exposed. As a direct consequence of this, dangers to the integrity of edge networks, such as botnet attacks, denial of service attacks, unauthorised access, packet sniffing, and man-in-the-middle assaults, are becoming increasingly prevalent. In order to address these problems and enhance edge network security, we describe an approach for the detection of intrusions. This article includes a mechanism for the detection of DDOS and Botnet attacks in Social Web of Things environments. In the first step of the process, a feature selection is determined using the PCA method. SVM, XGBoost, and AdaBoost are three algorithms that are utilised for the classification of malware data.

1 INTRODUCTION

In the Social Internet of Things (IoT), "physical things" may be changed into "smart things" by using the basic improvements that have been made in computing, communication, Internet protocols, and application development. The sensors, hardware, and connections have all been consolidated, which has made it simpler for us to use. Social IoT has led to an improved quality of life. The number of IoT devices and applications are growing at a rapid pace. It also demands development of robust security solutions (Al-Shabi, 2022).

To have a good grasp on the security of the Social internet of things, one need to be familiar with the threat, vulnerabilities and attacks that are involved. There is a possibility that a future act of vengeance may take place, which might be detrimental to an advantage. The degree to which an individual or item is exposed to risk is measured by what is known as its vulnerability. An activity that takes advantage of vulnerability or is granted permission to do so by such vulnerability is referred to as an attack. Attacks may take the form of making a malicious contribution to an application or flooding a device in an effort to block assistance from being provided (Ammar et al, 2018). Machine learning refers to the process of teaching a computer or other electronic device new skills using the information or data it has been given. There are three distinct categories of algorithms that

[a] https://orcid.org/ 0000-0003-1397-5133
[b] https://orcid.org/ 0000-0001-6894-2756
[c] https://orcid.org/ 0000-0003-4705-1799
[*] Associate Professor
[†] Assistant Professor

Ritonga, M., Jawarneh, M., Kaliyaperumal, K. and Anuradha, N.
PCA-SVM Enabled Intelligent Intrusion Detection System for Detection of DDoS and Botnet Attack in Social Web of Things.
DOI: 10.5220/0012615300003739

In *Proceedings of the 1st International Conference on Artificial Intelligence for Internet of Things: Accelerating Innovation in Industry and Consumer Electronics (AI4IoT 2023)*, pages 314-318
ISBN: 978-989-758-661-3

are used in machine learning: supervised learning, unsupervised learning, and reinforcement learning. The model is developed via supervised algorithms, which train themselves using a mapping of input data to output data, and then use this training to predict new output when they are given new input data. They are constructing the model. Unsupervised algorithms build models from incoming data using the distribution or pattern of a model as their starting point. These algorithms use the data as input. Models are developed through the process of reinforcement learning by identifying patterns in the data and applying what has been learned to the current circumstance.

Botnet infections are widespread over the Internet and may be found on a variety of different computer systems. It is shown in figure 1. Using a botnet, they might launch widespread, remote-controlled flood attacks on their targets. These attacks may be controlled by a botnet. We now have a variety of bots out in the wild at our disposal. As a consequence of this, they may replicate independently like worms, conceal themselves from detection like many viruses, and launch attacks independently like many stand-alone components. Because worms and viruses often leave back doors open, botnets may be able to exploit these backdoors in order to get access to networked components. Bots try to elude detection as much as possible by sneaking into networks and infecting them in a way that is undetected. This is done in an effort to avoid being discovered (Raghuvanshi et al, 2022).

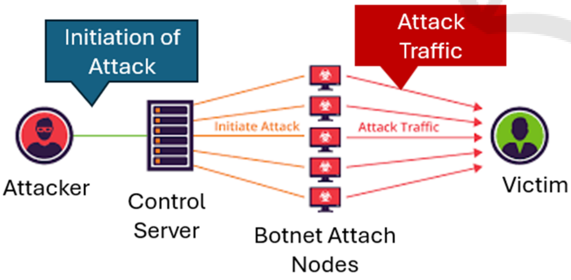

Figure 1: BOTNET DDOS Attack.

This article includes a mechanism for the detection of DDOS Botnet attacks in Social Web of Things environments. In the first step of the process, a feature selection is determined using the PCA method. SVM, XGBoost, and AdaBoost are three algorithms that are utilised for the classification of malware data.

