Modeling and Detection of Camouflaging Worm

[1]Mr. M.A.BASEER [2]Mr. POTU NARAYANA [3]Mr. SANDEEP BAJRANGRAO LAHANE

[1]Associate professor, CSE department HOD, SANA ENGINEERING COLLEGE
[2]Assistant Professor, Sana Engineering College
[3]Assistant Professor, CSE Department G.P.Beed, E-mail:sucsess711@gmail.com

ABSTRACT*-In this paper, we investigate a new class of active worms, referred to as Camouflaging Worm (C-Worm in short). C-Worm is different from traditional worms because of its ability to intelligently manipulate its scan traffic volume over time. By this means, the C-Worm camouflages its propagation from existing worm detection systems based on analyzing the propagation traffic generated by worms. We analyze the characteristics of C-Worm and conduct a comprehensive comparison between its traffic and non-worm traffic (background traffic). We observe that the two types of traffic are barely distinguishable in the time domain. But, their distinction is clear in the frequency domains, due to the recurring manipulative nature of the C-Worm. Motivated by our observation, we design a novel spectrum-based scheme to detect the C-Worm. Our scheme uses to Power Spectral Density (PSD) distribution of the scan traffic volume and its corresponding Spectral Flatness Measure (SFM) to distinguish the C-Worm traffic from background traffic. Then the performance data clearly demonstrates that our scheme can effectively detect the C-Worm propagation. Furthermore, we show that the generality of our spectrum-based scheme in effectively detecting not only the C-Worm, but the traditional worms as well.*

1. INTRODUCTION

In this system, the detection is commonly based on the self propagating behavior of worms that can be described as follows: after a worm-infected computer identifies and infects a vulnerable computer on the Internet, this newly infected computer will be automatically and continuously scan several IP addresses to identify and infect other vulnerable computer. As such, numerous existing detection schemes are based on a tacit assumption that each worm-infected computer keeps scanning the Internet and propagates itself at the highest possible speed.

Furthermore, it has been shown that the worm scan traffic volume and the number of worm-infected computers exhibit exponentially increasing pattern. Nevertheless, the attackers are crafting attack strategies that intend to defeat existing worm detection systems. In particularly, ‘stealth’ is one attack strategy used by a recently-discovered active worm called “Attack” worm and the “self-stopping” worm circumvent detection by hibernating (i.e., stop propagating) with a pre-determined period. Worm might also uses the evasive scan and traffic morphing technique to hide the detection In this paper, we conduct systematic study on a new class of such smart-worms denoted as Camouflaging Worm (C-Worm in short).

The C-Worm has a self-propagating behavior similar to traditional worm i.e., it intends to rapidly infect as many vulnerable computers as possible. But, the C-Worm is quite different from traditional worms in which it camouflages any noticeable trends in the number of infected computers over time. Then the camouflage is achieved by manipulating the scan traffic volume of worm-infected computers. Such, manipulation of the scan traffic volume prevents exhibition of any exponentially increasing trends or even crossing of thresholds that are tracked by existing detection schemes.

We note that the propagation controlling nature of the C-Worm (and similar smart-worms, such as “Atak”) cause a slowdown in the propagation speed. By carefully controlling its scan rate, the C-Worm can: (a) still achieve its ultimate goal of infecting as many computers as possible before being detected, and (b) position itself to launch subsequent attacks.

2. MODELING OF THEC-WORM

2.1 C-Worm

The C-Worm camouflages its propagation by controlling scan traffic volume during its propagation. Simplest way to manipulate the scan traffic volume is to randomly change the number of worm instances conducting port-scans. As other alternatives, worm attackers may use an open loop control (no feedback) mechanism by choosing a randomized and time-related pattern for the scanning and infection in order to avoid being detected. Nevertheless, a open-loop control approach raises some issues of the invisibility of the attacks. First, as we know, worm propagation over the Internet can be considered dynamic systems. When an attacker launches worm prolongations, it is very challenging for the attacker to know the accurate parameters for worm propagation dynamics over the Internet.

Given the inaccurate knowledge of worm propagation over the Internet, the open-loop control systems will not be able to stabilize the scan traffic. Consequently, the overall worm scan traffic volume in the open-loop control system will expose a much higher probability to show an increasing trend with the progress of worm propagations. Hence, we consider the C-worm as a worst case attacking scenario that uses a closed-loop control for regulating the propagation speed based on the feedback propagation status.

In order to effectively evade detections, the overall scan traffic for the C-Worm should be comparatively slow and variant enough to not show any notable increasing trends over time. On the other hand, very slow propagation of the C-Worm is also not desirable, since it delay rapid infection damage to the Internet. Hence, the C-Worm needs to be adjusting its propagation so that it is neither too fast to be easily detected, nor too slow to delays rapid damage on the Internet. To control the C-Worm scan traffic volume, we introduce the control parameter called attack probability P (t) for each worm-infected computers. P (t) is the probability that a C-Worm instance participates in the worm propagation (i.e., scans and infects other computers) at time t.

