JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Machine Learning Based Spam Detection over Internet Telephony

Mr. Kapu Sainath Reddy (M.C.A). Rajeev Gandhi Memorial college Of Engineering and Technology, Nandyal

Mr. Dr. K.E. Naresh Kumar (MTech, Ph.D.). Rajeev Gandhi Memorial college Of Engineering and Technology, Nandyal

Abstract

The greatest amount of SPIT devices is web-dependent, hence caution must be exercised when using web-based devices. It is typical in the workplace that the SPIT devices set up in an organization can be used to effectively carry out security and protection measures. For instance, wearable technology that gathers and transmits user health information to a connected smartphone should guard against data leaks to preserve privacy. According to market research, 25–30% of employees who are currently at work connect their own SPIT devices to the company network. The fact that SPIT is growing draws both the target audience, or the users, and the attackers as a result. However, when ML becomes more prevalent in various assault scenarios, SPIT devices decide on a defensive plan and the key. parameters for a trade-off between security, privacy, and computation in the security protocols. Instead of affecting a classification model, this work improves the technique such that it can affect a time-series regression model. It may also execute ML models concurrently. The purpose of this suggested research is to evaluate the SPIT device's reliability within the smart home network.

1. INTRODUCTION

1.1 Introduction

SPIT is viewed as a distributed, interconnected network of embedded systems that communicates via wired or wireless methods. Due to the Over Internet Telephony's (SPIT) explosive growth and development, SPIT devices are widely used in smart homes and smart cities. It is also defined by the network of physical things or objects that are endowed with modest computation, storage, and communication capabilities 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. The items in SPIT inquire about various aspects of our daily lives, starting with smart home appliances such a smart lamp, smart adaptor, smart meter, smart refrigerator, and smart oven. Several sensors are used in automobiles, ranging from basic ones like the AC, temperature sensor, smoke detector, IP camera, and accelerometers to more complex ones like frequency identification (RFID) devices, heartbeat detectors, parking zone sensors, and many other types. The SPIT offers a wide range of extensive applications and services, including those for critical infrastructure, agriculture, the military, household appliances, and individual health care. The number of abnormalities produced by SPIT devices likewise grows beyond what can be counted as their use expands. To address security problems including interruptions, spoofing attacks, DoS attacks, jamming, eavesdropping, spam, and malware, SPIT applications must offer information protection.

2. Literature Survey

The SPIT offers a wide range of extensive applications and services, including those for critical infrastructure, agriculture, the military, household appliances, and individual health care. The number of abnormalities produced by SPIT devices likewise grows beyond what can be counted as their use expands. To address security problems including interruptions, spoofing attacks, DoS attacks, jamming, eavesdropping, spam, and malware, SPIT applications must offer information protection. The greatest amount of SPIT devices are web-dependent, hence caution must be exercised when using web-based devices.

It is typical in the workplace that the SPIT devices set up in an organization can be used to effectively carry out security and protection measures. For instance, wearable technology that transmits user health information to a smartphone attached to the device should stop data leaks to protect privacy. According to market research, 25–30% of employees who are currently at work connect their own SPIT devices to the company network. The fact that SPIT is growing draws both the target audience, or the users, and the attackers as a result. In contrast, SPIT devices decide on a defensive approach and the critical security protocol parameters for a trade-off between security, privacy, and computation as ML emerges in various attack situations. Instead of affecting a classification model, this work improves the technique such that it can effect a time-series regression model. It may also execute ML models concurrently. The purpose of this suggested research is to evaluate the SPIT device's reliability within the smart home network.

By computing spam scores using several machine learning models, the system assigns a spam city score to an SPIT device in order to secure smart devices. Computer attackers frequently employ it to describe hosts or networks that they believe are being the target of hostile activities. The ability to recognize ports cans as potential precursors to a more serious attack is thus useful for system administrators and other network defenders. Network defenders frequently utilize it as well in order to comprehend and identify weaknesses in their own networks.

