Modeling and Detection of Camouflaging Worms-Using SDF

Aruna.A
Research Scholar, Department of M.C.A, VELS University, Pallavaram.

Perumal.S
Asst.Professor & HOD , Department of C.S, VELS University, Pallavaram..

Dr.A.Muthukumaravel
Asst.Professor, Department of M.C.A, VELS University, Pallavaram.

Abstract— Self-duplicating, self-propagating malicious codes known as computer worms spread themselves without any human interaction and launch the most destructive attacks against computer network. Active worm's is also one type of worm. Active worms major security threats to the Internet. This is due to the ability of active worms to propagate in an automated fashion as they continuously compromise computers on the Internet. Active worms evolve during their propagation, and so, pose great challenges to defend against them. In this paper, we investigate a new class of active worms, referred to as Camouflaging Worm (C-Worm in short). In this paper, we investigate a new class of active worms, referred to as Camouflaging Worm (C-Worm in short). The C-Worm is different from traditional worms because of its ability to intelligently manipulate its scan traffic volume over time. Thereby, the C-Worm camouflages its propagation from existing worm detection systems based on analyzing the propagation traffic generated by worms. We analyze characteristics of the C-Worm and conduct a comprehensive comparison between its traffic and nonworm traffic (background traffic). We observe that these two types of traffic are barely distinguishable in the time domain. However, their distinction is clear in the frequency domain, due to the recurring manipulative nature of the C-Worm. Motivated by our observations, we design a novel spectrum-based scheme to detect the C-Worm. Our scheme uses the 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. Using a comprehensive set of detection metrics and real-world traces as background traffic, we conduct extensive performance evaluations on our proposed spectrum-based detection scheme. The performance data clearly demonstrates that our scheme can effectively detect the C-Worm propagation. Furthermore, we show the generality of our spectrum-based scheme in effectively detecting not only the C-Worm, but traditional worms as well.

Keywords — Anomaly detection, C-worm, Containment, Code Red, Malicious code.

I. INTRODUCTION

Active worm's refers to malicious software program that propagates itself on the internet to infect other computers. Self-propagating malicious codes known as computer worms spread themselves without any human interaction and launch the most destructive attacks against computer networks. Being fully automated, a worm's behavior is usually repetitious and predictable, making it possible to be detected. A worm's life consists of the following phases - Worm's target finding, Worm's propagation scheme.Worm's Transmission scheme, Worm's payload format.After detecting any worm we will have to make containment which is known as recovery. The worms include various names Nimda, Code Red, Slammer, Witty and Sasser. These active worms will cause following infections,

- Launch massive distributed denial of service (DDoS) attacks against internet utilities.
- These attacks will access confidential information that can be misused through large scale traffic sniffing, key logging and identify theft etc.
- Destroy data that has high monetary value.
- Distribute large scale emails (spam) or software to other PC'S in a network.

Due to substantial damage caused by worms in past years, there have been significant efforts on developing detection and defense mechanism against worms. There are two detection techniques still now, Network based

detection scheme. IN this method we have to detect malwares which are present on the network. In a network, there is client, server and third party. When attacker attacks on client, client start executing its antivirus software which is installed on server, then server identifies whether malware is present or not. By using NBDS, only external attacks can be found out but if ant host connected inside.

LAN attacks, then these NBDS can't detect it.Examples are Honeypot, Polygraph, Hamsa. Another part of detection is Host based detection scheme .Here we detect worms by monitoring, collecting, analyzing worm behavior's on host.

By using HBDS, Detection of internal as well as external attacks can be done.

II. RELATED WORK

Worms are similar to biological viruses that cause damage to health of human beings and other animals. They have features like self-propagation and replication. These features are with malicious programs that cause problems to computers in the given network. Such malware is known as worms. The worm which is scanning traffic through IP address and also port number of systems and trying to propagate itself to new systems in the networking domain is known as active worm. As the worms are capable of damaging IT systems, the need for research to prevent the same has been felt. In accordance with this, researchers spend considerable time on this topic and still it needs further improvements [9], [16]. Active worms can use many ways in which they can propagate themselves from one system to another system. One such way is Pure Random Scan (PRS). This is a kind of scan in which the worms continuously and randomly find IP addresses an ports of other systems and propagate itself to those systems which IP addresses are known to the worm. Other ways in which worms can propagate include file sharing, email, network port scanning and instant messaging or chatting [17].