2 LITERATURE SURVEY

Attacks on computer systems that handle personal information without the appropriate authorisation are a danger to the right of consumers to privacy (Jia et al, 2017). This form of attack, in general, takes advantage of vulnerabilities in the authorization process of the model being attacked. According to the findings of some study, hackers might, for instance, steal from their victims or do damage to their property by exploiting holes in the idea of authorization for the administration of smart home applications.

Previous study looked on the ways in which SmartApps and SmartDevices carried out their functions. The protection of user information is the shared responsibility of both smart devices and smart applications. Through the use of events, SmartDevices are able to communicate critical information to apps; SmartApp is able to monitor and record these events as they occur. On the other hand, leaks may take place in the event security is poor, which would result in a substantially greater amount of damage to the consumer. As a consequence of this, the user's privacy can be jeopardised since the security of user input is insufficient. The framework was created to expose normal patterns of data transit so that sensitive data may be protected. This was done with the intention of improving data security. As a direct consequence of this, we were able to solve the issues that arose (Fernandes et al, 2016).

In a phishing attack, the perpetrator may seem to be a real person or a legitimate business organisation in order to get access to sensitive user information such as passwords and credit card data. Email is by far the most popular method of attack, and it may be used to obtain important information even before the recipient reads the message.

When it comes to protecting users' personal information and safety, the authentication function of IoT devices is very necessary. The new authentication methods are not capable of passing a fine-grained check (Unlu et al, 2020). For instance, once installed, malware payloads have the potential to be downloaded and used by attackers in order to spy on a computer remotely. At this moment, the authorization mechanism still has several flaws that make it susceptible to attack. Devices having an excessive number of access privileges may be able to obtain information without using all of the essential components. When it comes to the authorization source, the default choice has its own unique set of challenges. An adversary may potentially take use of this vulnerability to design a broad variety of attacks; the specifics of these attacks would depend on the level of access granted to a file or directory. There have been vulnerabilities discovered in remote authentication in a particular application scenario that makes use of the smart card. These vulnerabilities make it possible for user data to be leaked and manipulated. Because there isn't a completely foolproof security system, it's possible that a burglar may even get through the front door of the smart home.

Web browser instructions, such as login and authorisation directives, are used by remote cloud servers (Jensen et al, 2009). XML tokens, on the other hand, cannot be generated by the browser on its own. Attackers will take advantage of this vulnerability to get unauthorised access to the system. A cloud service that is accessible over the internet and that monitors the deployment of services and stores a large quantity of cloud-related data is able to potentially create this form of metadata. In the event that attackers get access to this information, the cloud may be put in jeopardy.

SQL injection attacks are ones that may be caused by inserting SQL instructions into the data that is being fed into badly built software (Zhang et al, 2009). These SQL instructions are used by attackers so that they may read, write, and destroy data. The ability to get access to personal information is one of the many benefits that come with a cyberattack of this kind; nevertheless, it is also one of the most significant advantages. On the current page, a number of findings made during SQL injection attacks on online applications are presented.

3 METHODOLOGY, RESULTS AND DISCUSSION

This section includes a mechanism for the detection of Botnet attacks in Social Web of Things environments. In the first step of the process, a feature selection is determined using the PCA method. SVM, XGBoost, and AdaBoost are three algorithms that are utilised for the classification of malware data. This framework is shown below in figure 2.

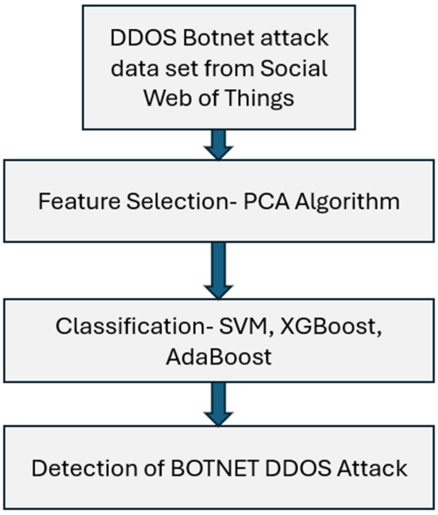
DDOS Botnet attack data set from Social Web of Things
Feature Selection- PCA Algorithm
Classification- SVM, XGBoost, AdaBoost
Detection of BOTNET DDOS Attack

Figure 2: Detection of Botnet Attack in Social Web of Things Using Feature Selection Enabled Machine Learning Technique.

