マルチメディア通信と分散処理 111−33
コンピュータセキュリティ 20−33
（２００３．２．２８）

DNS トラフィックとメールサーバのログ解析

武蔵 泰雄 †、松葉 龍一 †、杉谷 賢一†

概要: Frethem.K 等の大量メール送信型ワーム (MMW) の感染拡大が進行している時に、DNS 及び E-mail サーバ間のドメイン名前解決 UDP パケットの流量およびメールサーバにおける SMTP access log に関する統計的調査を行なった。我々の得た興味深い結果は以下の通りである: (1) MMW の感染した PC 端末が増加すると、配送延期された E-mail (stat=Deferred) 数が増加する。(2) 配送延期された E-mail 数が増加すると見掛け上 DNS 名前解決 ($D_{\rm q}$) 流量が増加する。これらは未知の MMW が検知されると、多くの E-mail サーバにおいて E-mail の受信を拒否することがしばしば行われるためであると考えられる。その結果、E-mail サーバの DNS サーバに対する $D_{\rm q}$ および SMTP syslog を監視することにより、MMW の感染の拡がりを検知可能であることが示唆された。

Statistical Analysis in Logs of DNS Traffic and E-mail Server

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

Abstract: The DNS query ($D_{\rm q}$) traffic between the DNS and E-mail servers of Kumamoto University was statistically investigated when a lot of PC terminal were infected by the mass mailing worm (MMW) like Frethem. K. The interesting results are: (1) The number of the deferred E-mail (stat=Deferred) increases when the MMW infected-PC terminal increases. (2) The $D_{\rm q}$ traffic increases in appearance when the number of the deferred E-mail increases. This is because a lot of E-mail servers are frequently closed to the E-mail receiving after detection of an unknown MMW. Therefore, we can detect an increase in MMW-infection by monitoring the $D_{\rm q}$ traffic from the E-mail server to the DNS server and the SMTP syslog of the E-mail server.

1. Introduction

Intrusion detection system, IDS,[1−7] is one of attractive solutions to keep security of the network servers such as the domain name system (DNS)[8] 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 (Misuse Intrusion Detection; MID),[4] which employs a database of the remote attacking pattern, to detect abnormality of the network servers, and the other is statistically to find abnormality of the network servers (Anomaly Intrusion Detection; AID).[4,5] The former needs to update frequently the signature file because of quick development of cracking technologies. However, the latter does not always need to update the signature files. Also, it detects change in profiles of the users and/or the systems in the network servers. In order to develop a new effective statistical IDS against future remote attack on the network servers, it is of considerable importance to get detailed profile/information for traffic of network packets like DNS query packets between a DNS server and a DNS client.[8]

We previously reported that the number of DNS query packets, $D_{\rm q}$, are predominantly generated from an E-mail server,[9]

$$\begin{aligned} D_{\rm q} &= m_{\rm S}N_{\rm S} + m_{\rm P}N_{\rm P} \qquad (1)\\ m_{\rm S} &= 2 + 4n(1-q) = 8 \sim 9\\ m_{\rm P} &= 1 \end{aligned}$$

where $N_{\rm S}$, $N_{\rm P}$, and n represent the numbers of the simple mail transfer protocol (SMTP)[10] access, the post office protocol version 3 (POP3)[11] access, and different domain hosts, respectively, and $m_{\rm S}$ and

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

$m_{\rm P}$ are linear coefficients. These results show that the DNS access from the E-mail server is mainly driven by the SMTP access. Here, a mail-receiving rate is $q = N_{\rm S}(r)/(N_{\rm S}(r) + N_{\rm S}(t))$, in which r and t show the received and the transfered E-mails, respectively.

In the present paper, we statistically investigated traffic of the DNS query packets between the DNS server (**1DNS**)[12] and the E-mail server (**1MX**)[13] when detecting an unknown MMW (Frethem. K).[14] We compare both logs of SMTP and POP3 accesses with that of DNS query access and show how the SMTP access by the MMW-infected PC terminals affects the DNS query access. Our purpose is to find out a detection method of the abnormality in the E-mail server with statistical analysis of the DNS traffic by the E-mail server.