Wearable technology, household appliances, and software now have the ability to share and transmit information online thanks to the Over Internet Telephony (SPIT) technology. Information security on the shared data is a crucial issue that cannot be ignored because the shared data contains a significant amount of private information. In this essay, we start with a broad overview of SPIT's information security history before moving on to the issues that SPIT will face in this area. Finally, we will highlight potential future research areas for the development of solutions to the security issues SPIT faces. requests are sent across regionally dispersed internet connections utilizing a network of zombie machines. Due to the network congestion and network component disablement caused by DDoS, SPIT is much more adversely affected. In order to analyze the interactive communication between various types of network nodes, a lightweight defensive method for DDoS attacks over SPIT network environments is presented and tested against various scenarios.

3. OVERVIEW OF THESYSTEM

3.1 Existing System

Denial of service (DDoS) attacks: To prevent SPIT devices from accessing various services, the attackers can flood the target database with erroneous requests. Bots are the term used to describe these malicious queries sent by a network of SPIT devices. DDoS has the ability to deplete every resource offered by the service provider. It has the power to disable legitimate users and disable network resources. Attacks on the SPIT device's physical layer are known as RFID attacks. The device's integrity is compromised as a result of this assault. Attackers make an effort to alter the data either at the storage node or during network transmission. Common attacks that could occur at the sensor node include assaults on confidentiality, availability, and authenticity. Keys for cryptography are brute-forced. Password protection, data encryption, and restricted access control are some of the countermeasures to ensure prevention of such assaults. Internet assaults The SPIT gadget can maintain an Internet connection to access a variety of resources. Spammers utilize spamming strategies when they wish to access information from other systems or keep getting visitors to their target website. Ad fraud is a typical method employed for the same. For financial gain, it creates the fake clicks at the targeted website. Cyber criminals are a group like this that practice online. Attacks using NFC: The major target of these attacks is fraud involving electronic payments. Unencrypted traffic, eavesdropping, and tag alteration are examples of potential assaults. The conditional privacy protection is the answer to this issue. As a result, the attacker is unable to generate the identical profile using the user's public key. The trusted service manager's random public keys form the basis of this concept.

3.1.1 Disadvantages of Existing System

In the existing work, the system is less effective due to lack of Spam Detection in SPIT using Machine Learning framework.

This system is less performance in which Supervised machine learning techniques is absence.

3.2 Proposed System

The smart devices are absolutely necessary for the digital age. These gadgets should only return information that is accurate and not spam. Because data is gathered from different domains, information retrieval from diverse SPIT devices is a significant difficulty. Due to the multiplicity of devices used in SPIT, a sizable amount of heterogeneous, diverse data is produced. This data may be referred to as SPIT data. Multiple characteristics of SPIT data include real-time, multi-source, rich, and sparse.

3.3 Methodology

4 Algorithm:

5 **THE DESCRIPTION OF MODULES SVMs**, or support vector machines Support vector networks and support vector machines are a family of closely related supervised learning techniques used for regression and classification. However, classification issues are where it's most frequently used. Each data point is plotted as a degree in n-dimensional space (n is the number of features you have) using the SVM algorithm, with each feature's value being equal to the value of a particular coordinate. The hyper-plane that separates the two classes can then be found to classify. So, we can say that the basic goal of SVM is to identify a hyperplane in an N-dimensional space that clearly divides the data points into categories. Both linear and non-linear data can be classified using SVM. classification of non-linear data extremely probabilistic categorization environment. Using a technique known as the kernel trick, SVMs can effectively conduct non-linear classification in addition to linear classification by implicitly translating their inputs into high-dimensional feature spaces. SVM also employs a different technique known as Soft Margin, which enables SVM to make a predetermined number of errors while maintaining the margin as broad as feasible to allow for the accurate classification of other points.

6 Architecture

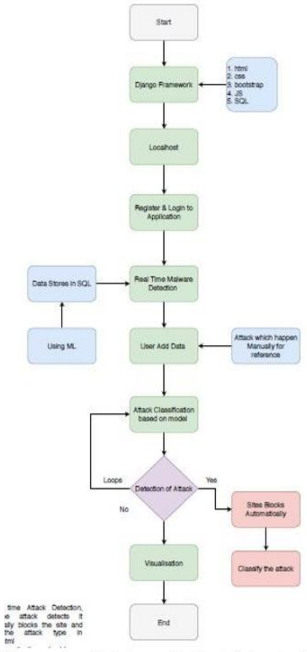
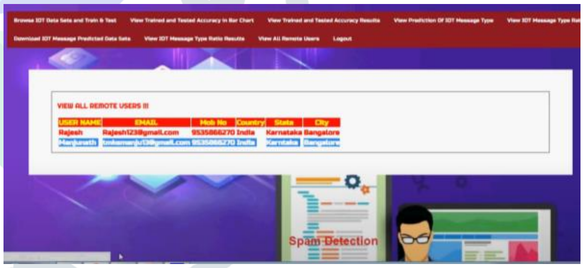


