Evaluation of Anti-Virus Application with KLEZ

Kentaro Arai s1070011 Supervised by Prof. Hiroshi Toyoizumi

Abstract

There are the gateway protection and the client protection anti-virus applications. The gateway protection said take more effective empirically [1]. This paper compare those two anti-virus applications using mathematical model and prove our saying is true.

1 Introduction

In recent years, the damage caused by worms is increasing. Worms, which are one of computer viruses, are self-contained programs or set of programs and are able to spread copies of themselves. Propagation usually takes place via network connections. Code-Red, Nimda and KLEZ are worms that have prevailed recently. These worms used security hole in Microsoft's Windows operating system. There are many anti-virus applications. Most of them are classified the client version that we can install on our computers and the gateway version that stops viruses at the entrance of our own network such as SMTP, HTTP, FTP server. We evaluate these two anti-virus applications using a mathematical model. Building and analyzing models of computer viruses has been done before. Christel and others [7] modeled a computer virus and got optimal rate of spreading the viruses. Toyoizumi and Kara [4] proposed a new method for reducing computer viruses and analyzed this method using a mathematical model. In this paper, We formulate and solve a mathematical model based on the Continuous-time Markov chain.

2 The virus model

2.1 Pure Birth Process

A pure birth process is a continuous-time Markov chain with states 0,1,...for which transitions from state i can only go to state i+1. We consider the state of the process as the number of machines infected by the virus, and when the state increases by 1 we say that a birth occurs. We suppose that whenever there is i infected machines in the network the time until the next birth is exponential with rate λ_n

2.2 Network Model

We assume that Network has only one entrance from outside. Except for one entrance, a Network is completely closed from outside. In addition, Network is large enough. (See Figure 1)

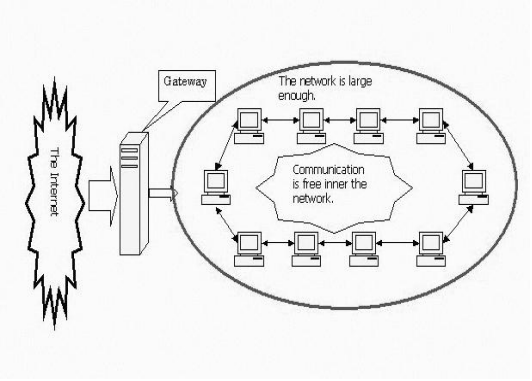


Figure 1: The network model

2.3 Model formula

We consider that no infected machines in the network at time 0 and once infected a machine is in that state forever. We suppose that each machines, inner the network, infected by viruses do infection activity independently and in accordance with the Poisson process having rate λ. That is, the times between successive infection activities are independent exponential random variables having mean $\frac{1}{\lambda}$. The gateway can protect the our network different rate. If the first birth gives at an exponential rate μ, then, if X (t) represents the number of machines at time t, $\{X(t), t \geq 0\}$ is a pure birth process with(See Figure 2)

$$\lambda_n = \begin{cases} \mu(n = 0). \\ n\lambda(n = 1, 2, 3....). \\ 0(otherwise). \end{cases} \quad (1)$$

Let $T_i, i \geq 0$, denote the time between the i st and (i+1)th birth. That is, T_i is the time it takes for the number of machines from i to i+1.

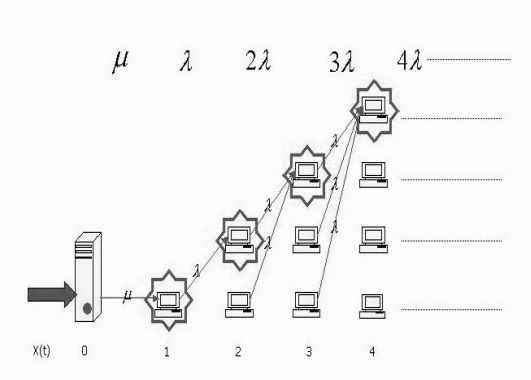


Figure 2: The infection model

$$P\{T_0 + T_1 + ... + T_j \le t\} \quad (2)$$

$$= \int_0^t P\{T_0 + T_1 + ...T_j | T_0 = x\} dP\{T_0 \le x\} dx. \quad (3)$$

$$= \int_0^t P\{T_1 + ... + T_j \le t - x\} dP\{T_0 \le x\} dx. \quad (4)$$

Now $P\{T_1 + ...T_j \le t\}$ is Yule Process and $P\{T_1 + ...T_j \le t\} = (1 - e^{-\lambda})^j$ [2] thus