For the C-Worm, P (t) needs not be a constant value and can be set as a time-varying function. In order to achieve its camouflaging behaviors, the C-Worm needs to obtain an appropriate P (t) to manipulate its scan traffic. Specifically, the C-Worm will regulate its overall scan traffic volume such that: 1) it is similar to non worm scan traffic in terms of the scan traffic volume over time. 2) it does not exhibits any notable trend, such as an exponentially increasing pattern or any mono-increasing pattern even when the number of infected hosts increases over time and 3) the average value of the overall scan traffic volume is sufficient to make the C-Worm propagate fast enough to cause rapid damage on the Internet.

2.2 Effectiveness of the C-Worm

We now demonstrate the effectiveness of the C-Worm in evading worm detection through controlling P(t). Given random selection of Mc, we generate three C-Worm attacks (viz., C-Worm 1, C-Worm 2, and C-Worm 3) that are characterized by different selections of mean and variance magnitudes for MC. In our simulations, we assume that the scan rate of the traditional PRS worm follow a normal distribution S_n= N(40, 40) (note that if the scan rate is generated by above distribution is less than 0, we set the scan rate as 0). We are also set the total number of vulnerable computers on the Internet as 360000, which is the total number of infected computer in “Code-Red” worm incident.

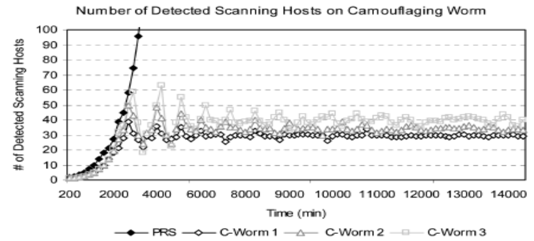

Fig 2.1 Observed infected instance number for the C-Worm and PRS worm.

Fig. 2.1 shows the observed number of worm-infected computers over time for the PRS worm and the above three C-Worm attacks. Fig. 2.2 shows the infection ratio (IR) for the PRS worm and the above three C-Worm attacks. These simulations are for a worm detection system discussed in Section 2.2 that covers the 2^{20} IPv4 address space on the Internet. Then the reason for choosing 2^{20} IP addresses as the coverage space of the worm detection system is due to the fact that the SANs ISC and a representative Internet Threat Monitoring (ITM) system, has similar coverage space.

In the ITM systems, a large number of monitors are commonly deployed all over the Internet and each monitor collects the traffic directed to a small set of IP address spaces, which are not commonly used (also called dark IP address). Therefore, the address space of ITM system is not a narrow range address space, rather than a large number of small chunks of addresses randomly spread across the global IP address space.

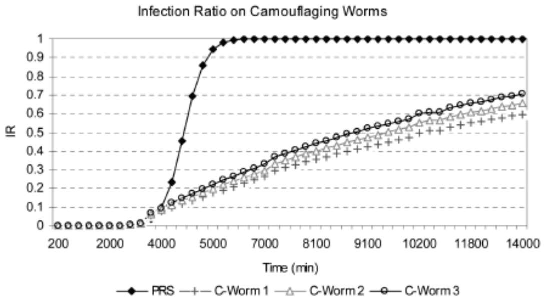

Fig 2.2 Infected ratio for the C-Worm and PRS worm.

For the C-Worm, the trend of observed number of worm instances over time (M_A (t)) ismuch different from that of the traditional PRS worm, as shown in Fig. 2.2. This clearly demonstrates how the C-Worm successfully camouflages its increase in the number of worm instances (M_A (t)) and avoids detection by worm detection systems that expect exponential increases in worm instance numbers during large-scale worm propagation. Fig. 2.3 shows the number of scanning computers from normal nonworm port-scanning traffic (background traffic) for several well-known ports, (i.e., 25, 53, 135, and 8080) obtained the over several months by the ISC. Comparing Fig. 2.3

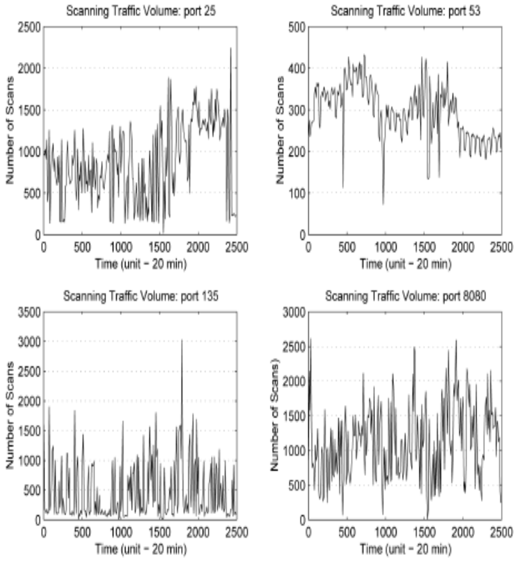

Fig 2.3 Observed infected instance number for background scanning reported by ISC.

