SPAM ZOMBIE DETECTION BY ANALYZING OUTGOING MESSAGES

[1]Mrs. Chaitrali Chaudhari, [2]Ms. Sonali G. Doiphode

[1,2] *Computer Engineering, Mumbai University,(India)*

ABSTRACT

Compromised machines on the Internet are generally referred as bots, and the set of bots controlled by an entity is called a botnet. Botnets are used for different purposes like mounting DDoS,generating click fraud, stealing user passwords and identities, and sending spam email. Compromised machines are one of the key security threats on the Internet. The compromised machines in a network that are involved in the spamming activities, are commonly known as spam zombies.We have developed an effective solution for detecting spam zombies named "SPOT". 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. We also study two spam zombie detection algorithms based on the number and the percentage of spam messages originated by the machine.

Keywords: Compromised Machines, Spam Filter, Dynamic IP, Spam Zombies, Spot Detection System.

I INTRODUCTION

E-mail spam, also known as unsolicited bulk e-mail or unsolicited commercial e-mail, is the practice of sending unwanted e-mail messages frequently with commercial content in large quantities to an indiscriminate set of recipients. Spam is technically delivered the same way as legitimate e-mail,utilizes the Simple Mail Transfer Protocol (SMTP). Currently, a large fraction of spam comes from botnets,with the implication that e-mail spam detection is an effective strategy for subsequent botnet detection. Botnet is the serious threat which occurs commonly in today's cyber-attacks and cybercrimes. Botnet are designed to perform predefined functions in an automated fashion, where these malicious activities ranges from online searching of data, accessing lists, moving files sharing channel information to DDoS attacks against click fraud,critical targets, phishing, etc. Existence of command and control(C&C) infrastructure makes the functioning of Botnet unique; in turn throws challenges in the mitigation of Botnet attacks.

In this paper, we focus on the detection of the compromised machines in a network that are used for sending spam messages, referred to as spam zombies. Two natures of the compromised machines on the Internet—sheer volume and widespread—render many existing security countermeasures less effective and defending attacks involving compromised machines extremely hard. A number of recent research efforts have studied the aggregate global characteristics of spamming botnets (networks of compromised machines involved in spamming) such as spamming patterns of botnets and the size of botnets.

Instead of studying aggregate global characteristics of spamming botnets, we develop a tool for system administrators to automatically detect the compromised machines in their networks in an online manner. In this paper, we develop a spam zombie detection system, called as SPOT, by monitoring outgoing messages. SPOT system is designed based on a statistical tool called Sequential Probability Ratio Test (SPRT), developed by Wald[1]. As a simple and powerful statistical method, SPRT has many desirable features. It minimizes the required number of observations for decision among all the sequential and non-sequential statistical tests less error rates. This means that the SPOT detection system can identify a compromised machine quickly. Both the false positive and false negative probabilities of SPRT can be bounded by user-defined thresholds.

II OBJECTIVE

The main objective of this paper is effectively identifying the Spam zombies using SPOT without any botnet spam signatures techniques in a network. Detection of Spam messages will minimize the network traffic and detect the compromised machines which produce the spam messages.

III LITERATURE SURVEY

In this section, we discuss related work in detecting compromised machines. Two recent studies [2], [3] investigated the aggregate global characteristics of spamming botnets including the size of botnets and the spamming patterns of botnets. These studies provide 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. These approaches are better suited for large e-mail service providers to understand the aggregate global characteristics of spamming botnets instead of being deployed by individual networks to detect internal compromised machines.Also they can not support online detection.

DBSpam tool developed by Xie et al. detect proxy-based spamming activities in a network relying on the packet symmetry property of such activities [4],Not only the spam proxies but we want to detect all types of compromised machines which are involved in spamming.

Here we have few botnet detection schemes. Gu et al., developed BotHunter [5] detects compromised machines by correlating the IDS dialog trace in a network. BotHunter which relies on the specifics of the malware infection process,while SPOT focuses on the economic incentive behind many compromised machines and their involvement in spamming.