Principal Component Analysis (PCA) is a strategy that makes use of a tool called principal component analysis (PCA), which reduces the high dimensionality of exploratory data and predictive models. By locating variables that are not connected with one another, principal component analysis (PCA) can condense a vast number of observations into a more manageable number. A covariant matrix is produced when these fundamental components are brought together. The Eigen values of the covariant matrix can be deconstructed for PCA once the normalising phase has been completed. Normalization requires two steps: first, separating each data point in a matrix of data from its observed mean (average) value, and then normalising the variance of each variable in turn. From the point of PCA that provides the most relevant information, one can derive a projection of the region of interest for detection. It is possible to improve the identification of pedestrians by using PCA and HOG in conjunction with one another. PCA acts differently in situations in which the initial variables are scaled in different ways (Salaria et al, 2020).

Identifying pedestrians can be accomplished through the use of a method known as the Support Vector Machine (SVM). Images are frequently represented as a non-linear matrix of pixels when computer vision techniques are being used to process them. From the image, we extract and categorise the features of the target object that we are looking for. In order to extract characteristics from photographs, one might utilise the conventional methods of feature extraction. SVM is a binary classifier that may be used to differentiate between pedestrians and other types of road users. A hyperplane is used in SVM to divide the non-linear data points of each image into two binary classes. This is done by placing the 250000 instances are used. 200000 records are used for training of model and remaining 50000 records are used for the testing of model. hyperplane between the non-linear data points of the images. The accuracy of the categorization increases proportionately with the distance between the hyperplane and the data point being analysed. The support vector machine (SVM) is another tool that may be used to choose healthy traits from the pictures (Sahho et al, 2020). The objective of the XGBoost technique, which is known as an Extreme Gradient Boosting method, is to classify regression tree models through the use of gradient lifting decision trees (Sun et al, 2020). In gradient boosting, new models are developed by progressively anticipating the errors of existing models in order to improve accuracy. When all of the individual models that were built earlier in the process are combined, the whole model is produced. XGBoost is a wonderful option for you to consider if you are seeking for an algorithm that functions competently with ordinary tabular data. AdaBoost is a gradient boosting approach that was created specifically for the purpose of binary classification. In the case of decision trees with a limited number of branches, the initial decision tree is constructed, and the performance of the tree is factored into each training sample (Dong et al, 2020).

DDOS Botnet attack data set is used for experimental work. This data set consists of 47 attributes and 1048575 instances. In this study To choose the features, the PCA approach is used. 18 features are selected by PCA. Accuracy, sensitivity, and specificity are the three measures that are used throughout this study to evaluate the performance of various algorithms. The output of the classifiers may be seen in Figures 3, 4, and 5. Because of its accuracy, sensitivity, and specificity, the SVM algorithm is the method of choice when it comes to the detection of malware on Social Internet of Things network.

Accuracy= (TP + TN) / (TP + TN + FP + FN)
Sensitivity = TP/ (TP + FN)
Specificity = TN/ (TN + FP)
Where
TP= True Positive
TN= True Negative
FP= False Positive
FN= False Negative

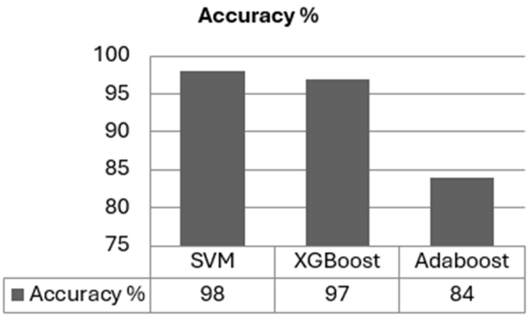

Figure 3: Accuracy of Classifiers for Botnet Attack Detection in Social Internet of Things.

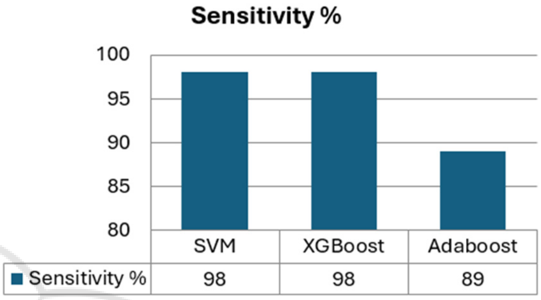

Figure 4: Sensitivity of Classifiers for Botnet Attack Detection in Social Internet of Things.

