コンピュータセキュリティ 19－4
（２００２．１２．２０）

Mass Mailing Worm と DNS/SMTP トラフィック解析

武蔵 泰雄 †、杉谷 賢一 †、松葉 龍一†

概要: Myparty、KLEZ、及び Yarner. A 等の大量メール送信型ワーム (MMW) の感染が拡大している時期に、DNS 及び E-mail サーバ間の名前解決 UDP パケットの流量について統計的に調査を行なった。我々の得た興味深い結果は以下の通りである: (1) SMTP アクセスが増加すれば、異常に大きなピークが DNS 流量 ($D_{\rm q}$) に現れる。(2) このピークの位置と SMTP アクセス量 ($N_{\rm SMTP}$) のピークの位置は一致する。(3) $N_{\rm SMTP}$ ピークの位置とある利用者の SMTP アクセス量のピークの位置は一致する。(4) 我々の調査によれば、その利用者の PC は MMW に感染していた可能性がある。以上のことから、E-mail サーバの DNS サーバに対する $D_{\rm q}$ を監視することにより、MMW に感染した PC 端末の利用者や IP アドレスなどを検知することが可能である。

Traffic Analysis on Mass Mailing Worm and DNS/SMTP

Yasuo Musashi,† Kenichi Sugitani,† and Ryuichi Matsuba†

Abstract: The name resolving UDP packet traffic between the domain name system (DNS) server and the electronic mail (E-mail) server of Kumamoto University was statistically investigated when several PC terminals were infected by the mass mailing worm (MMW), such as Myparty, KLEZ, or Yarner. A. The interesting results are: (1) An abnormally large peak of the number of DNS query access ($D_{\rm q}$) emerges when the number of the SMTP access ($N_{\rm SMTP}$) increases drastically. (2) The $N_{\rm SMTP}$ peak occurs at the same point of the $D_{\rm q}$ peak. (3) Also, this $N_{\rm SMTP}$ peak is taken to be as the same peak point as the number of the SMTP access for a user. (4) From our survey, the PC terminal of the user is infected by MMW. Consequently, we can detect an owner and/or an IP address of the MMW-infected PC terminal by observing the $D_{\rm q}$ traffic from the E-mail server to the DNS server.

1. Introduction

Intrusion detection system, IDS,[1−4] is one of attractive solutions to keep security of the network servers such as the domain name system (DNS)[5] server, the electronic mail (E-mail) server, and the web server. There are two ways of detection of abnormality of the network servers; one is a pattern-matching with a signature file, which is a database of the remote attacking pattern, to detect abnormality of the network server, and the other is statistically to find abnormality of the network server. The former needs to update frequently the signature file because of quick developing new signature files or new cracking technologies. However, the latter does not always need to update the signature files. To develop a new effective statistical IDS against future remote attacks on the network server, it is of considerable of importance to get detailed information of traffic of network packets like the DNS query packets (UDP packets) between the DNS server and the DNS clients.[5] In our previous paper,[6] the total number of DNS packets, $D_{\rm q}$, are predominantly generated from an E-mail server, as represented:

$$D_{\rm q} = (2 + 4n(1 - q))N_{\rm SMTP} + N_{\rm POP3} \qquad (1)$$

where $N_{\rm SMTP}$, $N_{\rm POP3}$, q, and n represent the number of the simple mail transfer protocol (SMTP)[7] access, the number of the post office protocol version 3 (POP3)[8] access, the mail-receiving rate, and the number of different domain hosts, respectively.

†熊本大学総合情報基盤センター・Center for Multimedia and Information Technologies, Kumamoto University.

If the q value is $0.50 \sim 0.75$, the n value is calculated to be $3.3 \sim 6.6$. These results show that the DNS access from the E-mail server is mainly driven by the SMTP access.

In the present paper, we statistically investigated traffic of the DNS query packets between the DNS server (**1DNS**)[9] and the E-mail server (**1MX**)[10] when several PC terminals were infected by the mass mailing worm (MMW),[17] such as Myparty, KLEZ, or Yarner. A. The traffic is schematically drawn in Scheme 1. Our purposes are (1) to compare both logs of SMTP and POP3 accesses with that of DNS query access, (2) to show how the DNS query packets depend on the SMTP access by the MMW-infected PC terminals, and (3) to find out methods to detect abnormality in E-mail server with statistical analysis of the DNS traffic by the E-mail server.

