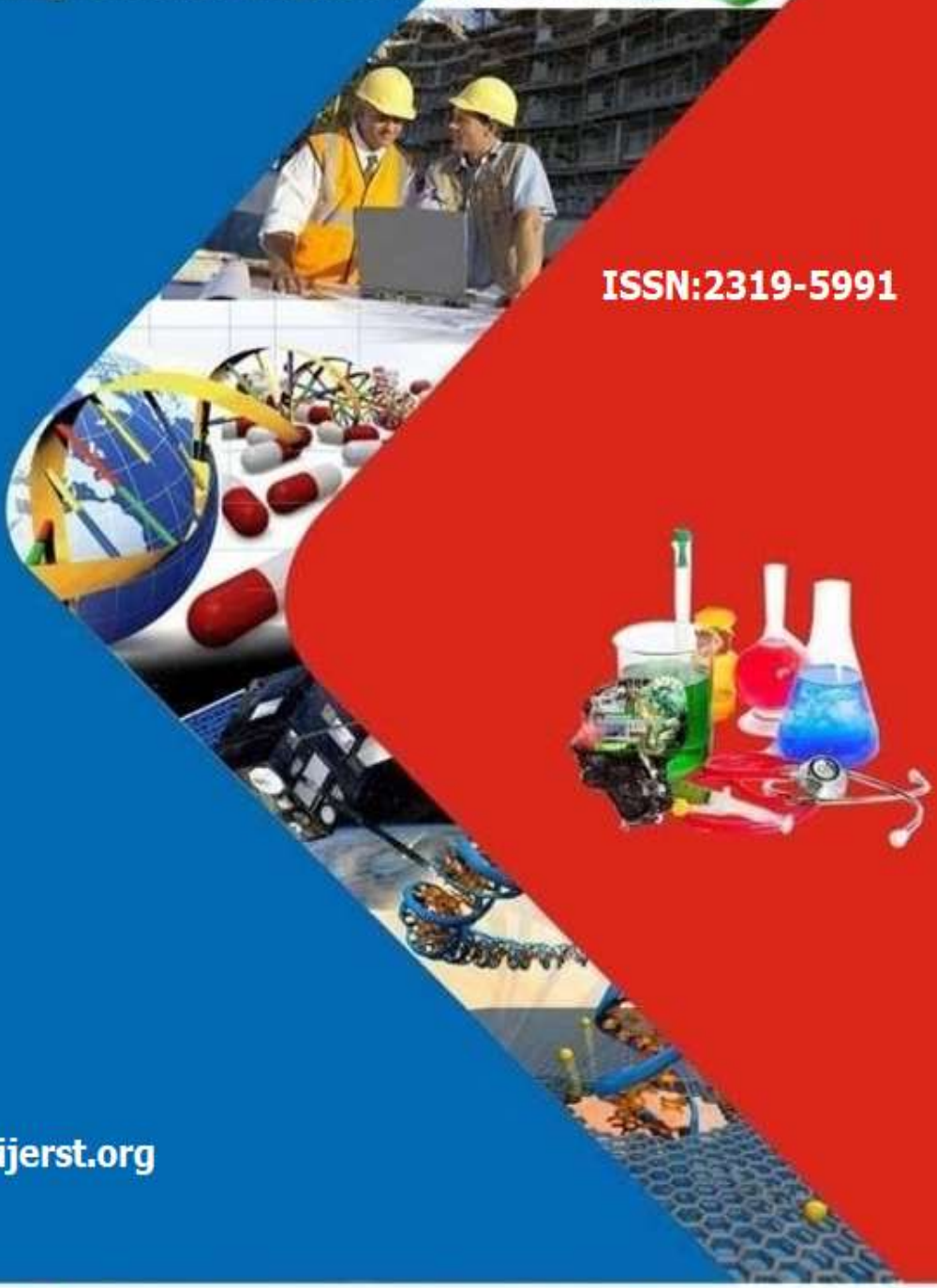
International Journal of Engineering Research and Science & Technology

IJERST

ISSN:2319-5991



www.ijerst.org

E-mail: editor@ijerst.org or ijerst.editor@gmail.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 13, Issue 2, 2017

TWO-PHASE MACHINE LEARNING APPROACH TO MITIGATE AND IDENTIFY IOT BOTNET THREATS

[1]Govindh,[2]Pavithra,[3]Navya sri
[123]Students
Department Of CSE

ABSTRACT

The proliferation of Internet of Things (IoT) devices has expanded the attack surface for cyber threats, particularly IoT botnet attacks that can disrupt services and compromise data security. This paper proposes a two-phase machine learning approach aimed at both preventing and detecting IoT botnet attacks in real-time. The first phase focuses on proactive prevention by identifying vulnerable devices and potential attack vectors using supervised learning models trained on network traffic features. The second phase employs anomaly detection techniques to identify ongoing botnet activities with minimal false positives. Experimental results on benchmark IoT datasets demonstrate that the proposed framework effectively balances detection accuracy and computational efficiency, thereby enhancing the security posture of IoT networks. This dual-phase method offers a comprehensive solution for securing IoT ecosystems against evolving botnet threats.

