Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 04, April-2017

Detection of Spam Zombies

Harsh Agrawal
Department of Information Technology
Sardar Patel Institute of Technology
Andheri(W), Mumbai.

Sneha Malani
Department of Information Technology
Sardar Patel Institute of Technology
Andheri(W), Mumbai.

Akriti Bhat
Department of Information Technology
Sardar Patel Institute of Technology
Andheri(W), Mumbai.

***Abstract*—This paper presents a 'Spam Zombie Detection' system which is an online system over the network that detects the spam and the sender of the spam (zombie) before the receiver receives it. Thus all the detection work is done at sender level itself. This paper focuses on a powerful statistical tool called Sequential Probability Ratio Test, which has bounded false positive and false negative error rates on which the Spam Zombie Detection system is based. This system is mainly implemented over the private mailing system. It also provides the enhanced security mechanism in which, if the system which has been hacked i.e. it has become a zombie, then it gets blocked within the network.**

***Keywords*— zombie, Spam Zombie Detection, sequential probability ratio test.**

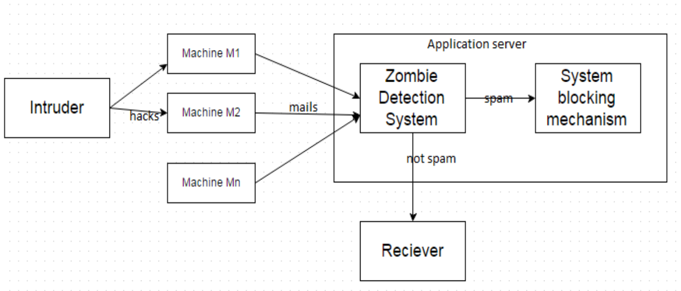
I. INTRODUCTION

In recent years, the increasing use of e-mail has led to the emergence and further escalation of problems caused by unsolicited bulk e-mail messages, commonly referred to as Spam. Evolving from a minor nuisance to a major concern, given the high circulating volume and offensive content of some of these messages, Spam is beginning to diminish the reliability of e-mail [1]. Personal users and companies are affected by Spam due to the network bandwidth wasted receiving these messages and the time spent by users distinguishing between Spam and normal (legitimate or ham) messages. Dealing with spam and classify it is a very difficult task. A single model for classifying spam is also a difficult task since new spams are constantly evolving. Further, spammers often actively tailor their messages to avoid detection adding further impediment to accurate detection. One such mechanism is to send the spams through some third party, which are often called as Zombies [2].

Zombie is defined as a compromised machine within the botnet. Compromised machines are one of the key security threats on the Internet; they are often used to launch various security attacks such as DDoS, spamming, and identity theft. Hence there is the need to detect such compromised machines and prevent such attacks. There are various algorithms present for detecting spam zombies.

In this paper, we have focused on detection of such spam zombies using the Sequential Probability Ratio Test (SPRT). Previously, the algorithms that were used for spam detection were Counter Threshold (CT) and Percentage Threshold (PT). But for these two mechanisms the false positive and false negative rates were quite high. Thus the proposed system uses the SPRT algorithm. This paper also presents a distinguished comparison between the three mechanisms.

The general architecture of the proposed system is given below:



II. DETECTING SPAM ZOMBIES

In this section, we have discussed the three approaches for detecting spam zombies –

1. Using SPRT algorithm,
2. Using Counter Threshold,
3. Using Percentage Threshold.

1. Sequential Probability Ratio Test

In its simplest form, SPRT is a statistical method for testing a simple null hypothesis against a single alternative hypothesis. In essence, SPRT is a variant of the traditional probability ratio tests for testing under what distribution (or with what distribution parameters), it is more likely to have the observed samples. However, unlike traditional probability ratio tests that require a pre-defined number of observations, SPRT works in an online manner and updates as samples arrive sequentially. Once sufficient evidence for drawing a conclusion is obtained, SPRT terminates [3].

As a simple and powerful statistical tool, SPRT has a number of compelling and desirable features that lead to the wide- spread applications of the technique in many areas. First, both the actual false positive and false negative probabilities of SPRT can be bounded by the user- specified error rates. This means that users of SPRT can pre-specify the desired error rates. A smaller error rate tends to require a larger number of observations before SPRT terminates. Thus users can balance the performance and cost of an SPRT test. Second, it has been proved that SPRT minimizes the average number of the required observations for reaching a decision for a given error rate, among all sequential and non-sequential statistical tests. This means that SPRT can quickly reach a conclusion to reduce the cost of the corresponding experiment, without incurring a higher error rate.

