International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 04 | Apr 2021 www.irjet.net p-ISSN: 2395-0072

DYNAMICS OF EMAIL MALWARE

G.Vivekanandan[1], Ashrutha B.N[2], Girisa T[3], Jenifer U[4], Keerthana S[5]

[1] *Assistant Professor, Department of IT, SRM Valliammai Engineering College, Tamilnadu, India*
[2-5] *UG Student, Department of IT, SRM Valliammai Engineering College, Tamilnadu, India*

Abstract - *In the present web mail worker, spam conveyance is the most well-known issue. In the beneficiary side just a large portion of the advanced spam-separating procedures are sent. They are acceptable at separating spam for end clients however spam messages actually continue to squander Internet transfer speed. This work is consequently expected to distinguish and stop spamming bots from the beginning. Because of the tremendous number of email tends to in the SMTP meetings, we store and oversee them proficiently through Bloom channels. We screen the SMTP meetings and track the number and the uniqueness of the recipient email addresses in the mail messages from every individual inside have as the highlights for recognizing spamming bots.*

Key Words: *Bloom filter, Word net, SMTP, Naïve Bayes, XG Boost, Spamming bots.*

1. INTRODUCTION

An email infection offers more in the PC infections. Email viruses may be of any kind such as malicious code, infected attachments etc. There are various kinds of malwares such as adware, key loggers, dialers, spywares etc. These malwares are been infused to the messages and are been shipped off to different clients. These malwares which are being an immense difficulty is been obstructed and makes the client helpful. The sender in this manner doesn't gets profited by sending these spam messages. This project is to identify email malware bots and square the spamming bots in sender side itself with the utilization of bloom filter techniques. The arrangement can be coordinated into existing email foundation, and the contribution of block chain is possibly required when issues during the message trade emerge. Consequently the sender no more sends any spam mail to the client and the client is profited.

2. LITERATURE SURVEY

1. **Understanding Localized-Scanning Worms**: Localized scanning is a simple technique used by attackers to search for vulnerable hosts. Localized scanning trades off between the local and the global search of vulnerable hosts and has been used by Code Red II and Nida worms.

2. **Characteristics of Spammers and Their Network Reachability Properties**: By analysing a two-month traceof more than 25 million emails received at a large US university campus network, of which more than 18 million are spam messages

3. **Fast Port scans Detection Using Sequential Hypothesis Testing**: Attackers routinely perform random "ports cans" of IP addresses to find vulnerable servers to compromise.

4. **An Effective Defence against Email Spam Laundering**: Laundering email spam through open-proxies or compromised PCs is a widely-used trick to conceal real spam sources and reduce spamming cost

3. EXISTING SYSTEM

The existing system of email malware performs the filtering of spam mails which is sent to the receiver. These sends are then put away in the spam envelope that is noticeable to the client. This may lead to infection of the system when the user accesses the malicious links or any other files unfortunately. The infection results in hacking the entire system or performing truncate process and so on. Hence a filtering technique called bloom filter was implemented to filter the spams by analyzing only the negative words. In any case, this strategy was not that much utilized progressively since it isn't effective to keep away from spammers. There are certain drawbacks in the existing system as follows:

1. There are an immense number of spams with more number of duplications in it, consequently the reflection can't identify the developing duplication in a successful manner.

2. If the user executes any links that are harmful to the system, then the botnet can infect in a rapid process.

3. There are countless bots which sends the huge measure of spam information from numerous assets, along these lines it can't be distinguished in a brief span.

