Modeling Penetration of Viruses at the Gateway

Keiichi Kato s1080063 Supervised by Hiroshi Toyoizumi

Abstract

This paper shows that viruses penetration seen at the gateway is Poisson process. We infected the machine with one virus, collected data and researched on the feature of activity, and looked up target viruses to know frequency of sending and receiving according to the mail server at the University of Aizu. On comparing them, we verified that viruses mail obey exponential distribution.

1 Introduction

Latest viruses cause damage to many computer users using internet by increasing packets sending mail and attacking specific http/ftp than doing damage to the each infected machine. As one way not to be infected and to repair machines if they're infected, everyone needs to buy and install one anti-virus software put on the market. This way is so effective, but people don't install the software because of needing much money, time and troubles for update. It's too difficult to prevent from being infected though all people must make an effort to decrease the damage.

As one way to eliminate malicious viruses, we can use good worm called *Predator* [1] such as CODEGREEN [2], and NACHI [3]. But, these viruses will exert a bad influence upon network unless they're the most suitable. Another way to eliminate viruses and not to increase packets is to stop mail servers from sending virus mails.

The purpose of this paper is to make useful when we work out new defense way from viruses. If we capture the feature of each virus algorithm for sending mails, expect equations based on exponential distribution and estimate the frequency of mail, we will be able to stop virus mails at servers and our anxiety that machines are infected with viruses would decrease, too.

2 Virus Activities

We chose three viruses as samples, SWEN, MIMAIL and LOVGATE, which have reached much. There are two reasons why we chose these. First, not to jump to a conclusion from result of one virus experiment. Second, to have better results by comparing with others.

We could look up the following information from Symantec webpage [4] and know the whole picture of virus activity. We didn't need the information without relating to mails, so we abridged these in this paper.

SWEN

Official name	SWEN.A
Type	Worm
Virus size	106,496 bytes
Platform	Windows 2000/95/98/Me/NT/ 2003 Server/XP
Discovered	September 18, 2003

MIMAIL

Official name	MIMAIL.R
Type	Worm
Virus size	11,520 bytes
Platform	Windows 2000/95/98/Me/NT/XP
Discovered	January 29, 2004

LOVGATE

Official name	LOVGATE.F
Type	Worm
Virus size	172,842 bytes
Platform	Windows 2000/95/98/Me/NT/XP
Discovered	March 12, 2003

Target viruses have two common points. One point is to send many mails by using own SMTP engine. The other point is to make own mailing list by retrieving many mail addresses from various files though the files to search for with each is difference.

The other side, there are some different points. For example, to connect news server for retrieving mail addresses, to use programs for having files jointly, to spread out through local network and so on.

3 Methods

3.1 Collection of the data

Upon experimenting, we prepared the following experiment environment and the software of *Sniffer* [5] for collecting packets.

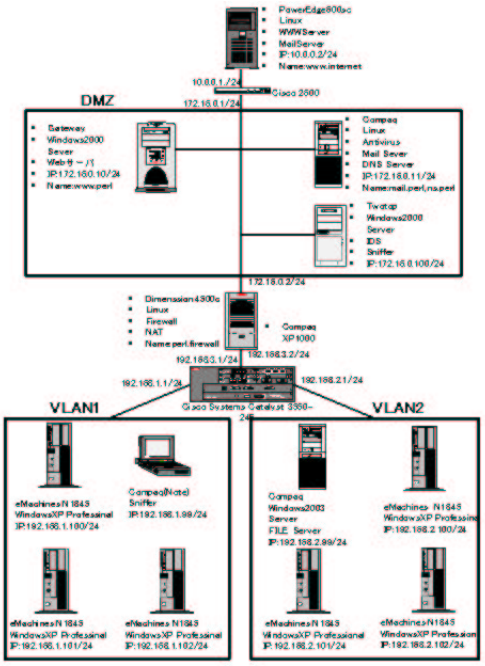

Figure 1: Experiment environment

We installed the software in a machine which we would infect with SWEN.A, after that, we infected to collect packets sent by SWEN.A and to know a feature of sending mail. We needed to know the feature of sending mails because of confirming that the algorithm doesn't obey Poisson distribution. As a result, we found the following features, which SWEN.A sent mails every several seconds and didn't send to all members of mailing list but some members chosen at random. And, SWEN.A wouldn't need 1 second to send a mail on circuit of the 100M base. SWEN.A will be able to send many mails in short time.

