International Journal of Research and Innovation in Multidisciplinary (www.ijrim.com)
|| Volume 2, Issue 2, February 2025 ||

Spam Detection Techniques for IoT Devices using AI Techniques

Akhilesh Kumar[1], Prof. Nitesh Kumar[2]

M.Tech Scholar, Dept. of ECE., Sagar Institute of Research Technology & Science, Bhopal, India[1]

Assistant Professor, Dept. of ECE., Sagar Institute of Research Technology & Science, Bhopal, India[2]

***Abstract*— With the rapid proliferation of Internet of Things (IoT) devices, the risk of spam and malicious attacks has significantly increased, posing severe threats to network security and data integrity. Traditional spam detection techniques struggle to handle the high volume, real-time processing, and resource constraints associated with IoT environments. Artificial Intelligence (AI) techniques, including Machine Learning (ML) and Deep Learning (DL) algorithms, provide robust and scalable solutions for detecting and mitigating spam in IoT networks. This paper explores various AI-driven spam detection approaches, including Supervised Learning (SVM, Decision Trees, Random Forest), Unsupervised Learning (Clustering, Anomaly Detection), and Deep Learning (CNN, LSTM, Autoencoders), highlighting their effectiveness in identifying spam patterns in IoT-based communication systems. Additionally, challenges such as low-power computing, real-time adaptability, data privacy, and adversarial attacks are discussed. The study concludes by emphasizing the importance of lightweight AI models, federated learning, and real-time anomaly detection techniques in securing IoT devices against spam threats.**

***Keywords*—Spam, Machine, IOT, Detection, Security, AI.**

I. INTRODUCTION

The Internet of Things (IoT) has revolutionized various domains, including smart homes, healthcare, industrial automation, and smart cities, by enabling seamless communication among interconnected devices [1]. However, as the number of IoT devices grows exponentially, security vulnerabilities have also increased, making them prime targets for cyber threats. One of the major security challenges in IoT networks is spam attacks, which involve the transmission of unwanted, malicious, or fraudulent messages that can lead to data breaches, denial-of-service (DoS) attacks, phishing scams, and system compromise. Traditional spam detection methods, such as rule-based filtering, blacklisting, and heuristic techniques, often fail to provide real-time, adaptive, and scalable solutions suitable for IoT networks due to their dynamic nature, heterogeneous architectures, and resource-constrained environments [2].

To effectively mitigate spam threats, Artificial Intelligence (AI) techniques have been increasingly adopted for real-time detection, pattern recognition, and automated mitigation of spam in IoT networks [3]. AI-driven approaches leverage Machine Learning (ML) and Deep Learning (DL) to analyze large volumes of IoT-generated data, identify spam patterns, and make intelligent decisions with minimal human intervention. AI-based spam detection methods offer several advantages, including:

Computational Constraints: Many IoT devices have limited processing power and memory, making it difficult to deploy complex AI models. Real-Time Processing: Spam detection must be instantaneous to prevent network slowdowns or disruptions. Data Privacy and Security: AI models require large datasets for training, raising privacy concerns in IoT applications. Adversarial Attacks: Attackers may manipulate AI models through poisoning attacks, evasion techniques, or adversarial samples to bypass spam detection systems. Scalability and Adaptability: AI models must efficiently scale across diverse IoT networks while adapting to emerging spam threats. To enhance AI-based spam detection in IoT networks, future research should focus on:

Lightweight AI Models: Developing energy-efficient AI architectures that can operate on low-power IoT devices. Federated Learning for Privacy-Preserving Spam Detection: Implementing distributed AI models that allow IoT devices to collaboratively detect spam without sharing raw data. Blockchain-Based AI Security: Using decentralized, tamper-proof blockchain frameworks to enhance spam

detection reliability. Explainable AI (XAI) for Trustworthy Decision-Making: Improving model interpretability to gain trust in AI-driven spam detection systems. Edge AI for Real-Time Spam Filtering: Deploying AI models at the network edge to minimize latency and improve real-time threat detection.

II. LITERATURE SURVEY

A. Makkar et al.,[1] suggest enhancing the safety of IoT devices by the use of ML for spam detection. This goal may be accomplished by implementing the suggested Spam Detection in IoT utilising a Machine Learning framework. This system uses a vast collection of input feature sets to assess five ML models using different criteria. All of the models take the improved input attributes into account when calculating the spam score. This score represents the reliability of an Internet of Things device according to a number of criteria. The suggested method is validated using the REFIT Smart Home dataset. The findings show that the suggested plan is more effective than the other strategies that are already in use.

