ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

Volume 3, Special Issue 3, March 2014

2014 International Conference on Innovations in Engineering and Technology (ICIET'14)

On 21st & 22nd March Organized by

K.L.N. College of Engineering, Madurai, Tamil Nadu, India

An Effective SPOT System by Monitoring Outgoing Messages

G.Mariammal, Dr.S.Allwin, M.E., Ph.D.

PG Scholar, Infant Jesus College of Engineering, Infant Jesus College of Engineering, Thoothukudi, India.

Associate Professor, Infant Jesus College of Engineering, Infant Jesus College of Engineering, Thoothukudi, India.

ABSTRACT- Develop an effective spam zombie detection system named SPOT. In the network SPOT can be used to monitoring outgoing messages. Using internet some attacker try to spread the spams or malware in order to collect the information about the network. The detection of the compromised machines in the network that are involved in the spamming activities is known as spam zombie detection system. The detection system can be used to identify the misbehavior of the person using Spam zombie detection system. We will create a framework to identify the message from the various persons. This system will record the information of the IP address using SPOT Detection Algorithm. We also compare the performance of SPOT with two other spam zombie detection algorithms based on the count and percentage of spam messages originated or forwarded by internal machines. Using these above techniques we will avoid and block the person who sends the spam's message.

Index term— SPOT System, SPOT Detection Algorithm, Count-threshold, Percentage-threshold.

I. INTRODUCTION

Existence of the large number of compromised machines is the major security challenge on the internet. Compromised machines have been increasingly used to launch various security attacks such as spamming and spreading malware, DDoS, and identity theft [6]. Then identifying and cleaning compromised machines in a network remain significant challenges for system administrators of networks of all sizes. Mainly 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. A Spam zombie is the detection of the compromised machines in the network that are involved in the spamming activities [6]. Given that spamming provides a critical economic incentive for the controller of the compromised machines to recruit these machines, it has been used to observe that many compromised machines are involved in spamming [9][10][12]. A number of recent research efforts have studied the aggregate global characteristics of spamming botnets such as the size of botnets and the spamming patterns of botnets, based on the sampled spam messages received at a large email service provider [12]. The main aim is to develop a tool for system administrators to automatically detect the compromised machines in their networks in an online manner. Normally in the network the local generated outgoing messages cannot provide the aggregate large-scale spam view required by these approaches [5]. These approaches cannot support the online detection requirement in the environment. The nature of sequentially observing outing messages gives rice to the sequential detection problem. We will develop a spam zombie detection system, named SPOT. The Spot can be used to monitoring outgoing messages. SPOT is designed based on a statistical method called Sequential Probability Ratio Test (SPOT).

SPRT is a powerful statistical method that can be used test between two systems sequentially in our case machine is compromised versus the machine is not compromised and another case based on outing messages. Both the false positive and false negative probabilities of SPRT can be bounded by user-defined thresholds. SPOT system can be used to select the desired thresholds to control the false positive and false negative rates of the system.

We develop the SPOT detection system, the system administrators can be used to automatically identifying the compromised machines in their networks. Evaluate the performance of the SPOT system based on a two-month e-mail trace collected in a large US campus network. Based on evaluation studies show that SPOT is an effective and efficient system in automatically detecting compromised machines in a network [11].

In addition, SPOT only needs a small number of observations to detect a compromised machine. Majority of spam zombies are detected with as little as three spam messages. At the time of comparison, we also design and

study two other spam zombie detection algorithms based on the number of spam messages and the percentage of spam messages originated or forwarded by internal machines. Also compare the performance of SPOT with the two other detection algorithms to explain the advantages of the SPOT system.

