Detecting and Blocking of Spam Zombies Mechanism

Prof. V M Patil[1] Mankar Aarti Rajendra[2] Sardeshpande Sandeep Padmakarrao[3] Shirohiya Mayur Ramesh[4] Thigale Shital Dasharath[5]

[1,2,3,4,5]Department of Computer Engineering

[1,2,3,4,5]SPPU, Pune

Abstract— A zombie is a computer connected to the internet that has been compromised by a hacker, computer virus or Trojan horse and can be used to perform malicious tasks of one sort or another under remote direction. Botnets of zombie computers are often used to spread e-mail spam and launch denial-of-service attacks. Most owners of zombie computers are unaware that their system is being used in this way. Because the owner tends to be unaware, these computers are metaphorically compared to zombies. These compromised machines send a lot of spam messages on the internet. Such machines result in spamming attacks, DDOS attacks, identity theft which result in different kind of losses to the victim. Spamming botnets is the network of compromised machines involved in spamming. The SPOT, the sequential probability ratio test is used for detecting the compromised machines. SPRT is used since the error rate produced is infinitesimally small and the number of observations required to deciding whether a machine is compromised or not is also small. It helps in observing the outgoing messages from a machine in a network. Out of a large number of machines in a network only a few of them are not compromised. For an instance, out of 440 internal IP addresses SPOT identifies 132 of them as being compromised. This system has been developed for system administrators for monitoring the machines in a network.

Key words: zombie, SPRT, percentage-threshold (PT) detection algorithm

I. INTRODUCTION

The roll cage (frame) is supporting component of automobile vehicle. It is the foundation for carrying the engine, transmission system and steering by means of spring, axles, rubber pads etc. The frame are made of box, tubular channels or U-shaped section, welded or riveted together. A roll cage is a safety feature installed in a vehicle used in environments where there is a high danger of rolling, such as race car driving as well as military and police use. Some cars are specifically designed with this feature installed, while others have had this device installed during a retrofit. The points which were considered while designing the roll cage were safety, ergonomics, market availability, Compromised machines(Machine use to send spam message) are one of the important security threats on the Internet. A major security challenge on the Internet is the existence of the large number of compromised machines. Such machines have been increasingly used to launch various security attacks including DDoS, spamming and identity theft. Identifying and cleaning compromised machines in a network remain a significant challenge for system administrators of networks of all sizes. we focus on the detection of the compromised machines in a network that are involved in the spamming activities, commonly known as spam zombies. We develop an effective spam zombie detection system by monitoring outgoing messages of machines. This system designed based on a powerful statistical tool called Sequential Probability Ratio Test. In this system we are monitoring the outgoing massages of sender machine and capture the IP Address of the sender machine. The nature of sequentially observing outgoing messages gives rise to the sequential detection problem. We designs detecting spam zombies, one based on the number of spam messages and another the percentage of spam messages sent from a machine, respectively. For simplicity, we refer to them as the count-threshold (CT) detection algorithm and the percentage-threshold (PT) detection algorithm respectively.

Also stores senders detail information such as Email Id of sender, number of files send by the sender, message text of sender etc. In this system we define the Spam messages threshold i.e. the number of Spam message or Spam files send from a particular machine. In this system we block the Spam senders Email Id reason behind that is one machine is used by many user. In this system we calculate the Count threshold and Percentage threshold .(value of CT and PT) CT is the number of Spam messages send from a machine by user.PT is the percentage of Spam message send by sender.PT is the ratio of Spam messages to the total number of messages into 100 ,send from a particular machine, from a particular user.

We calculate value of CT and PT for each sender i.e. for each Email Id we calculate value of CT and PT separately. we block senders Email Id if the PT value is greater or equals to 50%. Once we block user he/she can't do login. Instead of searching the aggregate global characteristics of spaming botnets (network), we aim to develop a tool for system administrators to automatically detect the compromised machines in their networks in an online manner.

II. CONCEPTS

Compromised machines (Machine use to send spam message

1) If the PT value is > 50% then the machine is consider as compromised machine.
2) In this project, we will develop a spam zombie (Compromised machines) detection system ,by monitoring outgoing messages. This system is designed based on a statistical method called Sequential Probability Ratio Test (SPRT).Spam Count and Percentage-Based Detection Algorithms:
3) We present two different algorithms in detecting spam zombies (Compromised machines), one based on the number of spam messages and another the percentage of spam messages sent from an machine, respectively. We refer to them as the count-threshold (CT) detection algorithm and the percentage-threshold (PT) detection algorithm.

