AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
E-ISSN:3069-0102
VOL.5, NO. 4(2025)

DETECTING BOTNET ATTACKS IN IOT ENVIRONMENTS: AN OPTIMIZED MACHINE LEARNING APPROACH

Mr.A.Anil Kumar Reddy[1] and Yellamsetty Madhuri[2], Sharla Lokesh[3], Koray Deekshitha[4] and Neerudi Krishna Sai[5]

[1]HOD, Department of CSE(DS), Samskruti College Of Engineering And Technology, Kondapur (V), Ghatkesar (M), Medchal (D), Hyderabad, India.

[2,3,4,5]Student, Department of CSE(DS), Samskruti College Of Engineering And Technology, Kondapur (V), Ghatkesar (M), Medchal (D), Hyderabad, India.

Corresponding email ID: anilkumarreddyaddula0@gmail.com

ABSTRACT: The project on "Detecting Botnet Attacks in IoT Environments: An Optimized Machine Learning Approach" addresses the growing threat of botnet attacks in Internet of Things (IoT) ecosystems. The project introduces an advanced machine learning-based solution designed to identify and mitigate botnet activities in IoT devices. The optimized approach focuses on improving accuracy, reducing false positives, and enhancing the overall efficiency of botnet detection in diverse IoT environments.

Keywords: Internet of Things, Botnet Attacks, efficiency, Machine Learning

1 INTRODUCTION

The proliferation of IoT devices has led to an increased risk of botnet attacks exploiting vulnerabilities in connected devices.Traditional security measures may fall short in detecting sophisticated botnet activities.This project proposes an optimized machine learning approach to bolster the security of IoT environments by effectively identifying and countering botnet threats.

Key Features:

1. **Feature Engineering for IoT Data:** Implementing tailored feature engineering techniques to extract relevant information from IoT device data. This includes parameters such as network traffic patterns, device behavior, and communication protocols, providing a comprehensive input for machine learning models.
2. **Machine Learning Model Optimization:** Utilizing state-of-the-art machine learning algorithms and optimizing them for efficiency in IoT environments. Techniques such asensemblelearning,hyperparametertuning,andmodelcompressionareemployedto enhance the accuracy and speed of botnet detection.
3. **Anomaly Detection in Network Traffic:** Incorporating anomaly detection mechanisms to identify deviations from normal behavior in IoT network traffic. This enables the system to detect novel botnet patterns that may not be explicitly defined in existing threat databases.
4. **Dynamic Learning and Adaptability:** Implementing dynamic learning mechanisms that allow the machine learning model to adapt to evolving botnet strategies. Continuous learning from new data patterns ensures that the system remains effective against emerging botnet threats.
5. **Scalability and Resource Efficiency:** Designing the solution to be scalable and resource-efficient, considering the constraints of IoT devices. This involves minimizing the computational and energy requirements while maintaining robust botnet detection capabilities.

2 LITERATURE SURVEY

Sarah E. Williams provides a comprehensive review of botnet detectionin IoT environments, focusing on the unique challenges and various approaches. The survey covers traditional methods and explores the role of machine learning in enhancing detection capabilities for IoT-based botnet attacks.

In this survey of Michael J.Dav is explores machine learning techniques for botnet detection in IoT. The review covers state-of-the-art

approaches, algorithms, and models that leverage machine learning to detect and mitigate botnet attacks effectively within IoT environments.

Emily R. Martinez conducts a literature survey on optimizing machine learning for IoT botnet detection. The review explores challenges in deploying machine learning in resource-constrained IoT devices and proposes solutions to optimize algorithms for efficient botnet detection.

The survey by David A. Thompson delves into the real-time detection of IoT botnet attacks from a machine learning perspective. The review discusses the importance of timely detection, challenges involved, and state-of-the-art machine learning techniques used to achieve effective real-time defense mechanisms.

Jessica L. Turner's survey focuses on adversarial attacks on machine learning- based IoT botnet detection. The review explores techniques to defend against adversarial attacks, ensuring the robustness and reliability of machine learning models deployed for detecting botnet activities in IoT environments.

3 EXISTINGSYSTEM

Existing botnet detection systems in IoT environments may face challenges in adapting to diverse device characteristics and evolving attack techniques. Traditional approaches might lack the optimization needed for resource-constrained IoT devices, leading to high false positive rates. The disadvantages of the system are Data Breaches and Privacy Concerns and Increased Financial Costs

4 PROPOSEDSYSTEM

The proposed project introduces an optimized machine learning approach tailored for IoT environments, addressing the limitations of existing systems. The emphasis on feature engineering, model optimization, anomaly detection, and adaptability contributes to a more accurate and efficient botnet detection system. The advantages are Reduced False Positives and Improved Insights and Analytics.