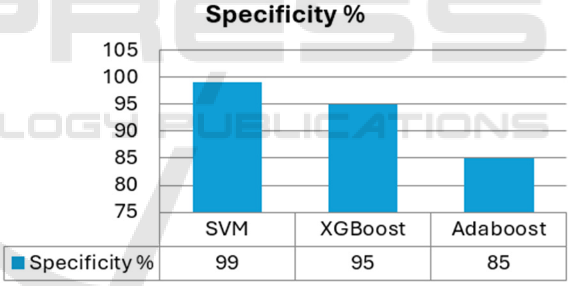

Figure 5: Specificity of Classifiers for Botnet Attack Detection in Social Internet of Things.

4 CONCLUSIONS

In essence, the Social Internet of Things (IoT) has revolutionized daily life, turning ordinary items into smart, interconnected entities. However, this surge in IoT devices necessitates robust security solutions. The article delves into the importance of understanding and addressing security challenges, emphasizing the role of machine learning in detecting threats like botnet attacks.

The literature survey highlights cyber threats such as unauthorized access and phishing attacks targeting IoT systems. Shared responsibility between smart devices and applications is crucial for protecting user information and privacy.

The proposed methodology introduces a practical framework for botnet attack detection, combining Principal Component Analysis (PCA) with SVM, XGBoost, and AdaBoost. The results demonstrate the effectiveness of this approach, showcasing high accuracy, sensitivity, and specificity.

In summary, the article navigates the transformative impact of the Social IoT, stressing the need for security in the face of escalating connectivity. It provides valuable insights into machine learning, cyber threats, and detection mechanisms, offering practical solutions for securing the evolving landscape of the Social IoT.

REFERENCES

Q. Zhang and X. Wang, “SQL injections through back-end of RFID system,” 2009 International Symposium on Computer Network and Multimedia Technology, 2009. doi:10.1109/cnmt.2009.5374533

M. Jensen, J. Schwenk, N. Gruschka, and L. L. Iacono, “On technical security issues in cloud computing,” 2009 IEEE International Conference on Cloud Computing, 2009. doi:10.1109/cloud.2009.60

T. Unlu, L. A. Shepherd, N. Coull, and C. McLean, “A taxonomy of approaches for integrating attack awareness in applications,” 2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security), 2020. doi:10.1109/cybersecurity49315.2020.9138885

E. Fernandes, J. Jung, and A. Prakash, “Security analysis of Emerging Smart Home Applications,” 2016 IEEE Symposium on Security and Privacy (SP), 2016. doi:10.1109/sp.2016.44

Y. J. Jia et al., “Contexiot: Towards providing contextual integrity to appified IOT platforms,” Proceedings 2017 Network and Distributed System Security Symposium, 2017. doi:10.14722/ndss.2017.23051

A. Raghuvanshi et al., “Intrusion detection using machine learning for risk mitigation in IOT-enabled smart irrigation in smart farming,” Journal of Food Quality, vol. 2022, pp. 1–8, 2022. doi:10.1155/2022/3955514

M. Ammar, G. Russello, and B. Crispo, “Internet of things: A survey on the security of IOT Frameworks,” Journal of Information Security and Applications, vol. 38, pp. 8–27, 2018. doi:10.1016/j.jisa.2017.11.002

M. A. Al-Shabi, “Design of a network intrusion detection system using complex deep neuronal networks,” International Journal of Communication Networks and Information Security (IJCNIS), vol. 13, no. 3, 2022. doi:10.17762/ijcnis.v13i3.5148

S. Salaria, S. Arora, N. Goyal, P. Goyal and S. Sharma, "Implementation and Analysis of an Improved PCA technique for DDoS Detection," 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA), 2020, pp. 280-285, doi: 10.1109/ICCCA49541.2020.9250912.

K. S. Sahoo et al., "An Evolutionary SVM Model for DDOS Attack Detection in Software Defined Networks," in IEEE Access, vol. 8, pp. 132502-132513, 2020, doi: 10.1109/ACCESS.2020.3009733.

L. Sun, "Application and Improvement of Xgboost Algorithm Based on Multiple Parameter Optimization Strategy," 2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE), 2020, pp. 1822-1825, doi: 10.1109/ICMCCE51767.2020.00400.

X. Dong, C. Dong, B. Chen, J. Zhong, G. He and Z. Chen, "Application of AdaBoost Algorithm Based on Decision Tree in Forecasting Net power of Circulating Power Plants," 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), 2020, pp. 747-750, doi: 10.1109/ITNEC48623.2020.9085000.

https://www.kaggle.com/datasets/siddharthm1698/ddos-botnet-attack-on-iot-devices?resource=download