Adaptive Spammer Detection at the Source Network
Kiran Jadhav#, Chetan Gawande, Yaqub Patel, Shrinivas Halhalli.
Department of Computer Engineering, G.S. Moze College of Engineering,
Ganeshkhind Pune, Maharashtra, India.

ABSTRACT

Identifying and fixing the affected machines is the key step to resolve any security threats in a network. Because, it becomes a route to launch several attacks such as Denial of service attacks, spamming, stealing user identities and spreading malware etc. Spamming is one of the major threats where attackers perform single attack and make multiple machines in a network as compromised machines. Even though few existing methods like spam signatures and spam behavior analysis resolved the problems to certain extent, it is still not applicable in large networks. Moreover, these methods lack online spam detection mechanism. Existing systems and its drawbacks are also discussed in this report. An effective design and implementation of a tool is important to monitor and detect spam attacks in a real time network. In this research, a tool is developed to differentiate spam affected and non-spam affected machines by the exchange of messages in a network in an online manner. The tool keeps track of IP addresses of each machine and records the spam percentage in a network. It maintains the privacy of the clients exchanging non-spam emails by encrypting its content from the view of the network administrator. A definite algorithm in this report is used to differentiate between spam and non-spam. The performance of this tool is based on the parameters like number of spam messages, percentage of spam detected and its efficiency to overcome the limitations of the existing systems.

I. INTRODUCTION

Due to the wide popularity of the internet and its communication with no cost, it was recognized as the premium tool for advertising and marketing. With respect to economic constraints, most number of people started sending emails to thousands of people across the world. This made internet, a commercial network with the association of electronic mail as one of the quick resources of communication. The major problem in today's internet world is sending bulk or unsolicited emails to numerous users. This adds an additional advantage of launching other attacks and wasting of resources. E-mail spam comes under the electronic spam which sends bulk of unnecessary or junk mail of duplicate emails to the recipients.

II. PROBLEM STATEMENT

Identifying and fixing the affected machines is the key step to resolve any security threats in a network. Because, it becomes a route to launch several attacks such as Denial of service attacks, spamming, stealing user identities and spreading malware etc. Spamming is one of the major threats where attackers perform single attack and make multiple machines in a network as compromised machines. Even though few existing methods like spam signatures and spam behavior analysis resolved the problems to certain extent, it is still not applicable in large networks. Moreover, these methods lack online spam detection mechanism. Existing systems and its drawbacks are also discussed in this report.

An effective design and implementation of a tool is important to monitor and detect spam attacks in a real time network. In this research, a tool is developed to differentiate spam affected and non-spam affected machines by the exchange of messages in a network in an online manner. The tool keeps track of
IP addresses of each machine and records the spam percentage in a network. It maintains the privacy of the clients exchanging non-spam emails by encrypting its content from the view of the network administrator. A definite algorithm in this report is used to differentiate between spam and non-spam. The performance of this tool is based on the parameters like number of spam messages, percentage of spam detected and its efficiency to overcome the limitations of the existing systems.

III. EXISTING SYSTEM

Many models have come into existence in order to reduce the amount of email spam transfer across the network. Despite the awareness of spam emails, productive efforts have not been developed for the network administrator to monitor the status of clients in the network. Each model has its own advantages and disadvantages. Following are few existing systems and their drawbacks.

The aim of this model is to analyze the behavior of spam in the network through its characteristics and properties. This behavioral analysis identifies the spam loop holes in the network thus helping the future strategies for the prevention of spam spread across the network. This mechanism is developed based on the framework AutoRE which identifies the spam attackers based on the signatures from the exchange of messages. This framework is based on the signature based worm and virus detection systems.

Figure 1 represents URL preprocessor, group selector and RegEx generator are the three modules used in this framework. The task of identifying URLs and other relative details from the emails of the end user are clustered based on the web domains. Based on the degree of burstiness, group selector selects the URL clusters that were performed by the URL preprocessor and transfers to RegEx generator. This module identifies the signatures by analyzing cluster behavior and generates for all the clusters. The signatures that match with the URLs are separated from the other URL groups to evade further progress. This is repeated until all the clusters are fulfilled.

