Enhanced Approach For Spam Detection In IOT Devices Using Machine Learning

Dr. K. Srinivasa Rao Professor, Dept of CSE, KSRM College of Engineering, Kadapa.

Machireddy Madhulita M Sahadeva PG SCHOLAR M.Tech.,(CSE), Dept of CSE,KSRM College of Engineering, Kadapa.

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.

1. INTRODUCTION

IoT is taken into account as an interconnected and distributed network of embedded systems communicating through wired or wireless communication technologies. Massive growth and rapid development in the field of the Internet of Things (IoT), makes the presence of IoT devices prevalent in smart homes and smart cities. It is also defined because the network of physical objects or things empowered with limited computation, storage, and communication capabilities is also embedded with electronics (such as sensors and actuators), software, and network connectivity that permits these objects to gather, sometimes process, and exchange data. The things in IoT ask the objects from our lifestyle starting from smart household devices like a smart bulb, smart adapter, smart meter, smart refrigerator, smart oven, AC, temperature sensor, smoke detector, IP camera, to more sophisticated devices like frequency Identification (RFID) devices, heartbeat detectors, accelerometers, sensors in the parking zone, and a variety of other sensors in automobiles, etc.

There are various large amounts of applications and services offered by the IoT ranging from critical infrastructure to agriculture, military, home appliances, and personal health care. As the usage of IoT devices increases the anomalies generated by these devices also grow beyond the count. IoT applications need to ensure information protection to fix security issues like interruptions, spoofing attacks, Dos attacks, jamming, eavesdropping, spam, and malware. The maximum care to be taken is with web-based devices as the maximum number of IoT devices are web-dependent. It is common in the work environment that the IoT devices introduced in an association can be utilized to execute security and protection includes proficiently. For example, wearable devices that collect and send user's health data to a connected smartphone should prevent leakage of data 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 therefore the attackers.

However, with the emergence of ML in various attack scenarios, IoT devices choose a defensive strategy and decide the key parameters in the security protocols for a trade-off between security, privacy, and computation. This work enhances the algorithm to affect the time-series regression model rather than a classification model and may also execute ML models in parallel. This proposed paper focuses on determining the trustworthiness of the IoT device within the smart home network. The algorithm scores an IoT device with a spamicity score to secure smart devices by calculating spam scores using different machine learning models.

II LITERATURE SURVEY

It is often used by computer attackers to characterize hosts or networks which they are considering hostile activity against. Thus it is useful for system administrators and other network defenders to detect portscans as possible preliminaries to a more serious attack. It is also widely used by network defenders to understand and find vulnerabilities in their own networks.

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 selfmanaged 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.

Fig 2.1: Flow Diagram of an efficient spam detection technique for iot devices using machine learning

A. Scope- Based

As part of the recommended approach, the spammy characteristics are detected. ML models are used in Internet of things. This is the IoT data. it is pre-processed with the aid of pattern development method. By playing around with the structure, each IoT device is rewarded with ML models. T he amount of spamming that has been detected As a result, the criteria for success have been refined. IoT equipment operating in a smart house As we go forward, will take into account meteorological conditions as well as the environment IoT devices more secured and reliable

B. Motivation

The use of new innovations give incredible advantages to people, organizations, and governments, be that as it may, messes some up against them. For instance, the protection of significant data, security of put away information stages, accessibility of information and so forth. Contingent upon these issues, digital fear based oppression is one of the most significant issues in this day and age. Digital fear, which made a great deal of issues people and establishments, has arrived at a level that could undermine open and nation security by different gatherings, for example, criminal association, proficient people and digital activists. Along these lines, Intrusion Detection Systems (IDS) has been created to maintain a strategic distance from digital assaults.

III. PROPOSED WORK

Blameless Bayes and Principal Component Analysis (PCA) were been used with the KDD99 dataset by Almansob and Lomte [9].Similarly, PCA, SVM, and KDD99 were used Chithik and Rabbani for IDS [10]. In Aljawarneh et al's. Paper, their assessment and examinations were conveyed reliant on the NSL-KDD dataset for their IDS model [11] Composing inspects show that KDD99 dataset is continually used for IDS [6]–[10].There are 41 highlights in KDD99 and it was created in 1999. Consequently, KDD99 is old and doesn't give any data about cutting edge new assault types, example, multi day misuses and so forth. In this manner we utilized a cutting-edge and new CICIDS2017 dataset [12] in our investigation.

Important steps of the algorithm are given in below. 1) Normalization of every dataset. 2) Convert that dataset into the testing and training. 3) Form IDS models with the help of using RF, ANN, CNN and SVM algorithms. 4) Evaluate every model's performances

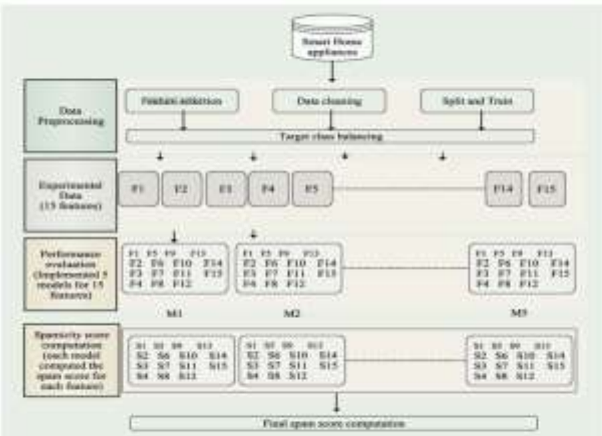


Fig 2: Architecture Diagram

A. Algorithms Used

Decision Tree: The goal of this algorithm is to create a model that predicts the value of a target variable, for which the decision tree uses the tree representation to solve the problem in which the leaf node corresponds to a class label and attributes are represented on the internal node of the tree.

