INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

SPAM DETECTION USNIG SPOT TOOL

SURAJ KUTE, DIPIKA MOHOD, PAYAL SHIRE, PRATIKSHA BARDE, DEEPIKA MACCHEWAR

Dept. of Information Tech, *D.B.N.C.O.E.T. Yavatmal-445001*

Accepted Date: 15/03/2016; Published Date: 01/05/2016

Abstract*:* In this paper we focus on detection of compromised machine in network that involve in spam zombie's activity. The detection system named SPOT used for monitoring outgoing messages of 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. It focuses to evaluate the performance of the developed SPOT system for large network in an effective and efficient system to automatically detecting compromised machines. In this paper we compare the SPOT with two other spam zombie detection algorithms based on the number and percentage of spam messages originated or forwarded by internal machines respectively and on basis of that we observe the SPOT is an efficient than other detection technique.

Keywords: SPOT, Compromised machines, CT, PT.

Corresponding Author: MR. SURAJ KUTE

Access Online On:

www.ijpret.com

How to Cite This Article:

Suraj Kute, IJPRET, 2016; Volume 4 (9): 747-757

PAPER-QR CODE

INTRODUCTION

Today's computing world the internet plays an important role in our daily life. Internet not only influences people for doing positive works but also secured the people from trouble by posing many attacks. These attacks are may be automatic attacks and other one i.e. manual attack. Most successful attacks are from automated generated code injected by the attackers. These are very dangerous for the user include spamming and spreading malware [2],

Denial of Service (DoS) [5], Distributed denial of Service (DDoS), E-mail Worms [3]. Rather than the aggregate global characteristics of spamming botnets, we study the tool for system administrators to automatically detect the compromised machines in networks. In this, a spam zombie detection system named SPOT, used to monitoring outgoing messages. SPOT focuses on the number of outgoing messages that originated or forwarded by one computer to another computer on a network to identify the presence of Zombies. SPOT is designed on the base of powerful statistical tool called Sequential Probability Ratio Test (SPRT) developed by Wald in his seminal work [1]. SPRT is a powerful statistical method that can be used to test between two hypotheses i.e. a machine is compromised versus the machine is not compromised as the events occur sequentially to outgoing messages.

2. RELATED WORK

We discuss related work based on email messages received at a large email service provider, the recent studies[11] gives aggregate global characteristics of spamming botnet including the size of botnet and the spamming patterns of botnet. These approaches are better suited for large e-mail service providers to understand the aggregate global characteristics of spamming botnet instead of being deployed by individual networks to detect internal compromised machines. A DBSpam, the effective tool developed by Xie et al. used detect proxy-based spamming activities in a network relying on the packet symmetry property of such activities [12]. Compared to general botnet detection systems such as BotHunter, BotSniffer, and BotMiner, SPOT is a lightweight compromised machine detection scheme, by exploring the economic incentives for attackers to recruit the large number of compromised machines. In the area of networking security, SPRT has been used to detect portscan activities, proxy-based spamming activities, anomaly-based botnet detection [7], and MAC protocol misbehavior in wireless networks [9].

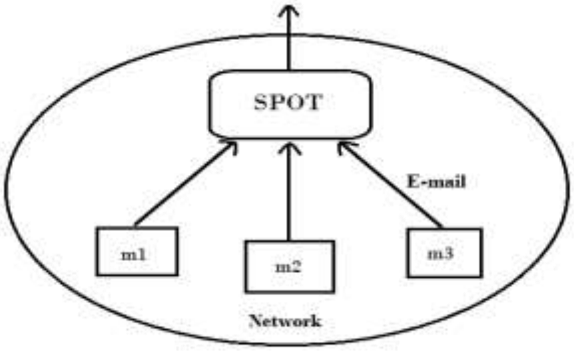

Fig.(1)- Network Model

3. ANALYSIS OF PROBLEM

These system launch various security attacks including spamming and spreading malware [2], DDoS [5], and identity theft by comparing to general botnet detection systems such as BotHunter [6], BotSniffer [7], and BotMiner [4], SPOT [1].