2. Used Server Daemon Programs and Estimation of $D_{\rm q}$, $N_{\rm S}$, $N_{\rm from}$, $N_{\rm to}$, and $N_{\rm P}$

In **1DNS**, the BIND-9.2.1 program package has been employed as a DNS server daemon.[15] The log of DNS query packet has been recorded by the iplog-1.2 program[16,17] with the syslog system.[18] In **1MX**, the sendmail-8.9.3 program package[19] and the Qualcomm qpopper-4.0 program package[20] were installed as SMTP and POP3 server daemons, respectively. The log of SMTP and POP3 accesses have been recorded in the syslog file. All of the syslog files are daily updated by the crond system.

The $D_{\rm q}$, $N_{\rm S}$, $N_{\rm from}$ $N_{\rm to}$, and $N_{\rm P}$ values are obtained, as follows: (1) The $D_{\rm q}$ value is given by the number "domain" lines of /var/log/messages in **1DNS** with the grep/wc commands. (2) The $N_{\rm from}$ value is as the same as $N_{\rm S}$ value which is the number of "from=" lines of /var/log/syslog in (**1MX**) with the grep/wc commands. (3) The $N_{\rm to}$ and $N_{\rm P}$ values were provided by the numbers of "to=" and "popper" lines of /var/log/syslog in **1MX**, respectively, with the grep/wc commands. Usually, the $N_{\rm to}$ value is represented as

$$N_{\rm to} \sim N_{\rm from} \tag{2}$$

This is because "to=" line includes information related to a E-mail destination corresponding to a domain host.

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, $D_{\rm q}$ access (**1DNS**), the SMTP accesses, $N_{\rm from}$ and $N_{\rm to}$ (**1MX**), and the POP3 access $N_{\rm P}$ (**1MX**) in Figure 1. The observation was performed at July 15th, 2002 when infection of Frethem. K starts to increase. In Figure 1, the $D_{\rm q}$ traffic curve rises strait upon going from 08:00 to 09:00, considerably increases up to 10:00 with small fluctuation, slightly decreases to a local minimum at 12:00, and repeats a local maximum twice at 14:00 and 17:30. These features are common so that almost users of **1MX** start to use an E-mail application in the morning, and start to return back to home from 17:00. The $D_{\rm q}$ curve resembles well $N_{\rm from}$ and $N_{\rm to}$ curves, indicating that both $N_{\rm to}$ and $N_{\rm from}$ values contribute to the $D_{\rm q}$ value in a much greater extent than that of the $N_{\rm P}$ value (see eq (1)).

It is of great interest that the $D_{\rm q}$ curve looks like the $N_{\rm to}$ curve after 18:00, while the $N_{\rm from}$ curve contributes to the $D_{\rm q}$ curve in a small scale manner. These features indicate that the $D_{\rm q}$ value is sometimes driven by the $N_{\rm to}$ value as compared with $N_{\rm from}$ value in a large scale manner. Thus, we should correct eq (2) as

$$N_{\rm to} \geq N_{\rm from} \tag{3}$$

As a results, although the $N_{\rm to}$ value is normally either equal to or slightly greater extent than the $N_{\rm from}$ value, the $N_{\rm to}$ value is sometimes abnormally much greater than the $N_{\rm from}$ value and mainly contributes to the $D_{\rm q}$ value in appearance.