Fig 2.1: Flow Diagram of an efficient spam detection technique for iot devices using machine learning

Fig 1: Frame work of proposed method

7 RESULTS SCREEN SHOTS

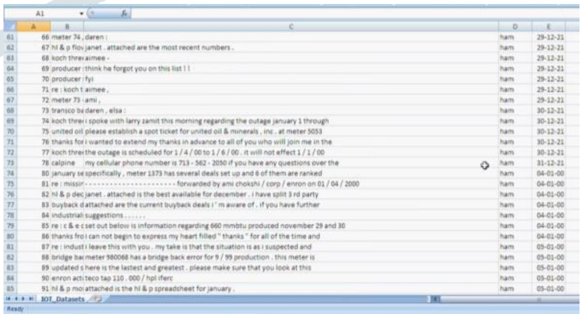
Home Page:


Browse IOT Data Sets and Train & Test
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View Prediction Of IOT Message Type
View IOT Message Type Rat
Download IOT Message Predicted Data Sets
View IOT Message Type Ratio Results
View All Remote Users
Logout
Spam Detection

VIEW ALL REMOTE USERS !!!

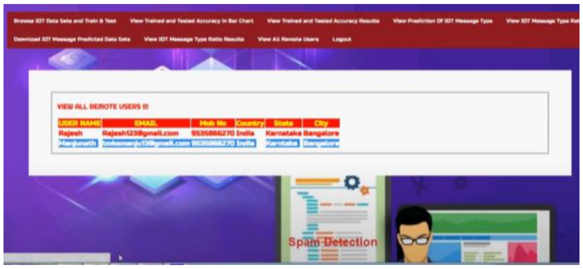
USER NAME	EMAIL	Mob No	Country	State	City
Rajesh	Rajesh123@gmail.com	9535866270	India	Karnataka	Bangalore
Manjunath	tmksmanju13@gmail.com	9535866270	India	Karntaka	Bangalore

Data Set:



	A	B	C	D	E
61	66	meter 74 ,	daren :	ham	29-12-21
62	67	hl & p flo	janet . attached are the most recent numbers .	ham	29-12-21
63	68	koch thre	aimee -	ham	29-12-21
64	69	producer :	think he forgot you on this list ! !	ham	29-12-21
65	70	producer :	fyi	ham	29-12-21
66	71	re : koch t	aimee ,	ham	29-12-21
67	72	meter 73 -	ami ,	ham	29-12-21
68	73	transco b	daren , elsa :	ham	30-12-21
69	74	koch thre	i spoke with larry zamit this morning regarding the outage january 1 through	ham	30-12-21
70	75	united oil	please establish a spot ticket for united oil & minerals , inc . at meter 5053	ham	30-12-21
71	76	thanks for	i wanted to extend my thanks in advance to all of you who will join me in the	ham	30-12-21
72	77	koch thre	the outage is scheduled for 1 / 4 / 00 to 1 / 6 / 00 . it will not effect 1 / 1 / 00	ham	30-12-21
73	78	calpine	my cellular phone number is 713 - 562 - 2050 if you have any questions over the	ham	31-12-21
74	80	january se	specifically , meter 1373 has several deals set up and 6 of them are ranked	ham	04-01-00
75	81	re : missir	- forwarded by ami chokshi / corp / enron on 01 / 04 / 2000	ham	04-01-00
76	82	hl & p dec	janet . attached is the best available for december . i have split 3 rd party	ham	04-01-00
77	83	buyback d	attached are the current buyback deals i ' m aware of . if you have further	ham	04-01-00
78	84	industrial	suggestions	ham	04-01-00
79	85	re : c & e c	set out below is information regarding 660 mmbtu produced november 29 and 30	ham	04-01-00
80	86	thanks fro	i can not begin to express my heart filled " thanks " for all of the time and	ham	04-01-00
81	87	re : indust	i leave this with you . my take is that the situation is as i suspected and	ham	05-01-00
82	88	bridge ba	meter 980068 has a bridge back error for 9 / 99 production . this meter is	ham	05-01-00
83	89	updated s	here is the lastest and greatest . please make sure that you look at this	ham	05-01-00
84	90	enron act	teco tap 110 . 000 / hpl iferc	ham	05-01-00
85	91	hl & p mo	attached is the hl & p spreadsheet for january .	ham	05-01-00