3.2 Server Time Data

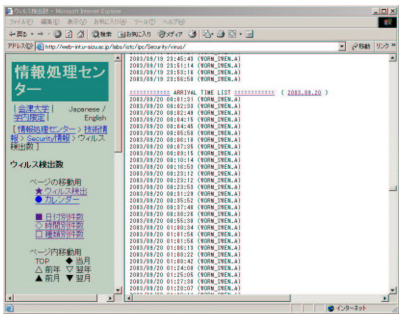

Figure 2: Webpage about Time Data

We checked information [6] about mail server at this university to know receiving or sending frequency, set up an equation and calculated based on this data.

About MIMAIL.R and LOVGATE.F, we led results from only this time data, because we haven't caught these and experimented.

3.3 Equations

E[X] is an average of the interval of every virus mails, Var[X] is a variance of the interval of every virus mails, *Dis* is a distribution through the average and the variance, $P\{t < X \leq t+1\}$ is probability of X from $t \times 30$ to $(t+1) \times 30$ seconds. X is the interval of every virus mails, and 30 is a unit of seconds, then we procured the following equations.

$$E[X] = \frac{1}{n}\sum_{i=1}^{n} X_{\mathrm{i}} \tag{1}$$

$$Var[X] = \frac{1}{n}\sum_{i=1}^{n}(X_{\mathrm{i}})^2 - (E[X])^2 \tag{2}$$

$$Dis = E[X] - \sqrt{Var[X]} \tag{3}$$

$$\begin{aligned} P\{t < X \leq t+1\} &= P\{X \leq t+1\} - P\{X < t\} \\ &= e^{\frac{-1}{E[X]} \times t \times 30} - e^{\frac{-1}{E[X]} \times (t+1) \times 30}. \end{aligned} \tag{4}$$

$$(t = 0, 1, 2...)$$

We knew all X*i* relating to target virus from server time data. Substituting all X*i* for the left side of formula (1) and (2), we obtained the average and the variance. Substituting the results of formula (1) and (2) for (3), we obtained distribution through these. This result shows difference between theory value and actual data. The closer the result is 0, the better. On calculating, we used the way to think *Interarrival and Waiting Time Distribution* [7]. Substituting formula (1) and *t* for the right side of formula (4), we obtained the probability from $30 \times t$ to $30 \times (t+1)$ seconds.

3.4 Calculations

The number of the mails which servers receive may be large, so we sampled the day when mails reached most abundantly.

The following data are quotation from mail server [6] and calculation result.

SWEN.A

Number of samples	373
Time the last mail came	23:49:59
E[X]	230.0241

MIMAIL.R

Number of samples	1003
Time the last mail came	23:57:57
E[X]	86.0199

LOVGATE.R

Number of samples	519
Time the last mail came	13:37:54
E[X]	94.5549

Substituting these E[X] for formula (4) and calculating, we get one line graph. Checking this graph with frequency graph of existing data, we can consider whether the graphs obey Poisson distribution or not.

4 Results

There are one graph a virus. The bar graph shows probability based on actual data and the line graph shows exponential distribution based on theory value. Checking the results to find whether graphs obey exponential distribution.

4.1 SWEN.A

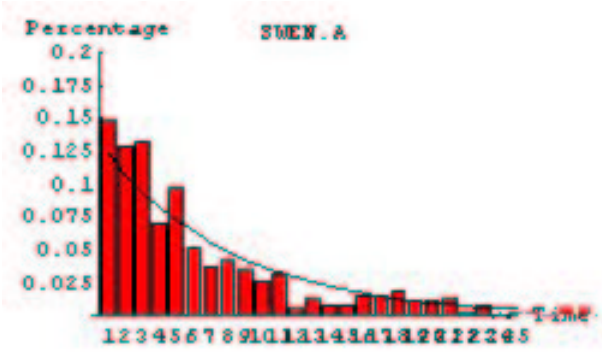

Figure 3:
Probability of the top of bar graph: 14.8%.
Probability of the top of line graph: 12.2%.