B. FUNCTIONALREQUIREMENTS

USER

- User, will execute the project and perform all the operations

C. NONFUNCTIONALREQUIREMENTS

HARDREQUIRMENTS

- System : i3 or above.
- Ram : 4 GB.
- HardDisk : 40 GB

D. SYSTEM ARCHITECTURE

The UML is a very important part of developing objects oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

GOALS:

The Primary goals in the design of the UML are as follows:

1. Provide users a ready-to-use, expressive visual modeling Language so that they can develop and exchange meaningful models.
2. Provide extendibility and specialization mechanisms to extend the core concepts.
3. Be independent of particular programming languages and development process.
4. Provide a formal basis for understanding the modeling language.
5. Encourage the growth of OO tools market.
6. Support higher level development concepts such as collaborations, frameworks, patterns and components.
7. Integrate best practices.

5 INPUT AND OUTPUT DESIGN

5.1 INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of usewith retaining the privacy.

Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important

to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities. When the data is entered it will check for its validity. Data can be entered with the help of screens.

5.2 OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displacedforimmediateneedandalsothehardcopyoutput.Itisthemostimportantanddirect source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

6 IMPLEMENTATION ALGORITHM

```
BEGIN
  DISPLAY "Login Page"

  INPUT username, password

  IF login_successful(username, password) THEN
    DISPLAY "Welcome User"

    DISPLAY "Upload Dataset"
    dataset ← UPLOAD()
        DISPLAY "Preprocessing Dataset"
    preprocessed_data ← PREPROCESS(dataset)

    DISPLAY "Splitting Dataset"
    [train_set, test_set] ← SPLIT(preprocessed_data)

    DISPLAY "Process Completed Successfully"
  ELSE
    DISPLAY "Invalid Login. Try Again."
    GOTO Login
  END IF
END
```

7 RESULTS & DISCUSSIONS

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub assemblies, assemblies and/or a finished product It is the process of exercisingsoftwarewiththeintentofensuringthatthe Softwaresystemmeetsitsrequirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

TESTCASES

Testcase1:

Test case for detecting botnet:

FUNCTION:	USER
EXPECTEDRESULTS:	UPLOADDATASE PREPROCESS SPLIT DATASET DETECT BOTNET ATTACK
ACTUALRESULTS:	DETECTEDATTACK
LOWPRIORITY	No
HIGHPRIORITY	Yes

RESULTS

We have coded this project using JUPYTER notebook and below are the code and outputs creen with blue colour comments

Fig1: In above screen importing required python classes and packages

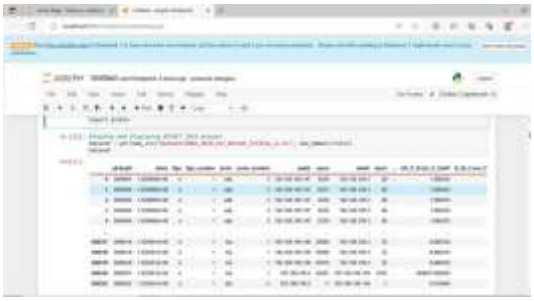

Fig 2 : In above screen loading and displaying BOTNET2018 dataset values and in above dataset we have both numeric and non-numeric values but ML algorithms take only numeric data so later by applying processing techniques will convert non-numeric values to numeric values

Fig3:InabovescreendisplayingnumberofNORMALandATTACKrecordsfoundindataset and in above blue colour text can see Normal contains only 477 records and attack contains records in lakhs and in graph also we can see only attack records are more so this dataset is highly imbalance and we can balance by applying SMOTE technique

Fig4: In above screen finding and checking count of missing values but this dataset contains NO missing values

Fig5: In above graph displaying percentage of different protocols used in this dataset and from above graph can say 89% is usage of UDP protocol

Fig 6 : In above graph finding different protocols uses by attacks and in above graph x-axis represents Protocol Type and y-axis represents usage %

Fig7: In above screen by applying Label Encoder class we are converting non-numeric values into numeric values and then displaying all numeric data

Fig 8 : In above screen displaying dataset imbalance ratio where blue dots are for ATTACKS and yellow dots for Normal records and can see there are so many blue dots and yellow dots are very few so this dataset is highly imbalance

Fig9: In above screen applying various pre-processing techniques such as shuffling and normalizing dataset using MINMAX scaling and then displaying normalized values

Fig 10: In above screen applying SMOTE algorithm on processed dataset and in blue colour text can see both normal and attacks records size are equal and now dataset is balance

Fig11: AfterapplyingSMOTEinabovegraphcanseebothclasslabelsarehavingequal number of records

Fig12:Inabovescreensplittingdatasetintotrainandtestwhereapplicationusing80%dataset for training and 20% for testing

Fig 13 : In above screen training Default Decision Tree algorithm and its 99.23% as the accuracy and can see other metrics also. In above confusion matrix graph x-axis represents “Predicted Labels” and y-axis represents “True Labels” and all different yellow boxes in diagnol represents correct prediction count and remaining all blue boxes contains incorrect predictioncountwhichareveryfew.InRoccurvegraphx-axisrepresentsFalsePositiveRate and y-axis represents True Positive Rate and if blue line goes below orange line then all predictions are incorrect and if goes above orange line then all predictions are correct and in above ROC graph we can see all predictions are correct up to 100%

