A Implementation Paper on Modeling & Detection of Camouflaging Worm

Minal Naik¹, Komal Ajgaonkar², Manisha Kumbharkar³, Sonal Nadarge⁴, Rohini Agawane⁵

Department of Computer Engineering, K J College of Engineering & Management Research, Pune, India

Email- minalnaik12@gmail.com,ajgaonkar.komal@gmail.com,sonal.nadarge@gmail.com,
kumbharkar.manisha@gmail.com,agawane.rohini@gmail.com

Abstract-

Worm is a malware computer program or harmful program code that spreads itself in order to infect other computers which uses mostly the network i.e. internet to spread. Worms possess a great threat to internet security. There are many worms on the internet which travel with data packets across the network along with their different classification and characteristics but we mainly focus on the one such traditional class of worm called as “Camouflaging worm”. Firstly the modeling of C-worm when worm identifies vulnerable computer by scanning several IP addresses and secondly detection is based on two types of traffic which are able to distinguish in time domain. But in detection we need to travel the data from time domain to frequency domain with the help of Power Spectral Density(PSD) which is a spectrum based detection scheme and other scheme to distinguish C-worm traffic from other background traffic. Our detection scheme is effective in detecting C-worms along with normal worms.

Keywords-Camouflage, Active worms, PSD, SFM

I. INTRODUCTION

We conduct a systematic study on our main point that is Camouflaging worm. This type of worm is a self-duplicating worm whose main characteristic is to camouflage meaning it hides itself in network traffic and propagation is based on a very high speed. The main term used for camouflage in the system is actual covering or hiding itself. In our system active worms and traditional worms can also be detected. There are certain type of worms which keeps on scanning the random lists of IP addresses to find new vulnerable
computers called as PRS worms. Other worms propagate themselves more effectively than PRS worms using various methods such as file sharing, email, port scanning and instant messaging. There are some systems whose detection schemes are based on a certain assumption which lead us to conclusion that each worm-infected computer keeps scans over the Internet and propagates itself at a very high range of speed.

Threshold based detection and trend-based detection have been developed to detect the large scale propagation of worms in the Internet. Mostly C-worm is invisible in the narrow range of frequency spectrum when observed under the time domain so it is not visible. But when observed under frequency domain when data is travelled with the help of PSD scheme, there is a wider range of frequency and it is easily visible for detection scheme. So we use technique of frequency domain technique. We note that the propagation controlling nature of the C-Worm cause a slowdown in the propagation speed. However, 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.

II. Background and Related Work:

A. Active Worms:

Many worms are used to infect the large number of computer in the form of bots and zombies. These worms have a great effect on the computer system and the network. Many real world worms have caused a notable damage on the internet. These worms include:

- Code-Red Worm
  From July 12, 2001, the Code-Red I worm began to exploit. The worm generates a random list of IP addresses. The 1st version of the Code-Red worm (Code-Red I v1) which is memory resistant. Began to infect hosts running unpatched versions of Microsoft’s IIS web server. The second version is Code- Red I v2 uses a random seed in its pseudo-random number generator.

B. Slammer Worm

Slammer was the fastest computer worm in history. This type of worm infected more than 90 percent of systems which were able to harm the systems within 10 minutes. Slammer's most important feature is its speed of propagation. By comparison, Slammer was two orders of magnitude faster than the Code Red worm. This worm used random scanning strategy for spreading. To make use of random-scanning worm more effectively, it needed a good source of random numbers to select new target system for attacks.

C. Witty Worm

The worm took advantage of a security flaw in these firewall applications such as Network telescope, ISS vulnerability, Witty worm details, Witty worm spread. The witty worm incident was very different because
this type of worm spread very rapidly after the announcement of the ISS vulnerability. On March 19, 2004, the “witty” worm started to spread to hosts connected to the Internet.

I. System Architecture:
In the proposed system we are modeling the Camouflaging worm (C-Worm), in which the its behavior is hidden and its action is implicitly kept secret. Our system is completely based on the detection scheme of the Active Worm and the C-Worm. The Detection scheme starts with the Pure Random Scan Method for detecting the C-worm. The Scanning starts with the selection of the folder, File, or the drive depending upon the choice of the user. Once the selection of the drive or folder is done by the User Scanning of the System starts According to the Pure Random scan. While Scanning the system if there are any worms or the C-worm present according to the behavior of the worm stored in the database the Detection scheme in our system gives the worm detection list to the user. Once the List is displayed to the user the analysis chart is also displayed to the user in which the propagation of the worms is seen.

System works as follows:

- User can select overall system or particular folder or drive or removable drives for scanning worm.
- After selection of any of one of four choices, scanning starts on selected drive or folder.
- When system is scanning file or .exe in system its progress is shown in scan progress according to each drive or folder contents.
- As well as infected files are also displayed.
- At completion of scanning of folder or drives total no of files which are scanned and total no of files infected and clean are displayed in file.
- At last analysis chart is displayed to user.

![System architecture diagram]

Figure: System architecture
II. Modeling of the C-Worm:

1. C-worm

In this paper we conduct a systematic study on a new class of active worm denoted as Camouflaging Worm. The C-worm has self-propagating behavior similar to traditional worm i.e it rapidly infect as many vulnerable computer(system that easily get affected by viruses or worms)as many as possible. The C-worm is very different from the traditional worm which hides or covers up in any trend which gets easily noticed in the number of infected computers over a time. The camouflage is achieved by managing the scan traffic volume of worms which infect other computers. C-worm has a behavioral property of remaining hidden in the time domain. Whereas it is easily visible in the frequency domain. So for the detection of the C-worm we are going to travel from the time domain to the frequency domain.