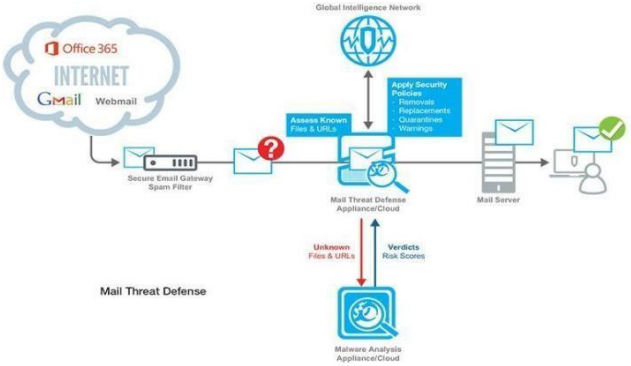

Fig:3.1 Existing system

4. PROPOSED WORK

In our project, we have enhanced the existing system by reporting and blocking the spam contents without getting received to the other side. This is implemented by giving a number of warnings to the spam sender. Even after getting warnings if the sender keeps sending malicious contents, the sources will get hindered. The enhancement executes the reporting the links and encrypted files like .exe extension which is not available in the existing concept. These are performed under two technologies namely Artificial Intelligence and Machine Learning. Under AI, Sentimental analysis and Text analysis are executed. Under ML, tracking of repeated events and analyzing spam links, encrypted files are executed.

4.1 REPORTING OF SPAM MESSAGES

WORD NET:

Word net is the large database which contains collection of words in it. The words in the word net will be of the malicious words, which are injected by the admin. If the user enters the words which is in the word net (Malicious words) those, contained in the mail will not be sent to the receiver.

The words will be spilt and get investigated; if the word in the message matches with the words in the word net the mail will be obstructed.

BLOOM FILTER:

Bloom filter is a Data structure and it is similar to the Hash table, Bloom filter will check whether the element is present or not present in the set. It is Memory efficient, which contains less storage than any other data structure and it also run faster The Bloom filter will check the words in the mail (Subject, Body) from the words that is in the word net.If the words get matches from the words in the word net, the sender will not be able to send that mail It will notify as a warning and then report will be generated to the sender.

In our Project we fix the count as 10, only 10 attempt or the warning will be given to the user, for every warning of reporting the count will be reducing once the count reaches 0 the user will be get blocked.

4.2 SPAM FILTERING

NAIVE BAYES

In Machine Learning we utilize a Naive Bayes calculation for email spam separating which is the probabilistic calculation dependent on the Bayes hypothesis. Supervised and unsupervised learning of Machine Learning process is used for spam analyzing In which we use the supervised learning of spam analyzing where the input data will be trained by Naïve Bayes algorithm which classify the spam and the Ham mails, the Ham mails also contains the repeated mails.

The repeated mail will be get analyzed with the events that occurred so for based on the occurrence of the events. Thus the repeated events will be get identified and sent as a warning to the sender of the mail.

XG BOOST

XG Boost Stands for Extreme Gradient Boosting, which is one, the Machine Learning algorithm that uses Gradient boosting framework. We use the boosting process to improvise the system, where the XG Boost is the Extension of the Gradient Boosting process.

Hyper parameter Tuning of the link and .exe file is been done, process such as the checking, analyzing and detection of the files will be done. The infected files and the corrupted files will be detected. The unstructured data such as the link and .exe file extension will also be get analyzed and detected, the extension .exe will include files that are only less than 1MB. If the link contains any malicious information or content, the link will get reported and will not be sent to the receiver.

Block Diagram:

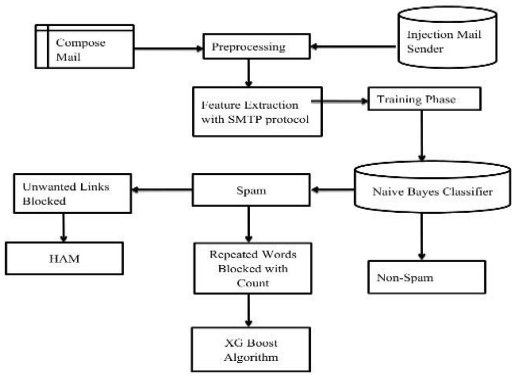

Fig 4: Block Diagram

5. SYSTEM ARCHITECTURE

Architecture diagram which shows the gateway and cloud storage in the sender side. The gateway is said to be the spam filter which checks and analyses the content of the mail from the sender. The technique used for filtering is Bloom filter technique. The algorithm used in the filtration process is called as Naive Bayes algorithm. This acts as an event tracker which analyses and tracks the occurrence of repeated events. The analyzed events are then classified as