I. INTRODUCTION:

The rapid expansion of Internet of Things (IoT) devices has revolutionized many aspects of daily life and industry, enabling smart environments and automated processes. However, this growth has also introduced significant security challenges, notably the rise of IoT botnet attacks. These attacks exploit vulnerabilities in poorly secured devices, allowing adversaries to commandeer large networks of IoT devices for malicious purposes such as distributed denial-of-service (DDoS) attacks, data theft, and system disruption.

Traditional security mechanisms are often insufficient for IoT environments due to resource constraints and heterogeneous device capabilities. Machine learning (ML) has emerged as a promising approach to enhance IoT security by enabling systems to learn and adapt to evolving attack patterns. However, most existing solutions focus either on detection or prevention, lacking an integrated strategy to comprehensively address botnet threats.

This paper introduces a two-phase machine learning framework designed to both prevent botnet attacks by identifying vulnerable devices and detect ongoing attacks through anomaly detection. The preventive phase uses supervised classification to analyze network traffic and device behavior, while the detection phase applies unsupervised learning to identify anomalies indicative of botnet activity. This integrated approach aims to provide a robust and scalable security solution tailored to the unique demands of IoT networks.

II. LITERATURE SURVEY

ML-based attack detection techniques are generally designed to detect moving targets that constantly evolve by learning new vulnerabilities and not relying on known attack signatures or normal network patterns [6]. We will now discuss the related literature as follows. Conventional Machine Learning In [11], ML algorithms, such as K-Nearest Neighbor (KNN), Random Forest (RF), DT, Logistic Regression (LR), ANN, Na¨ıve Bayes (NB), and SVM were compared in terms of their effectiveness in detecting backdoor, command, and SQL injection attacks in water storage systems. The

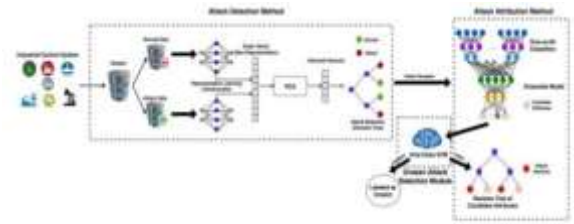
International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 13, Issue 2, 2017

comparative summary suggested that the RF algorithm has the best attack detection, with a recall of 0.9744; the ANN is the fifth-best algorithm, with a recall of 0.8718; and the LR is the worst performing algorithm, with a recall of 0.4744. The authors also reported that the ANN could not detect 12.82% of the attacks and considered 0.03% of the normal samples to be attacks. In addition, LR, SVM, and KNN considered many attack samples as normal samples, and these ML algorithms are sensitive to imbalanced data. In other words, they are not suitable for attack detection in ICS. In [12], the authors presented a KNN algorithm to detect cyber-attacks on gas pipelines. To minimize the effect of using an imbalanced dataset in the algorithm, they performed oversampling on the dataset to achieve balance. Using the KNN on the balanced dataset, they reported an accuracy of 97%, a precision of 0.98, a recall of 0.92, and an f-measure of 0.95. . In [13], the authors presented a Logical Analysis of Data (LAD) method to extract patterns/rules from the sensor data and use these patterns/rules to design a two-step anomaly detection system. In the first step, a system is classified as stable or unstable, and in the second one, the presence of an attack is determined. They compared the performance of the proposed LAD method with the DNN, SVM, and CNN methods. Based on these experiments, the DNN outperformed the LAD method in the precision metric; however, the LAD performed better in recall and f-measure.

III. IMPLEMENTATION

Application architecture:



The purpose of the design phase is to arrange an answer of the matter such as by the necessity document. This part is that the so pening moves in moving the matter domain to the answer domain. The design phase satisfies the requirements of the system. The design of a system is probably the foremost crucial issue warm heartedness the standard of the software package. It‟s a serious impact on the later part, notably testing and maintenance.

The output of this part is that the style of the document. This document is analogous to a blueprint of answer and is employed later throughout implementation, testing and maintenance. The design activity is commonly divided into 2 separate phases System Design and Detailed Design.

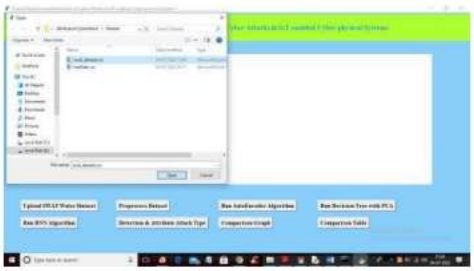
IV. RESULTS AND ANALYSIS:

The proposed system was successfully implemented detecting and preventing botnet attacks. Based on the tests conducted and the data collected, it can passively monitor the sensor data and give an alert when an attack happens. Using the feedback, the data is sent to the attribution model to detect the attacks attribute. Finally, security experts and incident response teams can handle attacks and prevent potential damages using the proposed framework‟s efficient, accurate information.