$$P\{T_0 + T_1 + ... + T_j \le t\} \quad (5)$$

$$= \int_0^t (1 - e^{-\lambda(t-x)})^j \mu e^{-\mu x} dx. \quad (6)$$

$$= \int_0^t \sum_{r=0}^{j} (-1)^r e^{-\lambda r(t-x)} \mu e^{-\mu x} dx. \quad (7)$$

$$= \sum_{r=0}^{j} \frac{(-1)^r}{\lambda r - \mu} \mu_r C_j e^{-\lambda r t} (e^{(\lambda r - \mu)t} - 1). \quad (8)$$

Hence, we see $P\{X(t) \ge j+1 | X(0) = 0\} = P\{T_0 + T_1 + ... + T_j \le t\}$, we can get the density function

$$P\{x(t) \ge j | X(0) = 0\} \quad (9)$$

$$= \sum_{r=0}^{j-1} \frac{(-1)^r}{\lambda r - \mu} \mu_r C_{j-1} e^{-\lambda r t} \{e^{(\lambda r - \mu)t} - 1\}. \quad (10)$$

If we let T denote the time until the j machines is infected, then T can be represented as

$$T = \sum_{i=0}^{j-1} T_i. \quad (11)$$

As the T_i are independent exponential random variables with respective rate $\lambda_0 = \mu$, $\lambda_i = i\lambda$, i=1,.....j-1, we can get mean time function

$$E[T] = \frac{1}{\mu} + \frac{1}{\lambda} \sum_{i=1}^{j-1} \frac{1}{i}. \quad (12)$$

E[T] can be approximated as follows

$$E[T] = \frac{1}{\mu} + \int_1^{j-1} \frac{1}{t} dt. \quad (13)$$

$$= \frac{1}{\lambda} \log[j-1]. \quad (14)$$

3 Analysis

We have analyzed KLEZ in the following way.

3.1 KLEZ

3.1.1 General Description

KLEZ was discovered in Oct. 26, 2001. There are a lot of variants such as KLEZ.A, KLEZ.H. All KLEZ variants, except KLEZ.B, are multi-threaded worms, where each thread performs a predefined task such as a network infection or a mass e-mailing. KLEZ.B spawns multiple copies of itself in memory. The KLEZ can implement on Microsoft's windows operating system. The worm exploits a vulnerability that opens an executable attachment even in Microsoft Outlook's preview pane, We have investigated KLEZ.H in detail. KLEZ.H obtains recipients from the entries in the default Windows Address Book (WAB), and also gathers addresses from the following files(See Table 1) in the infected computer.

Table 1: Files

MP8	EXE	SCR
PIF	BAT	TXT
HTM	HTML	WAB
DOC	XLS	CPP
C	PAS	MPQ
MPEG	BAK	MP3

KLEZ.H obtain a SMTP server from the registry as follows:
HKEY_LOCAL_MACHINE\Software\Microsoft \InternetAccountManagerAccounts\, SMTPServer.

KLEZ.H obtains email addresses to place in the FROM: field from the infected user's address book. This causes a non-infected user to appear as the person who has sent this worm's malicious email. It does this to hide the real

sender of the infected email.The subject of the email is composed in a complex manner, but also taken from a list in the worm's body. KLEZ.H generates a random mail body. The KLEZ.H is capable of spreading via shared drives or folders with read or write access. To accomplish this, it enumerates all the shared resources of an infected system. For each entry, it copies itself to files with randomly generated filenames. KLEZ.H has another thread. KLEZ.H kills running processes and occasionally deletes executable files of programs that associated with some anti-virus products. KLEZ.H drops the virus EL-KERN. [1]

3.1.2 The infection rate observation for λ

We investigated the propagating thread the following way. The network consists of the Windows98 operating system as the client and Windows 2000 Server operating system as the mail and file server. The client was infected by the KLEZ.H and captured packets inside the network and counted the number of mail transmitting packets and file transmitting packets. The numbers of these were different according to the contents of the client's hard-disk. However, observing these results with the unit of one hour, these results are almost the same, so we can choose default infection rate (λ) is 1.6708 as an average(See Figure 3).