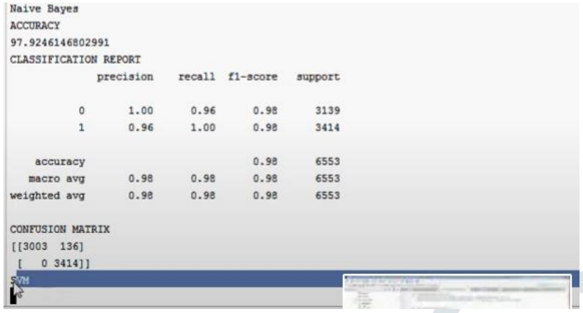
Register:


Browse IOT Data Sets and Train & Test
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View Prediction Of IOT Message Type
View IOT Message Type Rat
Download IOT Message Predicted Data Sets
View IOT Message Type Ratio Results
View All Remote Users
Logout
Spam Detection

VIEW ALL REMOTE USERS !!!

USER NAME	EMAIL	Mob No	Country	State	City
Rajesh	Rajesh123@gmail.com	9535866270	India	Karnataka	Bangalore
Manjunath	tmksmanju13@gmail.com	9535866270	India	Karntaka	Bangalore

Predict Result:



```
Naive Bayes
ACCURACY
97.9246146802991
CLASSIFICATION REPORT
              precision    recall  f1-score   support

           0       1.00      0.96      0.98      3139
           1       0.96      1.00      0.98      3414

    accuracy                           0.98      6553
   macro avg       0.98      0.98      0.98      6553
weighted avg       0.98      0.98      0.98      6553

CONFUSION MATRIX
[[3003  136]
 [   0 3414]]
SVM
```

7. CONCLUSION

✓ The suggested system uses machine learning models to find the SPIT devices' spam settings. The feature engineering approach is utilized to pre-process the SPIT dataset that was used in the tests. Each SPIT appliance receives a spam score as a result of the framework's machine learning model experiments.

Future Enhancement

✓ This clarifies the prerequisites needed for SPIT devices in a smart home to operate well. Future SPIT device design will consider environmental factors to increase security and dependability.

8. References

[1] Fatima Hussain,Rasheed Hussain,Syed Ali HassanHossain. Machine Learning in SPIT Security: Current Solutions and Future Challenges

[2] Choi, J.; Jeoung, H.; Kim, J.; Ko, Y.; Jung, W.; Kim, H.; Kim, J. Detecting and identifying faulty SPIT devices in smart homes with context extraction. In Proceedings of

The 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks, DSN 2018, Luxembourg, 25– 28 June 2018; pp. 610–621.

[3] Tang, S.; Gu, Z.; Yang, Q.; Fu, S. Smart Home SPIT Anomaly Detection based on Ensemble Model Learning from Heterogeneous Data. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 4185–4190.

[4] Makkar A.; Garg S.; Kumar, N.; Hossain, M.S.; Ghoneim, A.; Alrashoud, M. An Efficient Spam Detection Technique for SPIT Devices using Machine Learning. IEEE Trans. Ind. Inform. 2020.

[5] Ameema Zainab, Shady S. Refaat and Othmane Bouhali;Ensemble-Based Spam Detection in Smart Home SPIT Devices Time Series Data Using Machine Learning Techniques

[6] L. University, "Refit smart home dataset," https://repository.lboro.ac.uk/ articles/REFIT Smart Home dataset/2070091, 2019 (accessed April 26, 2019

.