AN EFFICIENT SPAM DETECTION TECHNIQUE FOR IOT DEVICES USING MACHINE LEARNING

[1]Daiwala Sirichandana,[2]MD. Sirajuddin,
[1]Mtech, Department of Computer Science, VAAGESWARI COLLEGE OF ENGINEERING Thimmapur, Karimnagar, Telangana, INDIA, H. No : 20S41D5806, siri.chandana.9237@gmail.com
[2]Department of Computer Science, VAAGESWARI COLLEGE OF ENGINEERING Thimmapur, Karimnagar, Telangana, INDIA, siraj569@hotmail.com

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 technique. The results obtained proves the effectiveness of the proposed scheme in comparison to the other existing schemes.

Keywords *– IOT systems, machine learning and smart home datasets*

1. INTRODUCTION

Internet of Things (IoT) enables convergence and implementations between the real-world objects irrespective of their geographical locations. Implementation of such network management and control make privacy and protection strategies utmost important and challenging in such an environment. IoT applications need to protect data privacy to fix security issues such as intrusions, spoofing attacks, DoS attacks, DoS attacks, jamming, eavesdropping, spam, and malware. The safety measures of IoT devices depends upon the size and type of organization in which it is imposed. The behavior of users forces the security gateways to cooperate. In other words, we can say that the location, nature, application of IoT devices decides the security measures [1]. For instance, the smart IoT security cameras in the smart organization can capture the different parameters for analysis and intelligent decision making [2]. The maximum care to be taken is with web-based devices as maximum number of IoT devices are web dependent. It is common at the workplace that the IoT devices installed in an organization can be used to implement security and privacy features efficiently. For example, wearable devices collect and send user's health data to a connected smartphone should prevent leakage of information to ensure privacy. It has been found in the market that 25-30% of working employees connect their personal IoT devices with the organizational network. The expanding nature of IoT attracts both the audience, i.e., the users and the attackers. However, with the emergence of ML in various attacks scenarios, IoT devices choose a defensive strategy and decide the key parameters in the security protocols for trade-off between security, privacy and computation. This job is challenging as it is usually difficult for an IoT system with limited resources to estimate the current network and timely attack status.

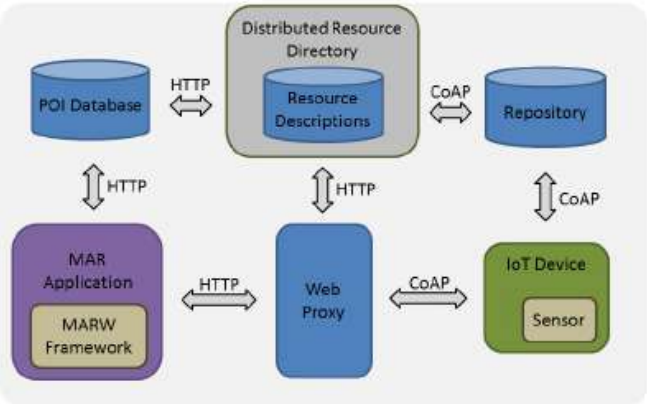

Fig.1: Example figure

1) The proposed scheme of spam detection is validated using five different machine learning models.
2) An algorithm is proposed to compute the spamicity score of each model which is then used for detection and intelligent decision making.
3) Based upon the spamicity score computed in previous step, the reliability of IoT devices is analyzed using different evaluation metrics.

2. LITERATURE REVIEW

Iot security: ongoing challenges and research opportunities

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.

Communication security in internet of thing: preventive measure and avoid ddos attack over iot network:

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 behaviors 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.

The dark side of the internet:Attacks, costs and responses

The Internet and Web technologies have originally been developed assuming an ideal world where all users are honorable. However, the dark side has emerged and bedeviled the world. This includes spam, malware, hacking, phishing, denial of service attacks, click fraud, invasion of privacy, defamation, frauds, violation of digital property rights, etc. The responses to the dark side of the Internet have included technologies, legislation, law enforcement, litigation, public awareness efforts, etc. In this paper, we explore and provide taxonomies of the causes and costs of the attacks, and types of responses to the attacks.

