ISSN:2320-0790

An International Journal of Advanced Computer Technology

# Spot- Zombie Filtering System

**Arathy Rajagopal, B. Geethanjali, Arulprakash. P**

**Abstract:** 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 spamming and spreading malware, DDoS, and identity theft. These compromised machines are called "Zombies". In general E-mail applications and providers uses spam filters to filter the spam messages. Spam filtering is a technique for discriminating the genuine message from the spam messages. The attackers send the spam messages to the targeted machine by exalting the filters, which causes the increase in false positives and false negatives. We develop an effective spam zombie detection system named SPOT by monitoring outgoing messages of a network. SPOT focuses on the number of outgoing messages that are originated or forwarded by each computer on a network to identify the presence of Zombies. SPOT is designed based on a powerful statistical tool called Sequential Probability Ratio Test, which has bounded false positive and false negative error rates.

## I. INTRODUCTION

In today's computing world, internet plays an important role in our daily lives (in almost every aspect).It is the place where we do lot of things just sitting at one place. Internet not only influences the people to do positive works but also influences the people to trouble others by posing many attacks. These attacks are posed by the attackers directly or indirectly. Attacks are broadly classified into two types, one is automatic attacks and other type is manual attacks. Most of the successful attacks are from the automated generated code injected by the attackers. These are very dangerous which includes Denial of Service (DoS), Distributed denial of Service (DDoS), E-mail Worms, Viruses, Worms, Trojan horses, phishing attacks etc.

Internet e-mail worms are very popular because they are very hard to track. After creating a worm, attacker uses one of the many anonymous e-mail services to launch it. Most of them are in huge size and the user is enticed to execute the worm. The worm first load into the machines main memory and it looks for additional email addresses to send itself to. Attackers get the control over the machines, to launch the attacks on targeted machine, which are formally known as drones, bots, zombies or compromised machines. In E-mail applications these are called as spam zombies because these zombies generate huge number of spam messages to launch the attack on the target machine. It is given that spamming is the major security challenge in the email communication.

In general E-mail applications and providers uses spam filters to filter the spam messages. Spam filtering is a technique for discriminating the genuine message from the spam messages. The attackers send the spam messages to the targeted machine by exalting the filters, which causes the increase in false positives and false negatives. False positive is the misclassification of good message as a spam message and false negative is the misclassification of spam message as a good message. Efficient spam filter aims to minimize the false positive and false negatives.

In this, we are developing a spam zombie detection system named SPOT, by monitoring outgoing messages. SPOT focuses on the number of outgoing messages that are originated or forwarded by each computer on a network to identify the presence of Zombies. SPOT is designed based on a powerful statistical tool called Sequential Probability Ratio Test (SPRT).

SPRT is a powerful statistical method that can be used to test between two hypotheses (in our case, a machine is compromised versus the machine is not compromised), as the events (in our case, outgoing

messages) occur sequentially. SPRT has a number of desirable features. It minimizes the expected number of observations required to reach a decision among all the sequential and nonsequential statistical tests with no greater error rates. This means that the SPOT detection system can identify a compromised machine quickly. Moreover, both the false positive and false negative probabilities of SPRT can be bounded by user-defined thresholds. Consequently, users of the SPOT system can select the desired thresholds to control the false positive and false negative rates of the system. SPOT only needs a small number of observations to detect a compromised machine. The majority of spam zombies are detected with as little as three spam messages.

## II. RELATED WORKS

Choi et.al proposed a technique to detect the bots based on the DNS queries generated. Based on the similarity in the group activity of the DNS traffic the bots are detected in this paper. In [6] the botnets are detected based on the passive analysis on flow data.

Xie et al. developed an effective tool named DBSpam to detect proxy-based spamming activities in a network relying on the packet symmetry property of such activities [7]. We intend to identify all types of compromised machines involved in spamming, not only the spam proxies that translate and forward upstream non- SMTP packets (for example, HTTP) into SMTP commands to downstream mail servers as in [5].

BotHunter uses the IDS trace [3] to detect the bots by comparing the inbound intrusion alarms with the outbound communication patterns. SPRT algorithm focuses on any spamming activity unlike BotHunter which depends on specifics of malware infection process.

An anomaly-based detection system named BotSniffer [4] identifies botnets by exploring the spatial-temporal behavioral similarity commonly observed in botnets. It focuses on IRC-based and HTTP-based botnets. In BotSniffer, flows are classified into groups based on the common server that they connect to. If the flows within a group exhibit behavioral similarity, the corresponding hosts involved are detected as being compromised.

