ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2015

A Study Paper on Detecting Spam Zombies

Jiten Shah, Satish Singh, Vikram Singh, Prof. Vijaya Sagvekar

Student Member (Comp), Atharva College of Engineering, Mumbai, India

Assistant Professor (Comp), Atharva College of Engineering, Mumbai, India

ABSTR*ACT:* In this paper we aim to detect the compromised machines in a network that are involved in the spamming activities, widely known as spam zombies. Given that spamming provides a key economic incentive for attackers to recruit the large number of compromised machine, we develop a spam zombie detection system named SPOT by keeping track of outgoing messages in a network. SPOT is designed based on a powerful statistical tool called Sequential Probability Ratio Test (SPRT), which has bounded false positive and false negative error rates.

KEYWORDS: spam zombies, compromised machine

I. INTRODUCTION

Nowadays the email spam problem becomes more and more serious issue. Spam not only causes the misuse of time and computational resources, thus leading to financial losses, but it is also often used to advertise illegal goods and services or to promote online frauds. The most popular way of anti-spam detection is Spam filtering. A Spam filter is a program that is used to detect unsolicited and unwanted email and prevent those messages from getting to user’s inbox. The primary challenge of spam detection lies in the fact that spammers will always find new ways to attack spam filters owing to the economic benefits of sending spams. Spammers have no choice to but to send out large quantities of identical or similar spam’s storage size of e-mail abstraction should be small. simultaneously to make profits.

The Internet is growing at a very rapid speed and multiplying its websites in every 30 days and the number of people using the Internet is growing day by day. Hence, Global communication is playing an important role in day-to-day life. At the same time, computer crimes are also increasing. Counter measures are developed to detect or prevent attacks - most of these measures are based on known facts, known attack patterns.

These types of the spam messages in the internet are generally stored in the form of the bots, it is controlled by the term is called as the botnet. The Botnet controllers make use of the technologies such as the IRC channels to handle and manage these bots. Botnets contain manifold wrong uses: growing DDoS attacks, theft password of the each user and individuality, create tick fraud [1], and basis spam email [2]. This technology source damages the spam email, everywhere spam is summarize kindly to include earliest public relations email messages.

II. RELATED WORK

Existing approach mainly depend upon two area One is effectively detecting the spam mails from the outgoing mails network/system. Commonly the Botnet attacks are showing some common characteristics. These studies provided important insights into the aggregate global characteristics of spamming botnets by clustering spam messages received at the provider into spam campaigns using embedded URLs and near-duplicate content clustering, respectively [3]. The common approaches to the botnet attacks are quite different they are not consider about healing from a single network they only consider about whole network. They try to only in detect the spamming activity at only in the receiving side. However, their approaches are better suited for large email service providers to understand the aggregate global characteristics. In this paper we mainly concentrate to make an administrative tool for detecting the compromised machine in that network SPOT is a light-weight spam zombie detection system; it does not need the support from the network intrusion detection system as required by Bot Hunter [4] As a simple and powerful statistical method, Sequential Probability Ratio Test (SPRT) [5] has been successfully applied in many areas. In the area of networking security, SPRT has been used to detect port scan activities proxy-based spamming activities and MAC protocol misbehavior in wireless networks.

10.15680/ijircce.2015.0302006

III. PROBLEM FORMULATION

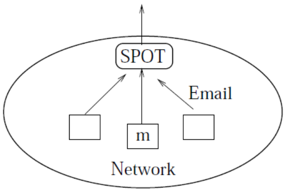


Figure 1: SPOT system design[1]

Problem formulation deals with training the machine in the network based on unsupervised machine learning to plot the different cluster among the email is zombie spam or not, here the sequential probability testing plays a vital role in classifying the test data in to zombie spam or not and then training the system to clearly identify the zombie spam [6]. Here we have a trained set of data and the emails which were successfully classified by spot using the sequential probability testing In this section we formulate the spam zombie detection problem in a network. In particular, we discuss the network model and assumptions we make in the detection problem. Giving the logical view of network this network that consists number of pc may be compromised or not. We assume that the messages are will be originated from this network. Each outgoing messages are passed through the spot detection system.

