International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

A TWO-FOLD MACHINE LEARNING APPROACH TO DETECT AND PREVENTIoT BOTNET ATTACKS

Author 1: Mr Mohammad Raziuddin, M.Tech, (Ph.D)

(Associates Professor, Department of Computer Science and Engineering, Sphoorthy Engineering College, Hyderabad.

Email: mohammedraziuddin@sphoorthyengg.ac.in

Author 2: Sukka Saiteja, B.Tech

(Student, Department of Computer Science and Engineering, Sphoorthy Engineering College, Hyderabad.

Email: 19n81a05qosaiteja10@gmail.com

Author 3: Korra Venkatesh, B.Tech

(Student, Department of Computer Science and Engineering, Sphoorthy Engineering College, Hyderabad.

Email: 19n81a05h1venkatesh@gmail.com

Author 4: Yelleni Mani Sai, B.Tech

(Student, Department of Computer Science and Engineering, Sphoorthy Engineering College, Hyderabad.

Email: 19n81a05j0yms@gmail.com

ABSTRACT:

The botnet attack is a multi-stage and the most prevalent cyber-attack in the Internet of Things (IoT) environment that initiates with scanning activity and ends at the distributed denial of service (DDoS) attack. The existing studies mostly focus on detecting botnet attacks after the IoT devices get compromised, and start performing the DDoS attack. Similarly, the performance of most of the existing machine learning based botnet detection models is limited to a specific dataset on which they are trained. As a consequence, these solutions do not perform well on other datasets due to the diversity of attack patterns. Therefore, in this work, we first produce a generic scanning and DDoS attack dataset by generating 33 types of scan and 60 types of DDoS attacks. In addition, we partially integrated the scan and DDoS attack samples from three publicly-available datasets for maximum attack coverage to better train the machine learning algorithms. Afterwards, we propose a two-fold machine learning approach to prevent and detect IoT botnet attacks. In the first fold, we trained a state-of-the-art deep learning model, i.e., ResNet-18 to detect the scanning activity in the premature attack stage to prevent IoT botnet attacks. While, in the second fold, we trained another ResNet-18 model for DDoS attack identification to detect IoT botnet attacks. Overall, the proposed two-fold approach manifests 98.89% accuracy, 99.01% precision, 98.74% recall, and 98.87% f1-score to prevent and detect IoT botnet attacks. To demonstrate the effectiveness of the proposed two-fold approach, we trained three other ResNet-18 models over three different datasets for detecting scan and DDoS attacks and compared their performance with the proposed two-fold approach. The experimental results prove that the proposed two-fold approach can efficiently prevent and detect botnet attacks as compared to other trained models.

Keyword: Machine Learning, Botnet Detection, Decision tree, AutoEncoder Algorithm, DNN, Comparison graph, Comparison Table.

I. INTRODUCTION:

The purpose of this document is to define and describe the requirements of the project and to spell out the system"s functionality and its constraints. Internet of Things (IOT) devices are increasingly integrated in cyber-physical systems, including in critical infrastructure sectors such as dams and utility plants. In these settings, IoT devices are Often part of an Industrial Control System (ICS), tasked with the reliable operation of the infrastructure. ICS can be broadly defined to include supervisory control and data acquisition (SCADA) systems, distributed control systems (DCS), and systems that comprise programmable logic controllers (PLC) and Modbus protocols. The connection between ICS or IIOT based systems with public networks, however, increases their attack surfaces and risks being targeted.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

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 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

[1] Application architecture:

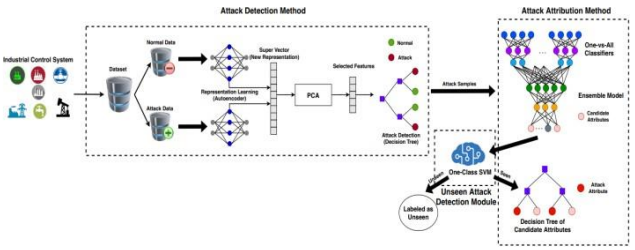


Fig. 1. Proposed attack detection and attribution framework

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 sopening 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.

