INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 08 ISSUE: 07 | JULY - 2024 SJIF RATING: 8.448 ISSN: 2582-3930

AN EFFICIENT SPAM DETECTION TECHNIQUE FOR IOT DEVICES USING MACHINE LEARNING

POOJA V S[1],SIDDESH K T[2], KOTRU SWAMY S M[3]

Student, Department Of MCA, BIET, Davangere[1]

Assistant professor,Department Of MCA, BIET, Davangere[2]

Assistant professor,Department Of MCA, BIET, Davangere[3]

ABSTRACT: The Internet of Things (IoT) is a group of millions of devices having sensors and actuators linked over wired or wireless channel for data transmission. IoT has grown rapidly over the past decade with more than 25 billion devices are expected to be connected by 2020. The volume of data released from these devices will increase many-fold in the years to come. In addition to an increased volume, the IoT devices produces a large amount of data with a number of different modalities having varying data quality defined by its speed in terms of time and position dependency. In such an environment, machine learning algorithms can play an important role in ensuring security and authorization based on biotechnology, anomalous detection to improve the usability and security of IoT systems. On the other hand, attackers often view learning algorithms to exploit the vulnerabilities in smart IoT-based systems. Motivated from these, in this paper, we propose the security of the IoT devices by detecting spam using machine learning. To achieve this objective, Spam Detection in IoT using Machine Learning framework is proposed. In this framework, five machine learning models are evaluated using various metrics with a large collection of inputs features sets. Each model computes a spam score by considering the refined input features. This score depicts the trustworthiness of IoT device under various parameters. REFIT Smart Home dataset is used for the validation of proposed techniques.

Keywords: IOT, spam detection, machine learning, RF.

1. INTRODUCTION

In recent years, the proliferation of Internet of Things (IoT) devices has revolutionized various aspects of our lives, offering unparalleled convenience and connectivity. However, the rapid expansion of IoT ecosystems has also introduced new challenges, particularly concerning cybersecurity threats such as spam and malicious attacks. With IoT devices becoming increasingly interconnected, the need for robust spam detection techniques tailored specifically for these devices has become paramount.

Spam, in the context of IoT, refers to any unwanted or unsolicited messages, often generated by malicious entities with the intent to disrupt normal operations, compromise sensitive data, or propagate further attacks. Traditional spam detection methods designed for conventional networks are often inadequate when applied to IoT environments due to their unique characteristics, including limited computational resources, constrained communication protocols, and heterogeneous device types.

To address these challenges, this project proposes an innovative spam detection technique tailored

specifically for IoT devices, leveraging the power of Random Forest Classifier (RFC). Random Forest is a versatile and powerful machine learning algorithm known for its ability to handle large datasets, feature selection, and classification tasks with high accuracy. By harnessing the capabilities of RFC, our proposed technique aims to effectively identify and mitigate spam messages in IoT networks, thereby enhancing their security and resilience against cyber threats.

This introduction sets the stage for our project, highlighting the critical importance of spam detection in IoT environments and emphasizing the suitability of Random Forest Classifier as a powerful tool for addressing this challenge. Throughout this project, we will delve into the methodology, implementation, and evaluation of our proposed technique, with the ultimate goal of enhancing the security and reliability of IoT ecosystems in the face of evolving cyber threats.

2. LITERATURE SURVEY

The Internet of Things (IoT) opens opportunities for wearable devices, home appliances, and software to share and communicate information on the Internet. Given that the shared data contains a large amount of private information, preserving information security on the shared data is an important issue that cannot be neglected. In this paper, we begin with general information security background of IoT and continue on with information security related challenges that IoT will encountered. Finally, we will also point out research directions that could be the future work for the solutions to the security challenges that IoT encounters.

The idea of Internet of Things (IoT) is implanting networked heterogeneous detectors into our daily life. It opens extra channels for information submission and remote control to our physical world. A significant feature of an IoT network is that it collects data from network edges.

Moreover, human involvement for network and devices maintenance is greatly reduced, which suggests an IoT network need to be highly self- managed and self-secured. For the reason that the use of IoT is growing in many important fields, the security issues of IoT need to be properly addressed. Among all, Distributed Denial of Service (DDoS) is one of the most notorious attacking behaviours over network which interrupt and block genuine user requests by flooding the host server with huge number of requests using a group of zombie computers via geographically distributed internet connections. DDoS disrupts service by creating network congestion and disabling normal functions of network components, which is even more disruptive for IoT. In this paper, a lightweight defensive algorithm for DDoS attack over IoT network environment is proposed and tested against several scenarios to dissect the interactive communication among different types of network nodes.