SPRT algorithm:

In the context of detecting spam zombies in SPRT, we consider two hypothesis, 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, let Xi = 1 if the message from the concerned machine in the network is a spam, and Xi = 0 otherwise [4]. SPRT requires four configurable parameters from users, namely, the desired false positive probability, the desired false negative probability, the probability that a message is a spam when H1 is true and the probability that a message is a spam when H0 is true. Users configure the values of the four parameters. Based on the user-specified values of α and β, the values of the two boundaries A and B of SPRT are computed.

Algorithm: Spam Zombie Detection System [3]

```
An outgoing message arrives at Spam Zombie
Detection System.
Get IP address of sending
machine m. Let n be the
message index.
Let Xn = 1 if message is spam, Xn = 0
otherwise if ( Xn == 1) then
   spam
   An+=ln(
   ɵ1/ɵ0)
else
   nonspam
   An+=ln(1-
   ɵ1/1-ɵ0)
end if

if (An ≥ B) then
   Machine m is compromised. Test terminates for
m. else if (An ≤ A) then
   Machine m is normal. Test is reset
   for m. An = 0
   Test continues with new
observations else
   Test continues with an additional
observation end if
```

When an outgoing message arrives at the spam zombie detection system, the sending machine’s IP address is recorded, and the message is classified as either spam or non spam by the spam filter. For each observed IP address, it maintains the logarithm value of the corresponding probability ratio “An” whose value is updated according to as message n arrives from the IP address. Based on the relation between An and A and B, the algorithm determines if the corresponding machine is compromised, normal, or that nothing can be concluded. To note that in the context of spam zombie detection, from the view point of network monitoring it is more important to identify the machines that have been compromised than the machines that are normal. After a machine is identified as being compromised it is added into the list of potentially compromised machines that system administrators can go after to clean. The message-sending behavior of the machine is also recorded, if in case further analysis is required. Before the machine is cleaned and removed from the list, the detection system does not need to further monitor the message sending behavior of the machine. A machine that is currently normal may get compromised at a later time. Therefore, it needs to continuously monitor the machines that are determined to be normal by SPRT. Once such a machine is identified, the records of the machine in SPRT are re-set, in particular, the value of An is set to zero, so that a new monitoring phase starts for the machine. It requires four user defined parameters: α, β, ɵ1, and ɵ0. [3] [5].

2. Counter Threshold (CT):

Here we need to set the threshold value Cs where Cs denotes the fixed length of spam mail. If counts are greater than or equal to the threshold value then the mails are spam mails and the machine is compromised. [3] [4].

3. Percentage Threshold (PT):

In this we need to set two thresholds values: Ca - specifies the minimum number of mail that machine must send and 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. Now check if this count of total mails are greater than or equal to Cs and the count of spam mails are greater than or equal to P. If it's true these mails are spam mail and the machine is compromised.[3] [4] [5].

III. VARIOUS SPAM FILTERING ALGORITHMS

There have been several approaches to identify if the incoming messages are spams or not. Some of them are defined below:

1. Whitelist/Blacklist:

This approach simply creates a list. A whitelist is a list which includes the email addresses or entire domains which the user knows. An automatic white list management tool is also used by user that helps in automatically adding known addresses to the whitelist. A blacklist is the opposite of whitelist. In this list we add addresses that are harmful for users [6][7].

2. Mail Header Checking:

This approach is a well known approach. It simply consists of set of rules that we match with mail headers. If a mail header matches, then it triggers the server and return the mails that have empty "From" field, that have too many digits in address, that have different addresses in "To" field from same source etc [6][7]

3. Content based SpamFiltering:

The basic format of e-mail generally consists of the following sections:

- Header section includes the sender email address, the receiver email address, the Subject of the email and
- The Content of the email includes the main body consisting of images, pictures, texts and other multimedia data

In content based spam filtering, the main focus is on classifying the email as spam or as ham, based on the data that is present in the body or the content of the mail. However, the header section is ignored in the case of content based spam filtering. There are number of techniques such as Bayesian Filtering, Word-based Filtering, Heuristic approach, AdaBoost classifier, Gary Robinson technique, KNN classifier, etc [6] [8].
The proposed system makes use of Bayesian Filtering technique for detecting and filtering the spam messages.

IV. BAYESIAN FILTERING

There are particular words used in spam emails and non spam emails. These words have particular probability of occurring in both emails. The filters that we used don't know these probabilities in advance; we must train them first so it can build them up. After training the word probabilities are used to compute the probability that an email having particular set of words in it belong to either spam or legitimate emails. Each particular word or only the most interesting words contribute to email's spam probability. This contribution is known as the posterior probability and is computed using Bayes' theorem. Then, the emails spam probability is computed all over the word in the emails. If this total value exceed over certain threshold then the filters will mark emails as spam.