Figure1: AutoRE modules and processing flow chart [8]

Regular expression signatures are more robust relative to the fixed string based signatures and able to identify more spam information. These signatures can highly bring down the false-positive rates in identifying the URLs. Moreover, this framework also uses these signatures to the large number of group emails, which are referred as spam campaigns. This framework used some email patterns from hot mail and monitored for three months in order to study the spam characteristics [8]. The success ratio of this framework is really high based on the studies it carried out on the email traces.

IV PROPOSED MODEL

Unique credentials are maintained for the master server and this user is solely responsible to monitor
and configure the network statistics. There are two fields containing User ID and password.

Testing is a very important module in the software development to verify, validate and provide quality and service for different components of software. It is used to minimize the risks by efficient use of resources in the development life cycle. This module can be employed at any point of the development process. It is efficient for the testing phase to be implemented at initial level to lower down the risks of defects and failures. Software is tested and implemented in various conditions and environments to examine different aspect of software. There exist various organizations for testing which is based on the type of software developed. This testing phase helps each and every module to work effectively and generate the optimum results. This Application Validating the client credential and IP address. This tool was also tested various email containing spam and non spam contains.

V FRAMEWORK AND OVERVIEW

This tool is really useful in any organization in which servers and clients are connected in a network. In reference to Figure 3, the clients (PC 1.1, PC 1.2 etc.) can communicate either in a network or even outside the network. All outgoing messages and internal messages sent by any client to other machines are routed through their individual servers (server1). The server scans the email and detects the spam that is being sent by the client. The server receives the email and delivers to its corresponding destination nodes if the email is free of spam.

VI MATHEMATICAL MODEL

Set Theory:

Let

\[ S = \{ U, MS, D \} \]

\[ MS = \{ IP, MAC \} \]

Where

\[ U \text{ is a set of users} \]
There can be number of users in our system which interacts with our developed system and uses features of our system. 

**MS is the MicroWeb Server**  
**D is set of Devices**  
\[ \{d_1, d_2, \ldots, d_n\} \in D \]

IV. There can be number of AC or DC devices which can be controlled by the system we developed.

V. **IP is an IP address of MicroWebServer (MS)**  
**MAC is MAC address of MicroWeb Server (MS)**

VI. Let  
**IP = (Server IP Address)**  
**MAC = (Server MAC Address)**

VII. Connection Begin = set(IP, MAC) \[ \{\text{IP, MAC}\} \in \text{MS} \]

MS is a web server which is started by setting corresponding IP & MAC & controlling the Devices from set D  
\[ \{\text{d}_1, \text{d}_2, \ldots, \text{d}_n\} \in D \]

**Figure 3. Function Mapping Between IC And MS**

**Figure 4. Function mapping between MS and D**
VII. CONCLUSIONS AND FUTURE WORK

Due to enormous usage of internet technology, there is a huge increase in the network attacks. Among them, spam is considered as one of the main attacks in launching various attacks like stealing user identities and spreading malware etc. In this project, a spam detector is developed, which can monitor and detect the machines involved in spam across the network. This tool is based on a spam filtering algorithm that has the efficiency of detecting high percentage of spam. It can differentiate spam and non-spam affect machines in a network of any size. To avoid network administrator to view non-spam emails and to maintain the privacy among the clients in a network, encryption technique is used to encrypt them. The performance is evaluated based on the functionality and results generated with respect to the drawbacks of existing systems using the algorithm. This tool is considered as the light-weight tool because of its minimal amount of time and observations to detect a spam. It can also be used in a network consisting of any number of clients by providing an aggregate large-scale view of the spam in an online manner.

In future work, this tool can be extended to image spam detection as this one is completely based on the content spam filtering. It can be further enhanced by incorporating the sending message service feature to personal contact numbers if the spam exceeds the assumed threshold value. And finally, apart from spam attacks several other attacks can also be focused along with the protective measures.

VII REFERENCES