When some IP addresses are known to the worms, they try to propagate themselves by maintaining a hit list and following the strategies to propagate themselves into those systems whose IP addresses are scanned by worm. The worms also split IP address space in order to avoid repetition of work and thus they divide and conquer in terms of scanning and propagating themselves into new networks. Some research also considered developing a new topology that is attack resilient with respect to works [18], [19].

There is a special category of worm that is quite different from the other worms described above. This worm is capable of manipulating its scan traffic and thus making it possible that the traditional worm detecting systems fail to help in this regard. A new camouflaging worm thus created is causing more damage to IT world as it is not detected by conventional anti-worm programs. Essentially the worms that hide their scan traffic are polymorphic in nature [20], [21]. Such worms are known as Camouflaging worms as they are hiding their presence and making the normal worm detection systems vulnerable. With respect to stealthiest, the normal worm and C-worm are having certain similarities. Both are generating same traffic and finding the similarities such as both can detect the difference between the normal traffic and worm's scan traffic. The other main difference between them is that the traditional worm detectors can't find difference while the proposed scheme can distinguish the traffic of the C-worm in the time domain. However, it is challenging to find such result from other schemes. The proposed scheme finds the difference in scan traffic of normal worm and systematic in frequency domain though in time domain it can't differentiate the existing worms and new kind of worm known in frequency domain. The new class of worm is named "C-Worm". Due to self propagation nature of C-worm and its ability to manipulate to hide its presence in the system by camouflaging technique. The actual detection of worms is provided in the next section.

III. MODELING OF C-WORM

A. C-Worm

The C-Worm modeling is based on our observations that have been made after some research. The C-worm block diagram is shown in fig. 1. The initial research revealed that the C-Worm is not same as other worms though it has similarities with normal worms. The normal worms perform scan traffic in order to replicate themselves and also propagate from one system to another system in a network environment. The same is followed by C-Worms also. However, there are two observations made clearly. The first observation is that, the C-Worm scan traffic involves IP addresses and port numbers and scan traffic is different from normal worms. The second observation is that the detection systems can't find the difference between scan traffic of C-worms and normal worms in terms of frequency domain. In time domain they appear to be same. The second observation also reveals that it is essential to differentiate the C-Worm traffic from other worm's traffic only in frequency domain. Based on these observations, our experiments are made. Our experiments focused on the

traffic analysis and frequency domain and the results revealed that our scheme is capable of detecting C-Worms. When our scheme launches, it analyzes the dynamics of C-Worm traffic in Internet. It follows a theory known as control system theory [27]. In order to demonstrate effectiveness of the proposed scheme, the overall traffic flow of C-Worm should be slow so as to show the detection process effectively. Control parameters are introduced to this effect such as attack probability on each infected computer. This indicates the probability in which C-Worm participates in the propagation of the worm. The control parameter in our model is generic in nature and its value is 1 indicating traditional worms and other value for C-Worms. In the process of modeling camouflaging worm, the following characteristics are followed.

- The traffic of C-worm is similar to non-worm traffic in terms of time domain. This means that over a period of time the scan traffic of the normal worm and C-worm is same.
- C-Worm does not show any trends while its propagation so as to hide its presence effectively.
- The average traffic of Worm is sufficient to model the C-Worm propagation model faster in order to cause rapid damage on the Internet.

We assume that the worm attacker manipulates scan traffic and the scan traffic of C-Worm follows different random distribution means.

