

RECOGNITION OF SPAM BALANCING FOR IOT NETWORK WITH MACHINE LEARNING TECHNIQUE

Dr Raghavender K V, Assoc Prof CSE department, G NARAYANAMMA INSTITUTE OF TECHNOLOGY AND SCIENCE, SHAIKPET, HYDERABAD Email: drkvraghavender@gmail.com

Abstract – The Internet of Things (IoT) is a group of millions of devices having sensors and actuators linked over the wired or wireless channel for data transmission. IoT is constantly growing day by day over the past few years. The data will increase tremendously in the upcoming years. In addition to an increased volume, the IoT devices produce 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. ML calculation pays a significant job insecurity and approval dependent on biotechnology unnatural recognition to improve the convenience and security of IOT frameworks. Aggressors frequently see learning calculations to abuse the weaknesses in shrewd IOT-based frameworks. The security of the IOT gadgets by distinguishing spam utilizing ML. Spam Detection in IOT utilizing Machine Learning structure is proposed.

Index terms – Internet of Things, Spamsity Score, Principle component analysis.

I. INTRODUCTION

Internet of Things (IoT) enables convergence and implementations between real-world objects irrespective of their geographical locations. Implementation of such network management and control makes privacy and protection strategies of utmost importance 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 measure of IOT devices depends upon the size and type of the 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, and application of IoT devices decide the security measures. For instance, the smart IOT security cameras in a smart organization can capture different parameters for analysis and intelligent decision-making. The maximum care to be taken is with web-based devices as maximum number of IOT devices are web dependent. It is common at workplaces that IOT devices installed in an organization can be used to implement security and privacy features efficiently. For example, collecting and sending user's health data to a connected smart phone that should prevent leakage of information 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 machine learning (ML) in various attacks scenarios, IOT devices choose a defensive strategy and decide the key parameters in the security protocols for tradeoff 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.

II. LITERATURE SURVEY

IOT systems are vulnerable to network, physical, and application attacks as well as privacy leakage, comprising objects, services, and networks. These attacks are presented. Let us have a look at some of the attack scenarios launched by the attackers. Denial of service (DDOS) attacks: The attackers can flood the target database with unwanted requests to stop IOT devices from having access to various services. These malicious requests produced by a network of IOT devices are commonly known as bots,

1) DDOS: Can exhaust all the resources provided by the service provider. It can block authentic users and can make the network resource unavailable.

2) RFID attacks: This is inflicted in physical layering in IOT devices. This leads to integrity of device. Attackers attempt to modify the data either at the node storage or while it is in the transmission within network. The common attacks possible at the sensor node are attacks on availability, attacks on authenticity, attacks on

confidentiality, and cryptography keys brute-forcing. The countermeasures to ensure prevention of such attacks includes password protection, data encryption, and restricted access control.

3) Internet attacks: A digital assault is an attack dispatched by cybercriminals utilizing at least one PCs against a solitary or numerous PCs or organizations. The spammers who need to take different frameworks data or need their objective site to be visited ceaselessly use spamming strategies. The basic strategy utilized for the equivalent is Ad extortion. It produces the artificial clicks at a focused on site for money related profit. Such rehearsing group is known as digital hoodlums.

4) Near field communication (NFC): The possible attacks are unencrypted traffic, eavesdropping, and tag modification. The solution for this problem is the conditional privacy protection. Thus, the attacker fails to create the same profile with the help of user's public key. This model is based on random public keys by trusted service manager. Various ML techniques such as supervised learning, unsupervised learning, and reinforcement learning have been widely used to improve network security. The existing ML technique, which helps in detection of above-mentioned attacks, is discussed. Each ML technique according to its type and role in detection of attacks is described as below.

1) Supervised ML techniques: The models such as support vector machines (SVMs), random forest, naive Bayes, K-nearest neighbor (K-NN), and neural networks are used for labeling the network for detection of attacks. In IOT devices, these models successfully detected the DOS, DDOS, intrusion, and malware attacks.

2) Unsupervised ML techniques: These techniques out perform 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.

3) Reinforcement ML techniques: These 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. ML techniques help to build protocols for lightweight access control to save energy and extend the IOT systems lifetime. The outer detection scheme as developed, for example, applies KNN store address the issue of unregulated outer detection in WSNs. The literature survey demonstrates the applications of ML in enhancing the network security. Therefore, in this article, the given problem of web spam is detected with an implementation of various ML techniques.

