International Journal of Advanced Engineering Science and Information Technology (IJAESIT)
ISSN: 2349-3216
Vol.6, No.6, June 2023

Enhancing IoT Security: Web Spam Detection with Machine Learning

Ms. Monisha R[1]
Department of Artificial Intelligence and Data Science
KGiSL Institute of Technology
monisha.r@kgkite.ac.in

Hariharan J[2],Mohammed Ishfaq F[3], Rathna Praba N[4],
Department of Artificial Intelligence and Data Science
KGiSL Institute of Technology
hariharan.j2020@kgkite.ac.in ,
mohammedishfaq.f2020@kgkite.ac.in ,
rathnapraba.n2020@kgkite.ac.in

Abstract— The Internet of Things (IoT) is a network of millions of sensors and actuators that are connected for data transfer across wired or wireless channels. In addition to a rise in volume, the IoT devices generate a large amount of data using a range of diverse modalities with varying data quality. Machine learning techniques can be used in this situation to detect anomalies, improving the usability and security of IoT devices as well as security. Along with the growth in IoT devices, there are more abnormalities now than ever before. Applications for the Internet of Things must address security challenges such interruptions, spoofing, DoS, jamming, eavesdropping, spam, and malware.However, to exploit the security holes in IoT-based smart systems, hackers frequently use learning algorithms. To secure IoT devices as a result of web spam, we advocate using machine learning to detect it. Our project's primary goal is to use machine learning techniques to categorise and find IoT attacks.

***Index Terms*— IoT devices, IoTsecurity,attack , Machine Learning, web spam**

I. INTRODUCTION

The Internet of Things (IoT) is viewed as a distributed, interconnected network of embedded systems that communicate via wired or wireless communication methods. Due to the Internet of Things' (IoT) explosive growth and rapid development, IoT devices are widely present in smart cities and smart households. It is also defined by the network of actual physical objects or things that are endowed with restricted computation, storage, and communication abilities as well as by the embedded electronics (such as sensors and actuators),software, and network connectivity that allow these things to gather,occasionally process, and exchange data. As the Internet of Things (IoT) technology has advanced, network security issues have gotten worse. The network is frequently the target of malicious assaults. The extensively scattered and numerously connected properties of IoT devices make it difficult to maintain their stability and reliability. The most frequent approach to make a network unreachable is through distributed denial-of-service (DDoS) attacks. When several systems saturate a targeted system's bandwidth or resources, which is often one or more web servers, it results in a distributed denial-of-service (DDoS) attact. A DDoS attack frequently makes use of thousands of malware-infected hosts from more than one distinct IP address or workstation. A DDoS attack can slow down or stop all of your online services, whether you run a small non-profit or a huge international conglomerate. This includes emails, websites, and anything else that connects to the internet. Distributed denial-of-service attacks have a significant financial impact, especially at a time when we are using web applications more and more. Therefore, it is critical to be able to identify such threats early and take 2 appropriate action to prevent serious financial losses. This project focuses on machine learning techniques for identifying these types of assaults, which integrate historical data with existing datasets, extracted features, and finally the methods themselves. The solutions discussed in this work are based on machine learning methods such as logistic regression, random forest, and KNN. This will simultaneously help to solve all the issues and improve the effectiveness and accuracy of recognizing and detecting DDos attacks.

II. RELATED WORK

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[1]. A framework is recommended for the detection of malicious network traffic with RF supervised machine learning algorithm achieving far better accuracy of 85.34% [2]. Cognitive spammer framework (CSF) for web spam detection is proposed with a accuracy of 97.3 percent[3]. A framework for agents operating in an IoT environment, called ResIoT, where the formation of communities for collaborative purposes is performed on the basis of agent reputation with an accuracy of not less than 11 percent[4]. Machine learning techniques used for spam filtering techniques used in email and IoT platforms by classifying them into suitable categories.A comprehensive comparison of these techniques is also made based on accuracy, precision, recall, etc[5]. Describes a focused literature survey of Artificial Intelligence (AI) and Machine Learning (ML) methods for intelligent spam email detection, which we believe can help in developing appropriate countermeasures[6]. Deep learning technique of spam detection for IOT devices application[7]. Two representation models for social interaction's graph-based datasets with high spam detection accuracy[8].