II. RELATED WORK

In the related work we discuss the detection of compromised machines. The characterizing spamming botnet by leveraging both spam payload and spam server traffic properties. We developed a spam signature generation framework called *AutoRE* to detect botnet-based spam emails and botnet membership [12]. Our in-depth analysis of the identified botnet revealed several interesting finding regarding the degree of email obfuscation, properties of botnet IP addresses, sending patterns, and their correlation with network scanning traffic [1]. To group bots into botnets we look for multiple bots participating in the same spam email campaign. We have applied our technique against a trace of spam email from Hotmail web mail services.

In this trace, we have successfully identified hundreds of botnet. We present new finding about botnet sizes and behavior while also confirming other researcher's observations derived by different methods. In addition, using this information combined with a three-month Hotmail email server log, we were able to establish that 97% of mail servers setup on dynamic IP addresses sent out solely spam emails, likely controlled by zombies [2]. Moreover, these mail servers sent out a large amount of spam- counting towards over 42% of all spam emails to Hotmail. These results highlight the importance of being able to accurately identify dynamic IP addresses for spam filtering, and we expect similar benefits of it for phishing site identification and botnet detection. To our knowledge, this is the first successful attempt to automatically identify and understand IP dynamics.

We reveal one salient characteristic of proxy-based spamming activities, namely packet symmetry, by analyzing protocol semantics and timing causality [6]. Based on the packet symmetry exhibited in spam laundering, we propose a simple and effective technique, DBSpam, to on-line detect and break spam laundering activities inside a customer network [8].

We provide the first comprehensive study on the received: header field of spam emails to investigate, among others, to what degree spammers can and do forge the trace information of spam emails. Also report our findings and discuss the implications of the findings on various spam control efforts, including email sender authentications and spam filtering [3].

We find that most spam is being send from a few regions of IP address space, and that spammers appear to be using transient "bots" that send only a few pieces of email over very short periods of time. Finally, a small, yet non-negligible, amount of spam is received from IP addresses that correspond to short-lived BGP routes, typically for hijacked prefixes. These trends suggest that developing algorithms to identify botnet membership, filtering email messages based on network-level properties, and improving the security of the internet routing infrastructure, may prove to be extremely effective for combating spam [9].

A. Problem Formation and Assumptions

In the network formulate the spam zombie detection problem. We discuss the network model and assumptions can be used to make in the detection problem. Fig.1 describes the logical view of the network model.
Assume that messages originated from machine inside the network. The message will pass the developed spam zombie detection system. This assumption can be achieved in a few different scenarios.

In the network assume that the machine has been either compromised or normal (that is, not compromised). The term compromised machine is denoted as spam zombie. The detection system assumes that the behavior of a compromised machine is different from that the normal machine based on the messages sending. Based on the higher probability the compromised machines are generating a spam message compare to the normal machine. Once a decision is reached, the detection system reports the result, and further action can be taken.

We assume that a content-based spam filter is developed at the detection system. The outgoing message can be classified as either a spam or nonspam using the detection system. None of existing spam filters can achieve perfect spam detection accuracy. They all suffer from both false positive and false negative errors. The false negative rate of spam filter measures the percentage of spam messages that are misclassified.

The false positive rate measures the percentage of nonspam message that are misclassified. We assume that a sending machine m as observed by the spam zombie detection system is an end-user client machine. It is not a mail relay server. .

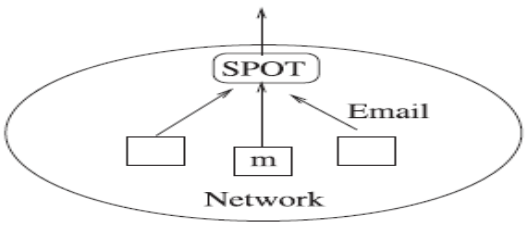

Fig 1 Network Model

B. Sequential Probability Ratio Test

SPRT can be used to monitor the network performance. The goal of the SPRT is to decide which hypothesis is correct as soon as possible. SPOT is designed based on a powerful statistical method called Sequential Probability Ratio Test. SPRT has bounded false positive and false negative error rates. The SPRT is the powerful statistical method that can be used to test between two systems sequentially. In our case machine is compromised versus the machine is not compromised another case based on the outgoing messages. Provide the necessary background on the Sequential Probability Ratio Test for understanding the proposed spam zombie detection system. SPRT is a statistical method for testing a simple null hypothesis against a single alternative hypothesis.