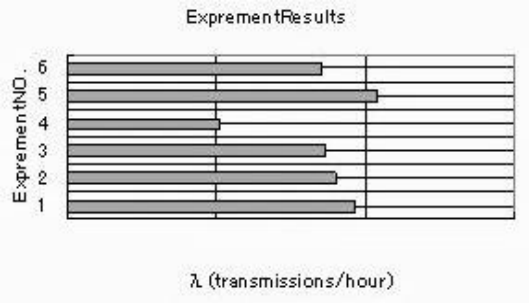


Figure 3: Experiment Results

3.1.3 The infection rate observation for μ

Mie University [8] detected 43 e-mails infected by KLEZ.H at their mail gateway in April, 2002. We can calculate default μ is following.

$$\frac{43}{14(days) \times 24(hours)} = 0.1279(mails/hour) \quad (15)$$

We can chose default infection rate (μ) is 0.1279.

3.2 The anti-virus applications

At present, anti-virus applications, that is carried out actually, can be classified into the gateway version and the client version. The gateway version, which is established at the entrance of the network, prevents the virus from invading the network. The client version, which we install on our machine, prevents the virus from infecting our machine. These anti-virus applications are required latest pattern file, which is the latest virus database, to discover the virus. The pattern file can be downloaded from anti-virus application venders' server. The gateway version must download the pattern file only one time every update, however the client version must download the pattern file once for each machines in the network, every update. As the result, the gateway version operates easily, however the client version is complicated. [1] [6]

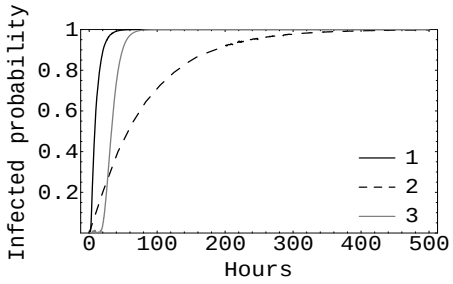
We assume that capacity of anti-virus applications follows. If the gateway anti-virus application's capacity is 90%, the anti-virus can eliminate 90% KLEZ.H that try to invade the network. If the client anti-virus application's capacity is 90%, anti-virus applications are installed on the 90% client machines in the network.

3.3 Analysis condition

We assume that all machines in the network can implements KLEZ.H. In addition, we can calculate λ depending on the gateway anti-virus application's capacity and the client anti-virus application's capacity. If the gateway anti-application's capacity is 90%, we can calculate μ is 0.1279×0.1. If client anti-virus application's capacity is 90% we can calculate 1.6078×0.1.

3.4 Results

Figure 4 shows the probability that more than 50 clients in the network infected by KLEZ.H. We suppose that the gateway anti-virus application's capacity and the client anti-virus application's capacity is 90%.

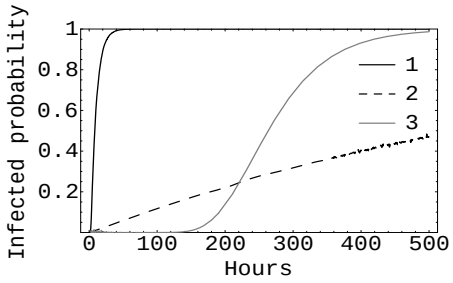


Probabilty that more than 50 machines are infected

Figure 4: Using a (10) expression,
1 is $\mu = 0.1279$ $\lambda = 1.6708$ $j = 50$(The network has no anti-virus applications),
2 is μ= 0.01279 $\lambda = 1.6708$ $j = 50$(The network has the gateway anti-virus application),
3 is $\mu = 0.1279$ $\lambda = 0.16708$ $j = 50$(The network has the client anti-virus application).

We can observe both the client anti-virus applicant and the gateway anti-virus application are effective in our network prevention, and the gateway anti-virus application more effective than the client anti-virus applications.

Figure 5 shows the probability that more than 50 clients in the network infected by KLEZ.H We suppose that the gateway anti-virus application's capacity and client anti-virus application's capacity is 99%.