III. MATHEMATICAL FORMULATION

- N: Total no of spam messages send from a machine.
- CT: number of spam messages.
- PT:CT/N

IV. LITERATURE SURVEY

In this chapter we provide the necessary background on the Sequential Probability Ratio Test (SPRT) for understanding the proposed spam zombie detection system. 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. 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. 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. We use the concept of SPRT and design a spam zombie detection system in which we define threshold i.e. number of spam messages send by machine. We design detecting spam zombies, one based on the number of spam messages and another the percentage of spam messages sent from a machine, respectively. For simplicity, we refer to them as the count-threshold (CT) detection algorithm and the percentage-threshold (PT) detection algorithm respectively. For each user we generate CT and PT and capture IP Address for each sender machine. Here we generate IP Address using random function.

V. SPAM COUNT AND PERCENTAGE

For comparison, in this section we present two different algorithms in detecting spam zombies, one based on the number of spam messages and another the percentage of spam messages sent from user machine, respectively. For simplicity, we refer to them as the count-threshold (CT) detection algorithm and the percentage-threshold (PT) detection algorithm, respectively. In CT, the time is partitioned into windows of fixed length *T*. A user-defined threshold parameter *Cs* specifies the maximum number of spam message that may be originated from a normal machine in any time. The system monitors the number of spam messages *n* originated from a machine in each window. If $n > Cs$, then the algorithm declares that the machine has been compromised. Similarly, in the PT detection algorithm the time is partitioned into windows of fixed length *T*. PT monitors two email sending properties of each internal machine in each time window: one is the percentage of spam messages sent from a machine, another the total number of messages. Let *N* and *n* denote the total messages and spam messages originated from a machine *m* within a time window, respectively, then PT declares machine *m* as being compromised if $N \geq Ca$ and $n/N > P$, where *Ca* is the minimum number of messages that a machine must send, and *P* is the user-defined maximum percentage of a normal machine. The first condition is in place for preventing high false positive rates when a machine only generates a small number of messages. For example, in an extreme case, a machine may only send a single message and it is a spam, which renders the machine to have a 100%spam ratio. However, it does not make sense to classify this machine as being compromised based on this small number of messages generated. In the following we briefly compare the two spam zombie detection algorithms CT and PT with the SPOT system. The three algorithms have the similar running time and space complexities. They all need to maintain a record for each observed machine and update the corresponding record as messages arrive from the machine. However, unlike SPOT, which can provide a bounded false positive rate and false negative rate, and consequently, a confidence how well SPOT works, the error rates of CT and PT cannot be a priori specified. In this system value of CT and PT calculate dynamically for respective IP Address.

VI. EXISTING SYSTEM

There are some existing systems for the spam Zombie detection which are explained as follows:

A. Detecting Spam Zombies by Monitoring Outgoing Messages:

There is need to control the existing compromised systems over the network that perform the various security attacks. This paper mainly focuses on the detection of the compromised machines that send the spam messages which are also known as spam zombies. This system does not require the spamming global characteristics such as the size of the botnets and the spamming patterns of the botnets. This system has tool with the help of which an administrator can detect the compromised machines automatically. Thus this system is known as an online botnet detection system. Here the name given to this spam zombie detection system is SPOT system which monitors the outgoing messages. The statistical method called Sequential Probability Ratio Test (SPRT) is used to design the SPOT system. The SPRT method is used to test the two hypotheses which the machine is compromised and the machine is not compromised [1]. This tool helps to minimize the expected number of observations used to take the decision. Here the user can define the threshold limit for the false positive and false negative probabilities required by the SPRT method. Thus the SPOT system can quickly identify the spam zombies within the network [1].

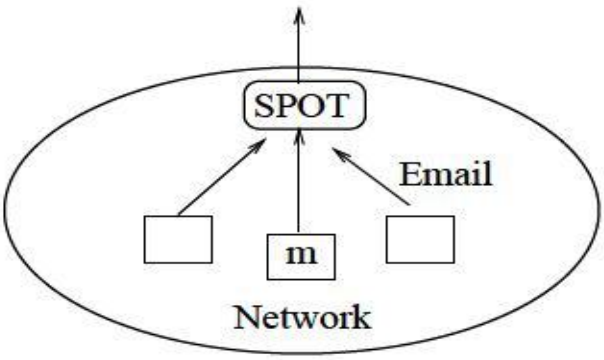

Fig. 1: Network Model for SPOT system

B. Botminer: Data Adaptive Clustering Analysis For Online Botnet Detection

It is an online botnet detection technique. As the offline botnet detection strategies depend upon some historical data that may change over the time leading to the difficulties in the detection process. So the Botminer system was developed which gathers the current botnet activities. First the system gathers the network traffic and then forms the multidimensional features streams from them. Then these feature streams are grouped with high similarity by novel data adaptive cluster operation. No re clustering operation is required as all the clusters are updated periodically. The clusters are considered to be as bots whose feature streams are highly similar [2].