SPRT can be considered as a one-dimensional random walk with two user-specified boundaries corresponding to the two hypotheses. Based on simple and powerful statical tool, SPRT has a number of compelling and desirable features that lead to the widespread applications of the technique in many areas. Before the SPRT terminates smaller error rate tends to require a large number of observations. The user can balance the performance and cost of an SPRT test. In second, has been provide 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.

III. SPAM ZOMBIE DETECTION ALGORITHMS

In this section we develop three spam zombie detection algorithm. First one is SPOT, which utilizes the Sequential Probability Ratio Test. We discuss the impacts of SPRT parameters on SPOT in the content of spam zombie detection. The other two spam zombie detection algorithms are developed based on the number of spam messages and the percentage of spam messages sent from an internal machine.

A. Spot Detection Algorithm

SPOT is designed based on the powerful statistical tool called SPRT.In the below, we describe the SPOT detection algorithm. When an outgoing messages arrives at the SPOT detection system. After the outgoing message reach to the SPOT detection system the sending machine's IP address is recorded.

Based on the recorded IP address, then the message is classified as either spam or nonspam by the content- based spam filter. For each observed IP address, SPOT maintains the logarithm value of the corresponding probability ratio Λ_n. A and B the algorithm determines if the corresponding machine is compromised, normal, or a decision cannot be reached and additional observations are needed.

Algorithm 1:

Step 1: Outgoing message arrives at SPOT
Step 2: Get IP address of sending machine *m*
Step 3: //all following parameters specific to machine *m*
Step 4: Let *n* be the message index
Step 5: Let $X_n = 1$ if message is spam, $X_n = 0$ otherwise
Step 6: **if** ($X_n = = 1$) **then**
Step 7: // spam, 3
Step 8: $\Lambda_n + = \ln \theta_1 / \theta_0$
Step 9: **else**
Step 10: // nonspam
Step 11: $\Lambda_n + = \ln (1-\theta_1) / (1-\theta_0)$
Step 12: **end if**
Step 13: **if** ($\Lambda_n \geq B$) **then**
Step 14: Machine *m* is normal. Test is reset for *m*.
Step 15: **else if** ($\Lambda_n \leq A$) **then**
Step 16: Machine *m* is normal. Test is reset for *m*.
Step 17: $\Lambda_n = 0$
Step 18: Test continues with new observations
Step 19: **else**
Step 20: Test continues with an additional observation
Step 21: **end if**

From the viewpoint of network monitoring, it is more important to identify the machine that has been compromised than the machines that are normal. After a machine has been identified as compromised, then these compromised machines are added into the list of potentially compromised machines that system administrators can go after to clean.

Also record the message-sending behavior of the machine. Before the machine is cleaned and removed from the list, the SPOT detection system does not need to further monitor the message-sending behavior of the machine.

Currently the machine has been normal may get compromised at a later time. We need to continuously monitor machines that are determined to be normal by SPOT. Once such a machine is identified by SPOT, the records of the machine in SPOT are reset, in particular, the value of Λ_n is set to zero, so that a new monitoring phase starts for the machine.

B. Spam Count and Percentage-Based Detection Algorithm

In this section, we present two different algorithms in detecting spam zombies. First one is based on the number of spam messages and another the percentage of spam messages send from an internal machine. We refer to them as the count-threshold (CT) detection algorithm and the percentage-threshold (PT) detection algorithm.

In CT, the time is partitioned into fixed length *T*. A threshold parameter C_s specifies the maximum number of spam message that be originated from a normal machine in

any time. The system monitors the number of spam messages *n.* That message can be originated from a machine. If $n > C_s$ then the algorithm declares that the machine has been compromised.