Conditional privacy preserving securityprotocol for nfc applications
In recent years, various mobile terminals equipped with NFC (Near Field Communication) have been released. The combination of NFC with smart devices has led to widening the utilization range of NFC. It is expected to replace credit cards in electronic payment, especially. In this regard, security issues need to be addressed to vitalize NFC electronic payment. The NFC security standards currently being applied require the use of user's public key at a fixed value in the process of key agreement. The relevance of the message occurs in the fixed elements such as the public key of NFC. An attacker can create a profile based on user's public key by collecting the associated messages. Through the created profile, users can be exposed and their privacy can be compromised. In this paper, we propose conditional privacy protection methods based on pseudonyms to solve these problems. In addition, PDU (Protocol Data Unit) for conditional privacy is defined. Users can inform the other party that they will communicate according to the protocol proposed in this paper by sending the conditional privacy preserved PDU through NFC terminals. The proposed method succeeds in minimizing the update cost and computation overhead by taking advantage of the physical characteristics of NFC 1 .

Neural network basedsecure media access control protocol for wireless sensor networks
This paper discusses an application of a neural network in wireless sensor network security. It presents a multilayer perceptron (MLP) based media access control protocol (MAC) to secure a CSMA-based wireless sensor network against the denial-of-service attacks launched by adversaries. The MLP enhances the security of a WSN by constantly monitoring the parameters that exhibit unusual variations in case of an attack. The MLP shuts down the MAC layer and the physical layer of the sensor node when the suspicion factor, the output of the MLP, exceeds a preset threshold level. Backpropagation and particle swarm optimization algorithms are used for training the MLP. The MLP-guarded secure WSN is implemented using the Vanderbilt Prowler simulator. Simulation results show that the MLP helps in extending the lifetime of the WSN.

3. METHODOLOGY

Unsupervised machine learning techniques outperform their counterparts techniques in the absence of labels. It works by forming the clusters. In IoT devices, multivariate correlation analysis is used to detect DoS attacks.

- Reinforcement machine learning technique models Enable an IoT system to select security protocols and key parameters by trial and error against different attacks. Q-learning has been used to improve the performance of authentication and can help in malware detection as well.

Disadvantages:

- This job is challenging as it is usually difficult for an IoT system with limited resources to estimate the current network and timely attack status.
- Prone to attacks

The digital world is completely dependent upon the smart devices. The information retrieved from these devices should be spam free. The information retrieval from various IoT devices is a big challenge because it is collected from various domains. As there are multiple devices involved in IoT, so a large volume of data is generated having heterogeneity and variety. We can call this data as IoT data. IoT data has various features such as real-time, multi-source, rich and sparse.

- The proposed scheme of spam detection in IOT is validated using machine learning model. An algorithm is proposed to compute the spamicity score of the model which is then used for detection and intelligent decision making. Based upon the spamicity score computed in previous step, the reliability of IoT devices is analyzed using different evaluation metrics.
- To protect the IoT devices from producing the malicious information, the web spam detection is targeted in this proposal. We have considered the machine learning algorithm for the detection of spam from the IoT devices.
- The dataset used in the experiments, contains the data recorded for the span of eighteen months. For better results and accuracy, we have considered the data of one month. Considering the fact, the climate is the

important parameter for the working of IoT device, the month with maximum variations has been taken into the consideration.

Advantages:

- ❖ Machine learning techniques help to build protocols for lightweight access control to save energy and extend the IoT systems lifetime.
- ❖ The efficiency IoT data increases, if stored, processed and retrieved in an efficient manner. This proposal aims to reduce the occurrence of spam from these devices.

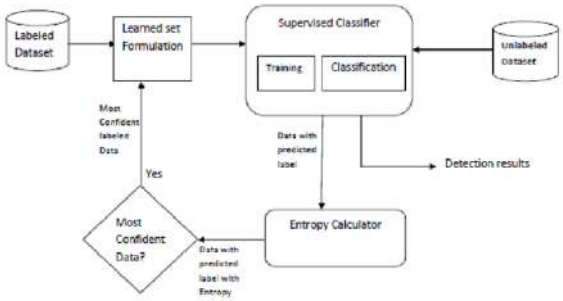

Fig.2: System architecture

4. IMPLEMENTATION

MODULES:

- ❖ Data Collection
- ❖ Dataset
- ❖ Data Preparation
- ❖ Model Selection
- ❖ Analyze and Prediction
- ❖ Accuracy on test set
- ❖ Saving the Trained Model

MODULES DESCSRIPTION:

Data Collection:

This is the first real step towards the real development of a machine learning model, collecting data. This is a critical step that will cascade in how good the model will be, the more and better data that we get, the better our model will perform.

There are several techniques to collect the data, like web scraping, manual interventions and etc.

REFIT Smart Home datasetLink: https://www.refitsmarthomes.org/datasets/

Dataset:

In this data set we are taken32 columns and 503910 rows in the dataset, which are described below.