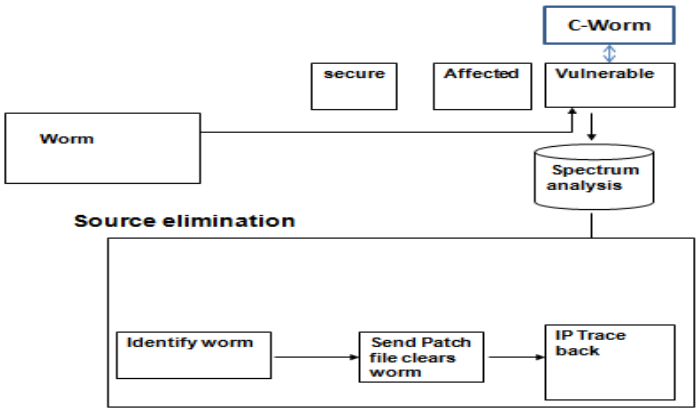

Figure-1: General Architecture

B. Propagation model of the C-Worm

To analyze the C-Worm, we adopt the epidemic dynamic model for disease propagation, which has been extensively used for worm propagation modeling .This model matches the dynamics of real worm propagation over the Internet quite well. Since our investigated C-Worm is a novel attack, we modified the original Epidemic dynamic formula to model the propagation of the C-Worm by introducing the p(t) the attack probability that a worm-infected computer participates in worm propagation at time t. We note that there is a wide scope to notably improve our modified model in the future to reflect several characteristics that are relevant in real- world practice Particularly, the epidemic dynamic model assumes that any given computer is in one of the following states: immune, vulnerable, or infected. An immune computer is one that cannot be infected by a worm; a vulnerable computer is one that has the potential of being infected by a worm; an infected computer is one that has been infected by a worm. The simple epidemic model for a finite population of traditional PRS worms can be expressed as4,

$$\frac{dM(t)}{dt} = \beta \cdot M(t) \cdot [N - M(t)]$$

where M(t) is the number of infected computers at time t; N(= T · P1 · P2) is the number of vulnerable computers on the Internet; T is the total number of IP addresses on the Internet; P1 is the ratio of the total number of computers on the Internet over T ; P2 is the ratio of total number of *vulnerable* computers on the Internet over the total number of computers on the Internet; β = S/V is called the pair wise infection rate, S is the scan rate defined as the number of scans that an infected computer can launch in a given time interval. We assume that at t = 0, there are M(0) computers being initially infected and N−M (0) computers being susceptible to further worm infection.

C-Worm has a different propagation model compared to traditional PRS worms because of its P(t) parameter[13]. Consequently, Formula (1) needs to be rewritten as,

$$\frac{dM(t)}{dt} = \beta \cdot M(t) \cdot p(t) \cdot [N - M(t)]$$

C. Effectiveness of the C-Worm

The system is highly efficient in its working and detecting the C-Worm propagation. We get the accurate result in the form of graph and observes the infected instance number for the C-Worm and PRS Worm. It gives a detailed report of infected ratio of the C-Worm and PRS Worm and also gives observed result of infected instance number for background scanning report by ISE in the form of graphs. In contrast, the C-Worm tries to manipulate the scan traffic pattern to avoid detection(refer section 4)

D. Detecting the C-Worm

In this section, we develop a novel spectrum-based detection scheme[13]. Recall that the C-Worm goes undetected by detection schemes that try to determine the worm propagation only in the time domain. Our detection scheme captures the distinct pattern of the C-Worm in the frequency domain, and thereby has the potential of effectively detecting the C-Worm propagation. In order to identify the C-Worm propagation in the frequency domain, we use the distribution of *Power Spectral Density* (*PSD*) and its corresponding *Spectral Flatness Measure* (*SFM*) of the scan traffic. Particularly, *PSD* describes how the power of a time series is distributed in the frequency domain. Mathematically, it is defined as the *Fourier* transform of the auto-correlation of a time series. In our case, the time series corresponds to the changes in the number of worm instances that actively conduct scans over time. The *SFM* of *PSD* is 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.

We now present the details of our spectrum-based detection scheme. Similar to other detection schemes [14][15], we use a “source count” as the basis for worm detection in our spectrum-based detection scheme. The ”source count” is the number of the unique sources that launch scans during worm propagation. To understand how the source count data is obtained[13].To illustrate *SFM* values of both the C-Worm and normal non-worm scan traffic, we plot the *Probability Density Function* (PDF) of *SFM* for both C-Worm and normal non-worm scan traffic as shown in Figs. 3 and Fig. 4(in section 4) , respectively. To obtain the *PSD* distribution for worm detection data, we need to transform data from the time domain into the frequency domain[13].

i. Power Spectral Density (PSD)

To obtain the PSD distribution for worm detection data, data in time domain is transformed into the frequency domain. The PSD function of the scan Transform (DFT) of its autocorrelation function. As the PSD inherently captures any recurring pattern in the frequency domain, the PSD function shows a comparatively even distribution across a wide spectrum range for the normal non-worm scan traffic. The PSD of C-Worm scan traffic shows spikes or noticeably higher concentrations at a certain range of the spectrum.

$$\Phi(R[L],k) = \Sigma_{n=0}^{N-1} (E[X(t)X(t+L)]). e^{-j2\pi Kn/N}$$

Where k=0,1,…,N-1.X(t) is the random process to model the worm detection data and R[L] is the correlation of worm detection data in an interval.PSD tells that at which frequency ranges variations are strong and that might be quite useful for further analysis. The concept and use of the power spectrum of a signal is fundamental in electrical engineering, especially in electronic communication systems, including radio communications, radars, and related systems, plus passive [remote sensing] technology.

