ISSN : 2494-9150 Vol-01, Issue 11, FEB 2016.

# Eradication of Spam Mails and Comparison using Percentage and Counter Threshold

**[1]Ashwin Parekh, [2]Mohit Bhat, [3]Kunal Katira**

**[1,2,3]Department of IT, K J Somaiya Institute of Engineering & IT, Sion, Mumbai, Maharashtra, India.**

**[1]ashwin.p@somaiya.edu, [2]mohit.bhat@somaiya.edu, [3]kunal.katira@somaiya.edu**

***Abstract* — In the Real time Internet world ,the infected or Compromised Machine are a real threat to the legitimate user's Account since Hackers are always in a prey of looking into the loop holes or any flaws in the system so that, they can intrude via spam emails to affect legitimate user and steal his confidential data. We have developed an effective tool named "Spam Zombie Eradication System "to deal with such spam problems. This tool uses SPRT algorithm (Sequential Probability Ratio Test) for detecting the spam mails.**

***Keywords— Spam Mails, IP, Filter, Hackers, Eradication System, Zombie.***

## I. INTRODUCTION

In recent times there is a great threat to legitimate user's Email account because of increase in attacks to the account such as Spamming, Phishing and other attacks such DdOS. These attacks are mainly carried out by Hackers, Intruders etc. Spamming is one of such attack, where the attacker (Hackers) sends emails containing spam to legitimate user's account for malicious purposes such confidential data stealing, Hacking etc. In order to prevent user's account from these attacks the tool called "Eradication of spam mails ", this tool mainly detects and monitors the mails of user and detects whether the Mail is legitimate one or whether it is infected (spam) one. This tool mainly makes use of SPRT (Sequential Ratio Probability Test) Algorithm to check whether the mail is a spam or not. Further the tool also compares the results with CT (Counter Threshold) & PT (Percentage Threshold).

## II. RELATED WORK

Zhenhai Duan, Kartik Gopalan, Xin Yuan In this Paper, Zh, KG focused on to develop the antispam techniques. To develop antispam techniques, need the understanding of behavioral characteristics of spammers that distinguishes it from it from senders of non spam messages. The feasibility and effectiveness of CI anti spam techniques affected by the behavioural characteristics of the spammers such as distributions of spam and non-spam messages by spam ratios, statistics of spam messages from different spammers etc. The spam ratio of message sender is a fraction of messages sent by the sender that is spam. Three types of mail servers are defined: spam only mail server which sends only spam messages, non spam mail server only ,sends non spam messages only, mixed mail server sends spam and non spam both messages.

Disadvantages:

1. Majority of spammers are only active for a short period of time. Guofei Gu, Junjie Zhang, and Wenke Lee To identify botnet CC channels in a local area network without any prior knowledge of signature or CC server addresses propose an approach that uses network based anomaly detection. CC servers and infected host in the network identify by this detection approach. This approach is based on the observation that, because of the pre-programmed activities related to CC, bots within the same botnet will likely demonstrate spatial-temporal correlation and similarity. They engage in coordinated communication,

 

propagation, and attack and fraudulent activities. Spatial-temporal correlation in network traffic and utilize statistical algorithms to detect botnets with theoretical bounds on the false positive and false negative rates capture by BotSniffer. The bots of a botnet demonstrate spatial-temporal correlation and similarities due to the nature of their pre-programmed response activities to control commands this observation help use to identify CC within network traffic. Message response and activity response are the two types of responses observable in network traffic.

