Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 1, January 2014, pg.106 – 110

RESEARCH ARTICLE

An Efficient Methodology for Detecting Spam Using Spot System

MRS. SARANYA.S[1], MRS. R.BHARATHI[2]

[1]M.TECH (Computer Science & Engineering), PRIST UNIVERSITY, Pondicherry
[2]Assistant Professor (Computer Science & Engineering), PRIST UNIVERSITY, Pondicherry
Email: [1] saranya.sekar@yahoo.com, [2] prist2009cse@gmail.com

***Abstract*—Major Security challenge on the Internet is the existence of the large number of compromised machines. Compromised machines on the Internet are generally referred to as bots, and the set of bots controlled by a single entity is called a botnet. Botnets have multiple nefarious uses: mounting Distributed denial of service attacks, stealing user passwords and identities, generating click fraud, and sending spam email. Compromised machines are one of the key security threats on the Internet. Given that spamming provides a key economic motivation for attackers to recruiting the large number of compromised machines, and focus on the detection of the compromised machines in a network that are involved in the spamming activities, commonly known as spam zombies. Develop an effective spam zombie detection system named SPOT by monitoring outgoing messages of a network. SPOT is designed based on a powerful statistical tool called Sequential Probability Ratio Test, which has bounded false positive and false negative error rates. The number and the percentage of spam messages originate by spam detection technique.**

Keywords—Compromised machines; spam zombies; spam detection techniques; spot detection system.

I. Introduction

Compromised machines involved in the spamming activities are referred as spam zombies. Network of spam zombies are recognized as one of the most serious security threats today. This paper presents the study of the discovery of e-mail spam using the spam detection technique.

Compromised machines are generally referred as bots and the set of bots that are controlled by a single entity are called botnets. In this, identifying and cleaning of compromised machines in a network remain a significant challenge for system administrators of network of all sizes. Botnet have multiple nefarious uses such as generating click fraud, stealing user passwords and identities and sending spam email. To aggregate the global characteristics of spamming botnets, we developed a tool for the administrators to automatically detect the compromised machines in a network in an online manner. Such machines have been increasingly used to launch various security attacks including spamming and spreading malware, DDoS, and identity theft .Compromised machines are one of the key security threats

on the Internet. On the other hand, identifying and cleaning compromised machines in a network remain a significant challenge for system administrators of networks of all sizes.

In this paper, 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 project, we will develop a spam zombie detection system, named SPOT, by observing outgoing messages. SPOT is planned based on a statistical method called Sequential Probability Ratio Test (SPRT), as a simple and powerful statistical method, SPRT has a number of attractive features. It reduces the expected number of observations required to reach a decision among all the sequential and non-sequential statistical tests with no greater error rates. This means that the SPOT finding system can identify a compromised machine quickly. SPOT can be used to test between two hypotheses whether the machine is compromised or not. SPOT is an effective and efficient system in automatically detecting compromised machines in a network. SPOT has bounded false positive and false negative error rates. It also decreases the number of required observations to detect a spam zombie.

II. Related work

Unsolicited commercial email, commonly known as spam, has become a pressing problem in today's Internet. In this paper we re-examine the architectural foundations of the current email delivery system that are responsible for the proliferation of email spam. We argue that the difficulties in controlling spam stem from the fact that the current email system is fundamentally sender-driven and distinctly lacks receiver control over email delivery. Based on these observations we propose a Differentiated Mail Transfer Protocol (DMTP).Botnets are now the key platform for many Internet attacks such as spam, distributed denial-of-service (DDoS), identity theft, and phishing. Most of the current botnet detection approaches work only on specific botnet command and control (C&C) protocols (e.g., IRC) and structures (e.g., centralized), and can become ineffective as botnets change their C&C techniques. In this paper, we present a general detection framework that is independent of botnet C&C protocol and structure, and requires no a priori knowledge of botnets (such as captured bot binaries and hence the botnet signatures, and C&C server names/addresses).

BotHunter is an application designed to track the two-way communication flows between internal assets and external entities, developing an evidence trail of data exchanges that match a state-based infection sequence model. In contrast to previous malware, botnets have the characteristic of a command and control (C&C) channel. Botnets also often use existing common protocols, e.g., IRC, HTTP, and in protocol-conforming manners. This makes the detection of botnet C&C a challenging problem. In this paper, we propose an approach that uses network-based anomaly detection to identify botnet C&C channels in a local area network without any prior knowledge of signatures or C&C server addresses. This detection approach can identify both the C&C servers and infected hosts in the network. Our approach is based on the observation that, because of the pre-programmed activities related to C&C, bots within the same botnet will likely demonstrate spatial-temporal correlation and similarity. Analysis of real world botnets indicates the increasing sophistication of bot malware and its thoughtful engineering as an effective tool for profit-motivated online crime.

We initial focus on the studies that utilize spamming activities to detect bots and then briefly discuss a number of efforts in detecting general botnets. Based on email messages received at a large email service provider, two recent studies [5], [6] investigated the aggregate global characteristics of spamming botnets including the size of botnets and the spamming patterns of botnets. These studies present 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. But, 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. In addition, their approaches cannot support the online detection requirement in the network environment considered in this project. We aim to develop a tool to assist system administrators in automatically detecting compromised machines in their networks in an online manner.