2. Used Server Daemon Programs and Estimation of $D_{\rm q}$, $N_{\rm SMTP}$, and $N_{\rm POP3}$

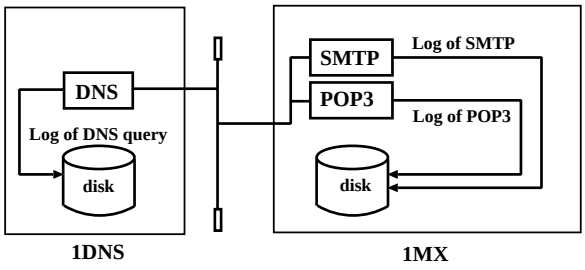
In **1DNS**, the BIND-9.1.3 program package have been employed as a DNS server daemon.[11] The log of DNS query packet have been recorded by the iplog-1.2 program[12,13] with the syslog system.[14] In **1MX**, the sendmail-8.9.3 program package[15] and the Qualcomm qpopper-4.0 program package[16] were installed as SMTP and POP3 server daemons, respectively. The log of SMTP and POP3 accesses have been observed in the syslog file. All of the syslog files are daily updated by the crond system.

The $D_{\rm q}$, $N_{\rm SMTP}$, and $N_{\rm POP3}$ values are estimated, as follows: (1) We connect to the DNS server (**1DNS**) by a ssh client, and then change into the "/var/log" directory. We enter the following commands:

```
% grep domain messages.1 >/tmp/1dns
```

After writing its output into a file at the "/tmp" directory, we count lines of the file by a "wc" command:

```
% grep "133.95.xx.yy:" /tmp/1dns | wc
```



Scheme 1

The $D_{\rm q}$ value is given as an output of the wc command. (2) We connect to the E-mail (**1MX**), and then we change into the "/var/log" directory. We enter the following commands:

```
% grep "sendmail" syslog.0 >/tmp/1smtp
```

After using this command, we enter the next commands:

```
% grep "from=" /tmp/1smtp | wc
```

The $N_{\rm SMTP}$ value is given as an output of the wc command. (3) We enter the following commands to estimate the $N_{\rm POP3}$ value:

```
% grep "poppe\[" syslog.0 | wc
```

3. Results and Discussion

3.1 Analysis of Traffic between the DNS server and the E-mail Server

We plot observed traffic curves of the DNS query access $D_{\rm q}$ (**1DNS**), the SMTP access (**1MX**), and the POP3 access (**1MX**) in Figure 1. The observation was performed at February 16th, 2002.

In Figure 1, the traffic curve of $D_{\rm q}$ exhibits three significantly large local maximums in the early morning (00:00-07:00). In previous work,[6] no such large peak was found in the $D_{\rm q}$ traffic curve. These features indicate that several incidents take place at **1MX**. After these peaks, have a lunch at noon, the $D_{\rm q}$ traffic curve rises strait upon going from 08:00 to 09:00, considerably increases up to 11:00 with small fluctuation, slightly decreases to a local minimum at 12:00, and repeats a local maximum twice. It is common features because almost users of **1MX** start to use an E-mail application in the