Advantage:

1. Botsniffer has very promising detection accuracy with very low false positive rate. Disadvantage:

2. Botnet CC traffic is difficult to detect.

Yingfei Dong, Kartik Gopalan, Zhenhai Duan In this paper focusing on unwanted commercial email known as spam. Spam messages cause loss to industry, so many antispam techniques are proposed, email spam filters and sender authentication scheme are included in the technique. DMTP allows receiver, greater control over the message delivery mechanism. It acts as a content filter at the receiver side that scan the contents of message after it has been delivered DMTP is designed based on different form of Receiver-pull model. For designing the DMTP two models are proposed: Sender-push and Receiver-pull. In Sender-push, delivery of traffic is controlled by a sender and receivers just accept whatever sender had sent. In Receiver-pull, it allow receiver to control over system and when they want any data from sender. DMTP (Differentiated mail transfer protocol) which is a pull based model as a counterpart to the spam problem, which grants the control over the message delivery to the receiver.

In the push-based white list and black list are defined along with receiver to determine whether to accept the message.

Advantages:

1. Spam retrieval behaviors of receivers determine the delivery rates of spam.

2. DMTP can easily deploy on internet. Nicholas Ianelli, Aaron Hackworth this paper describes the capabilities present in bot malware and motivations for operating system. To create botnets there are three primary motivators such as communication, resource sharing and curiosity.

The botnets are created by the following techniques: 1.Building from scratch,

2.Social Engineering-which forces user to take the action he or she would not otherwise take,

3.Email attack-In email attack, the user is forced to open an attachment or any link and after opening that link, system is directly infected with malware.

This paper also describes control technologies for botnet such as Web based Command and Control, P2P Command and Control and DNS Command and Control etc. Disadvantage: 1. Botnets are easily built on always on broadband connection.

## III. EXISTING SYSTEM

In this paper, our Aim will be to develop a spam mail detection system, named “Eradication an Detection of spam mails” of spam. It is designed based on a statistical method called Sequential Probability Ratio Test (SPRT) which has bounded false positive and false negative error rates and further results are compared using CT & PT.

## IV.ALGORITHM

### A. Sprt Detection Algorithm:

As the name sprt detection these algorithm is mainly usec for detecting the mail as spam or non-spam. Here we consider H1 as a detection and H0 as normality. That is, H1 is true if the concerned machine is compromised, and H0 is true if it is not compromised. In addition, we let Xi =1 if the ith message from the concerned machine in the network is a spam, and Xi = 0 otherwise.

 

Algorithm:

1: An outgoing message arrives at SPOT

2: Get IP address of sending machine m

3: // all following parameters specific to machine m

4: Let n be the message index

5: Let Xn = 1 if message is spam, Xn = 0 otherwise

6: if (Xn == 1) then

7: // spam, Eq. 3

8: $A_n$ += $ln(Ø_1/Ø_2)$

9: else

10: // nonspam

11: $A_n$ += $ln(1-Ø_1/1-Ø_2)$

12: end if

13: if $(A_n \geq B)$ then

14: Machine m is compromised. Test terminates for m.

15: else if $(A_n \leq A)$ then

16: Machine m is normal. Test is reset for m.

17: $A_n = 0$

18: Test continues with new observations

19: else

20: Test continues with an additional observation

21: end if

### B. CT and PT Algorithms

1) Counter threshold and Percentage threshold are two more algorithms are used based on the number of spam messages received at the legitimate user's account and the percentage of spam messages sent from the internal machine, referred as Count Threshold detection algorithm and Percentage Threshold detection algorithm.

2) To check if machine is compromised, we have to monitor messages. Hence, here we are using CT to calculate total no of spam messages. If this count exceeds the value 30 then will calculate the percentage in PT. And if we get the result above 50 percent then that machine will be declared as compromised.

## V. SYSTEM ARCHITECTURE

The system architecture of "Eradication of spam Mails System" consists of various components such as:

- Spam filter
- A compromised or infected machine
- IP address capture tool

The block diagram of "Eradication of spam Mails system" with its various components is shown below.