C. BotSniffer: Detecting Command and Control Channels in the Network

BotSniffer is a network based anomaly detection technique which can be also used as online botnet detection technique. It identifies the botnet C&C channels in Local Area Network. It does not require any prior knowledge of signatures or C&C server addresses. This approach identifies both C&C servers and infected hosts in network. It detects the Bots within the botnets that have spatial temporal correlation and similarity. Thus the BotSniffer technology is developed as a system with many real world network traces. This technology follows the normal protocol usage and is similar to the normal traffic. Sometimes the traffic that it handles may have low volume. If the number of the bots within the network goes on increasing there it may find difficulty in detection process [3].

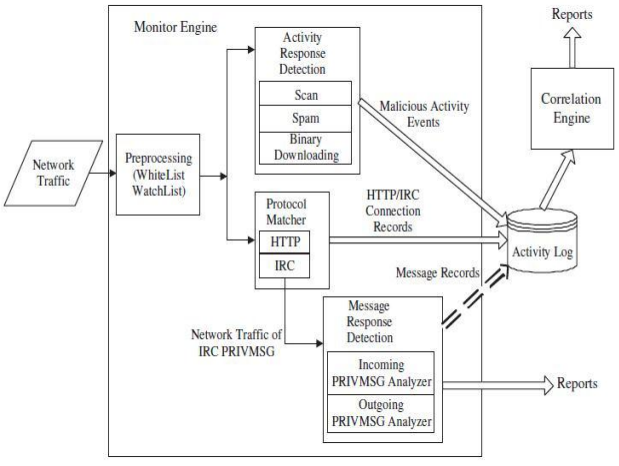

Fig. 2: BitSniffer Architecture

D. BotHunter: Detecting Malware Infection through IDS-Driven Dialog Correlation

BotHunter is a "Dialog Correlation Strategy" which initially recognizes the infection and then coordinates the dialog that occurs during the successful malware infection. It manages the internal assets and the external entities by tracking the two way communication flow between them. Then the trail of data exchanges is developed matching with the state based infection sequence model. This system mainly contains the correlation engine which is involved in detecting the stages of the malware infection process with the help of the three malware focused network packet sensor. Finally the dialog trails of the inbound intrusion alarms with outbound communication patterns are tied together by the correlation enginewhich determine whether the local host is infected or not. Thus is a passive bot detection strategy which generates a consolidated report to capture the relevant events and email sources [4].

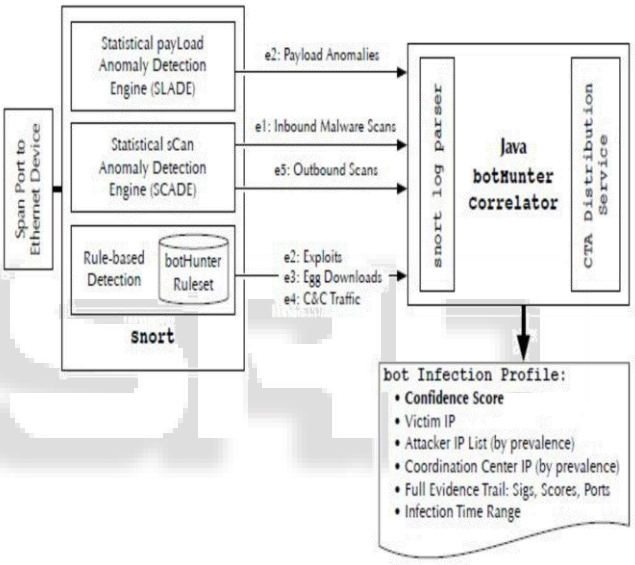

Fig. 3: BotHunter Architecture

VII. PROPOSED SYSTEM

Compromised machines(Machine use to send spam message) are one of the important security threats on the Internet. In this system ,we focus on the detection of the compromised machines in a network that are used for sending spam messages, which are commonly referred to As spam zombies.The nature of sequentially observing outgoing messages gives rise to the sequential detection problem. In this system, we will develop a spam zombie detection system by monitoring outgoing messages. This detection system can identify a compromised machine quickly. In proposed system to develop an effective spam zombie detection system (SPOT).It is used to monitoring outgoing messages of a network. Its designed based on a statistical method called sequential probability ratio test (SPRT).