IV. SPRT

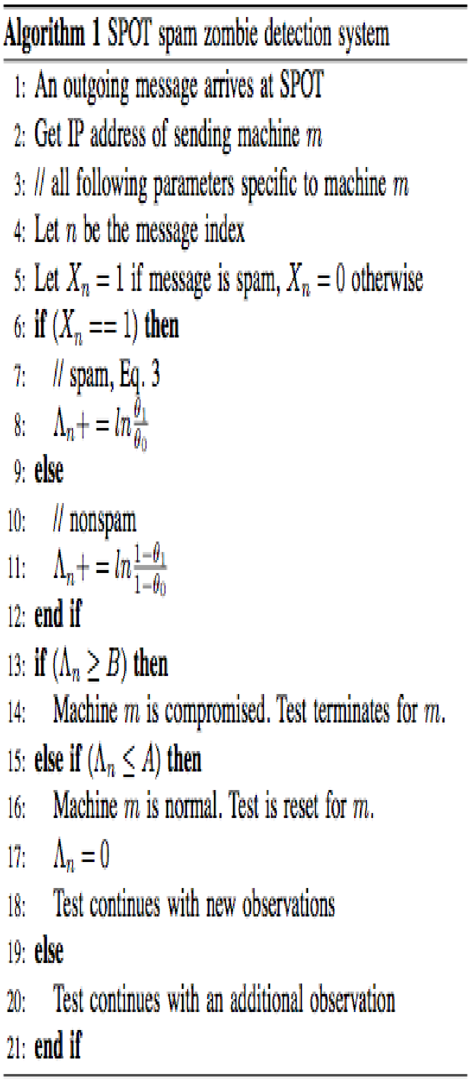
In this section we provide the necessary background on the Sequential Probability Ratio Test (SPRT) for understanding the proposed spam zombie detection system. Interested readers are directed to [7] for a detailed discussion on the topic of SPRT. In its simplest form, SPRT is a statistical method for testing a simple null hypothesis against a single alternative hypothesis. Intuitively, SPRT can be considered as an one- dimensional random walk with two user-specified boundaries corresponding to the two hypotheses. As the samples of the concerned random variable arrive sequentially, the walk moves either upward or downward one step, depending on the value of the observed sample. When the walk hits or crosses either of the boundaries for the first time, the walk terminates and the corresponding hypothesis is selected. 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. 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 [8]. 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 (in terms of false positive and false negative rates) and cost (in terms of number of required observations) 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. In the following we present the formal definition and a number of important properties of SPRT. The detailed derivations of the properties can be found in [7].

V. SPOT DETECTION ALGORITHM

SPOT is designed based on the statistical tool SPRT we discussed in the last section. In the context of detecting spam zombies in SPOT, 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. SPRT requires four configurable parameters from users, namely, the desired false positive probability α, the

desired false negative probability β, the probability that a message is a spam when H1 is true (θ1), and the probability that a message is a spam when H0 is true (θ0). We discuss how users configure the values of the four parameters after we present the SPOT algorithm. Based on the user-specified values of α and β, the values of the two boundaries A and B of SPRT are computed.

VI. WORKING MODEL OF PROPOSED SYSTEM



Algorithm 1 SPOT spam zombie detection system

```
An outgoing message arrives at SPOT
Get IP address of sending machine $m$
// all following parameters specific to machine $m$
Let $n$ be the message index
Let $X_n = 1$ if message is spam, $X_n = 0$ otherwise
if ($X_n == 1$) then
    // spam, Eq. 3
    $\Lambda_n + = ln\frac{\theta_1}{\theta_0}$
else
    // nonspam
    $\Lambda_n + = ln\frac{1-\theta_1}{1-\theta_0}$
end if
if ($\Lambda_n \geq B$) then
    Machine $m$ is compromised. Test terminates for $m$.
else if ($\Lambda_n \leq A$) then
    Machine $m$ is normal. Test is reset for $m$.
    $\Lambda_n = 0$
    Test continues with new observations
else
    Test continues with an additional observation
end if
```

Architecture and working proposed system has following which help to train the machine by collecting the data. system has a spot detection circuit which helps to classify the mail in to a Zombie or not which is a machine learning environment which is unsupervised .each mail that is outgoing from the mail server on which the tool is deployed is being captured .Each mail is being cross checked by the spam detection system .spam findings are recorded in to a database for the use of training the statistical machine in the mail server .then mail count is being retrieved to find out the count of the number of spam in fact present in the total number of outgoing email ,then calculation of value is being performed which help to analyze the threshold value ,the trained system with the help of SPRT and SPOT detection compares the email with the threshold value and the trained value of the system to predict it as a spam zombie or not. Comparison is performed and the machine is compromised with the learning process.