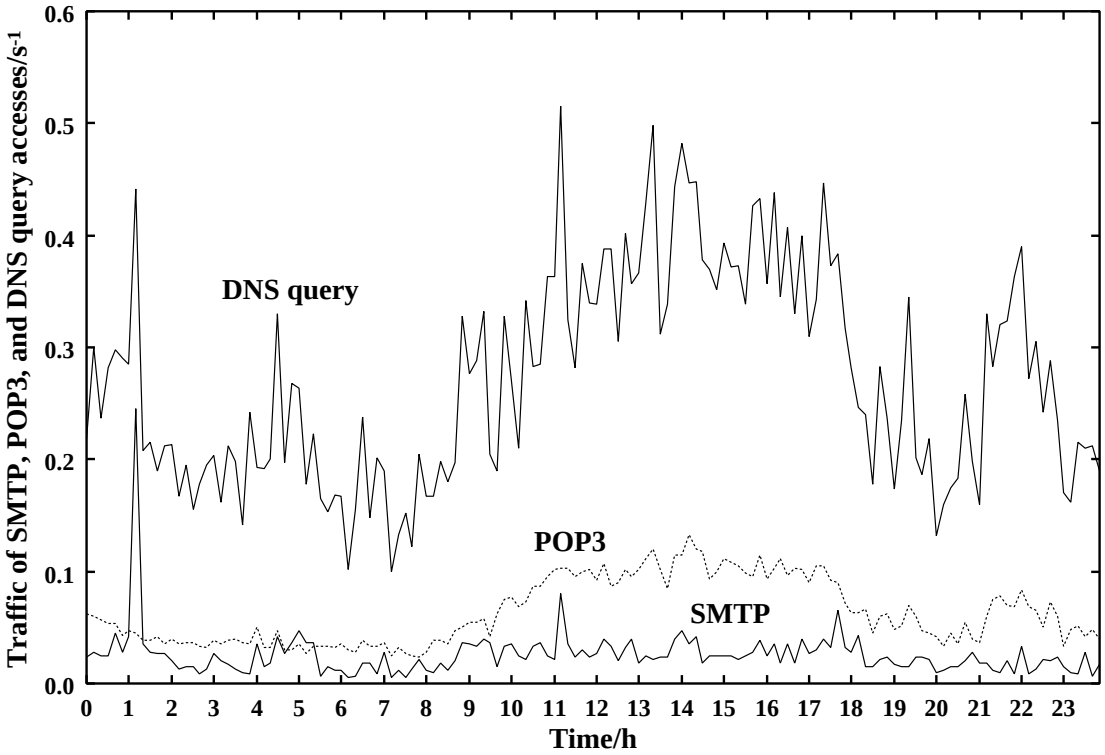


Figure 1. Traffic of the SMTP, POP3, and DNS query accesses in February 16th, 2002. The upper real line shows the DNS query access, the middle broken line means the POP3 access, and the bottom real line indicates SMTP access (s^{-1} unit).

morning, and start to return back to home from 18:00. Usually, the D_q traffic curve decreases gradually to 08:00. However, it repeats a large local maximum at 22:00. As shown in Figure 1, we can find four local large maximums. These maximums mean that the network incidents take place at least four times in the day. Thus, we need to investigate further on the local maximus.

Interestingly, the traffic curve of $N_{\rm SMTP}$ resembles well that of D_q in a small scale manner; for instance, (1) the first abnormally large local maximum of the $N_{\rm SMTP}$ curve is almost the same point as that of the D_q curve, (2) the other D_q local maximums is taken to be almost the same points as those of the $N_{\rm SMTP}$ ones, respectively, and (3) the rippled part of the D_q curve at 16:30-17:00 is significantly similar to that of the $N_{\rm SMTP}$ curve. This is because the contribution of $N_{\rm SMTP}$ to D_q is a much greater extent than that of $N_{\rm POP3}$ (see eq (1)). On the other hand, the traffic curve of $N_{\rm POP3}$ changes in a mild manner and slightly resembles that of D_q because of the small contribution of $N_{\rm POP3}$ to D_q. These results show that the $N_{\rm SMTP}$ traffic provides more important information of the abnormally large local maximums in the D_q curve than the $N_{\rm POP3}$ traffic.

3.2 Detection of Strange SMTP Access of the E-mail User

The number of the SMTP access ($N_{\rm SMTP}$) is represented by the sum of the number of SMTP access of the E-mail user, $N_{\rm SMTP}(i)$, as follows

$$N_{\rm SMTP} = \sum_i N_{\rm SMTP}(i) \qquad (2)$$

where the index i means an account name of the E-mail user. From this reason, it is worthwhile to compare the $N_{\rm SMTP}$ curve of **1MX** with several $N_{\rm SMTP}(i)$ curves. Figure 2 demonstrates $N_{\rm SMTP}$, $N_{\rm SMTP}$(A), $N_{\rm SMTP}$(B), $N_{\rm SMTP}$(C), $N_{\rm SMTP}$(D), $N_{\rm SMTP}$(E), and $N_{\rm SMTP}$(F) curves through the day, February 16th, 2002, in which the users A to F are six top users of SMTP access **1MX** in the day.[18]