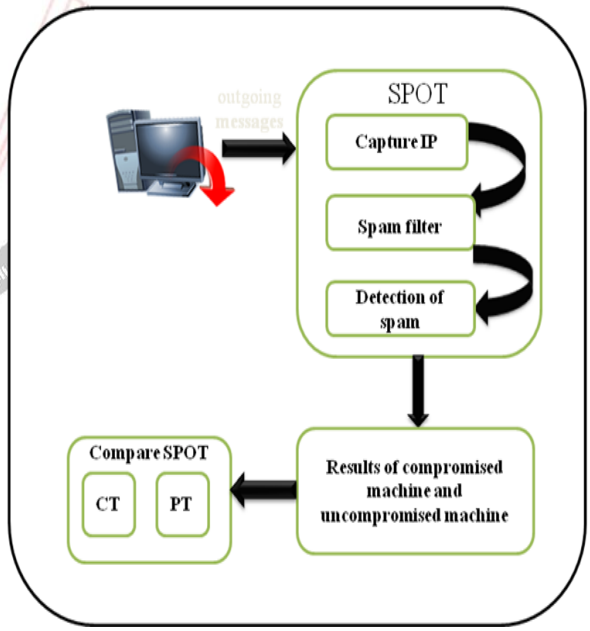


**Fig. 1 Eradication of Spam Mails System**

 

## VI . MODULE DESCRIPTION

**LIST OF MODULES:**

1) Account authentication
2) Sending mails
3) SPOT detection
   i.capture IP
   ii.SPOT filter
   iii.SPOT results
4) CT detection.
5) PT detection

**MODULES DESCRIPTION:**

***1) Account authentication***

In this module to check the mail id and password. If these two fields are valid, the account is authenticated. Otherwise is not valid.

***2) Sending mails***

In this module a single person to send one or more mails to other person. This mails either spam or non spam. Spam means the more copies of the single message are send. And it contains more than 20 lines.

3) ***SPOT detection***

In this module to capture the IP address of the system. That system mails are applied to filtering process. In this process, the mail content is filtered. Finally to produce the result of filter.

***4) CT detection***

In this module to set the threshold value $C_s$ .$C_s$ denotes the fixed length of spam mail. Also to count the number of lines in each mail. If the each mail, counts are greater than equal to threshold value. So, these mails are spam mail.

5) ***PT detection***

In this module to set two threshold values. $C_a$- specifies the minimum number of mail that machine must send. 2) P- specifies the maximum spam mail percentage of a normal machine. This algorithm is used to compute the count of total mails and the count of spam mails of machine. To check this count of total mails are greater than equal to Cs and the count of spam mails are greater than equal to P. If it's true these mails are spam mail.

## VII. CONCLUSION

We are currently working on the tool "Eradication of spam mails" .This tool will enable legitimate user to access his/her mail account effectively without the interference of spam mail that are sent by the intruder for malicious purposes. This tool will make use of strong test tool named SPRT (Sequential Probability Ratio Test) for detecting the spam mails .Also the observations are compared with the help of CT(counter threshold) and PT (percentage threshold).

## REFERENCES

[1] Zhenhai Duan, Kartik Gopalan, and Xin Yuan: *"An Empirical Study of Behavioral Characteristics of Spammers: Findings and Implications**".

[2] G.Gu, J.Zhang, and W.Lee. *"BotSniffer: Detecting botnet command and control channels in network traffic"*. In Proceedings of the 15th Annual Network and Distributed System Security Symposium (NDSS 2008), SanDiego,CA, Feb. 2008.

[3] Zhenai Duan, Y. Dong, and Kartik. Gopalan. *"DMTP: Controlling spam through message delivery differentiation*". Computer Networks (Elsevier), July 2007

[4] N.Ianelli and A.Hackworth. "*Botnets as a vehicle for online crime*". In Proc. of First International Conference on Forensic Computer Science, 2006.

[5] P.Bacher,T.Holz,M.Kotter and G.Wicherski."Know your enemy: Tracking botnets".http://www.honeynet.org/paper/bots.

[6] J.Markoff.Russian gang hijacking PCs in vast sceme."The New Yark Times, Aug 2008" http://www.nytimes.com/2008/08/06/technology/06hack.htmll

 