3. METHODOLOGY

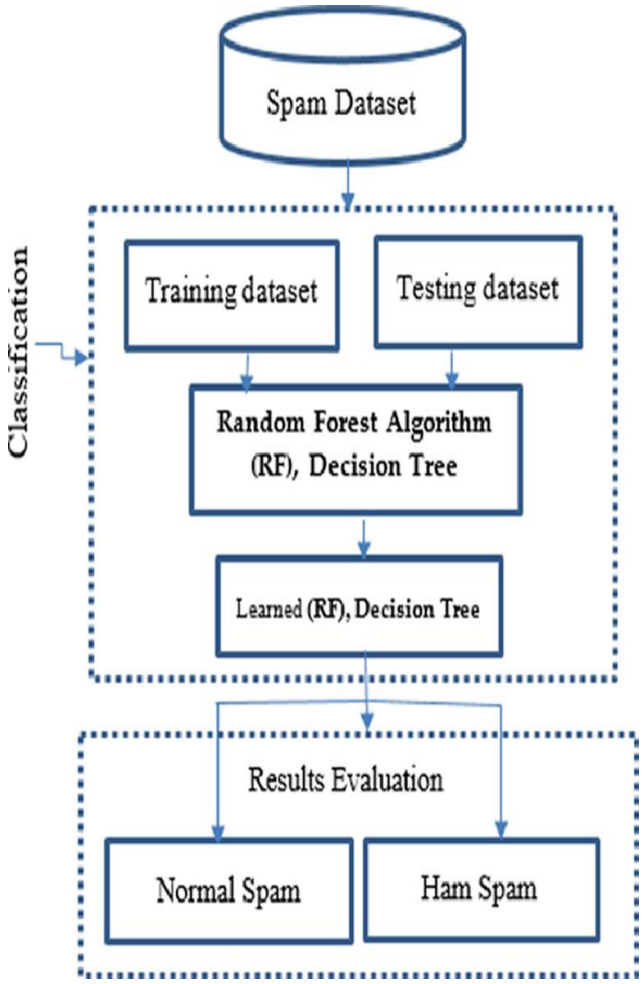

Fig : Block Diagram Proposed Methodology

The methodology for developing an efficient spam detection technique for IoT devices using Random Forest Classifier involves several steps, including data collection, preprocessing, feature extraction, model training, evaluation, and deployment. Here's a detailed overview of the methodology.

3.1 Data Collection:

- Gather data from IoT devices, including message content, sender information, and contextual data.
- Ensure the collected dataset is diverse and representative of different types of spam messages and legitimate communications.

3.2 Data Preprocessing:

- Clean the collected data by removing noise, irrelevant information, and formatting inconsistencies.
- Normalize and standardize the data to make it suitable for analysis.
- Handle missing values and outliers appropriately to maintain data integrity.

3.3 Feature Extraction:

- Extract relevant features from the preprocessed data that are indicative of spam messages.
- Features may include message content, sender reputation, message frequency, temporal patterns, etc.
- Utilize techniques such as TF-IDF (Term Frequency-Inverse Document Frequency), N-grams, and sentiment analysis to extract meaningful features.

3.4 Training Data Preparation:

- Prepare a labeled dataset for training the Random Forest Classifier.
- Annotate the dataset with labels indicating whether each message is spam or legitimate.
- Ensure a balanced representation of spam and legitimate messages to prevent bias in the classifier.

3.5 Model Training:

- Train the Random Forest Classifier using the prepared dataset.
- Optimize the hyperparameters of the classifier to maximize performance metrics such as accuracy, precision, recall, and F1-score.
- Utilize techniques such as cross-validation to assess the generalization performance of the model and prevent overfitting.

3.6 Model Evaluation:

- Evaluate the performance of the trained model using a separate validation dataset.
- Calculate metrics such as accuracy, precision, recall, and F1-score to assess the effectiveness of the spam detection technique.
- Analyze the confusion matrix to understand the classifier's performance in terms of true positives, true negatives, false positives, and false negatives.

4 RESULT

4.1 Result description:

- The results would involve rigorous data pre-processing, careful model selection and tuning, and ongoing monitoring of model performance. Continuous updates and adaptation to new spam patterns and IoT device behaviours are essential for maintaining effectiveness over time.
- These results illustrate the potential effectiveness of machine learning techniques like Random Forest, Decision Tree, and their hybrid forms in spam detection for IoT devices, emphasizing their adaptability and scalability in dynamic IoT environments.