Similarly, in PT detection algorithm, the time is partitioned into fixed length *T*. In each internal machine in each time PT monitors two e-mail properties. The first one is based on the percentage of spam messages send from a machine. Then the second one is based on the total number of messages. Let *N* and *n* denote the total messages and spam messages originated from a machine *m* within a time. Then PT declares machine m as being compromised if $N \geq C_a$ and $n/N > P.C_a$ is the minimum number of messages that a machine must send. Then *P* is the user-defined maximum spam percentage of a normal machine.

C. Dynamic IP Addresses

For simplicity ignored the potential impact of dynamic IP addresses and assumed that an observed IP corresponds to a unique machine. In the following, we discuss how well the three algorithms fair with dynamic IP addresses. Formally evaluate the impacts of dynamic IP addresses on detecting spam zombies using a two-month e-mail trace collected on a large US campus network. Extremely the SPOT can work in the environment of dynamic IP addresses. We have noted three or four observations are sufficient for SPRT to reach a decision for the vast majority of cases.

If a machine is compromised, more than three or four spam messages will be sent before the user shutdowns the machine and the corresponding IP address gets reassigned to a different machine. Therefore, the dynamic IP addresses will not have any significant impact on the SPOT.

Dynamic IP addresses can have a greater effect on the other two detection algorithm Ct and PT. In first, both required the continuous monitoring of the sending behavior of a machine for at least a specified time.

IV. PERFORMANCE EVALUATION

In this section, we evaluate the performance of the three detection algorithm based on performance of SPOT, performance of count threshold and the performance of percentage threshold.

A. Performance of SPOT

In this section, evaluate the performance of SPOT based on the collected e-mails. The infected messages are only used to confirm if a machine is compromised in order to study the performance of SPOT. Infected messages are not used by SPOT. SPOT relies on the spam messages instead of infected messages to detect if a machine has been compromised to produce the result. Infected messages are more likely to be observed during the spam zombie detection system. To improve the performance the infected messages can be easily incorporated into the SPOT system. Table 1 shows the performance of SPOT detection system.

B. Performance of Count Threshold

Table 2 shows the performance of count threshold which include the machine IP addresses, count threshold value and the machine status. Use the same methods to confirm detection or identify a missed IP address as we have done with the SPOT detection algorithm. In the machine IP address status has denote the machine IP addresses. In the count threshold value status the value of the count threshold value can be defined. Then in the machine status can be define, if the machine is compromised or uncompromised, based on the performance.

C. Performance of Percentage Threshold

Table 3 shows the performance of Percentage Threshold which includes the machine IP address, count threshold, percentage threshold and also the machine status. First note that the methods to confirm detection or identify a missed IP address are different from the ones used in SPOT, CT and PT. From the table we can see that, CT and PT performance. In the machine IP address status has denote the performance of the machine IP address. In the count and the percentage threshold define the threshold value in the table. In the machine status has been defined, if the machine is compromised or the machine is uncompromised.

TABLE 1
SPAM SENDING MACHINE DETAIL

From IP	Total	Non Spam	Spam
127.0.0.1	3	3	0
127.0.0.1	1	1	0
124.0.2.1	20	2	18
124.0.2.2	15	12	3
124.0.2.1	8	7	1

TABLE 2
NORMAL SPAM'S COUNT FOR THRESHOLD

FROM IP	COUNT THRESHOLD VALUE	MACHINE STATUS
127.0.0.1	0	UNCOMPROMSED
127.0.0.1	2	COMPROMSED
124.0.2.1	0	UNCOMPROMSED
124.0.2.2	5	COMPROMSED
124.0.2.1	0	UNCOMPROMSED