- These are major security challenge on the internet for existence of the large number of compromised machines.
- Their approaches are better suited for large e-mail service providers but cannot support the online detection requirement in the network environment considered in this paper.
- The existing spam detection algorithm is less effective as compare to SPOT.
- Identifying and cleaning compromised machines in a network remain a significant challenge for system administrators of networks of all sizes.

4. ARCHITECTURE

4.1 Working:-

SPOT is designed based on a statistical method called Sequential Probability Ratio Test (SPRT), which illustrate in fig.(2) simple and powerful statistical method, SPRT has a number of desirable features. The proposed SPOT system can handle the case where an outgoing message is forwarded by a few internal mail relay servers before leaving the network. An IP address corresponds to a unique machine can be detecting by applying algorithms. We will investigate the potential impacts of dynamic IP addresses on the detection algorithms in technical working term.

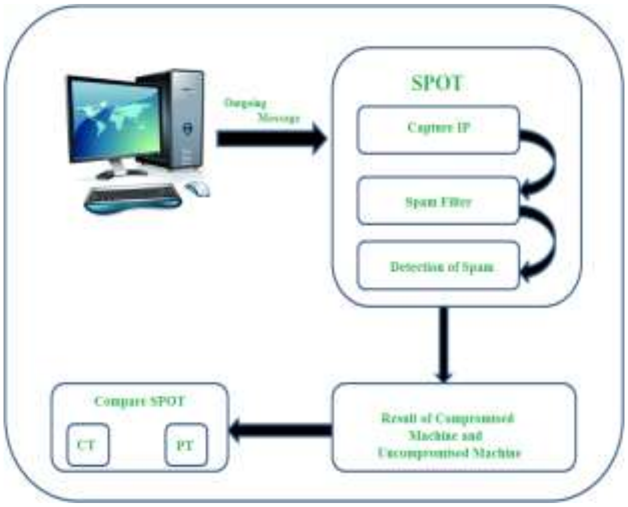

Fig.(2)- System Architecture

An practically these technique implement as: -

I. User Interface Module

In these 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.

II. 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.

III. Count Threshold (CT) Module

The count threshold module is counting the number of the spam messages sent by the compromised system in the network. 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.

IV. 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.

V. Spam Zombie Detection Module

Here the SPOT monitor system will clean the details about the Spam zombie system then reset the values of the corresponding compromised system details from the monitoring process.

4.2 Spam Detection Technique:-

In this section, we see how spam zombies and spam messages can be identified using SPRT [1] technique. Spam Zombie produces the spam messages and the spam message enters into the network. Server, first identifies the which message is Spam by-

i. Detecting the Compromised Machines :-

- Compromised machines are the machines that involved in spamming activities on the internet that are generally referred to as bots.
- These set of bots controlled by a single entity is called a botnet.
- To detecting and black listing such individual bots is commonly regarded as difficult, due to both the transient nature of attack and fact that each bot may send.

i. Spam Detection :-

- Here by capture the IP address of the system the mails are applied to filtering process where the mail content is filtered.
- Spam filter is deployed at the detection system so that an outgoing message can be classified as either a spam or non spam.

ii. Spot Detection Algorithm :-

- Intuitively, SPRT can be considered as a one dimensional random walk with two user-specified boundaries corresponding to the two hypotheses. Which concerned with random variable arrive sequentially, the walk moves either upward or downward one step, depending on the value of the observed sample.
- As a simple and powerful statistical tool, SPRT has a number of compelling and desirable features that lead to the widespread applications of the technique in many areas. These

both are actual false positive and false negative probabilities of SPRT can be bounded by the user-specified error rates.

- 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.
- Algorithm :

```
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 Xn = 1 if message is spam,
Xn = 0 otherwise
if (Xn = = 1) then // spam
Δn+=ln ɵ1/ ɵ2
else // nonspam
Δn+=ln 1-ɵ1/1-ɵ0
end if
if (Δn <=B)
Machine m is compromised. Test terminates for m.
lse if (Δn <=A) then
Machine m is normal. Test is reset for m.
Δn = 0
Test continues with new observations
else
Test continues with an additional observation
```

end if.

iii. Spam Count and Percentage Based

Detection Algorithm:-

- For simplicity, we refer these count-threshold (CT) detection algorithm and percentage-threshold (PT) detection algorithm respectively.
- SPOT can provide bounded false positive rate and false negative rate and confidence how well SPOT works also the error rates of CT and PT cannot be a priori specified.
- The proper values for the four user defined parameters (α, β, θ1, θ2) in SPOT is relatively straightforward which select the "right" values for parameter of CT and PT which is much more challenging and tricky.
- The performance of the two algorithms is sensitive to the parameters used in the algorithm.