Bayesian inference uses aspects of the scientific method, which involves collecting evidence that is meant to be consistent or inconsistent for a given hypothesis. Bayesian inference uses a numerical estimate of the degree of belief in a hypothesis before evidence has been observed and calculates a numerical estimate of the degree of the belief in the hypothesis after evidence has been observed [9] [10].

Bayes theorem as shown in Eq.1, relates the conditional and marginal probabilities of stochastic events A and B [11] :

$$P(A/B) = P(B/A)\,P(A) / P(B) \;\ldots\ldots\ldots\ldots\; (1)$$

P (A) is the prior probability or marginal probability of A. Prior in the sense that it doesn't take into account any information about B.

P(A/B) is conditional probability of A, given B P(B/A) is conditional probability of B, given A P(B) is the prior or marginal probability of B

```
for all arriving messages m do-
Class = classification of m by auxiliary detection
method; sc = find cluster for m.sender;
Update spam probability for sc using
mClass; Ps(m) = spam probability for
sc;
Pr(m) = 0;
for all recipient r in m,
recipients do rc = find cluster
for r;
```

```
Update spam probability for rc using mClass;
Pr(m) = Pr(m) +spam probability for rc;
end for
Pr(m) = Pr(m)/size (m. recipients)
SP (m) = compute spam rank based on Ps(m) &
Pr(m); if SP (m) > w then
classify m as spam;
else if SP(m) < 1 − w
then classify m as
legitimate; else
classify m as
mClass; end if
end for
```

V. IMPACT OF DYNAMIC IP ADDRESSES

However in all these three zombies detecting mechanisms, we have explicitly assumed the IP address to be constant. From the view of dynamic IP addresses, each of the zombie detection as well as the spam filtering algorithms performs differently. But for simplicity, we have ignored the potential impact of dynamic IP addresses and did assume that an observed IP corresponds to a unique machine.

VI. CONCLUSION

The proposed system detects the spam mails by monitoring the outgoing mails using Bayesian Filtering. In order to detect the spam zombies, the proposed system uses the Sequential Probability Ratio Test algorithm. This proposed system also provides the blocking mechanism in which if the system is identified as the spam zombie then the system gets blocked so that it cannot send the spam messages further.

VII. FUTURE WORK

The current system works only for non-dynamic IP addresses. In future, a spam detection system considering the dynamic nature of IP addresses can be implemented or capturing MAC address when used for an internal network.

Also, a feedback mechanism is possible to implement, where the system will notify the compromised machine that it has become a zombie.

VIII. REFERENCES

[1] B. Hoanca, “How good are our weapons in the spam wars”, IEEE Technology and Society Magazine, pp 22–30, 2006.

[2] Geerthik S.,” Filtering Spam: Current Trends and Techniques”, International Journal of Mechatronics, Electrical and Computer Technology Vol. 3(8), Jul, 2013, pp 208-22.

[3] Chen, Peng, "Detecting Spam Zombies By Monitoring Outgoing Messages" (2008). Electronic Theses, Treatises and Dissertations. Paper 3844.

[4] “Detecting Spam Zombies Using Sprt Tool By Monitoring Outgoing Messages”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 4, April 2013, ISSN: 2277 128X.

[5] “Identifying SPAM Information by Screening Outgoing Messages”, International Journal of Computer Science And Technology, IJCST Vol. 4, Issue Spl - 4, Oct - Dec 2013.

[6] “A Review on Different Spam Detection Approaches”, International Journal of Engineering Trends and Technology (IJETT) – Volume 11 Number 6 - May 2014, ISSN: 2231-5381.

[7] Amol H. “Survey Paper on Intelligent System for Text and Image Spam Filtering”, International Journal of Computer Engineering and Applications, Volume IX, Issue I, January15, www.ijcea.com, ISSN 2321- 3469

[8] “Content-Based Spam Filtering and Detection Algorithms- An Efficient Analysis & Comparison”, International Journal of Engineering Trends and Technology (IJETT) – Volume 4 Issue 9- Sep 2013, ISSN: 2231-5381.

[9] Leonard and Hsu, 2001. Bayesian methods: an analysis for statisticians and interdisciplinary researchers. Cambridge University Press, Cambridge.

[10] Krushna P, Savyasaachi P, “Blog-Spam Detection Using intelligent Bayesian Approach”, International Journal of Engineering Research and General Science, Volume 2, Issue 5, August-September, 2014.

[11] “Email classification for Spam Detection using Word Stemming”, 2010 International Journal of Computer Applications (0975 – 8887), Volume 1 – No. 545.