3. DETECTING THE C-WORM

3.1 Design Rationale

In this section, we develop a novel spectrum-based detection scheme. Recall the C-Worm goes undetected by detection schemes that try to determine the worm propagation only in the time domains. Our detection scheme captures the distinct pattern of the C-Worm in the frequency domains and thereby has the potential of effectively detecting the C-Worm propagations. In order to identify the C-Worm propagation in the frequency domain, then we use the distribution of PSD and its corresponding SFM of the scan traffic. Particularly, PSD describes the, how the power of a time series is distributed in the frequency domain.

Mathematically, it is defined as the Fourier transform of the autocorrelation of a time series. In our case, the time series correspond to the changes in the number of worm instances that actively conduct scans over time. The SFM of PSD was defined as the ratio of geometric mean to arithmetic mean of the coefficients of PSD. The range of SFM values is [0, 1] and a larger SFM value implies flatter PSD distribution and vice versa. To illustrate the SFM values of both the C-Worm and normal nonworm scan traffics. We plot the PDF of SFM for both C-Worm and normal nonworm scan traffic, as shown in Figs. 3.1 and 3.2, respectively. The normal nonworm scan traffic data shown in Fig. 3.2 is based on real-world traces collected by the ISC.6 Note that, we only shows the data for port 8080 as an example, and other port show similar observations.

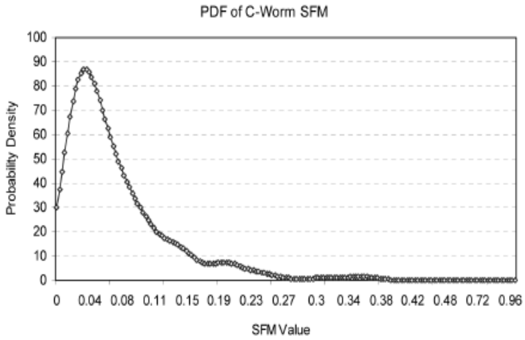

Fig3.1 PDF of SFM on C-Worm traffic.

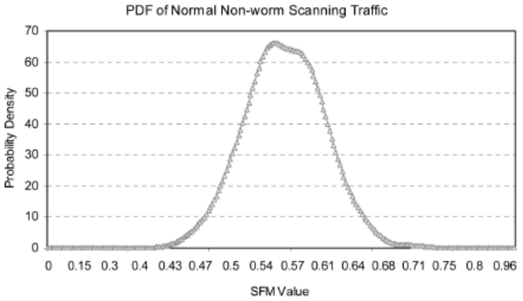

Fig3.2 PDF of SFM on normal nonworm traffic

From this figure, we know that the SFM value for normal nonworm traffic is very small. The C-Worm data shown in Fig. 3.1 is based on 800 C-Worms attacks generated by varying attack parameters defined in Section 2, such as P (t) and M_c (t). From this figure, we know the SFM value of the C-Worm attacks is high. From the above two figures, we can observe that there is a clear demarcation range of SFM ∈ (0.3, 0.38) between the C-Worm and normal nonworm scan traffic. As such, the SFM can be used to sensitively detect the C-Worm scan traffic.

The large SFM values of normal nonworm scan traffic can be explained as follows: The normal nonworm scan traffic does not tend to concentrate at any particular frequencies, since its random dynamics is not caused by any recurring phenomenons. The small value of SFM can be reasoned by fact that the power of C-Worm scan traffic is within a narrowband frequency range. Such concentrations within a narrow range of frequencies are unavoidable, since the C-Worm adapts to the dynamics of the Internet in a recurring manner for manipulating the overall scan traffic volume. In certainly, the above recurring manipulations involve steady increase followed by a decrease in the scan traffic volume.

Notice that the frequency-domain analysis will require more samples in comparison with the time-domain analysis, since the frequency-domain analysis techniques. Such as the Fourier transform need to derive power spectrum amplitude for different frequencies. In order to generate the accurate spectrum amplitude for relatively high frequencies and a high granularity of data sampling will be required. In our case, we rely on ITM systems to collects traffic traces from monitors (motion sensors) in a timely manner. Enabling the ITM system with timely data collection will benefit worm detection in real time.