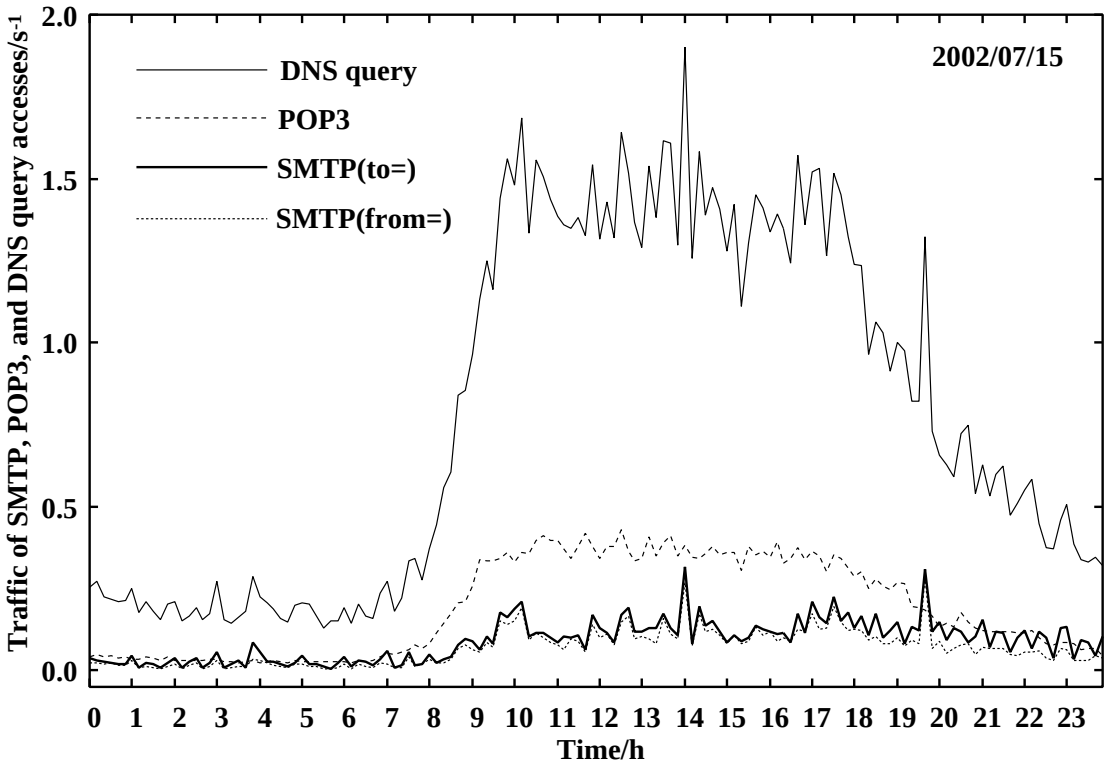

Figure 1. Traffic of the SMTP, POP3, and DNS query accesses in July 15th, 2002. The first curve shows the DNS query access, the second curve means the POP3 access, the third and fourth curves indicate SMTP accesses where the real and broken curves demonstrate numbers of "to=" and "from=" lines, respectively (s^{-1} unit).

3.2 Deferred E-mail and DNS access

As is well-known,"to=" line in syslog file mainly consists of a "stat=Sent" line which means that the E-mail is sent, successfully. The "to=" line consists of a "stat=Deferred" line which shows that the E-mail is deferred. Thus, we can approximate

$$N_{\rm to} = N_{\rm SS} + N_{\rm SD} \tag{4}$$

where $N_{\rm SS}$ and $N_{\rm SD}$ represent the numbers of the "stat=Sent" and "stat=Deferred" lines. We illustrate observed $N_{\rm to}$, $N_{\rm SS}$, and $N_{\rm SD}$ curves in Figure 2. Both $N_{\rm SS}$ and $N_{\rm to}$ curves correspond to each other and $N_{\rm SD}$ curve takes a characteristically rippled curve through 00:00-17:00. In normal situation, the E-mail is sent to its destination so that the $N_{\rm to}$ value is nearly equal to the $N_{\rm SS}$ value. As described above, $N_{\rm to}$ is nearly equal to $N_{\rm from}$ *i.e.*,

$$N_{\rm to} \sim N_{\rm from} \sim N_{\rm SS} \tag{5}$$

The $N_{\rm to}$ value is, however, to be the sum of $N_{\rm SS}$ and $N_{\rm SD}$ values when the deferred E-mail is accumulated to the E-mail server;

$$N_{\rm to} \sim N_{\rm SS} + N_{\rm SD} > N_{\rm from} \tag{6}$$

As a result, we can easily detect how much the E-mail is deferred in the E-mail server by comparing the $N_{\rm to}$ value with the $N_{\rm from}$ one in syslog file. Such situation gradually appears increasing in the amplitude of the $N_{\rm SD}$ curve (see Figure 2, after 17:00).