Ten fold cross validation approach is used to improve the accuracy of model, i.e., 98.2%. The results obtained demonstrate that the proposed scheme has the power of preventing the spam web pages in Cognitive Internet of Things (CIoT) environment[9]. Out of five different clustering algorithms investigated in this work, OPTICS produced the optimum clustering demonstrating a 0.26% higher average efficacy than its nearest performer DBSCAN. The average balanced accuracy for OPTICS and DBSCAN was found to be ≈75.76%[10].

III. THE PROPOSED MECHANISM

The fact that data is gathered from different domains makes it difficult to get it from different IoT devices. IoT involves a wide range of devices, which results in a big volume of data that is heterogeneous and varied. This data can be classified as IoT data. Real-time, multi-source, rich, and sparse IoT data are just a few of its many features. The use of smart gadgets is essential to the digital age. These devices should not retrieve any spam-filled information. IoT device failure is primarily the result of attacks, where online spam is more prevalent. The method will assign a spamicity score to each IoT device based on the feature importance and the root mean square error score of the machine learning models to assess the device's reliability in the home network. The suggested algorithm is used to determine the network's linked IoT devices' spamicity score. Different evaluation metrics are used to analyse the reliability of IoT devices based on the spamicity score computed in the previous step. We employ the Random Forest, Logistic Regression, and KKN algorithms to visualise the reliability of an IoT device under different conditions.

IV. PERFORMANCE EVALUATION

The model is trained using the dataset, and it then receives an algorithm to utilise when predicting outcomes from the data. Once trained, the model may be used to assess new data and make intelligent predictions about it. Due to this, the DDoS prediction model in this study was used to look for indications of a TCP, UPD, or ICMP flood attack in order to evaluate whether a DDoS attack is imminent or not. The performance evaluation for this project involves assessing the accuracy, precision, recall, and F1 score of the developed machine learning models, comparing them with existing methods, and conducting real-world testing to determine the system's effectiveness in identifying and mitigating web spam related to IoT.

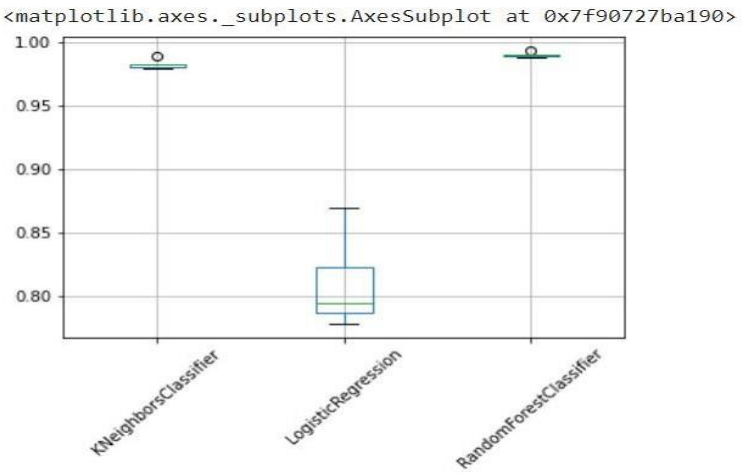

Fig.1 Output

	duration	protocol_type	service	flag	src_bytes	dst_bytes	land	wrong_fragment	urgent	hot	...
0	0	udp	other	SF	146	0	0	0	0	0	...
1	0	tcp	private	S0	0	0	0	0	0	0	...
2	0	tcp	http	SF	232	8153	0	0	0	0	...
3	0	tcp	http	SF	199	420	0	0	0	0	...
4	0	tcp	private	REJ	0	0	0	0	0	0	...

Fig.2 Database Schema

V. CONCLUSION

Further, a three different .txt datasets are used in this study which rely on the training phase to learn from a given dataset and develop a learning profile to find patterns, restricting the applicability of the methods used. Various attributes and features of the data was considered for test and train case.Using machine learning models, the framework that was provided in this paper was able to identify the spam parameters of IoT devices. The IoT dataset utilised for the studies is preprocessed using a feature engineering approach. Each Internet of Things (IoT) is given a spam score by the framework using machine learning experiments. In order to evaluate the dependability of IoT devices, this study uses the spamicity score. With the help of numerous tests and analyses, various ML models were used to evaluate the time-arrangement data generated by keen metres. For the purpose of detecting spam and correctly classifying DDoS attacks, KNN, Random Forest, and Logistic Regression were used. Performance of the models are evaluvated using parameters such as precision, recall, F-score, accuracy and confusion matrix.An accuracy score of 97.22 percent was obtained by the test results. In order to increase the security and dependability of IoT devices, we intend to take into account the time series data and weather features which will help to analyse spam detection and classification in future.