4.3 Performance Analysis :-

We study the potential impact of dynamic IP addresses on detecting spam messages. Moreover, in certain environment where user feedback is reliable, for example, feedback from users of the same network in which SPOT is deployed, SPOT can rely on classifications from end users in addition to the spam filter.

User feedbacks may be incorporated into SPOT to improve the spam detection rate of the spam filter. As we see discussion in the previous section, trying to send spam at a low rate will also not evade the SPOT system. SPOT relies on the number of (spam) messages, not the sending rate, to detect spam zombies. The current performance analysis study over the SPOT was carry out using FSU e-mails collected in the year of 2005. However, based on the above discussion, we expect that SPOT will work equally well in today's environment. Indeed, as long as the spam filer deployed together with SPOT can provide a reasonable spam detection rate. Various studies recently years have shown that spam messages sent from botnet accounted for above 80 percent of all spam messages on the Internet. For example, the Message Labs Intelligence annual security report showed that approximately 88.2 percent of all spam sent from botnets.

➢ Overview of the E-Mail Trace and Methodology:-

The e-mail trace was collected at a mail relay server deployed in the Florida State University (FSU) campus network between 8/25/2005 and 10/24/2005, excluding 9/11/2005 (we do not have trace on this date). During the course of the e-mail trace collection, the mail server relayed messages destined for 53 sub domains in the FSU campus network. The mail relay server ran Spam Assassin to detect spam messages [10]. The e-mail trace contains the following information for each incoming message: the local arrival time, the IP address of the sending machine i.e., the upstream mail server that delivered the message to the FSU mail relay server and cheque whether message is spam or not. In addition, if a message has a known virus/worm attachment, it was so indicated in the trace by antivirus software. The antivirus software and Spam Assassin [10] were two independent components deployed on the mail relay server.

We refer to this set of messages as the FSU e-mails and perform our evaluation of the detection algorithms based on the FSU e-mails and if a message has a known virus/worm attachment, we refer to such a message as an infected message.

- Performance of Spot

In this section, we evaluate the performance of SPOT based on the collected FSU e-mails. There are few FSU internal IP addresses observed in the e-mail trace. Out of these some IP addresses identified by SPOT, we can confirm them to be compromised in this way. For remaining IP addresses, we manually examine the spam sending patterns from the IP addresses and the domain names of the corresponding machines. If the fraction of the spam messages from an IP address is high (if greater than 98percent), we also claim that the corresponding machine has been confirmed to be compromised. We can confirm some of them to be compromised in this way.

- Performance of CT and PT

CT is a detection algorithm based on the number of spam messages originated or forwarded by an internal machine, and PT based on the percentage of spam messages originated or forwarded by an internal machine. For comparison, it includes a simple spam zombie detection algorithm that identifies any machine sending at least a single spam message as a compromised machine. In this, suppose we set the length of time windows to be 1 hour, that is, T ¼ 1 hour, for both CT and PT. For CT, we set the maximum number of spam messages that a normal machine can send within a time window to be 30 (Cs=3), that is, when a machine sends more than30 spam messages within any time windows, CT concludes that the machine is compromised. The simple detection algorithm can detect more machines as being compromised than CT, and PT. It also has better performance than CT and PT in terms of both detection rate (89.7 percent) and false negative rate (10.3 percent).

- Dynamic IP Addresses

In this section, we group messages from a dynamic IP address into clusters with a time interval threshold of 30 minutes. Messages with a consecutive inter arrival time no greater than allocate minutes are grouped into the same cluster. From this we consider all the messages from the same IP address within each cluster as being sent from the same machine i.e., the corresponding IP address has not been reassigned to a different machine within the concerned cluster.