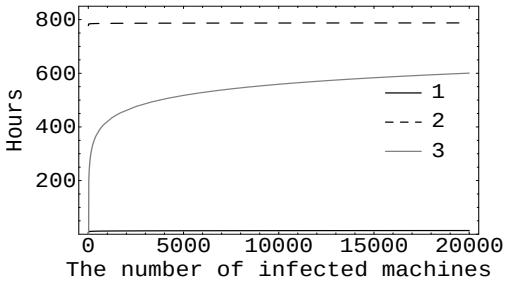


Probabilty that more than 50 machines are infected

Figure 5: Using a (10) expression,
1 is $\mu = 0.1542$ $\lambda = 0.016708$ $j = 50$(The network has no anti-virus applications),
2 is μ= 0.001542 $\lambda = 1.6708$ $j = 50$ (The network has the gateway anti-virus application),
3 is $\mu = 0.1542$ $\lambda = 1.6708$ $j = 50$(The network has the client anti-virus application).

We can observe, when the time have not passed, the client anti-virus applications are more effective than the gateway anti-virus application. However, when the time have passed, the gateway anti-virus application is more effective than the client anti-virus applications. In addition, the client anti-virus applications reach faster probability 1.

Figure 6 shows the Mean time. We suppose that gateway anti-virus application's capacity and client anti-virus application's capacity is 99%.

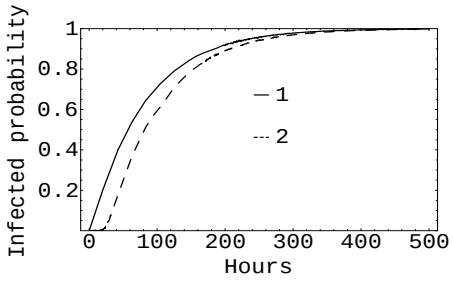


Mean time

Figure 6: Using a (15) expression,
1 is $\mu = 0.1279$ $\lambda = 0.016708$ (The network has no anti-virus applications),
2 is μ= 0.01279 $\lambda = 1.6708$(The network has the gateway anti-virus application),
3 is $\mu = 0.1279$ $\lambda = 0.16708$(The network has the client anti-virus application).

In case anti-virus applications capacity is 99%, we can observe the gateway application is more effective than the client anti-virus applications from Figure 5 and Figure 6.

Figure 7 compare in the case that the network has the gateway anti-virus application that has 90% capacity with the network has the client anti-virus applications at the gateway that capacity is 90% by probability that more than 50 machines infected by KLEZ.H.



Probabilty that more than 50 machines are infected

Figure 7: Using (10) expression,
1 is $\mu = 0.01279$ $\lambda = 1.6708$ (The network has onlt the gateway anti-virus application),
2 is μ= 0.01279 λ=1.6708 (The network has the gateway anti-virus application and the client anti-virus application).

We can observe, when the time have not passed, the network has two anti-virus applications are more effective than only the gateway anti-virus application. However, those two cases reach probability 1 almost same. So, The gateway anti-virus application is more effective than client anti-virus application.

4 Conclusion

I have modeled the worm propagation, and have analyzed the gateway and the client anti-virus application with the KLEZ.H. We prove the gateway anti-virus application is more effective than the client anti-virus applications by our solving mathematical model with KLEZ.H.

5 Acknowledgment

I thank you to Prof. Toyoizumi who always advised me. Moreover, I want to thank to people of laboratory who helped my research.

References

[1] Trend Micro. *http://www.trendmicro.com*.

[2] Sheldon M.Ross. STOCHASTIC PROCESS. JOHN WILEY SONS,INC.,1996.

[3] Messagelabs. *http://www.messagelabs.com/*.

[4] Toyoizumi and Kara. Predators: Good Will Mobile Codes Combat against Computer Viruses ASM Sigsac New Security Paradigms Workshop 2002.

[5] Okamot and Ishida, An analysis on a Model of Computer Viruses via Electronic Mail.

[6] Symantec *http://www.symantec.com/*

[7] Christel Kamp, Claus O. Wilke, Christoph Adami, Stefan Born holdt. Viral evolution under the pressure of an adaptive immune system optimal mutation rates for viral escape. arXiv:Cond-mat/0209613 v1, 2002.

[8] MieUniversity *http://www.cc.mieu.ac.jp*