2. Modeling

Modeling of C-worm is to affect the system but won’t reduce the traffic level, this is done by dividing the system using the three methods: 1) Protected 2) Affected 3) Vulnerable. We won’t do anything to the system which are already protected and affected. We will predict the target system and model the few worms for the testing purpose. The modeling of the worms which includes, increases the cpu load and for the removal of the C-worm corresponding patches are created by us.

3. Propagation of C-worm

C-worm uses the epidemic dynamic formula to model the propagation of C-worms. Propagating of the C-worm assumes that the computers are : 1) Immune 2) vulnerable 3) infected. Based on the existing result the simple epidemic model for the propagation of the traditional worm is given by the formula:

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

Where $M(t)$ is the number of infected computer at time $t$, $N$ equals to number of vulnerable computer on the internet, $t$ is the total number of IP address. Simple epidemic model for a finite population of C worms can be expressed as follows

$$\frac{dM(t)}{dt} = \beta_0 M(t) [N - M(t)]$$

Where $M(t) =$number of infected computers, $N =$ number of vulnerable computers. $\beta =$ pairwise infection rate.
4. C-worm Detection

- Design rationale-

In this system we are going to develop the novel spectrum based detection scheme. As discussed in the C-worm section the C-worm goes undetected in the time domain. Our detection scheme captures the distinct pattern in the frequency domain, and thereby has the potential of detecting the C-worm propagation.

In order to identify the C-worm propagation in the frequency domain we use the distribution of the PSD and the corresponding SFM of the scan traffic. Particularly, PSD describes how the power of the time series is distributed in the frequency domain.

- Spectrum-Based detection scheme-

In our spectrum based detection scheme, the distribution of PSD and its distribution of PSD, and its corresponding SFM are used to distinguish the C-worm scan traffic from the non-worm scan traffic.

- Power Spectral Density-

By obtaining the distribution of PSD for worm detection data, there is a need to transform data from time domain to frequency domain. Random process $X(t)$ is used to model the worm detection data. Assuming $X(t)$ is source count in time period $[t-1, t]$. Autocorrelation of $X(t)$ is given by:

$$R_{x}(L) = E[X(t)X(t+L)]$$

$R_{x}(L)$ is the correlation of worm detection data in an interval $L$. If a recurring behavior exists, Fourier Transform of the autocorrelation function of $R_{x}(L)$ can reveal such behavior. The next auto-correlation function of $R_{x}(L)$ can such behavior. Thus, PSD function is determined by Discrete Fourier Transform (DFT) of its autocorrelation function as follows:

$$\varphi(R_{x}[L], K) = \sum_{n=0}^{N-1} (R_{x}[L]) e^{\frac{j2\pi Kn}{N}}$$

Where $K=0,1,\ldots,N-1$.

As PSD inherently captures any recurring pattern in the frequency domain, the PSD function shows a comparatively even distribution across a wide spectrum range for normal non-worm scan traffic. The PSD of C-Worm scan traffic show spikes or noticeably higher concentrations at a certain range of a spectrum.

- Spectral Flatness Measure-

Here we measure flatness of PSD to distinguish the scan traffic of C-Worm from the normal worm scan traffic. For that, we introduce the SFM, which can capture anomaly behavior in certain range of

[Type text]
frequencies. The SFM is defined as the ratio of geometric mean to the arithmetic mean of PSD coefficients. It can be expressed as

\[
SFM = \frac{\prod_{k=1}^{n} s(f_k)^{1/n}}{1/n \sum_{k=1}^{n} s(f_k)}
\]

where \(S(f_k)\) is an PSD coefficient for the PSD obtained. SFM is a widely existing measure for discriminating frequencies in various applications, such as voiced frame detection in speech recognition. The SFM value for the C-worm is very small i.e. approximately 0.075.

### 5. Analysis Report

There are two metrics used for the measuring of files scanned by the worm.

a) Precision: It will measure the percentage of the infected files by worm which are present in the folder and are relevant to give result.

b) Recall: It will measure the percentage of the infected files which were present in the folder were retrieved.

### 6. Result

![Fig.C-worm creation](image)

**Step 1:**

This is the first GUI, we first click Create C-Worm, then executable files with .exe extension is created and this file is replicated in each and every five seconds which are present in the selected drive i.e D:/C:
declared in the code. Then warning box will be prompted displaying message that C-Worm is created and system will be restarted soon.

Step 2:

Select drive from the list drives options, if not selected properly, it will display a warning message to select the drive and if selected scanning will start. In the status window, it will display the total number of

a) Scan files present in the user defined drive.

b) Clean files present in the given drive.

c) Infected files present in the drive.
Fig. Selecting Drive or system

Fig. Error if no drive selected

Fig. Scanning of Drives
Step 3:

Now it will show the progress of the scanned files showing the number of status of scanned files.

**Fig. Infection Found are displayed**

**Fig. Total file Scan displayed**
7. Conclusion

In this paper we studied the new class of active worm called as the camouflaging worm (C-worm). The Camouflaging worm has the behavior of hiding itself in the time domain and propagates in the faster manner without any noticeable change. For the detection of the Camouflaging worm (C-worm) we are going to travel from time domain to frequency domain. Since the C-worm has the drawback that it is visible in the frequency domain. So in this paper we are detecting the C-worm in the frequency domain using the spectral based detection in which we are going to use the Power Spectral Density (PSD) and the corresponding Spectral Flatness Measure (SFM).
8. References

1. Modeling and Detection of Camouflaging Worm

2. A Novel Approach of Detecting the Camouflaging Worm

3. Modeling and Detection of Camouflaging worm using IP Traceback

4. On detecting Camouflaging worm
   Dept. of Computer Science, Texas A&M University, College Station, TX 77843.

5. Discrete Mathematical Model Of Computer Worm Propagation