positive and negative mails. Secondly the cloud storage which contains the word net dictionary performs under Sentimental analysis technique in AI. Here the technique is implemented by NLP (Natural Language Processing) where the texts in the mail are analyzed whether it is meaningful or not. The NLP keeps a count on the malicious texts and other contents such as links, encrypted data in .exe extension, etc. These are collected as a word count data in the dictionary. Now after the filtering and analyzing techniques these mails are being categorized as positive mails and negative mails. The negative mails are been verified with their content and reported from the sender side. The positive mails are been sent directly to the receiver's server and it's been delivered to the receiver. In that case, if there are any repeated mails in the receiver side then a technique called XG boost (Extreme Gradient Boosting) is being implemented to check the repeated mails or spams and these are then reported as spam content and eventually blocked after a certain number of counts.

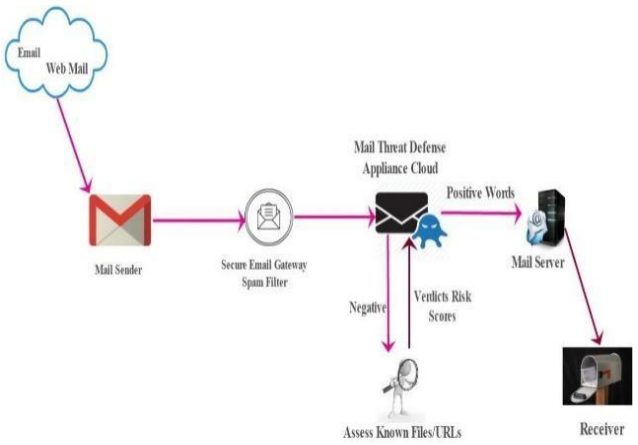

Fig 5: System Architecture

6. IMPLEMENTATION OF THE SYSTEM

An email application is been developed. The user cancreate his/her own email account or can sign in to their existing account with username and password. The new user has to sign up by filling his/her own details such as first name, username, password, contact details, gender, Date of Birth, etc. Now, an individual email ID will be generated. If in case the user forgets his/her own username or password they can reset the password by filling out the specified details. Once the user receives the mail, the mail is been checked whether it contains any malicious words or not. When certain malicious are been found in the mail, a report message is been sent to the sender. The sender is been reported regarding the usage of words and eventually blocked.

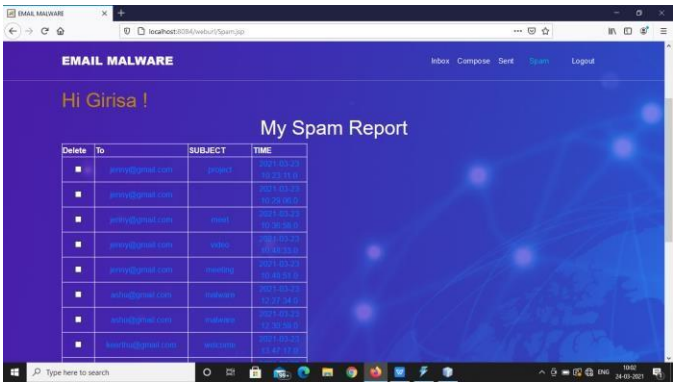

Fig 6.1: Spam Report

Sentimental analysis uses natural language processing (NLP), text analysis to identify, analyze, quantify subjective information in the email. Machine Learning (ML) techniques are combined to sentiment analyze to the topics, categories or entities within a phrase. The user input will be get analyzed and detected. The text that you have sent through mail is been further analyzed whether they are positive or negative and analyzing it making sure that no negative contents or malicious contents are been sent to the receiver.

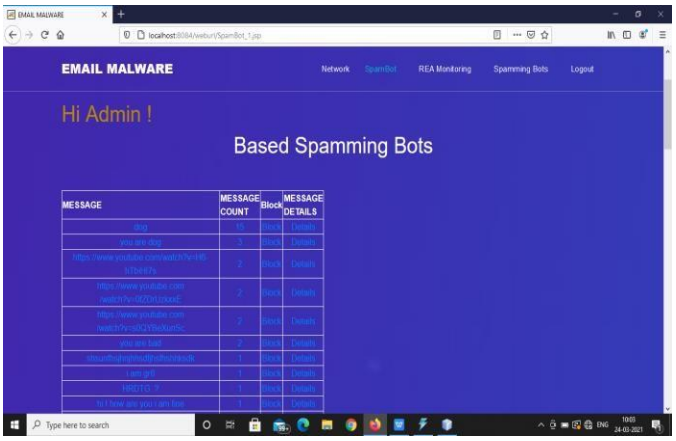

Fig 6.2: Spamming bots

Spam filtration process is done by checking and separating the spam and Non spam email The (HAM) Non spam emails are further categorized as new non spam (HAM) mails and repeated non spam mails. The occurrence of events will be analyzed by which the repeated events will be identified.