III. PROPOSED SYSTEM

A. System Model

Framework displaying is the interdisciplinary investigation of the utilization of models to conceptualize and build frameworks in business and IT advancement. The data recovered from these gadgets ought to be sans spam. The data recovery from different IOT gadgets is a major test since it is gathered from different areas. As there are numerous gadgets engaged with IOT, huge volumes of information are produced having heterogeneity and assortment. We can call these information as IOT information. IOT information have different highlights like constant, multisource, rich, and meager.

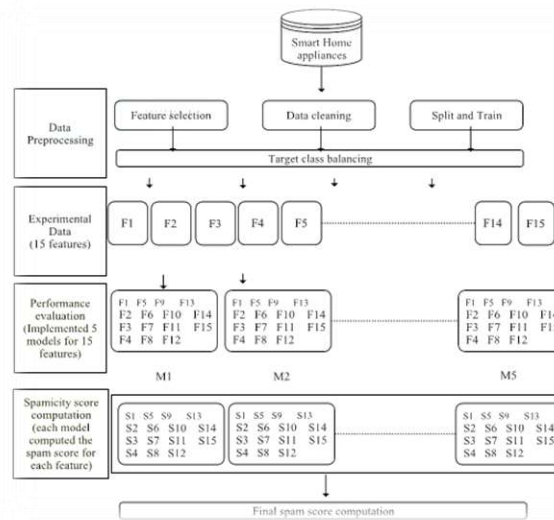


Fig. 1: System Overview

B. Proposed Methodology

To protect the IOT devices from producing the malicious information, the web spam detection is targeted in this proposal. We have considered various ML algorithms for the detection of spam from the IOT devices. The target is to resolve the issues in the IOT devices deployed within home. However, the proposed methodology considers all the parameters of data engineering before validating it with ML models. The procedure used to accomplish the target is presented and discussed in various steps as follows.

1. **Feature Engineering:** The ML algorithms works accurately with the appropriate instances and their attributes. We all know that the instances are the real data world value, gathered from the real world smart objects deployed across the globe. Feature extraction and feature selection is the core of feature engineering process.

a. **Feature reduction:** This method is used to reduce the dimension of data. In other words, feature. This Technique reduces the issues such as overfitting, large memory requirement, and computation power. There are various feature extraction techniques. Among these, principal component analysis (PCA) is the most popular. However, the method used in this proposal is PCA along with following IOT parameters.

2. **Feature Selection:** It is the process of computing the most important subset of features. It works by computing the importance of each feature. Entropy-based filter is used as a feature selection technique in this proposal.

a. **Entropy-based filter:** This algorithm uses the correlation among the discrete attributes with continuous attributes to find out the weights of discrete attributes. There are three functions using this entropy-based filter, namely, information gain, gain ratio, and symmetrical uncertainty. The syntax for these functions is as follows. Information Gain (formula, data, unit), gain Ratio (formula, data, unit) symmetrical. Uncertainty (formula, data, unit). The arguments used in the function definition are described here.

Formula: It is the description of the working behind the algorithm.

Data: It is the set of training data with the defined attributes for which the selection is to be made.

C. ML Models

The proposed technique is validated by finding the spam parameters using ML technique. The ML models used for experiments are summarized

- 1) First, prior information is incorporated. In general, prior information is quantitatively specified in the form of a distribution and represents a distribution of probability for a co-efficient.
- 2) Second, the prior is paired with a function of likelihood. The function of probability represents the results.
- 3) Third, the combination of the prior and the probability function results in a subsequent distribution of co-efficient values being formed.
- 4) Fourthly, simulations are taken from the posterior distribution to construct an empirical distribution for the population parameter of probable values.
- 5) Fifth, to sum up the statistical distribution of simulates from the posterior, simple statistics are used.

Boosted Linear Model: For the data elements, multiple decision trees are created, with the decision tree models by dividing the data series into a plurality of data classes. Therefore, as a linear function, each of the data groups is modeled. From the modeling modules, the boosted models are formed.

D. Spam City Score

After the evaluation of ML models, we computed the spam city score of each appliance. This score indicates the trust worthiness and reliability of the device. It is defined using (2) as follows:

$$e[i] = \sqrt{\frac{\sum_{i=1}^n (pi - ai)^2}{n}}$$

$$S \leftarrow RMSE[i] * Vi$$

In the above equations, e[i] is the error rate computed with the predicted and actual arrays. S is the spam city score, which is computed with the support of attribute importance score and error rate. The complete procedure of spam city score computation is described in the Algorithm 1.

Algorithm 1: Spamicity Score Computation.