III. SPAM DETECTION TECHNIQUE

In this section, thus the spam zombies and spam messages can be identified. Normal Machine generates the original message. Original message enter in to the network and received by the server. Spam Zombie produces the spam messages and the spam message enters into the network. Server, first identifies the which message is Spam.

A. Detecting the Compromised Machines

Compromised machines are the machines that are involved in spamming activities. Compromised machines on the Internet are generally referred to as bots, and the set of bots controlled by a single entity is called a botnet. Botnets have been widely used for sending spam emails at a large scale, by programming a large number of distributed bots; spammers can effectively transmit thousands of spam emails in a short duration. To date, detecting and blacklisting individual bots is commonly regarded as difficult, due to both the transient nature of the attack and the fact that each bot may send only a few spam emails. Furthermore, despite the increasing awareness of botnet infection and their associated control process little effort has been devoted to understanding the aggregate behaviors of botnet from the perspective of large email servers that are popular targets of botnet spam attacks.

B. Spam Detection

The complete Spam Detection System is introduced here. In this section, first to capture the IP address of the system. Then the system mails are applied to filtering process. In this process, 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.

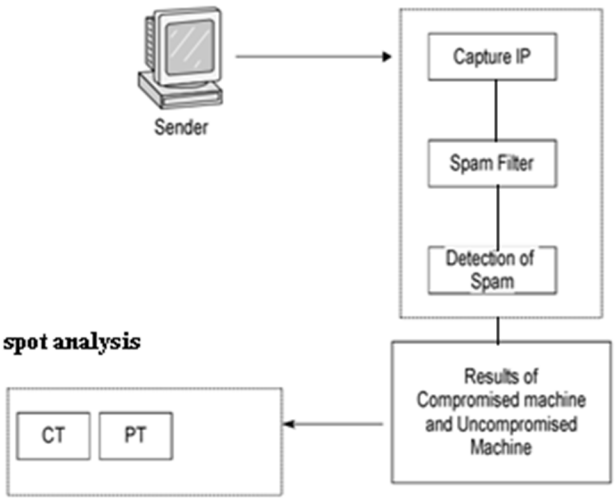

Figure: System Architecture

C. Spot Detection Algorithm

The SPOT detection algorithm, 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 non-spam by the (content-based) spam filter.

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, 3
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.
else 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

D. Percentage Count and Spam Based Technique

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 an internal machine, respectively. For simplicity, we refer to them as the count-threshold (CT) detection algorithm and the percentage-threshold (PT) detection algorithm respectively. 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 addition, choosing the proper values for the four user defined parameters (α, β, θ1, θ2) in SPOT is relatively straightforward. In contrast, selecting the "right" values for the parameters of CT and PT is much more challenging and tricky. The performance of the two algorithms is sensitive to the parameters used in the algorithm.

IV. PERFORMANCE EVALUATION

In this section, we evaluate the performance of the spam detection techniques based on a two-month e-mail trace collected on a large US campus network. We also study the potential impact of dynamic IP addresses on detecting spam messages.

Performance of Spot

In this section, we evaluate the performance of SPOT based on the collected FSU e-mails. In all the studies, we set α=0.01, β=0.01, θ1=0.9, and θ0=0.2. For example, there are FSU internal IP addresses observed in the e-mail trace. Out of the 132 IP addresses identified by SPOT, we can confirm 110 of them to be compromised in this way. For the remaining 22 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 (greater than 98 percent), we also claim that the corresponding machine has been confirmed to be compromised. We can confirm 16 of them to be compromised in this way. We note that the majority (62.5percent) of the IP addresses confirmed by the spam percentage are dynamic IP addresses, which further indicates the likelihood of the machines to be compromised. For the remaining six IP addresses that we cannot confirm by either of the above means, we have also manually examined their sending patterns.

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,, 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 (210) as being

compromised than SPOT, 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).

V. CONCLUSION

We proposed an 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. It also minimizes the number of required observations to detect a spam zombie. Our evaluation studies based on a two-month e-mail trace collected on the FSU campus network showed that SPOT is an effective and efficient system in automatically detecting compromised machines in a network. In addition, it showed that SPOT outperforms two other detection algorithms based on the number and percentage of spam messages sent by an internal machine, respectively.

We develop SPOT to assist system administrators in automatically detecting compromised machines in their networks in an online manner. Our main future objective is to extend these ideas to detect spam in sender itself and stop the user emails.

REFERENCES

[1] SpamAssassin, "The Apache SpamAssassin Project," http://spamassassin.apache.org, 2011.

[2] Zhenhai Duan, Senior Member,Peng Chen, Fernando Sanchez,Yingfei Dong, Member,Mary Stephenson, and James Michael Barker"Detecting Spam Zombies By Monitoring The Outgoing Messages"2012.

[3]L. Zhuang, J. Dunagan, D.R. Simon, H.J. Wang, I. Osipkov, G.Hulten, and J.D. Tygar, "Characterizing Botnets from Email Spam Records," Proc. First Usenix Workshop Large-Scale Exploits and Emergent Threats, Apr. 2008.

[4] K.V.SrinivasaRaoS.SrinivasuluandA.Amrutavallli"Detection of Spam through E-mail Abstraction Scheme" IJERT june 2012

[5] P. Bacher, T. Holz, M. Kotter, and G. Wicherski, "Know YourEnemy: Tracking Botnets," http://www.honeynet.org/papers/bots, 2011.

[6] "Botnet Analysis Using Command and ControlChannels" 15 December 2011 Carleton University

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

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