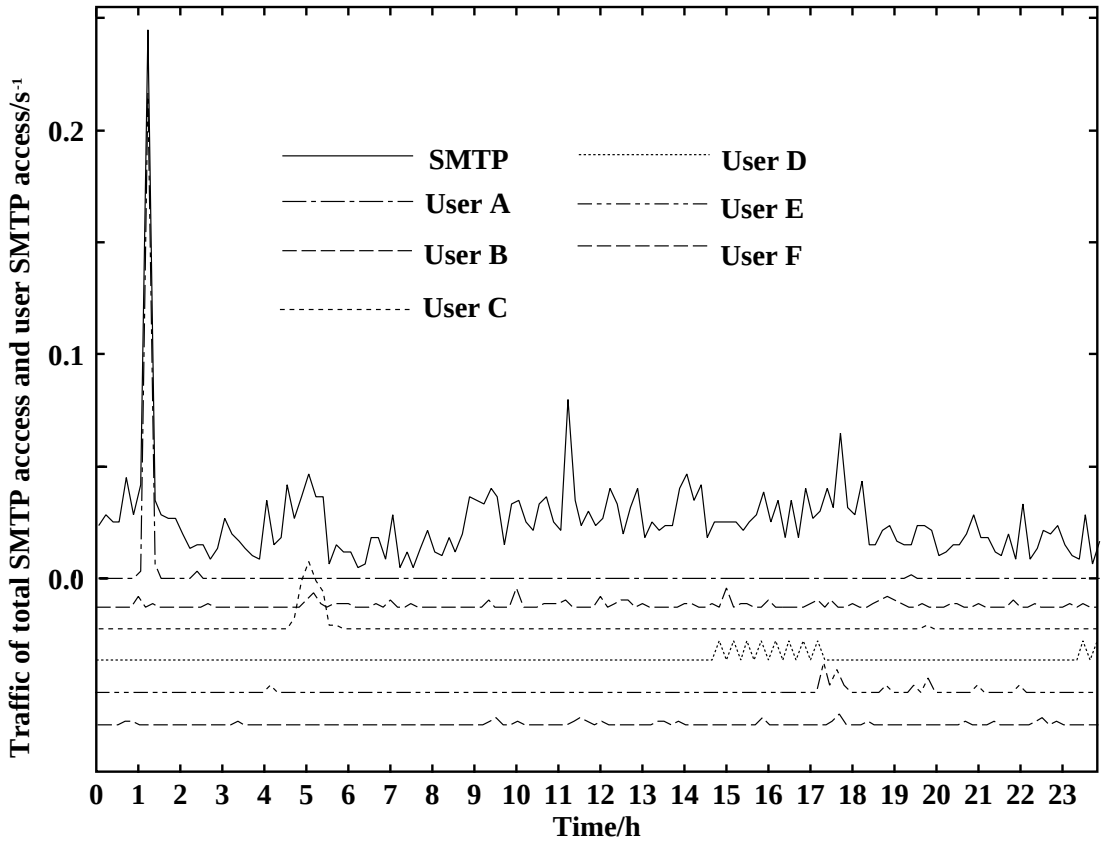


Figure 2. Traffic of total SMTP access and the top SMTP access of the E-mail users in February 16th, 2002. The real line indicates the traffic of the total SMTP access, and the others are the top SMTP accesses of each E-mail users (User A-F) (s^{-1} unit).

Apparently, the local maximum in the $N_{\mathrm{SMTP}}(\mathrm{A})$ curve occurs at the same point as that in the N_{SMTP}. This indicates that the D_{q} curve at 01:00-01:30 is driven by the SMTP access of the user A. This user is a staff of our university and provided us important information that user A sent a lot of E-mails at once mailing function of the mailer. Thus, it is clearly concluded that the first local maximum of the $N_{\mathrm{SMTP}}(\mathrm{A})$ curve is unrelated to the MMW-infection.

The local maximums of the $N_{\mathrm{SMTP}}(\mathrm{B})$ and $N_{\mathrm{SMTP}}(\mathrm{C})$ curves are a mailing list (ML) SMTP account, ML-B and ML-C, respectively. The former local maximums emerge in day and night. The latter, on the other hand, local maximum occurs at the same point of the secondary local maximum of the N_{SMTP} curve, indicating that the submitted mail is delivered simultaneously with its submitting in ML-B, and that the submitted mail is regularly delivered at 05:00 in ML-C.

In the $N_{\mathrm{SMTP}}(\mathrm{D})$ curve, the rippled part is found through 14:30-17:30 and takes almost the same point as that of N_{SMTP} curve. This result indicates that the D_{q} curve at 14:30-17:30 is mainly driven by the SMTP access of the user D. In the SMTP logs, the strange "User Unkonw" messages are found before and after the SMTP access of the user D and repeats the same pattern of access are also found. These strange pattern have been detected in the usual IDS logs by mass mailing worm (MMW).[19] Thus, it is possible to detect MMW by comparing $N_{\mathrm{SMTP}}(i)$ with N_{SMTP}. In other words, we can detect MMW by only analysis of the D_{q} curve, since the D_{q} curves resembles well the N_{SMTP} one. Furthermore, it is also possible to identify the MMW-infected E-mail user or PC terminal by analysis of the SMTP logs.

The users E and F are ML SMTP accounts and their local maximums of $N_{\mathrm{SMTP}}(\mathrm{E})$ and $N_{\mathrm{SMTP}}(\mathrm{F})$ curves slightly contribute to two peaks after the

rippled part (17:30) and the fourth local maximum (22:00) of the N_{SMTP} curves. The fourth local maximum of the D_{q} is rather similar to that of the N_{POP3} curve than that of N_{SMTP} one (see Figure 1), indicating that the dial-up or PPPonE users starts to get receiving E-mail by the POP3 access before and after 22:00.