The bar graph isn't ideal type falling down as it goes to the right, so the bar graph hardly matches with the line. We found that the one of the causes is lack of samples. These observational errors may disappear by taking time and collecting data. We will not judge that the graph obey exponential distribution.

4.2 MIMAIL.R

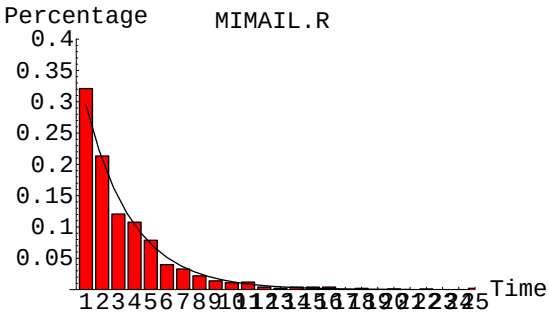

Figure 4:
Probability of the top of bar graph: 32.1%.
Probability of the top of line graph: 29.4%.

There are a few parts appearing errors. But unlike a graph of SWEN.A, the bar graph is ideal type and almost matches with the line graph. The larger the number of samples become, the easier we will judge whether mails obey exponential distribution or not.

4.3 LOVGATE.F

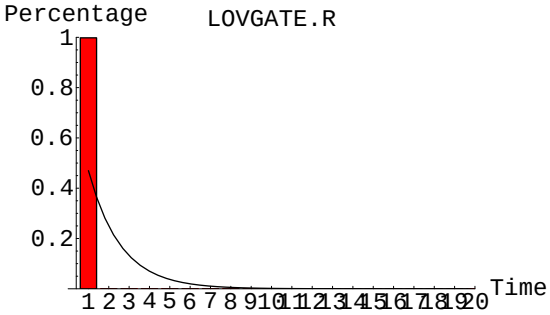

Figure 5:
Probability of the top of bar graph: 99.9%.
Probability of the top of line graph: 47.0%.

There is large difference between the bar graph and the

line. Anyone in University of Aizu sent all mails, so the feature of algorithm for sending mails showed squarely. This result is wrong and exceptional on this paper. Judging mails which servers or routers receive in such frequency viruses, and we may work out a system to stop mails.

	Dif
SWEN.A	-58.7469
MIMAIL.R	-17.4021
LOVGATE.F	-2021.19

All *Dif* are far from 0. We felt a good impression on the graph of MIMAIL.R. So there may be no matter even if *Dif* is a little far.

5 Conclusion

Depending on the network environment and the number of samples, we found that viruses mail obey Poisson process even if an algorithm for sending mails has special feature. And from the result of LOVGATE.F, we found to need to sort out the sending mails from the receiving to obtain correct results.

We may be able to construct better defense system applying this result to others virus.

Acknowledgment

I specially thank Prof. H Toyoizumi for precise advise to write this paper ,all members of the Performance Evaluation Lab at University of Aizu for building the experiment network and all sorts of seminar, and *ISTC* who gives the information of time data for writing this paper.

References

[1] H Toyoizumi and A Kara, Predator: Good Will Mobile Codes Combat against Computer Virus, ASM Sigsac New Security Paradigms Workshop, 2002.

[2] Yuuzou Kobayashi, Modeling of Computer Viruses, University of Aizu, Graduation Thesis. March, 2003.

[3] Satoshi Onoda, Optimization of Nachi spreads, University of Aizu, Graduation Thesis. March, 2004.

[4] Symantec, http://www.symantec.com/

[5] TOYO Corporation, http://www.toyo.co.jp/sniffer/

[6] InterScan VirusWall http://web-int.u-aizu.ac.jp/labs/istc/ipc/Security/virus/

[7] Sheldon M.Ross, Stochastic Processes Second Edition, 1996