A model based on deep learning and machine learning is presented by F. Hossain et al., [2] to conduct a comparative study. To include the ensemble technique into machine learning applications, Multinomial Naïve Bayes (MNB), Random Forest (RF), K-Nearest Neighbour (KNN), and Gradient Boosting (GB) are used. To aggregate the output of several classifiers, an ensemble approach is created. When applied to prediction tasks, ensemble approaches outperform standalone classifiers. Based on an email spam base dataset acquired from the UCI machine learning repository, our suggested model achieved a 100% accuracy, AUC=100, MSE error = 0 and RMSE error = 0 for machine learning implementation, and an accuracy of 99%, loss value= 0.0165 for deep learning implementation.

In terms of accuracy and overhead, the WEBSPAM-UK 2007 dataset is presented by A. Makkar et al., [3]. When compared to other methods already published in the literature, the findings show that CSF significantly improves accuracy (by 97.3%).

The ResIoT framework, proposed by G. Fortino et al.,[4], allows agents to cooperate together in an Internet of Things (IoT) setting by forming communities based on agents' reputations. We conducted an experimental campaign in a simulated framework to validate our technique. We were able to confirm that devices do not have any economic convenience to execute false behaviours using our approach. Additionally, additional experimental results have demonstrated that our method can distinguish between honest and malicious active agents in systems, outperforming the best competitor by 11% and demonstrating exceptional resilience against certain malicious actions.

For datasets based on social interactions, K. A. Al-Thelaya et al. [5] propose two methods for representation. User interactions and relations analysis forms the backbone of the representation model development process. Both models are built on top of each other, however one uses graph-based analysis while the other uses sequential processing of user interactions. We find that both representation models exhibit great accuracy in spam identification based on the trials that were done. In contrast to models that analyse interaction sequences, graph-based analysis models provide more accurate results.

Only the characteristics of the source IP, the destination IP, the timestamp of the connection, and the amount of connections are considered in the study by T. Y. Ho et al., [6]. This study uses the deep learning paradigm and suggests a form of VGG16 to analyse traffic characteristics in order to crack the code of complex network traffic. Lastly, this research suggests a way to enhance traffic behaviour explanations using a learning model.

An approach to creating malicious PDF files that seem harmless enough to bypass harmful file detection systems was suggested by J. Zhang et al., [7] and is based on the Wasserstein Generative Adversarial Network (WGAN). Our method's adversarial examples are able to circumvent the PDF classifier-PDFrate of 100%, according to the testing data. Our suggested technique is able to elude the classifiers of several machine learning algorithms, including Support Vector Machine (SVM), Linear Regression, Decision Tree, and Random Forest, according to the results of our tests of their performance in various classifiers.

In order to automatically identify spammy websites, A. Makkar et al.[8] offer a webpage filtering algorithm. Before search engines' ranking modules scan the spam sites, they are recognised. The suggested technique is validated using the decision tree machine learning model. In order to achieve the desired level of accuracy (98.2%), the tenfold cross validation method is used. Based on the findings, it is clear that the suggested technique may successfully block spam web pages in a CIoT setting.

In their experiment with the dataset, A. K. Singh et al. [9] applied each classifier without choosing any characteristics to see what happened. The next step is to use a number of classification algorithms and the best initial feature selection method to choose the characteristics that will be most useful. Using a feature selection technique in our experiments led to a significant improvement in accuracy.

To provide the reader a fast overview of the challenges, T. Lange et al. [10] discuss botnet development, trends, and mitigations, and they provide relevant instances and studies.

III. CHALLENGES

Despite the effectiveness of Artificial Intelligence (AI) techniques in detecting spam in Internet of Things (IoT) networks, several challenges hinder their seamless implementation. Some of the key challenges include:

- Computational Constraints

Most IoT devices operate with limited processing power, memory, and energy resources, making it difficult to deploy complex AI models that require high computational capabilities.

Deep Learning models, such as CNNs and LSTMs, are often too resource-intensive for real-time spam detection on low-power IoT devices.

- Real-Time Detection and Latency Issues

IoT networks require instantaneous spam detection to prevent delays, disruptions, or network congestion.

AI models must process vast amounts of streaming data in real time, which can be challenging given the constraints of edge computing and cloud infrastructure.

- Scalability and Dynamic IoT Environments

IoT networks are highly dynamic, with devices continuously joining and leaving the network. AI-based spam detection models must adapt in real-time to evolving spam threats.

The heterogeneity of IoT devices and communication protocols makes it difficult to design a one-size-fits-all spam detection mechanism.

- Data Privacy and Security Concerns

AI models require large datasets for training and continuous learning, which may include sensitive user information.

Privacy-preserving techniques, such as federated learning and homomorphic encryption, must be integrated to ensure secure AI model training while preventing data leaks.