Data Preparation:

we will transform the data. By getting rid of missing data and removing some columns. First we will create a list of column names that we want to keep or retain.

Next we drop or remove all columns except for the columns that we want to retain.

Finally we drop or remove the rows that have missing values from the data set.

Model Selection:

While creating a machine learning model, we need two dataset, one for training and other for testing. But now we have only one. Solets split this in two with a ratio of 80:20. We will also divide the dataframe into feature column and label column.

Here we imported train_test_split function of sklearn. Then use it to split the dataset. Also, *test_size = 0.2*, it makes the split with 80% as train dataset and 20% as test dataset.
The *random_state* parameter seeds random number generator that helps to split the dataset.
The function returns four datasets. Labelled them as *train_x, train_y, test_x, test_y*. If we see shape of this datasets we can see the split of dataset.
We will useRandomForestClassifier, which fits multiple decision tree to the data. Finally I train the model by passing *train_x, train_y* to the *fit* method.

Once the model is trained, we need to Test the model. For that we will pass *test_x* to the predict method.
Random Forest is one of the most powerful methods that is used in machine learning for classification problems. The random forest comes in the category of the supervised classification algorithm. This algorithm is carried out in two different stages the first one deals with the creation of the forest of the given dataset, and the other one deals with theprediction from the classification.

Analyze and Prediction:
In the actual dataset, we chose only 22features:

Accuracy on test set:
We got a accuracy of 99.1% on test set.

Saving the Trained Model:
Once you're confident enough to take your trained and tested model into the production-ready environment, the first step is to save it into a .h5 or .pkl file using a library like pickle .
Make sure you have pickle installed in your environment.
Next, let's import the module and dump the model into.pkl file

5. EXPERIMENTAL RESULTS

Fig.3: Home screen

Fig.4: login

ISSN: 0972-2750
Vol-12 Issue-11 No.01: 2022

Fig.5: Upload file

Fig.6: Preview

Fig.7: Prediction result

Fig.8: Prediction result

Fig.9: Prediction result

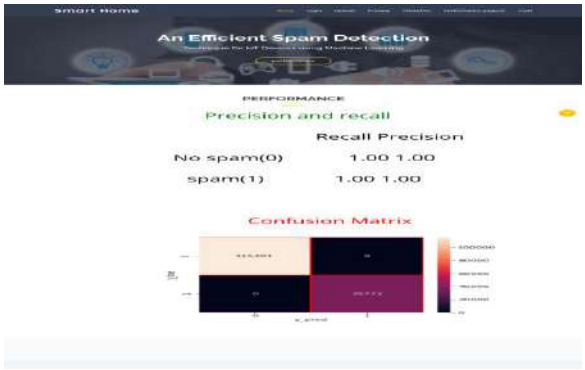

Fig.10: Confusion matrix

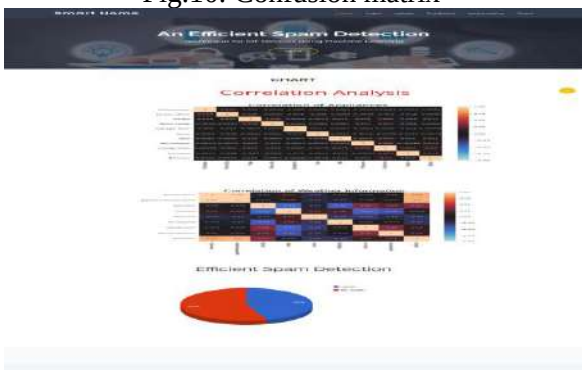

Fig.11: Correlation analysis

6. CONCLUSION

The proposed framework, detects the spam parameters ofIoT devices using machine learning models. The IoT datasetused for experiments, is pre-processed by using feature engineeringprocedure. By experimenting the framework withmachine learning models, each IoT appliance is awarded witha spam score. This refines the conditions to be taken for successfulworking of IoT devices in a smart home. In future, weare planning to consider the climatic and surrounding featuresof IoT device to make them more secure and trustworthy.

REFERENCES

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

[2] A. Dorri, S. S. Kanhere, R. Jurdak, and P. Gauravaram, “Blockchain foriot security and privacy: The case study of a smart home,” in 2017 IEEEinternational conference on pervasive computing and communicationsworkshops (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 Proceedingsof the 18th Symposium on Communications & Networking. Societyfor 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 securityprotocol 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 basedsecure media access control protocol for wireless sensor networks,” in2009 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.