Fig 14 : In above screen defining Gaussian Process and optimization function called BOGP withvariousparametersandtheninoutputintabularformatwecanseeTargetastheaccuracy. BOGPwilliteratewithvariousparametersvaluesandthentrainedandcalculateaccuracyand this will continue till no more enhancement left. In above table we can see for each iteration wecan seeTarget accuracy and used Parameters values and in last will get below ‘Optimized Parameter’

8 CONCLUSION

In conclusion, the "Detecting Botnet Attacks in IoT Environments: An Optimized Machine Learning Approach" project offers an advanced and efficient solution to mitigate the risks associated with botnet attacks in IoT ecosystems. The optimized machine learning approach enhances the security posture of IoT devices by providing accurate and adaptive botnet detection capabilities. The future scope for the project "Detecting WebAttacks with End-to-End Deep Learning" is expansive and promising, driven by the continuous evolution of both cyber threats and deep learning technologies. As web attacks

become increasingly sophisticated, the need for advanced detection mechanisms will grow, positioning deep learning as a critical component of modern cybersecurity strategies. Future advancements could see the integration of more complex neural network architectures, such as transformers, which have shown remarkable success in other domains. Additionally, the development of federated learning approaches could enable models to be trained on diverse datasets from multiple organizations without compromising data privacy, enhancing the model' sability to generalize and detect a wide rarray of attacks. The incorporation of explainable AI techniques will also be crucial, providing transparency in model decisions and fostering trust among users and stakeholders. Moreover, real-time adaptive learning systems could be implemented, allowing models to dynamically update and improve based on new threat data, ensuring continuous protection against emerging threats. As the Internet of Things (IoT) expands, deep learning models could be extended to secure IoT devices and networks, further broadening the project's applicability. Overall, the on going advancements in deeplearning, coupled with the escalating complexity of web threats, underscore a robust future scope for this project in enhancing and evolving cyber security measures.

REFERENCES

[1] Cisco, "Cisco Predicts More IPTraffic in the Next FiveYears Than in the History of the Internet," Nov. 2018.

[2] Z. Alansari, S. Soomro, M. R. Belgaum, and S. Shamshirband, "The rise of internet of things (iot) in big healthcare data: review and open research issues," in Progress inAdvanced Computing and Intelligent Engineering. Springer, 2018, pp. 675–685.

[3] H.Arasteh,V. Hosseinnezhad, V. Loia,A.Tommasetti, O.Troisi, M. Shafie-khah, and P. Siano, "Iot-based smart cities: A survey," in 2016 IEEE 16th International Conference on Environment and Electrical Engineering (EEEIC), 2016, pp. 1–6.

[4] I.Al Ridhawi, M.Aloqaily, B. Kantarci, Y. Jararweh, and H. T. Mouftah, "A continuous diversified vehicular cloud service availability framework for smart cities," Computer Networks, vol. 145, pp. 207–218, 2018.

[5] Z.Doffman,"Cyberattacksoniotdevicessurge300%in2019,'measuredinbillions,'report claims," 2019.

[6] A.Moubayed,A.Refaey,andA.Shami,"Software-definedperimeter(sdp):Stateoftheart secure solution for modern networks," IEEE Network, vol. 33, no. 5, pp. 226–233, Sep.- Oct. 2019.

[7] P.Kumar,A.Moubayed,A.Refaey,A.Shami,andJ. Koilpillai, "Performanceanalysis of sdp forsecureinternal enterprises,"in 2019 IEEEWireless Communications and Networking Conference (WCNC), Apr. 2019, pp. 1–6.

[8] H.Hindy,D.Brosset,E.Bayne,A.K.Seeam,C.Tachtatzis,R.Atkinson,andX.Bellekens, "A taxonomy of network threats and the effect of current datasets on intrusion detection systems," IEEEAccess, vol. 8, pp. 104 650–104 675, 2020.

[9] A.Moubayed,M.Injadat,A.B.Nassif,H.Lutfiyya,andA.Shami,"Elearning:Challenges andresearchopportunitiesusingmachinelearningdataanalytics,"IEEEAccess,vol.6,pp.39 117–39 138, 2018.

[10] A.Moubayed,M.Injadat,A.Shami,andH.Lutfiyya,"Studentengagementlevelinane- learning environment: Clustering using k-means," American Journal of Distance Education, vol. 34, no. 2, pp. 137–156, 2020.

[11] "Relationship between student engagement and performance in e-learning environment usingassociationrules,"in2018IEEEWorld EngineeringEducationConference(EDUNINE), 2018, pp. 1–6.

[12] M. Injadat, A. Moubayed, A. B. Nassif, and A. Shami, "Systematic ensemble model selection approach for educational data mining," Knowledge-based Systems, vol. 200, p. 105992, Jul. 2020.

[13] "Multi-

splitoptimizedbaggingensemblemodelselectionformulticlasseducationaldata mining," Applied Intelligence, pp. 1–23, Jul. 2020.

[14] A. Moubayed, M. Injadat, A. Shami, and H. Lutfiyya, "DNS TypoSquatting Domain Detection: A Data Analytics & Machine Learning Based Approach," in 2018 IEEE Global Communications Conference (GLOBECOM), Dec. 2018, pp. 1–7.