TABLE 3
NORMAL SPAM PERCENTAGE -40%

FROM IP	COUNT THRESHOLD	PERCENTAHGE THRESHOLD	MACHINE STATUS
127.0.0.1	0	0%	UNCOMPROMSED
127.0.0.1	7	95%	COMPROMSED
124.0.2.1	0	0%	UNCOMPROMSED
124.0.2.2	3	100%	COMPROMSED
124.0.2.1	0	0%	UNCOMPROMSED

V. EXPERIMENTAL AND RESULT

A mail system machines are involved in the mail transactions. The machine which is entering into the network will be monitored by the SPOT. It will monitor the spam messages sent by the system. If the message exceeded the level in the sense SPOT will do some process and decide that system as Spam Zombie. This detection is based on the outgoing messages. SPOT is a lightweight compromised machine detection system.

SPOT detection can be used to identify the compromised machine quickly. It also minimizes the number of required observations to detect a spam zombie. System administrators can automatically detect the compromised machines in their network in an online manner.

VI. CONCLUSION

In this paper, we developed an effective spam zombie detection system named SPOT. In the network the SPOT can be used to monitoring outgoing messages. SPOT was designed based on a simple and powerful statistical method named as Sequential Probability Ratio Test (SPRT). SPRT can be used to detect the compromised machines that are used to involve in the spamming activities. SPRT can be used to minimize the number of required observations to detect a spam zombie. SPOT is an effective and efficient system in automatically detecting compromised machines in a network. Also the SPOT outperforms two other detection algorithm based on the number and percentage of spam messages sent by an internal machine.

REFERENCES

[1] Chen.Z, Chen.C, and Ji.C, (2007) "Understanding Localized-Scanning Worms," Proc. IEEE Int'l Performance, Computing, and Comm. Conf. (IPCCC '07), 2007.

[2] Duan.Z, Gopalan.K, and Yuan.X, (2007) "Behavioral Characteristics of Spammers and Their Network Reachability Properties," Proc.IEEE Int'l Conf. Comm. (ICC '07), June 2007.

[3] Gu.G, Perdisci.R, Zhang.J, and Lee.W, (2008) "BotMiner: Clustering Analysis of Network Traffic for Protocol- and Structure-Independent Botnet Detection," Proc. 17th USENIX Security Symp. July 2008.

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

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

[6] John.J.P, Moshchuk.A, Gribble.S.D, and Krishnamurthy.A, (2009) "Studying Spamming Botnets Using Botlab," Proc. Sixth Symp.Networked Systems Design and Implementation (NSDI '09), Apr. 2009.

[7] Jung.J, Paxson.V, Berger.A, and Balakrishnan.H, (2004) "Fast Portscan Detection Using Sequential Hypothesis Testing," Proc. IEEE Symp.Security and Privacy, May 2004.

[8] Radosavac.S, Baras.J.S, and Koutsopoulos.I, (2005) "A Framework for MAC Protocol Misbehavior Detection in Wireless Networks,"Proc. Fourth ACM Workshop Wireless Security, Sept. 2005.

[9] Ramachandran.A and Feamster.N, (2006) "Understanding the Network- Level Behavior of Spammers," Proc. ACM SIGCOMM, pp. 291-302, Sept. 2006.

[10] Sanchez.F, Duan.Z, and Dong.Y, (2010) "Understanding Forgery Properties of Spam Delivery Paths," Proc. Seventh Ann. Collaboration, Electronic Messaging, Anti-Abuse and Spam Conf. (CEAS '10), July 2010.

[11] Xie.M, Yin.H, and Wang.H, (2008) "An Effective Defense against Email Spam Laundering," Proc. ACM Conf. Computer and Comm. Security, Oct. /Nov. 2006.

[12] Xie.Y, Xu.F, Achan.K, Panigrahy.R, Hulten.G, and Osipkov.I, (2008) "Spamming Botnets: Signatures and Characteristics," Proc. ACMSIGCOMM, Aug. 2008.