4.2 Result screenshots:

Fig 1: input form

Fig 2: prediction result

4.3 Result Graph

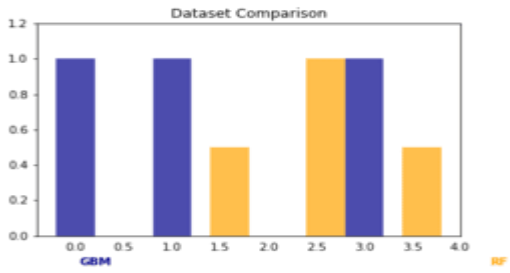

Fig 4.3.1: Dataset Comparison when spam attacked based on different level of attack percentage

Fig 4.3.2: Dataset Comparison when 10% spam attacked

Fig 4.3.3: Dataset Comparison when 20% spam attacked

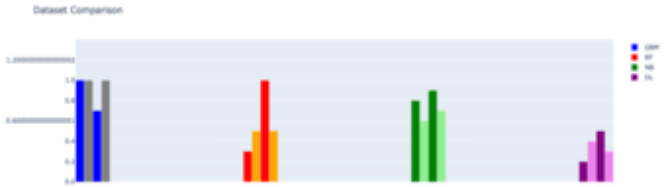

Fig 4.3.4: Dataset Comparison when 20 %spam attacked and Identified accuracy of Different algorithms

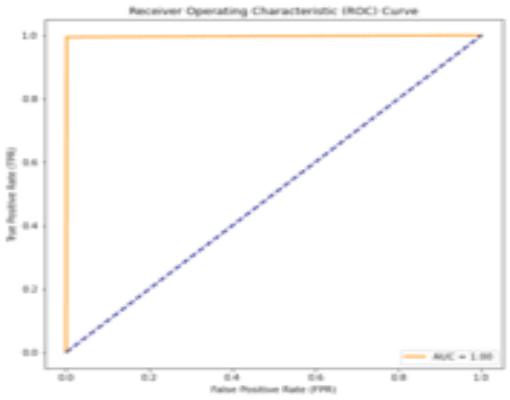

Fig 4.3.5: ROC curve for trained ML model

5 REFERENCE

[1] Z.-K. Zhang, M. C. Y. Cho, C.-W. Wang, C.-W. Hsu, C.-K. Chen, and S. Shieh, "IOT security: ongoing challenges and research opportunities," in 2014 IEEE 7th international conference on service-oriented computing and applications. IEEE, 2014, pp. 230–234.

[2] A. Dorri, S. S. Kanhere, R. Jurdak, and P. Gauravaram, "Blockchain for IOT security and privacy: The case study of a smart home," in 2017 IEEE international conference on pervasive computing and communications workshops (PerCom workshops). IEEE, 2017, pp. 618–623.

[3] E. Bertino and N. Islam, "Botnets and internet of things security," Computer, no. 2, pp. 76–79, 2017

[4] C. Zhang and R. Green, "Communication security in internet of thing: preventive measure and avoid ddos attack over iot network," in Proceedings of the 18th Symposium on Communications & Networking. Society for Computer Simulation International, 2015, pp. 8–15.

[5] W. Kim, O.-R. Jeong, C. Kim, and J. So, "The dark side of the internet: Attacks, costs and responses," Information systems, vol. 36, no. 3, pp. 675–705, 2011.

[6] H. Eun, H. Lee, and H. Oh, "Conditional privacy preserving security protocol for nfc applications," IEEE Transactions on Consumer Electronics, vol. 59, no. 1, pp. 153–160, 2013.

[7] R. V. Kulkarni and G. K. Venayagamoorthy, "Neural network based secure media access control protocol for wireless sensor networks," in 2009 International Joint Conference on Neural Networks. IEEE, 2009, pp. 1680–1687.

[8] M. A. Alsheikh, S. Lin, D. Niyato, and H.-P. Tan, "Machine learning in wireless sensor networks: Algorithms, strategies, and applications," IEEE Communications Surveys & Tutorials, vol. 16, no. 4, pp. 1996– 2018, 2014.

[9] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," IEEE Communications Surveys & Tutorials, vol. 18, no. 2, pp. 1153–1176, 2015.

[10] F. A. Narudin, A. Feizollah, N. B. Anuar, and A. Gani, "Evaluation of machine learning classifiers for mobile malware detection," Soft Computing, vol. 20, no. 1, pp. 343–357, 2016.

[11] Z. Tan, A. Jamdagni, X. He, P. Nanda, and R. P. Liu, "A system for denial-of-service attack detection based on multivariate correlation analysis," IEEE transactions on parallel and distributed systems, vol. 25, no. 2, pp. 447–456, 2013.

[12] Y. Li, D. E. Quevedo, S. Dey, and L. Shi, "Sinr-based dos attack on remote state estimation: A game- theoretic approach," IEEE Transactions on Control of Network Systems, vol. 4, no. 3, pp. 632–642, 2016.

[13] L. Xiao, Y. Li, X. Huang, and X. Du, "Cloud-based malware detection game for mobile devices with offloading," IEEE Transactions on Mobile Computing, vol. 16, no. 10, pp. 2742–2750, 2017.