In this system we provide registration to user once user registered he/she get user id and password. when user makes login user must use same user id and password , if user id or password is wrong user can't do login

In this system we are monitoring the outgoing massages of sender machine and capture the IP Address of

the sender machine. Also stores senders detail information such as Email Id of sender, number of files send by the sender ,message text of sender etc.In this system we define the Spam messages threshold i.e. the number of Spam message or Spam files send from a particular machine . In this system we block the Spam senders Email Id reason behind that is one machine is used by many user. In this system we calculate the Count threshold and Percentage threshold .(value of CT and PT) CT is the number of Spam messages send from a machine by user.PT is the percentage of Spam message send by sender. PT is the ratio of Spam messages to the total number of messages into 100 ,send from a particular machine, from a particular user.

We calculate value of CT and PT for each sender i.e. for each Email Id we calculate value of CT and PT separately we block senders Email Id if the PT value is greater or equals to 50%. Once we block user he/she can't do login. Blocked user cant access self-account due his/her access is denied. If user want to access self-account he/she make request to Admin to unblock self. If user want to access self-account it needs to pay fine for sending spam file after pay fine charges admin unblock the user. When admin Unblock the user ,user can access self-account .Admin have spam word file containing spam words it compare user attached file with this spam file and find out whether the message is spam or not, Admin can update this file it performs various operations on this file update ,delete and display the content of spam file. Admin stores spam file in database.

VIII. BLOCKED DIAGRAM

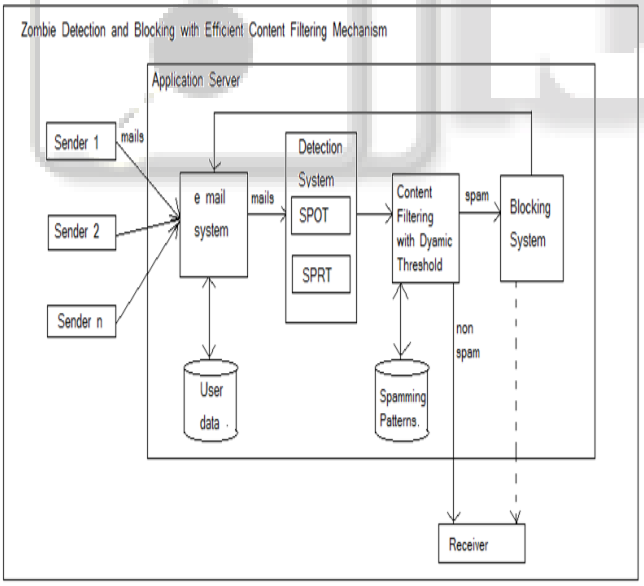

Fig. 4: Architecture of Proposed System

IX. RESEARCH METHODOLOGY (MODULES/ALGORITHMS)

- Account authentication
- Sending mail
- CT detection
- PT detection

A. Account Authentication:

In this system we provide registration to user until and unless user make registration user can't get unique User Id and password. When user make registration user get unique User Id and password .when user make login if User Id and password is valid then user can get access otherwise sue can't access account i.e user can't login.

B. Sending Mail:

User can login to account and user can send mail .User can attached number of files single person can send mail to many times to many pepole. This mails are spam or non-spam.

C. CT Detection:

In this we find out spam file per mail for each user also find out Spam words from message body for each mail and respective sender.

D. PT Detection:

In this we find out total attached file how many files are spam and find out percentage of spam file.

X. PROBLEM DEFINITION

They are many security related issue problems in existing system. often used to lunch various security attacks such as spamming, and spreading malware, DDoS, and identity theft. To find out compromised machines in network which are involved in such activities and try to reduce such security attacks. It helps to reduce cybercrime and increase reliability. In this System, we focus on the detection of the compromised machines in a network.