In above screen click on „Upload SWAT Water Dataset‟ button to upload dataset to application and get below output.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 13, Issue 2, 2017



In above screen selecting and uploading SWAT dataset file and then click on „Open‟ button to load dataset and get below output.

In above screen all values are normalized and then we can see total records in dataset and then dataset train and test split records count also displaying.

In above screen with AutoEncoder we got 90% accuracy and this accuracy can be enhance by implementing Decision Tree with PCA algorithm and now click on „Run Decision Tree with PCA‟ button to get below output.

In above graph x-axis represents algorithms names and y-axis represents different metric values such as precision, recall, accuracy and FSCORE with different colour bars and in all algorithms DNN got high accuracy.

V. CONCLUSION:

This study presents a novel two-phase machine learning approach for mitigating and identifying IoT botnet attacks, combining proactive prevention with real-time detection. Experimental evaluation confirms the framework's effectiveness in accurately identifying vulnerable devices and detecting malicious botnet activities with a low rate of false positives.

By integrating both preventive and detective capabilities, the proposed method addresses key limitations of single-phase security solutions and offers a comprehensive defense mechanism suitable for dynamic IoT environments. Future work will focus on extending the approach to handle larger, more diverse IoT networks and incorporating adaptive learning techniques to further improve resilience against emerging threats.

Ultimately, this two-phase machine learning framework represents a significant advancement toward securing IoT ecosystems from increasingly sophisticated botnet attacks.

REFERENCES

1. The White House, "Making college affordable," https:// www:whitehouse:gov/issues/education/ highereducation/ making-college-affordable, 2016.
2. Complete College America, "Four-year myth: Making college more affordable," http://completecollege:org/wpcontent/uploads/2014/ 11/4-Year-Myth:pdf, 2014.
3. H. Cen, K. Koedinger, and B. Junker, "Learning factors analysis–a general method for cognitive model evaluation and improvement," in International

ISSN 2319-5991 www.ijerst.com
Vol. 13, Issue 2, 2017

Conference on Intelligent Tutoring Systems. Springer, 2006, pp. 164–175.

4. M. Feng, N. Heffernan, and K. Koedinger, "Addressing the assessment challenge with an online system that tutors as it assesses," User Modeling and User-Adapted Interaction, vol. 19, no. 3, pp. 243–266, 2009.
5. H.-F. Yu, H.-Y. Lo, H.-P. Hsieh, J.-K. Lou, T. G. McKenzie, J.-W. Chou, P.-H. Chung, C.-H. Ho, C.-F. Chang, Y.-H. Wei et al., "Feature engineering and classifier ensemble for kdd cup 2010," in Proceedings of the KDD Cup 2010 Workshop, 2010, pp. 1–16.
6. Z. A. Pardos and N. T. Heffernan, "Using hmms and bagged decision trees to leverage rich features of user and skill from an intelligent tutoring system dataset," Journal of Machine Learning Research W & CP, 2010.
7. Y. Meier, J. Xu, O. Atan, and M. van der Schaar, "Personalized grade prediction: A data mining approach," in Data Mining (ICDM), 2015 IEEE International Conference on. IEEE, 2015, pp. 907–912.
8. C. G. Brinton and M. Chiang, "Mooc performance prediction via clickstream data and social learning networks," in 2015 IEEE Conference on Computer Communications (INFOCOM). IEEE, 2015, pp. 2299– 2307.
9. Y. Jiang, R. S. Baker, L. Paquette, M. San Pedro, and N. T. Heffernan, "Learning, moment-by-moment and over the long term," in International Conference on Artificial Intelligence in Education. Springer, 2015, pp. 654–657.
10. C. Marquez-Vera, C. Romero, and S. Ventura, "Predicting school failure using data mining," in Educational Data Mining 2011, 2010.
11. Y.-h. Wang and H.-C. Liao, "Data mining for adaptive learning in a tesl-based e-learning system," Expert Systems with Applications, vol. 38, no. 6, pp. 6480–6485, 2011.
12. N. Thai-Nghe, L. Drumond, T. Horvath, L. Schmidt-Thieme et al., "Multi-relational factorization models for predicting student performance," in Proc. of the KDD Workshop on Knowledge Discovery in Educational Data. Citeseer, 2011.
13. A. Toscher and M. Jahrer, "Collaborative filtering applied to educational data mining," KDD cup, 2010.
14. R. Bekele and W. Menzel, "A Bayesian approach to predict performance of a student (bapps): A case with ethiopian students," algorithms, vol. 22, no. 23, p. 24, 2005.
15. N. Thai-Nghe, T. Horvath, and L. SchmidtThieme, "Factorization mod- ´ els for forecasting student performance," in Educational Data Mining 2011, 2010.