3.2 Spectrum-Based Detection Scheme

We now present the details of our spectrum-based detection scheme. To understand how the destination count data is obtained, we recall that an ITM system collect logs from distributed monitors across the Internet. On a side note, ITM systems are widely deployed facility to analyze, detect, and characterize dangerous Internet threats such as worms. In general, an ITM system consists of one centralized data center and a number of monitors distributed across the Internet. Each monitor record traffic that addressed to a range of IP addresses (which are not commonly used IP address also called the dark IP addresses) and periodically sends the traffic logs to the data center. The data center analyzes the collected traffic LOGS and publishes reports (e.g., statistics of monitored traffic) to ITM system users. Hence, the baseline traffic in our study is scan traffic. With reports in a sampling window Ws and the source count X(t) is obtained by counting the unique source IP addresses in received logs. To conducts spectrum analysis, we consider the detection sliding window W_d in the worm detection system. W_d consists of q(>1) continuous detection sampling windows and each sampling window lasts Ws. Then the detection sampling window is the unit time interval to sample the detection data (e.g., the destination count). Hence, at time i, within a sliding window Wd, there are q samples denoted by (X(i – q- 1), X(i- q- 2), . . .X(i)), where X(i- j- 1) (j ∈ (1, q) is the jth destination count from time (i – j- 1) to (i- j).

In our spectrum-based detection schemes, the distribution of PSD and its corresponding SFM are used to distinguish the C-Worm scan traffic from the nonworm scan traffic. Recall the definition of PSD distribution and its corresponding SFM are introduced in Section 3.1. In our worm detection scheme, the detection data (e.g., destination counter) is advance processed in order to obtain its PSD and SFM. In the following, we detail how the SFM and PSD are determined during the processing of the detection data.

4. CONCLUSION

We studied a new class of smart-worm called a C-Worm, which has the capability to camouflage its propagation and further avoid the detection. Our research showed that, although the C-Worm successfully camouflages its propagation in that time domain, its camouflaging nature inevitably manifests as a distinct pattern in the frequency domain. Based on the observation, we developed a novel spectrum-based detection scheme to detect the C-Worm. Our evaluation of data showed that our scheme achieved superior detection performance against the C-Worm in comparison with existing representative detection scheme. This paper lays the foundation for ongoing studies of “smart” worms that intelligently adapt their propagation patterns to reduce the effectiveness of countermeasures.

5. REFERENCES

[1] Modeling and Detection of Camouflaging Worm IEEE TRANSACTIONS ON EPENDABLE AND SECURE COMPUTING, VOL. 8, NO. 3, MAY/JUNE 2011.

[2] D. Moore, C. Shannon, and J. Brown, “Code-Red: A Case Study on the Spread and Victims of an Internet Worm,” Proc. Second Internet Measurement Workshop (IMW), Nov. 2002.

[3] D. Moore, V. Paxson, and S. Savage, “Inside the Slammer Worm,” Proc. IEEE Magazine of Security and Privacy, July 2003.

[4] CERT, CERT/CC Advisories, http://www.cert.org/advisories/, 2010.12 IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, VOL. 8, NO. 3, MAY/JUNE 2011

[5] P.R. Roberts, Zotob Arrest Breaks Credit Card Fraud Ring, http:// www.eweek.com/article2/0,1895,1854162,00. asp, 2010.

[6] W32/MyDoom.B Virus, http://www.us-cert.gov/cas/techalerts/ TA04-028A.html, 2010.

[7] W32.Sircam.Worm@mm, http://www. symantec.com/avcenter/venc/data/w32.sircam. worm@mm.html, 2010.

[8] Worm.ExploreZip, http://www.symantec. com/avcenter/venc/ data/worm.explore. zip. html, 2010.

[9] R. Naraine, Botnet Hunters Search for Command and Control Servers, http://www.eweek.com/article2/0,1759,1829347,00.asp, 2010.

[10] T. Sanders, Botnet Operation Controlled 1.5m PCs Largest Zombie Army Ever Created, http://www.vnunet.com/vnunet/news/ 2144375/botnet-operation-ruled-million, 2005.

Author Profile’s

Mr. M.A.Baseer is currently working as Associate professor & HOD In the department Of C.S.E at SANA ENGINEERING COLLEGE, kodad, Nalgonda Dist, AndhraPradesh. HE Completed his M.Tech degree JNTU–Hyderabad in 2010. He has more than eight years of experience in teaching. He assisted more number of M.tech students for successful Completion of their project works.

Mr. Potu Narayana received B.Tech degree was completed from Sri Venkateswara Eng College and M.Tech degree was completed from unvi of hyder.

Mr. Sandeep Bajrangrao Lahane received B.E(CSE) degree was completed from Babasaheb Ambedkar University(BAMU), Aurangabad and M.Tech(CSE) degree was completed from SANA Engineering College, kodad, nalgonda Dist., Hyderabad. Working as Lecture (CSE) at Govt. Polytechnic Beed, Dist Beed, Maharashtra.