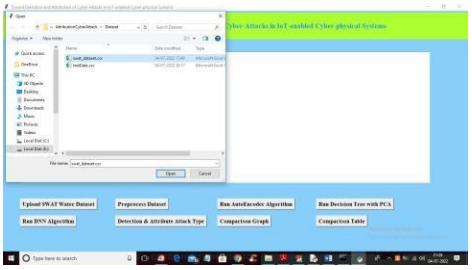
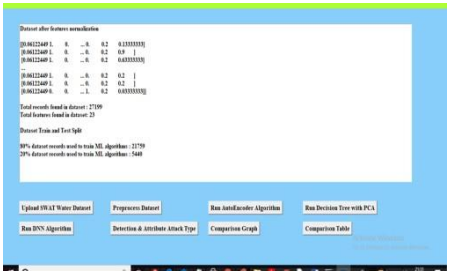
The output of this part is that the style of the document. This document is analogous to a blueprint of answer and is employed later throughout implementation, testing and maintenance. The design activity is commonly divided into 2 separate phases System Design and Detailed Design.

IV. RESULTS AND ANALYSIS:

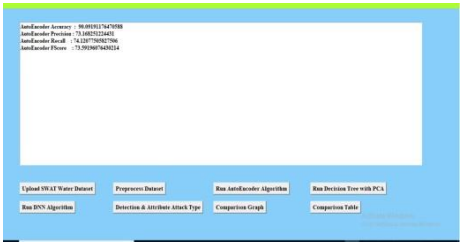
The proposed system was successfully implemented detecting and preventing botnet attacks. Based on the tests conducted and the data collected, it can passively monitor the sensor data and give an alert when an attack happens. Using the feedback, the data is sent to the attribution model to detect the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

.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

UI DESIGN	Design Description (functions, operations etc)
	In above screen click on „Upload SWAT Water Dataset" button to upload dataset to application and get below output
	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.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

	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:

In this paper, we proposed a novel method for predicting students" future performance in degree programs given their current and past performance. A latent factor model-based course clustering method was developed to discover relevant courses for constructing base predictors. An ensemble-based progressive prediction architecture was developed to incorporate students" evolving performance into the prediction. These data-driven methods can be used in conjunction with other pedagogical methods for evaluating students" performance and provide valuable information for academic advisors to recommend subsequent courses to students and carry out pedagogical intervention measures if necessary. Additionally, this work will also impact curriculum design in degree programs and education policy design in general. Future work includes extending the performance prediction to elective courses and using the prediction results to recommend courses to students.

FUTURE SCOPE: It is not possible to develop a system that makes all the requirements of the user. User requirements keep changing as the system is being used. Some of the future enhancements that can be done to this system are:

- As the technology emerges, it is possible to upgrade the system and can be adaptable to desired environment.
- Based on the future security issues, security can be improved using emerging technologies like single sign-on.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

VI. REFERENCES

I. The White House, "Making college affordable," https:// www:whitehouse:gov/issues/education/higher-education/ making-college-affordable, 2016.

II. Complete College America, "Four-year myth: Making college more affordable," http://completecollege:org/wp-content/uploads/2014/ 11/4-Year-Myth:pdf, 2014.

III. H. Cen, K. Koedinger, and B. Junker, "Learning factors analysis–a general method for cognitive model evaluation and improvement," in International Conference on Intelligent Tutoring Systems. Springer, 2006, pp. 164–175.

IV. 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.

V. [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.

VI. 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.

VII. 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.

VIII. 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.

IX. 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.

X.] C. Marquez-Vera, C. Romero, and S. Ventura, "Predicting school failure using data mining," in Educational Data Mining 2011, 2010.

XI. 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.

XII. 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.

XIII. A. Toscher and M. Jahrer, "Collaborative filtering applied to educational data mining," KDD cup, 2010.

XIV. 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.

XV. N. Thai-Nghe, T. Horvath, and L. Schmidt-Thieme, "Factorization mod- ´ els for forecasting student performance," in Educational Data Mining 2011, 2010.

VII. APPENDIX

I. ML- Machine Learning.
II. DNN -Deep Neural Network.
III. DOS-Denial Of Service.