- Adversarial Attacks and Model Evasion

Attackers can manipulate AI-based spam detection models through adversarial attacks, model poisoning, or data manipulation, making it difficult to maintain high detection accuracy.

Evasion techniques, such as modifying spam content to bypass detection filters, pose significant risks to AI-driven security mechanisms.

- False Positives and False Negatives

AI models may struggle with high false positive rates, incorrectly classifying legitimate messages as spam, leading to communication disruptions.

Conversely, false negatives (failure to detect spam) allow malicious content to bypass security measures, increasing IoT vulnerability to cyberattacks.

- Limited Labeled Data for Training AI Models

Effective spam detection models require large, labeled datasets for supervised learning, which can be difficult to obtain in IoT environments.

The lack of standardized spam datasets for IoT makes it challenging to train and benchmark AI-based models effectively.

- Energy Efficiency and Power Consumption

IoT devices often run on battery power or operate in low-energy environments, making energy-efficient AI models essential for practical deployment.

Lightweight AI techniques, such as TinyML or edge computing, must be explored to balance performance and energy consumption.

IV. CONCLUSION

The integration of Artificial Intelligence (AI) techniques into IoT spam detection systems has significantly improved real-time security, threat mitigation, and network efficiency. AI-driven approaches, including Machine Learning (ML), Deep Learning (DL), and hybrid models, offer enhanced accuracy and adaptability in identifying and preventing spam in IoT environments. However, challenges such as computational constraints, real-time adaptability, data privacy, and adversarial threats remain critical concerns. Future research should focus on developing lightweight AI models, federated learning frameworks, and blockchain-based security mechanisms to enhance spam detection in next-generation IoT networks. By addressing these challenges, AI-powered spam detection can ensure secure, efficient, and scalable IoT communications in an increasingly interconnected world.

REFERENCES

1. A. Makkar, S. Garg, N. Kumar, M. S. Hossain, A. Ghoneim and M. Alrashoud, "An Efficient Spam Detection Technique for IoT Devices Using Machine Learning," in IEEE Transactions on Industrial Informatics, vol. 17, no. 2, pp. 903-912, Feb. 2021, doi: 10.1109/TII.2020.2968927.
2. F. Hossain, M. N. Uddin and R. K. Halder, "Analysis of Optimized Machine Learning and Deep Learning Techniques for Spam Detection," 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), 2021, pp. 1-7, doi: 10.1109/IEMTRONICS52119.2021.9422508.
3. A. Makkar, U. Ghosh, P. K. Sharma and A. Javed, "A Fuzzy-based approach to Enhance Cyber Defence Security for Next-generation IoT," in IEEE Internet of Things Journal, doi: 10.1109/JIOT.2021.3053326.
4. G. Fortino, F. Messina, D. Rosaci and G. M. L. Sarne, "ResIoT: An IoT social framework resilient to malicious activities," in IEEE/CAA Journal of Automatica Sinica, vol. 7, no. 5, pp. 1263-1278, September 2020, doi: 10.1109/JAS.2020.1003330.
5. K. A. Al-Thelaya, T. S. Al-Nethary and E. Y. Ramadan, "Social Networks Spam Detection Using Graph-Based Features Analysis and Sequence of Interactions Between Users," 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), 2020, pp. 206-211, doi: 10.1109/ICIoT48696.2020.9089509.
6. T. Y. Ho, W. Chen, M. Sun and C. Huang, "Visualizing the Malicious of Your Network Traffic by Explained Deep Learning," 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), 2020, pp. 687-692, doi: 10.1109/ICAIIC48513.2020.9065247.
7. J. Zhang, Q. Yan and M. Wang, "Evasion Attacks Based on Wasserstein Generative Adversarial Network," 2019 Computing, Communications and IoT Applications (ComComAp), 2019, pp. 454-459, doi: 10.1109/ComComAp46287.2019.9018647.
8. A. Makkar, N. Kumar and M. Guizani, "The Power of AI in IoT : Cognitive IoT-based Scheme for Web Spam Detection," 2019 IEEE Symposium Series on Computational Intelligence (SSCI), 2019, pp. 3132-3138, doi: 10.1109/SSCI44817.2019.9002885.
9. A. K. Singh, S. Bhushan and S. Vij, "Filtering spam messages and mails using fuzzy C means algorithm," 2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU), 2019, pp. 1-5, doi: 10.1109/IoT-SIU.2019.8777483.
10. T. Lange and H. Kettani, "On Security Threats of Botnets to Cyber Systems," 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN), 2019, pp. 176-183, doi: 10.1109/SPIN.2019.8711780.