4.4 Advantages :-

The advantages of this technique as compare to other spam detection scheme are –

i. Capture the fundamental invariants of botnet behavior rather than symptoms i.e. scanning.

ii. It provides the complementary techniques and cover multiple stages, dimensions and perspectives.

iii. Our solution should be general and extensible.

iv. By general, we mean the solution should not be restricted to a specific botnet instance or dependent on a specific symptom.

v. The solution should provide an open and flexible framework that can easily incorporate new user-provided components/plug-ins.

5. APPLICATION

1. SPOT is used to monitoring outgoing messages of a network.
2. SPOT can be used for testing two hypotheses whether the machine is compromised or not.
3. In the area of networking security, SPRT has been used to detect portscan activities, proxy-based spamming activities, anomaly-based botnet detection, MAC protocol mis behaviour in wireless networks.
4. SPOT has bounded false positive and false negative error rates which minimizes the number of required observations to detect a spam zombie.
5. SPOT is a lightweight compromised machine detection scheme as compare to other.
6. Their approaches are better suited for large e-mail service providers.
7. The existing spam detection algorithm is less effective as compare to SPOT.
8. It identifies and cleans compromised machines in a network for system administrators of networks of all sizes.

6. CONCLUSION

We study the effective spam zombie detection system named SPOT by monitoring outgoing messages in a network. 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 which minimizes the number of required observations to detect a spam zombie. Our main future objective is to extend these ideas to detect spam in sender itself and stop the user emails.

7. FUTURE WORK

In the future, we plan to study the following directions:

- We plan to study new techniques to improve the efficiency and increase the coverage of existing monitoring and correlation components which intended to be more robust against evasion attempts.
- We plan to develop a new generation of real-time detection systems combining vertical and horizontal correlation techniques seamlessly, using a layered design, a flexible sampling strategy, and a highly scalable distributed scheme.
- More robust and less controversial active techniques with wider applicable areas.
- Cooperative detection combining host and network-based systems.
- Botnet mitigation and defense.

ACKNOWLEDGEMENT:

We express our sincere gratitude to *Prof. S. C. Chumble* for his valuable guidance and encouragement for the fulfillment of these work.

REFERENCES:

1. Zhenhai Duan, Peng Chen, Fernando Sanchez, Yingfei Dong, Mary Stephenson and J ames Michael Barker “Detecting Spam Zombies by Monitoring Outgoing Messages”IEEE Transaction On Dependable and Secure Computing, VOL. 9, NO. 2, MARCH/APRIL 2012.
2. P. Bacher, T. Holz, M. Kotter, and G. Wicherski, “Know Your Enemy: Tracking Botnets,”http://www.honeynet.org/papers/bots/, 2011.
3. Z. Chen, C. Chen, and C. Ji, “Understanding Localized-Scanning Worms,” Proc. IEEE Int’l Performance, Computing, and Comm. Conf.(IPCCC ’07), 2007.
4. G. Gu, R. Perdisci, J. Zhang, and W. Lee, “BotMiner: Clustering Analysis of Network Traffic for Protocol- and Structure-Independent Botnet Detection,” Proc. 17th USENIX Security Symp., July 2008.
5. Z. Duan, K. Gopalan, and X. Yuan, “Behavioral Characteristics of Spammers and Their Network Reachability Properties,” Proc. IEEE Int’l Conf. Comm. (ICC ’07), June 2007.

6. 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.
7. 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.
8. J. Markoff, “Russian Gang Hijacking PCs in Vast Scheme,” The New York Times, http://www.nytimes.com/2008/08/06/technology/ 06hack.html, Aug. 2008.
9. S. Radosavac, J.S. Baras, and I. Koutsopoulos, “A Framework for MAC Protocol Misbehaviour Detection in Wireless Networks,” Proc. Fourth ACM Workshop Wireless Security, Sept. 2005.
10. SpamAssassin,“The Apache SpamAssassin Project,”http://spamassassin.apache.org, 2011.
11. Y. Xie, F. Xu, K. Achan, E. Gillum, M. Goldszmidt, and T. Wobber, “How Dynamic Are IP Addresses?” Proc. ACM SIGCOMM, Aug. 2007.
12. M. Xie, H. Yin, and H. Wang, “An Effective Defense against Email Spam Laundering,” Proc. ACM Conf. Computer and Comm. Security, Oct./Nov. 2006.