BotMiner [4] is one of the first botnet detection systems that are both protocol and structure independent. In BotMiner, flows are classified into groups based on similar communication patterns and similar malicious activity patterns, respectively. The intersection of the two groups is considered to be compromised machines.
Compared to general botnet detection systems such as BotHunter, BotSniffer, and BotMiner, SPOT is a light weight compromised machine detection scheme, by exploring the economic incentives for attackers to recruit the large number of compromised machines.

## III. PROBLEM FORMULATION

A machine in the network can be either normal or compromised. Compromised machines in the network are called spam zombies. Let $X_i$ for i = 1, 2 . . . denote the successive observations of a random variable X corresponding to the sequence of messages originated from machine m inside the network. We let $X_i$ = 1 if message i from the machine is a spam, and $X_i$ = 0 otherwise. The detection system assumes that the behavior of a compromised machine is different from that of a normal machine in terms of the messages they send. Specifically, a compromised machine will with a higher probability generate a spam message than a normal machine.

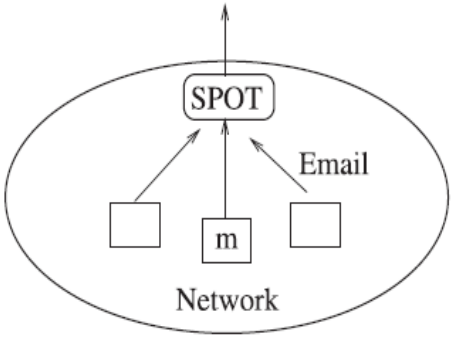



Fig: Network model

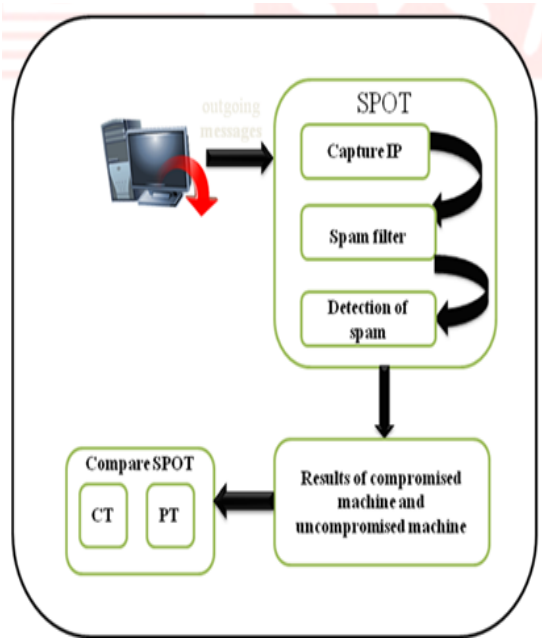
Formally,

$$Pr(X_i = 1/H_1) > Pr(X_i = 1/H_0)$$

Where $H_1$ represents the machine m is compromised and $H_0$ represents the machine is normal. We have included a content based spam filter so that it can identify the outgoing message is spam or not. The spam zombie detection problem is: as $X_i$ arrives sequentially at the detection system, the system determines with a high probability if machine m has been compromised. Once a decision is reached, the

detection system reports the result, and further actions can be taken, like disconnecting the system from the network.

## IV. SYSTEM ARCHITECTURE




## V. ALGORITHM

### *5.1 SPOT Detection Algorithm*

SPOT is designed based on the statistical tool SPRT. In the context of detecting spam zombies in SPOT, we consider $H_1$ as a detection and $H_0$ as normality. That is, $H_1$ is true if the concerned machine is compromised, and $H_0$ is true if it is not compromised. In addition, we let $X_i$ =1 if the $i^{th}$ message from the concerned machine in the network is a spam, and $X_i$ = 0 otherwise.

Let X denote a Bernoulli random variable under consideration with an unknown parameter $\Theta$, and $X_1$, $X_2$ . . . the successive observations on X. As discussed above, SPRT is used for testing a simple hypothesis $H_0$ that $\Theta=\Theta_0$ against a single alternative H1 that $\Theta=\Theta_1$. That is,

$$Pr(X_i =1/ H_0) =1\text{-}Pr(X_i = 0/ H_0) =\Theta_0$$

$$Pr(X_i =1/ H_1) =1\text{-}Pr(X_i = 0/ H_1) = \Theta_1$$

For any positive integer n=1, 2…

$$\Lambda_n = \ln\left[\frac{Pr(X_1, X_2 \ldots X_n /H_1)}{Pr(X_1, X_2 \ldots X_n /H_1)}\right]$$