Logistic Regression: Logistic regression uses the concept of predictive modeling as regression; therefore, it is called logistic regression, but is used to classify samples; Therefore, it falls under the classification algorithm.

Support Vector Machine: Support Vector Machine(SVM) is a supervised machine learning algorithm that can be used for both classification or regression challenges. However, it is mostly used in classification problems.

K-Nearest Neighbour: K-Nearest Neighbour. It is a supervised machine learning algorithm. The algorithm can be used to solve both classification and regression problem statements. The number of nearest neighbours to a new unknown variable that has to be predicted or classified.

Recursive Feature Elimination: Recursive Feature Elimination, or RFE for short, is a popular feature selection algorithm. RFE is popular because it is easy to configure and use and because it is effective at selecting those features (columns) in a training dataset that are more or most relevant in predicting the target variable.

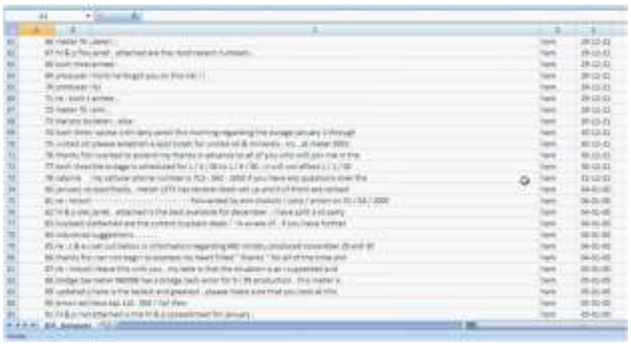
Neural Network Model: An artificial neural network learning algorithm, or neural network, or just neural net. , is a computational learning system that uses a network of functions to understand and translate a data input of one form into a desired output, usually in another form.

IV METHODOLOGY

An algorithm uses training data and feedback from humans to learn the relationship of given inputs to a given output. For instance, a practitioner can use marketing expense and weather forecast as input data to predict the sales of cans. You can use supervised learning when the output data is known. The algorithm will predict new data. There are two categories of supervised learning: □ Classification task □ Regression task Classification Imagine you want to predict the gender of a customer for a commercial. You will start gathering data on the height, weight, job, salary, purchasing basket, etc. from your customer database. You know the gender of each of your customer, it can only be male or female. The objective of the classifier will be to assign a probability of being a male or a female (i.e., the label) based on the information (i.e., features you have collected). When the model learned how to recognize male or female, you can use new data to make a prediction. For instance, you just got new information from an unknown customer, and you want to know if it is a male or female. If the classifier predicts male = 70%, it means the algorithm is sure at 70% that this customer is a male, and 30% it is a female. The label can be of two or more classes. The above Machine learning example has only two classes, but if a classifier needs to predict object, it has dozens of classes (e.g., glass, table, shoes, etc. each object represents a class)

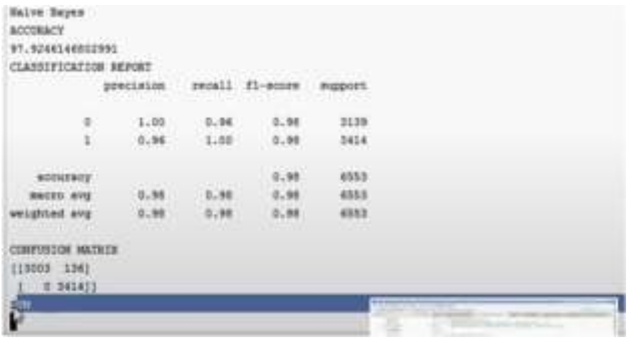
Regression When the output is a continuous value, the task is a regression. For instance, a financial analyst may need to forecast the value of a stock based on a range of feature like equity, previous stock performances, macroeconomics index. The system will be trained to estimate the price of the stocks with the lowest possible error. Unsupervised learning Unsupervised learning algorithms take a set of data that contains only inputs, and find structure in the data, like grouping or clustering of data points. The algorithms, therefore, learn from test data that has not been labeled, classified or categorized. Instead of responding to feedback, unsupervised learning algorithms identify commonalities in the data and react based on the presence or absence of such commonalities in each new piece of data. A central application of unsupervised learning is in the field of density estimation in statistics, such as finding the probability density function. Though unsupervised learning encompasses other domains involving summarizing and explaining data features.

Results



Screen 1: Data Set Related Spam Detection

Screen 2: Home Page of Project



Screen 3: ML Algorithms showing various Precision, recall and f1score values

Screen 4: Chart Showing Accuracy values using Various ML Methods

Screen 5: Chart Showing Accuracy values using Various ML Methods

V. CONCLUSION

The proposed framework, detects the spam parameters of IoT devices using machine learning models. The IoT dataset used for experiments, is pre-processed by using feature engineering procedure. By experimenting the framework with machine learning models, each IoT appliance is awarded with a spam score. This refines the conditions to be taken for successful working of IoT devices in a smart home. In future, we are planning to consider the climatic and surrounding features of IoT device to make them more secure and trustworthy.

REFERENCE

[1] REFERENCES 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] [3] 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. E. Bertino and N. Islam, "Botnets and internet of things security," Computer, no. 2, pp. 76–79, 2017. [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] 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. 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. 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. 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. 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. 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. 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. 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. Y. Li, D. E. Quevedo, S. Dey, and L. Shi, "Sinr-based dos attack on remote state estimation: A gametheoretic approach," IEEE Transactions on Control of Network Systems, vol. 4, no. 3, pp. 632–642, 2016. 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.