4. Concluding Remarks

We statistically investigated traffic between the DNS server and the E-mail server. Conclusions presented in this work are summarized as follows: (1) The abnormally large DNS query traffic from the E-mail server emerges when the mass mailing worm (MMW)-infected the PC terminals increase. (2) Since the DNS query traffic is mainly driven by the total number of SMTP access in the E-mail server, it is easy to detect the abnormality of the E-mail server by only watching the DNS query traffic from the E-mail server. (3) The abnormally large traffic of the total SMTP access is mainly driven when the abnormally large SMTP traffic of the E-mail users emerge. This is because the total number of the SMTP access are represented as a sum the SMTP access number of the E-mail user. As a result, we can reasonably conclude that it is of considerable importance to investigate the DNS query traffic generated by the E-mail server.

It is well-known that MMW diffuses through an attachment file of the E-mail and that MMW uses the SMTP access to send worm-included E-mail to the next victim PC terminal. The DNS traffic increases by the MMW-SMTP access. As a result, the DNS query traffic from the E-mail server or the MMW-infected PC terminal provides us important information of MMW, Therefore, we can statistically detect infection of MMW and can know quickly a location of the MMW-infected PC terminals by only watching traffic between the DNS server and the E-mail server/PC terminals. To get further information to develop a new statistics-based IDS (SIDS), a direct/indirect traffic between the DNS server and the DNS clients is under further investigation.

Acknowledgement. All the calculations were carried out with AMD Athlon, Intel Pentium III, and Sun Microsystems Ultra-Sparc machines in our center.

References and Notes

(1) S. Northcutt and J. Novak, "Network Intrusion Detection", 2nd ed; New Riders Publishing: Indianapolis, 2001.

(2) M. Bauer, "Stealthful Sniffing, Intrusion Detection and Logging", *LINUX Journal*, No.102, pp.34-40, 2002.

(3) D. Jones, "Building an E-mail Virus Detection System for Your Network", *LINUX Journal*, No.92, pp.56-65, 2001.

(4) K. Yamamori, "An Improvement of Network Security Using an Intrusion Detection Software", *Journal for Academic Computing and Networking*, No.4, pp.3-13, 2000.

(5) Z. S. Su and J. B. Postel, "The Domain Naming Convention for Internet User Applications", RFC819, Network Information Center,SRI International, Menlo Park, California, 1982.

(6) Y. Musashi, R. Matsuba, and K. Sugitani, "Traffic Analysis on a Domain Name System Server. SMTP Access Generates Many Name-Resolving Packets to a Greater Extent than Does POP3 Access", *Journal for Academic Computing and Networking*, No.6, pp.21-28, 2002.

(7) J. B. Postel, "Simple Mail Transfer Protocol", RFC821, Network Information Center, SRI International, Menlo Park, California, 1982.

(8) M. T. Rose, "Post Office Protocol - Version 3", RFC1081, The Wollongong Group, Palo Alto, California, 1982.

(9) **1DNS** is the secondary DNS server of the Kumamoto Univerity (kumamoto-u) which

is run by our center. The OS is Linux OS (kernel-2.4.16), and the AMD Athlon 1.4 GHz.

(10) **1MX** is our mail server of the generic domain name of the Kumamoto Univerity (kumamoto-u). The OS is Solaris 2.6 (Ultra-SPARC 300MHz, Sun Microsystems Inc.).

(11) http://www.isc.org/products/BIND/

(12) eric@ojnk.nu, http://tower.zot.nu/%7Eeric/

(13) http://www.st.ryukoku.ac.jp/~kjm/security/-memo/1999/07.html

(14) M. Bauer, "syslog Configuration", *LINUX Journal*, No.92, pp.32-39, 2001.

(15) http://www.sendmail.org/

(16) http://www.eudora.com/qpopper/

(17) http://www.symantec.com/region/jp/sarcj/re-fa.html

(18) A is the user A of the top SMTP user, B is the user B of the secondary top SMTP user, C is the user C of the third top SMTP user, D is the user D of the fourth top SMTP user, E is the user E of the fifth top SMTP user, and F is the user F of the sixth top SMTP user for **1MX** at February 16th, 2002.

(19) M. Yamaguchi, "Countermeasure for Computer-Virus", *Journal for Academic Computing and Networking*, No.6, pp.47-52, 2002.