Naive Bayes classification is a simple algorithm in which all features of the model are independent. In the context of the spam filter, every word in the message is independent of all other words and we count them with the ignorance of the context.

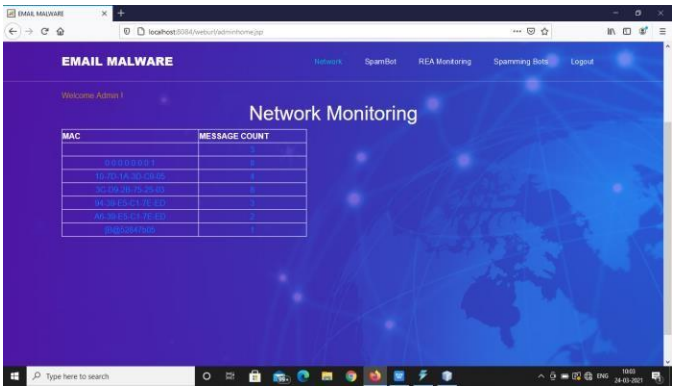

Fig 6.3: Monitoring

In REA monitoring, the recipient mail address and the count of messages are recorded. The email address of the sender is been monitored and analyzed and made sure that no malicious content reaches the receiver. The Messages are recorded with their mac address. The message type is analyzed. If they contain any malicious contents the count will get included in the server the count level will be eventually monitored. When the count reaches 0 the system will be get blocked. That particular email Id is been blocked and the user belonging to the email address can no longer use it to send the emails.

7. TECHNICAL SCOPE

1. HOSTING: Facilitating the web application in the Linode server. It is a cloud server where information can be put away. It is simpler and less expensive than other cloud administrations.

2. AUTHENTICATION: Authentication can be in two ways. They are single factor verification and Multi factor validation.

SINGLE FACTOR AUTHENTICATION: In single factor confirmation, just one stage check is required**.**

MULTI FACTOR AUTHENTICATION: In multifactor validation two stage confirmation is finished by giving consent in the portable and furthermore by OTP.

3. LOCATION: The area from where the email is been sent is been shown here thus it is a lot of simpler to discover the sender.

8. RESULT

Thus our project Email malware of spam reporting and blocking of the network as the system based blocking and REA as a email Id based blocking has been implemented.

9. CONCLUSION

In this project, we presented the repetitious spreading processes caused by the reinfection and the self-start. The experiments showed that the result close to the 0simulations. For the future work, there are also some problems needed to be solved, such as the independent assumption between users in the network and the periodic assumption of email checking time of users.

REFERENCES

[1] M. Fossi and J. Blackbird, “Symantec Internet Security Threat Report 2010,” technical report Symantec Corporation, Mar. 2011.
[2] P. Wood and G. Egan, “Symantec Internet Security Threat Report 2011,” technical report, Symantec Corporation, Apr. 2012.
[3] C.C. Zou, D. Towsley, and W. Gong, “Modeling and Simulation Study of the Propagation and Defense of Internet E-Mail Worms,” IEEE Trans. Dependable and Secure Computing, vol. 4, no. 2, pp. 105- 118, Apr.-June 2007.
[4] Z. Chen and C. Ji, “Spatial-Temporal Modeling of Malware Propagation in Networks,” IEEE Trans. Neural Networks, vol. 16, no. 5, pp. 1291-1303, Sept. 2005.
[5] C. Gao, J. Liu, and N. Zhong, “Network Immunization and Virus Propagation in Email Networks: Experimental Evaluation and Analysis,” Knowledge and Information Systems, vol. 27, pp. 253-279, 2011