An anomaly-based detection system named BotSniffer [6] identifies botnets by exploring the spatial-temporal behavioral similarity commonly observed in botnets. It focuses on HTTP-based and IRC-based botnets.

BotMiner [7] is both structure and protocol independent. In BotMiner, flows are classified into groups based on similar malicious activity patterns and similar communication patterns. 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 lightweight compromised machine detection system.

IV SEQUENTIAL PROBABILITY RATIO TEST (SPRT)

SPOT is designed based on the statistical tool Sequential Probability Ratio Test(SPRT). SPRT can be considered as an one-dimensional random walk with two user-specified boundaries corresponding to the two hypotheses. When the samples of the concerned random variable arrive sequentially, it moves either upward or downward direction, depending on the value of the observed sample. When the walk reaches or crosses either of the boundaries for the first time, it terminates and the corresponding hypothesis is selected.SPRT has many desirable features. One is both the actual false positive and false negative probabilities of SPRT can be bounded by the user-specified error rates. A smaller error rate tends to require a larger number of observations before SPRT ends. Thus, users can balance the performance and cost (in terms of number of required observations) of an SPRT test. Another is, SPRT minimizes the average number of the required observations for reaching a conclusion for a given error rate.

V SPAM ZOMBIE DETECTION ALGORITHMS

5.1 Spot Detection Algorithm

SPOT is designed based on the statistical tool SPRT. In SPOT, we consider H_1 as a compromised and H_0 as normality. That is, H_1 is true if the machine is compromised, and H_0 is true if it is normal. In addition, we let X_i =1 if the ith message from the concerned machine in the network is a spam, and X_i = 0 otherwise. When an outgoing message arrives at the SPOT system, sender machine IP address is recorded. Then using content-based spam filter message is classified as either ham or spam. Let X_i =1 if the ith message from the concerned machine in the network is a spam, and X_i = 0 otherwise. Spot maintains the logarithm value of the corresponding probability ratio Λn per IP address from where messages are received. Once a machine is identified as being compromised it is added into the list of potentially compromised machines that system administrators can go further for cleaning. The SPOT detection system does not need to further monitor the message sending behavior of the compromised machine. On the other hand, a machine which is currently normal may get compromised at a later time. Therefore, machines that are detected to be normal by SPOT need continuous monitoring. Once such a machine is identified by SPOT, the records of the machine in SPOT are reset so that a new monitoring phase starts for the machine.

5.2 Spam Count and Percentage-Based Detection Algorithms

We present two different algorithms in detecting spam zombies, first one based on the number of spam messages and another the percentage of spam messages sent from an internal machine, respectively. We call them as the count-threshold (CT) detection algorithm and the percentage-threshold (PT) detection algorithm.

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.The system observes 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 e-mail sending properties of each internal machine in each time window: one is the percentage of spam messages sent by a machine, another is the total number of messages.Let n and N denote the spam messages and total messages originated from a machine m within any time window, respectively. If N >=Ca and n/N > P ,then PT declares machine m as being compromised, where Ca is the minimum number of messages machine must send and P is the user-defined maximum spam percentage of a normal machine.

VI RESULT

Thus the system is properly detected and analyzed using various techniques mentioned above. An effective and efficient system in automatically detecting compromised machines in the network is achieved successfully. The machine which is entering into the network will be observed 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 detection system can identify a compromised machine using minimum number of observations. It also minimizes the number of required observations to detect a spam zombie. . SPOT is a lightweight compromised machine detection system

VII CONCLUSION

In this paper, we have discussed an effective spam zombies detection system called SPOT for detecting an compromised machine in a network. SPOT is designed based on the statistical tool Sequential Probability Ratio Test(SPRT).It also minimizes the number of required observations to detect a spam zombie.In addition we also studied two other spam zombie detection algorithms based on number of spam message and percentage of spam message forwarded by internal machines.

REFERENCES

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

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

[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] M. Xie, H. Yin, and H. Wang, “An Effective Defense against Email Spam Laundering,” Proc. ACM Conf. Computer and Comm. Security, Oct./Nov. 2006.

[5] 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.

[6] 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.

[7] 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.