Input :

Output : Computed spamicity score

1: **procedure** FUNCTION(PageRank)

2: **for** i = 1 to n **do**

3: **for** j = 1 to 15 **do**

4: Matrix representation z_i

5: Set $j \leftarrow j + 1$

6: Set $i \leftarrow i + 1$

7: **endfor**

8: **end for**

9: **for** i = 1 to 15 **do**

10: Set $V_i \leftarrow x$

11: **endfor**

12: $p[i] \leftarrow Y$

13: **for** i = 1 to 15 **do**

14: Compute $RMSE[i] = \sqrt{\frac{\sum_{i=1}^n (pi - ai)^2}{n}}$

15: **end for**

16: **for** I = 1 to 15 **do** $S \leftarrow RMSE[i] * Vi$

17: **end for**

18 **end procedure**

IV. RESULTS

The below figure shows original data of sensors it represents the normal behavior of sensors that recorded over a 18 months.

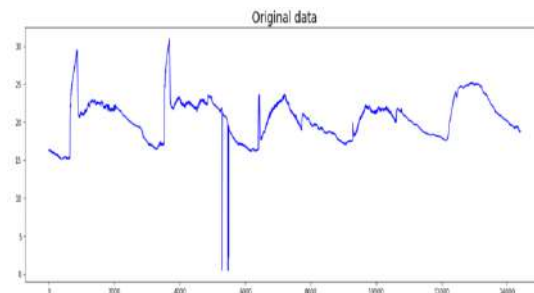


Fig 2. Original Score of Sensors

The below figure shows the overall spam data of sensors that means the abnormal behavior of sensors in the recorded data.

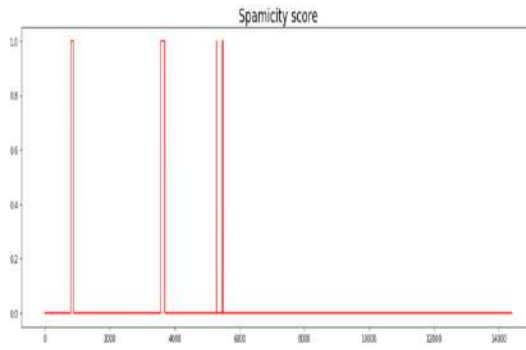


Fig 3. Spamicity Score of Sensors

The figure shows abnormal behaviour of sensors in different time intervals

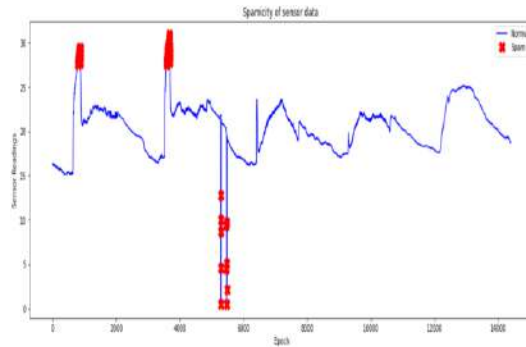


Fig 4. Abnormality Behavior of Sensors

It represents the overall spam and nonspam data that occurred in the given data

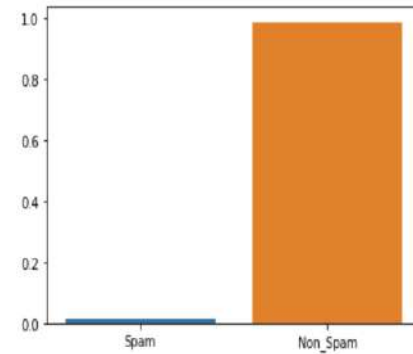


Fig 5. Bar Graph Representation of Spam and Non-Spam

The distribution scores of sensors shows the percentage of spam and non-spam data that we calculated

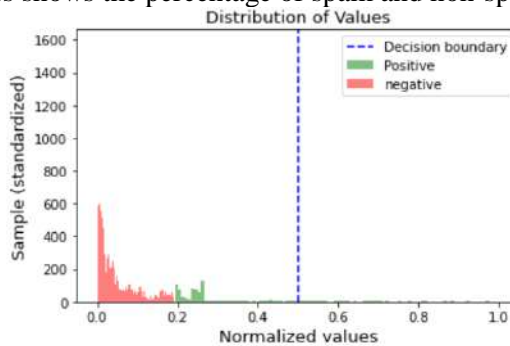


Fig 6. Distribution Scores of Sensors

V. CONCLUSION

The proposed framework detected the spam parameters of IOT devices using ML models. The IOT data set used for experiments was pre-processed by using feature engineering procedure. By experimenting the framework with ML models, each IOT appliance was awarded with a spam score. This refined the conditions to be taken for successful working of IOT devices in a smart home. In the future, we are planning to consider the climatic and surrounding features of 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, 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.