XI. ALGORITHM 1 SPOT SPAM ZOMBIE DETECTION SYSTEM

1) An outgoing message arrives at SPOT
2) Get IP Address of sending machine
3) // all following parameters specific to machine
4) Let n be the message index
5) If(Xn==1 && CT==1) then
6) {
    a. Compromised_machine=CT/n
7) }
8) Else
9) {
    a. Non_compromised=Nonspam_message
10) }
11) End If
12) If(PT>=Thersold Value) Then
13) {
    a. Machine is compromised
    b. Block User Account;
14) }
15) Else
16) {
    a. Machine is normal machine
    b. Allow user to access account
17) }
18) End If
19) If(PT==0) then
20) {
    a. Test continues with additional observation
    b. If(penalty==paid) then

c. {
   i. Unblock user account;
   ii. Allow access to account;
d. }
e. Else
f. {
   i. Keep block user account;
   ii. User can't access account;
g. }
h. End If

21) }
22) End If
23) Stop

XII. SYSTEM REQUIREMENTS

A. *Hardware Requirements:*

- System : Pentium IV 2.4 GHz.
- Hard Disk : 40 GB.
- Floppy Drive : 1.44 Mb.
- Monitor : 15 VGA Colour.
- Mouse : Logitech.
- Ram : 512 Mb.

B. *Software Requirements:*

- Operating system : - Windows XP.
- Coding Language : JAVA
- Technologies : Swing
- DATABASE : MYSQL

XIII. RELATED WORK

In this chapter we discuss related work, focusing on the studies that utilize spamming activities to detect bots. Based on email messages received at a large email service provider, two recent studies investigated the aggregate global characteristics of spamming botnets including the size of botnets and the spamming patterns of botnets. These studies provided important insight 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. However, their approaches are better suited for large email service providers to understand the aggregate global characteristics of spamming botnets instead of being deployed by individual networks to detect internal compromised machines. Moreover, their approaches cannot support the online detection requirement in the network environment considered in this thesis. We aim to develop a tool to assist system administrators in automatically detecting compromised machines in their networks in an online manner.

XIV. MOTIVATION AND CHALLENGES

In existing system there is many security related issues are present. In this System we capture IP address of machine, we also find out botnet in network. we focus on find out compromised machines and block users.

XV. CONCLUSION

The Spam Zombie Detection and blocking Mechanism detects the spam mails by monitoring the outgoing mails. The Spam Zombie Detection and blocking Mechanism uses the Sequential Probability Ratio Test algorithm to detect the spam zombies. The system also provides the blocking mechanism in which if the system is identified as the spam zombie then the user account gets blocked so that he cannot send the spam messages further. The system also provides the virus detector and attachment scanning mechanism.

REFERENCES

[1] Zhenhai Duan, Peng Chen,Fernando Sanchez, Yingfei Dong, Mary Stephenson, Jamnes Barker, "Detecting Spam Zombies by Monitoring Outgoing Messages," IEEE Transactions Dependable And Secure Computing Vol.9, No.2, Year 2012
[2] Xiaocong. Yu1,2, Xiaomei Dong1, Ge Yu1, Yuhai Qin2, Dejun Yue1, "Botminer: Data-adaptive Clustering Analysis for Online Botnet Detection," Computational Science and Optimization (CSO), 2010 Third International Joint Conference on (Volume:1), 2010.
[3] Guofei Gu, Junjie Zhang, and Wenke Lee, "BotSniffer: Detecting Botnet Command and Control Channels in Network Traffic," School of Computer Science, College of Computing Georgia Institute of Technology.
[4] Guofei Gu, Phillip Porras, Vinod Yegneswaran, Martin Fong, Wenke Lee, "BotHunter: Detecting Malware Infection through IDS-Driven Dialog Correlation," Proc. 16th USENIX Security Symposium, Boston, MA, Aug. 2007.
[5] Guofei Gu,Junjie Zhang, and Wenke Lee "BotSniffer: Detecting Botnet Command and Control Channels in Network Traffic.
[6] "Ahmed Khorsi "An Overview of Content-Based Spam Filtering Techniques."
[7] Yinglian Xie, Fang Yu, Kannan Achan, Rina Panigrahy "Spamming Botnets: Signatures and Characteristics.
[8] J. Markoff. Russian gang hijacking pcs in vast scheme. The New York Times,http://www.nytimes.com/2008/08/06/technology/06hack.html, Auguest 2008.
[9] G. B. Wetherill and K. D. Glazebrook. Sequential Methods in Statistics. Chapman and Hall, 1986.
[10] A. Wald. Sequential Analysis. John Wiley & Sons, Inc, 1947.
[11] Spam Assasian, "The apache spamassiaan project", http://spamassassin.apache.org 2011
[12] J.Klensin, "simple mail transfer protocol"IETF RFC 2821,apr 2001
[13] P.Wood et al,/ "messagelabs intelligence : 2010 annunal security report",2010
[14] P.Resnick "Internet Message Format" IETF RFC 2821,apr 2001.