ii. Spectral Flatness Measure (SFM)

We measure the flatness of PSD to distinguish the scan traffic of the C-Worm from the normal nonworm scan traffic. For this, we introduce the SFM, which can capture anomaly behavior in certain range of frequencies. The SFM is defined as the ratio of the geometric mean to the arithmetic mean of the PSD coefficients. In statistical signal processing and physics, the spectral density, power spectral density (PSD), or energy spectral density (ESD), is a positive real function of a frequency variable associated with a stationary stochastic process, or a

deterministic function of time, which has dimensions of power per hertz (Hz), or energy per hertz.

$$SFM = [\pi_{k=1}^{N} PSD(f)]^{1/n} / {}_{[(1/n)(} \Sigma_{k=1}^{N} PSD(f))]$$

where f is frequency. It is often called simply the spectrum of the signal. Intuitively, the spectral density measures the frequency content of a stochastic process and helps identify periodicities. SFM is a widely existing measure for discriminating frequencies in various applications, such as voiced frame detection in speech recognition. In general, small values of SFM imply the concentration of data at narrow frequency spectrum ranges.

IV. PERFORMANCE EVALUATION

Performance of the proposed scheme is evaluated using some evaluation metrics known as IR, DT, and MIR. The detection time is the time taken to detect C-Worm. MIR provides ratio of number of infected computers and total number of vulnerable computers. The higher the values of these metrics, the more effective the attacks are. The lower these values are, the lower the effectiveness of attacks.

i. Simulation Setup

The experiments are made both for normal and C-Worm traffic. The total number of vulnerable computers is assumed to be around 30000. By varying parameters C-Worm attacks are simulated. The detection involved port scan traffic and also non worm traffic. Logs and traffic traces are used to observe the behavior of worms. The detection results of C-Worm are provided in Table1.

Schemes	**VAR**	**TREND**	**MEAN**	**SPEC(W)**	**SPEC**
Detection Rate(DR)	48%	0%	14%	96.4%	99.3%
Maximal Infection Ratio(MIR)	14.4%	100%	7.5%	4.4%	2.8%
Detection Time(DT) in Minutes	2367	∞	1838	1707	1460

Table 1: Detection results for C-Worm

Table 1 shows the results of detection with various parameters and also with various evaluation schemes such as DR (Detection Rate), MIR (Maximal Infection Ratio) besides providing the detection time in minutes.

ii. Detection Performance for Traditional PRS Worms

The detection performance of traditional PRS worms is presented in fig. 3 and 4. The results use evaluation metrics such as MIR and DR respectively.