Also, we select the most deferred E-mail domain (A) and draw it in Figure 2(A). In Figure 2, the domain (A) curve is considerably similar to the $N_{\rm SD}$ curve after 17:00. This means that the $N_{\rm SD}$ value is mainly driven by the domain (A) value. After 17:00, the unknown MMW-infection started to expand so that an administrator outside our university had closed in emergency to the SMTP access from our university. Hence, the domain (A) curve gradually increased. We clearly conclude that the E-mail server accumulate the deferred E-mail when

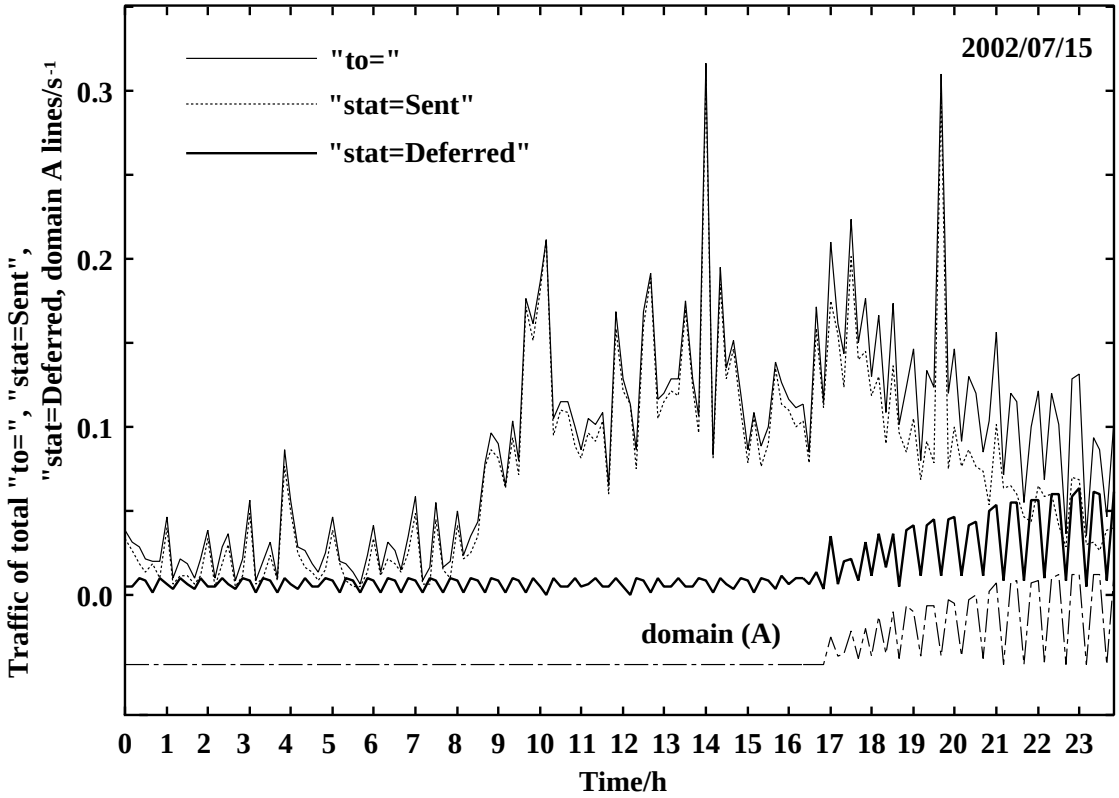

Figure 2. Traffic of total "to=", total "stat=Sent", total "stat=Deferred" lines, and top user domain of "stat=Deferred" lines in syslog file for sendmail of the E-mail server at July 15th, 2002. The first curve shows "to=", the second curve means "stat=Sent", the third and fourth curves indicate "stat=Deferred" where the real and broken curves demonstrate total "to=Deferred" and "stat=Deferred" for top domain of users, respectively (s^{-1} unit).

infection of the unknown MMW like Frethem. K increases.