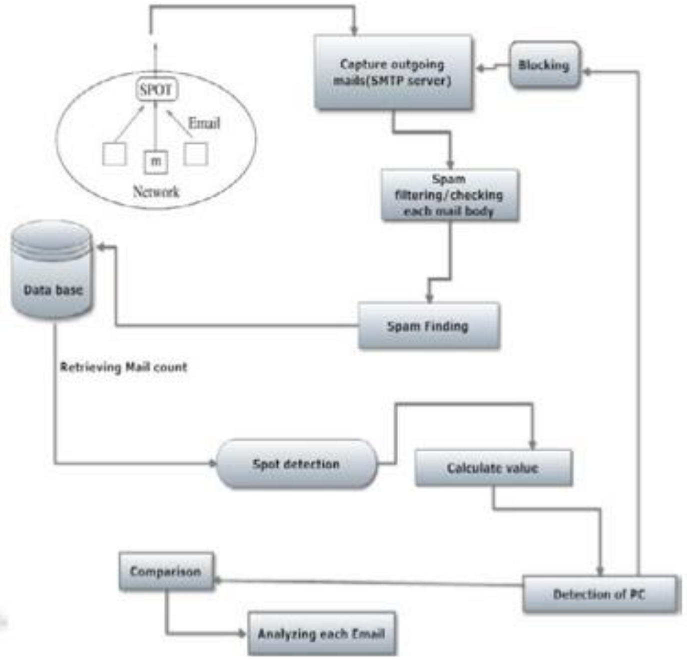


Figure 2: Working Model

VII. CONCLUSION

Spam zombie becomes one of the most important internet security problems in nowadays, it ever-increasing daytime extremely quickly. So the detection of the spam messages also plays major important to detect spam zombies for email communication, it becomes important to classify the messages into spam and non-spam messages In order to perform this we developed a system for detecting compromised machines in a network named SPOT by monitoring outgoing message. SPOT was designed based on a simple and powerful statistical tool named Sequential Probability Ratio Test to detect the compromised machines that are involved in the spamming activities. SPOT has bounded false positive and false negative error rates. It also minimizes the number of required observations to detect a spam zombie.

REFERENCES

[1] Daswani, N., Stoppelman, M., “The Google click quality and security teams”, The anatomy of clickbot.a. In HotBots'07, 2007.

[2] Ramachandran, A., and Feamster, N,” Understanding the network-level behavior of spammers”,[In SIGCOMM'06, 2006.

[3] M. Xie, H. Yin, and H. Wang. An effective defense against email spam laundering. In ACM Conference on Computer and Communications Security, Alexandria, VA, October 2006. 2

[4] G. Gu, P. Porras, V. Yegneswaran, M. Fong, and W. Lee. Bothunter: Detecting malware infection through ids-driven dialog correlation. In Proc. 16th USENIX Security Symposium, Boston, MA, Auguest 2007. 2

[5] Y. Xie, F. Xu, K. Achan, R. Panigrahy, G. Hulten, and I. Osipkov. Spamming botnets:signatures and characteristics. In Proc. ACM SIGCOMM, Seattle, WA, August 2008.1, 2

[6] S. Radosavac, J. S. Baras, and I. Koutsopoulos. A framework for mac protocol misbehavior detection in wireless networks.In Proceedings of the 4th ACM workshop on Wireless security, Cologne, Germany, September 2005. 2

[7] A. Wald. *Sequential Analysis.* John Wiley & Sons, Inc, 1947.

[8] G. B. Wetherill and K. D. Glazebrook. *Sequential Methods inStatistics.*