REFERENCES

[1] A. Makkar, S. Garg, N. Kumar, M. S. Hossain, A. Ghoneim and M. Alrashoud, "Efficient Spam Detection Technique for IoT Devices Using Machine Learning," in IEEE Transactions on Industrial,2020

[2] Maryam Anwer, Muhammad Umer Farooq, Shariq Mahmood Khan and Waseemullah," Attack Detection in IoT using Machine Learning", Vol. 11, No. 3, 2021, 7273-7278 -

[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,2021

[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, Volume: 7, Issue: 5, September 2020 -

[5] Naeem Ahmed ,Rashid Amin ,Hamza Aldabbas,Deepika Koundal,Bader Alouffi,and Tariq Shah,"Machine Learning Techniques for Spam Detection in Email and IoT Platforms: Analysis and Research Challenges",Volume 2022

[6] Asif Karim , Sami Azam , Bharanidharan Shanmugam , Krishnan Kannoorpatti , And Mamoun Alazab," A Comprehensive Survey for Intelligent Spam Email Detection", Volume 7, 2019

[7] Naman Nema, Dr. Vikas Gupta." Deep Learning Technique of Spam Detection for IoT Devices Application", Volume 28, Issue 7, 2022

[8] 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

[9] 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

[10] Asif Karim , (Member, IEEE), Sami Azam , (Member, IEEE), Bharanidharan Shanmugam , and Krishnan Kannoorpatti," An Unsupervised Approach for Content-Based Clustering of Emails into Spam and Ham Through Multiangular Feature Formulation", VOLUME 9, 2021

[11] Hongyu Gao, Yan Chen," Towards Online Spam Filtering in Social Networks",2020

[12] G. Kumar and V. Rishiwal, "Statistical Analysis of Tweeter Data Using Language Model With KLD",2018

[13] Motlagh, N.H.; Khajavi, S.H.; Jaribion, A.; Holmstrom, J. An IoT-based automation system for older homes: A use case for lightingsystem,2018

[14] F. Hossain, M. N. Uddin and R. K. Halder, "Analysis of Optimized Machine Learning and Deep Learning Techniques for SpamDetection",2021

[15] Ala Mughaid,Shadi AlZu'bi, Adnan Hnaif, Salah Taamneh,Asma Alnajjar,Esraa Abu Elsoud," An intelligent cyber security phishing detection system using deep learning techniques", 22April2022

Authors Profile

I'm Monisha R, an Assistant Professor at KGiSL Institute of Technology. With over 8 years of experience in the field of academia, I have focused my research efforts for 6 years on topics such as Machine Learning, Data Analytics, Software Engineering, Data Mining, and Design Thinking. My expertise lies in exploring the intersection of these areas to drive innovation and develop practical solutions. I am passionate about leveraging data-driven approaches to solve complex problems and enable informed decision-making. As an Assistant Professor, I am dedicated to imparting knowledge, mentoring students, and fostering a collaborative learning environment. I thrive on staying updated with the latest advancements in my field and continuously strive for personal and professional growth.

I am Hariharan J, a third-year B.Tech student in Artificial Intelligence and Data Science at KGiSL Institute of Technology. In addition to my academic pursuits, I have acquired proficiency as a Full Stack developer.With expertise in front-end technologies such as HTML, CSS, and JavaScript, I am able to create visually appealing and user-friendly interfaces using frameworks like React.js and Angular. On the back-end, I'm skilled in server-side languages like Node.js, Python, and Java, and has experience working with databases such as MySQL and MongoDB. I'm committed to expanding my skills and contributing to the development of innovative software solutions.

I am Mohammed Ishfaq F, currently pursuing my third year of B.Tech in Artificial Intelligence and Data Science at KGiSL Institute of Technology. I am an enthusiastic individual with a keen interest in cybersecurity, possessing knowledge in ethical hacking and networking. I have successfully completed several projects in cybersecurity, integrating artificial intelligence into my work. Notably, one of my projects was showcased in an article, highlighting its innovative approach and contributions.

I am Rathna Praba N, a passionate B.Tech student specializing in Artificial Intelligence and Data Science at KGiSL Institute of Technology. With a Keen interest in Machine Learning (ML) and Artificial Intelligence (AI), I have successfully completed courses, projects, and internships in these fields. I possess a solid foundation in ML algorithms and techniques, and has applied my knowledge to develop innovative solutions. I'm is dedicated to continuously expanding my expertise, staying updated with the latest advancements, and contributing to the growth of AI and ML technologies.