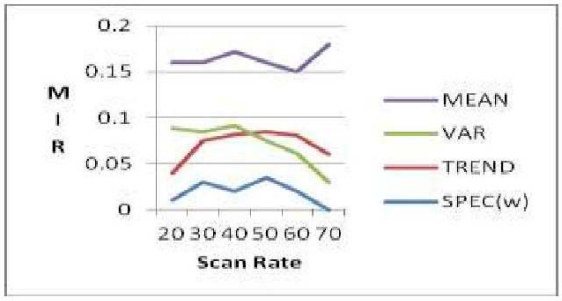

Fig. 2: Maximal Infection Ratio of PRS Worm

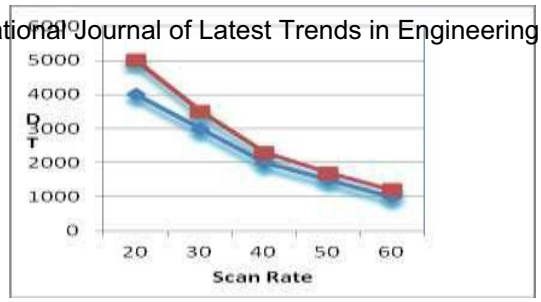

Fig. 3: Detection Time of PRS Worm

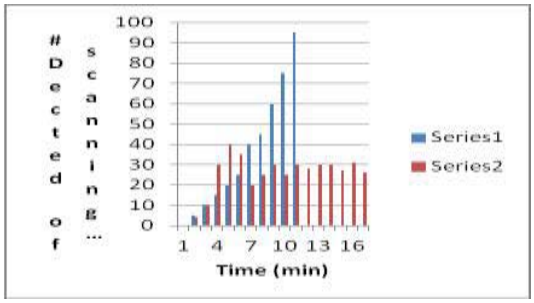

Fig. 4: Number of Detected Scanning Hosts on Camouflaging Worm

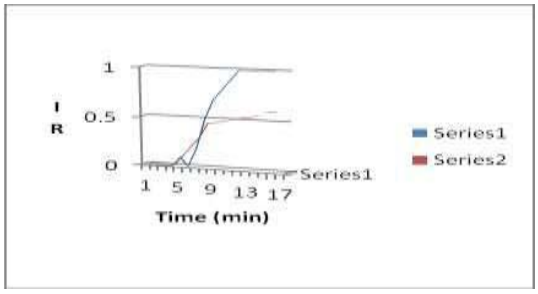

Fig. 5: Infected Ratio for the C-Worm and PRS Worm

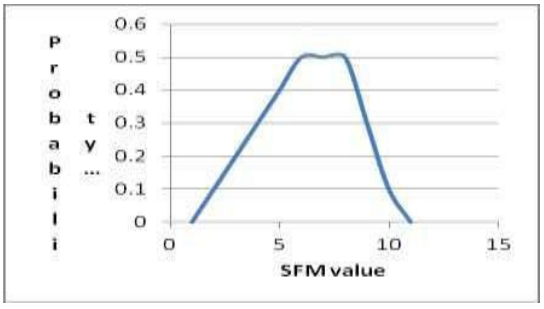

Fig. 6: PDF of SFM on normal non-worm traffic

IV. CONCLUSION

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

REFERENCES

[1] The Jargon file lexicon." http://www.catb.org/~esr/jargon

[2] National Institute of Standards and Technology Special Publication 800-83 Natl. Inst. Stand. Technol. Spec. Publ. 800-83, 101 pages (November 2005)

[3] IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING ,VOL. 8, NO.3, MAY-JUNE 2011.

[4] J. Postel. Rfc 792: Internet control message protocol. Volume 792 of *Request for Comments.*

[5] Z. S. Chen, L.X. Gao, and K. Kwiat, "Modeling th e spread of active worms," in *Proceedings of the IEEE Conference on Computer Communications (INFOCOM)*, San Francisco, CA, March 2003

[6] C. C. Zou, W. Gong, and D. Towsley, "Code-red w orm propagation modeling and analysis," in *Proceedings of the 9-th ACM Conference on Computer and Communication Security (CCS)*, Washington DC,November 2002

[7] C. C. Zou, W. Gong, and D. Towsley, "Worm propagation modeling and analysis under dynamic quarantine defense," in *Proceedings of the 1-th ACM CCS Workshop on Rapid Malcode (WORM)*, Washington DC,October 2003

[8] C. Zou, Don Towsley, and Weibo Gong, "Email wor m modeling and defense," in *Proceedings of the 13-th International Conference on Computer Communications and Networks (ICCCN)*, Chicago, IL,October 2004

[9] W. Yu, S. Chellappan C. Boyer, and D. Xuan, "Pe er-to-peer system based active worm attacks: Modeling and analysis," in *Proceedings of IEEE International Conference on Communication (ICC)*, Seoul, Korea,May 2005

[10] W. Yu, S. Chellappan C. Boyer, and D. Xuan, "Pe er-to-peer systembased active worm attacks: Modeling and analysis," in *Proceedings of IEEE International Conference on Communication (ICC)*, Seoul, Korea,May 2005

[11] Dshield.org, *Distributed Intrusion Detection System*, http://www.dshield. org/, 2005.

[12] SANS, *Internet Storm Center*, http://isc.sans.org/ [11]Wei Yu, Xun Wang, Prasad Calyam, Dong Xuan, and Wei Zha " On Detecting Camouflaging Worm"Proceedings of the 22nd Annual Computer Security Applications Conference @IEEE 2006 C.

[13] Zou, W. B. Gong, D. Towsley, and L. X. Gao, "Monitoring and early detection for internet worms, " in *Proceedings of the 10-th ACM Conference on Computer and Communication Security (CCS)*, Washington DC, October 2003

[14] J. Wu, S. Vangala, and L. X. Gao, "An effective architecture and algorithm for detecting worms with various scan techniques," in *Proceedings of the 11-th IEEE Network and Distributed System Security Symposium (NDSS)*, San Diego, CA, Febrary 2004.

.