Algorithm: SPOT spam zombie detection system

1: An outgoing message arrives at SPOT
2: Get IP address of sending machine, m
3: Let n be the message index
4: Let $X_n$ =1, if message is spam,

$X_n$ =0 otherwise

5: if ($X_n$ ==1) then
6: $\Lambda_n$ += $\ln\left[\frac{\Theta_1}{\Theta_0}\right]$
7: else $\Lambda_n$ += $\ln\left[\frac{1-\Theta_1}{1-\Theta_0}\right]$
8: end if
9: if ($\Lambda_n \geq B$) then
Machine m is compromised. Test terminates for m.
10: if ($\Lambda_n \leq A$) then
Machine m is normal. Test is rest for m.
11: $\Lambda_n$ =0
Test continues with new observations.
12: else
Test continues with an additional observation.
13: end if

### *5.2 Count-Threshold Algorithm*

In this time is partitioned into windows of fixed length, T. User defined threshold parameter Cs, maximum number of spam messages that may be originated from a normal machine in any window. If the number of message send, n>Cs then the machine is compromised.

### *5.3 Percentage-Threshold Algorithm*

In this also time is partitioned into windows of fixed length, T. Here N is the total messages and n is the spam messages originated from a machine, m within a time window. If N>Ca and n/N>P, then the message is compromised. Where Ca is the minimum number of messages that a machine must send and P is the user-defined maximum spam percentage of a normal machine.

## V. MODULE SPLIT-UP

### *6.1 User Interface Module*

In the user interface module we are creating the end user login page for the mailing system. Each and every machine in the network will get login to the mailing system then only it will forward the mail through the network. Here we are creating the user interface module using the JSP.

### *6.2 Spot Module*

In the SPOT Module when an outgoing message arrives at the SPOT detection system, the sending machine's IP address is recorded, and the message is classified as either spam or no spam by the (content-based) spam filter. The machines which are all sending the spam message are treated as the compromised System.

### *6.3 Count Threshold (CT) Module*

The count threshold module is counting the number of the spam messages sent by the compromised system in the network. In the SPOT Monitoring process the IP of the Spam spreading systems are monitored. The number of message sent by the machine in a time interval is counted here. If the one machine count gets increased with it then it will be decided as Spam system.

### *6.4 Percentage Threshold (PT) Module*

In this module we are monitoring the machines messages. Here we are calculating the number of messages sent by the system and counting the number of the spam messages sent by the compromised system then we are calculating the percentage of spam message sent by the compromised system.

### *6.5 Spam Zombie Detection Module*

In the spam zombie detection module the SPOT method will give the details about the compromised systems. Here the SPOT monitor system will clean the details about the Spam zombie system. Reset the values of the corresponding compromised system details from the monitoring process.

## VII. CONCLUSION

Spam messages are the main problem faced by internet users. Our system Spot detects the origin of spam messages called Spam Zombie using a powerful statistical tool, Sequential Probability Ratio Test (SPRT). SPOT has a bounded false positive and false negative error rate. It also minimizes the number of observations to detect spam zombies. We also design and study two other spam zombie detection algorithm based on number of spam message and percentage of spam message forwarded by internal machines.

## VIII. REFERENCES

[1] Z. Duan, Y. Dong, "Detecting Spam Zombies by monitoring outgoing Messages"

[2] A. Wald, Sequential Analysis. John Wiley & Sons, 1947.

[3] G. Gu, P. Porras, V. Yegneswaran, M. Fong, and W. Lee, "BotHunter: Detecting Malware Infection through Ids-Driven Dialog Correlation," Proc. 16th USENIX Security Symp., Aug. 2007.

[4] G. Gu, J. Zhang, and W. Lee, "BotSniffer: Detecting Botnet Command and Control Channels in Network Traffic," Proc.15th Ann. Network and Distributed System Security Symp. (NDSS '08),Feb. 2008.

[5] Y. Xie, F. Xu, K. Achan, R. Panigrahy, G. Hulten, and I. Osipkov, "Spamming Botnets: Signatures and Characteristics," Proc. ACM SIGCOMM, Aug. 2008.

[6] Botnet Detection by Monitoring Group Activities in DNS Traffic Hyunsang Choi, Hanwoo Lee, Heejo Lee, Hyogon KimKorea University.

[7] M. Xie, H. Yin, and H. Wang, "An effective defense against email spam laundering," in ACM Conference on Computer and Communication Security, Alexandria, VA, October 30 - November 3 2006.