Why is the D_q curve similar to the $N_{SD}(N_{to})$ curve after 17:00? We carried out regression analysis between D_q and N_{SD} values. We define R_{SD},

$$R_{SD} = m_{SD} N_{SD} \tag{7}$$

where R_{SD} is the number of the DNS query access by N_{SD}. From eqs (1) and (7), the relation between calculated and observed the DNS query accesses, D_q^{obs} and D_q^{calc}, are

$$D_q^{obs} - D_q^{calc} = m_{SD} N_{SD} \tag{8}$$

The correlation coefficient (R^2) is, unexpectedly, calculated to be almost zero.

The results described above can be interpreted in terms of the following reasons: A lot of E-mails are deferred so that E-mail users would repeat to send E-mail. In other words, the deferred E-mail causes to regenerate a newly deferred E-mail. This is one reason that the N_{from} value increases by the sending E-mail so as to regenerate DNS query packets (the D_q access increases). The another reason is that other E-mail servers using the E-mail server as a SMTP relay may retry to send the E-mail at stated periods. We expect that the latter is reasonable because the N_{SD} curve after 17:00 repeats at intervals of an approximately half of hour and the amplitude of the N_{SD} curve gradually increases.

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 "to=" line in syslog file includes important information of SMTP access and the DNS query access increases when the number of "to=" line is

larger than that of "from=" one. (2) The number of "to=" line ($N_{\rm to}$) is represented as $N_{\rm to} = N_{\rm SS} + N_{\rm SD}$, where $N_{\rm SS}$ and $N_{\rm SD}$ are the numbers of "stat=Sent" and "stat=Deferred" lines, respectively. The $N_{\rm to}$ value is usually driven by $N_{\rm SS}$ value. However, if infection of the unknown MMW, like Frethem. K, increases, the administrator of the E-mail server usually stops/closes the SMTP access to prevent further expansion of the MMW-infection. We can concluded that not only the DNS query traffic generated by the E-mail server but also the number of "to=" line in syslog file of the E-mail server give us important information of the MMW-infection.

It is well-known that MMW expands 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 identify quickly IP addresses 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) Northcutt, S. and Novak, J., *Network Intrusion Detection*, 2nd ed; New Riders Publishing: Indianapolis (2001).

2) Sato, I.,Okazaki, Y., and Goto, S.: An Improved Intrusion Detecting Method Based on Process Profiling, *IPSJ Journal*, Vol. 43, No.11, pp.3316-3326 (2002).

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

4) Denning, D. E.: An Intrusion-detection model, *IEEE Trans. Soft. Eng.*, Vol. SE-13, No.2, pp.222-232 (1987).

5) Mukherjee, B., Todd, L., and Heberlein, K. N.: Network Intrusion Detection, *IEEE Network*, Vol. 8, No.3, pp.26-41 (1994).

6) Hofmeyr, S. A., Somayaji, A., and Forrest, S.: Intrusion Detection Using Sequences of System Calls, *Computer Security*, Vol. 6, No.1, pp.151-180 (1998).

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

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

9) (a) Musashi, Y., Sugitani, K., and Matsuba, R.: Traffic Analysis on Mass Mailing Worm and DNS/SMTP, *IPSJ SIG Notes, Computer Secuirty 19th*, No.2002, pp.19-24 (2002). (b) Musashi, Y., Matsuba, R., and Sugitani, K.: 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).

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

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

12) **1DNS** is the secondary DNS server of the Kumamoto Univerity (kumamoto-u) which is run by our center. The OS is Linux OS (kernel-2.4.19), and the AMD Athlon 1.4 GHz.

13) **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.).

14) http://www.symantec.com/region/jp/sarcj/refa.html

15) http://www.isc.org/products/BIND/

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

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

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

19) http://www.sendmail.org/

20) http://www.eudora.com/qpopper/

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