Statistical Analysis in Syslog Files in DNS and Spam SMTP Relay Servers

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

[†]Center for Multimedia and Information Technologies, Kumamoto University, Kurokami, Kumamoto-City, 860-8555, JAPAN

Abstract: The syslog files of the subdomain E-mail(**sdMX**), the subdomain DNS(**sdDNS**), and the top domain DNS (**tDNS**) servers in Kumamoto University were statistically investigated when **sdMX** was a spam relay. **sdMX** worked as a spam relay becomes the worst DNS query client to **tDNS**. The main contents of the DNS query access from **sdMX** to **tDNS** are MX records. This is because the resources of **sdMX** is consumed by only the spamming SMTP relay accesses. Therefore, we can detect the subdomain E-mail server whether or not is a spam relay by only monitoring the DNS query traffic from the subdomain E-mail server to its top domain DNS server.

1. Introduction

Intrusion detection system (IDS) is one of attractive solutions to keep security of the network servers.[1−15] There are two ways of detection of abnormality in the network servers; one is a pattern-matching with a signature file which is a database of remote attacking patterns (Misuse Intrusion Detection; MID),[4,6] and the other is direct detection of abnormality in the network servers (Anomaly Intrusion Detection; AID).[4−12] With use of MID, we get a pertinent information on an attack to our network. However, it needs to update signature files frequently because of quick development of cracking technologies. On the other hand, AID does not require such signature files. It is a disadvantage in AID that we can not unambiguously differentiate an attack from a spontaneous increase in traffic.

In order to develop a new useful AID-based IDS against future remote attacks 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. We have shown that DNS query packets are predominantly generated from an E-mail server, and that the DNS access from the E-mail server is mainly driven by SMTP accesses.[16−19] Moreover, we have found a relation between the number of the DNS query packets $D_{\rm q}$ and those of the SMTP $N_{\rm S}$ and POP3 $N_{\rm P}$ accesses; $D_{\rm q} = (2 + kn(1 - q))N_{\rm S} + N_{\rm P}$,[16] where n is the numbers of different domain hosts and $q = N_{\rm S}(r)/(N_{\rm S}(r)+N_{\rm S}(t))$.[16] Notation that r and t indicate received and transfered E-mails, respectively.[‡]

The present paper is in a series of correlation analyses between DNS query packets and SMTP accesses.[16] Particularly, we focus on the case where a E-mail server relays masses of spam mails to other sites. By comparing a syslog file of subdomain SMTP accesses with that of DNS query accesses in the top domain DNS server, we show how the subdomain SMTP accesses affect the top domain DNS queries.

2. Observations

2.1 Network systems

We investigate traffic of DNS query accesses between the top domain DNS server (**tDNS**) [¶] and a

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

[‡]The k value is assigned to 4 where we use an old E-mail server program package like sendmail-8.9.3[20] or 2 in which a new one like postfix-2.0.6[21] is used.[19]

[¶]**tDNS** is a primary DNS server in Kumamoto University (kumamoto-u). The OS is Linux OS (kernel-2.4.21), and an AMD Athlon 2000MP machine.

subdomain E-mail server (**sdMX**).† Figure 1 shows a schematic diagram of a network observed in the present study. **tDNS** is one of the top level domain name (kumamoto-u) server and plays an important role of subdomain delegation. **sdMX** is one of subdomain network servers that operates as DNS (**sdDNS**) and SMTP (**sdSMTP**) servers. In **sdMX**, the /etc/resolv.conf file is configured to access only to 127.0.0.1 *i.e.* the configuration of resolver is directed only to **sdDNS**. The IP address of **tDNS** is only written in a root cache file in **sdDNS** so that the updating the DNS cache in **sdDNS** only depends on **tDNS**.

The zone data file in **sdDNS** is described only for the subdomain related host domain names, IP addresses, and two MX records; the former is a fully qualified domain name to the subdomain E-mail address and the latter is a generic domain name for the subdomain E-mail address. **sdSMTP** is set to allow an open relay for the local subdomain and top domain in our university, i.e., the third-party relay is omitted. The A, PTR, and MX records are always checked whenever a SMTP client accesses because of the network security. A POP before SMTP system is installed in **sdSMTP**.

2.2 A Method of Analysis

In **tDNS** and **sdMX**, BIND-9.2.2 program package has been employed as DNS and DNS cache server daemons.[22] The DNS query packets and their contents have been recorded by the query logging option (see man named.conf), as follows:

```
logging {
        channel qlog {
                syslog local1;
        };
        category queries { qlog; };
}
```

In **sdMX**, the program package of Postfix-2.0.6[21] was installed as server daemon of SMTP. The log of SMTP access has been recorded in the syslog file.[23] All of the syslog files are daily updated by the crond system.

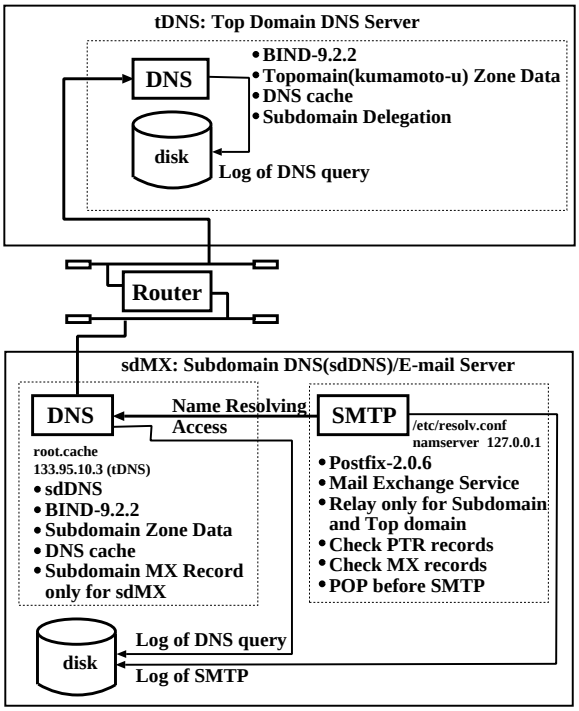

Figure 1. A schematic diagram of a network observed in the present study.

We extract lines described DNS query accesses by **sdMX** from the syslog file in **tDNS**. To check whether the DNS query traffic is in an abnormal phase or not, we need to get the values $D_{\rm q}$, $N_{\rm c}$, and $N_{\rm f}$. The steps of the procedure are as follows: The $D_{\rm q}$ value is given by the number of lines of /var/log/qlog/querylog in tDNS (grep and wc commands). The $N_{\rm c}$ value is as the same as $N_{\rm S}$ value, which is the number of "connect from" lines of /var/log/smtp/maillog in (**sdMX**) (grep and wc commands). The $N_{\rm f}$ value is provided by the number of "from=" line of /var/log/smtp/maillog in **sdMX** (grep and wc commands).

3. Results and Discussion

We observed DNS query access traffic from a subdomain E-mail server (**sdMX**) to the top domain name server (**tDNS**) for October 10th-13th,

†**sdMX** with the DNS cache **sdDNS** is a mail server in the subdomain name in Kumamoto University (sub.kumamoto-u). The OS is Linux and an AMD Athlon 2000XP machine.

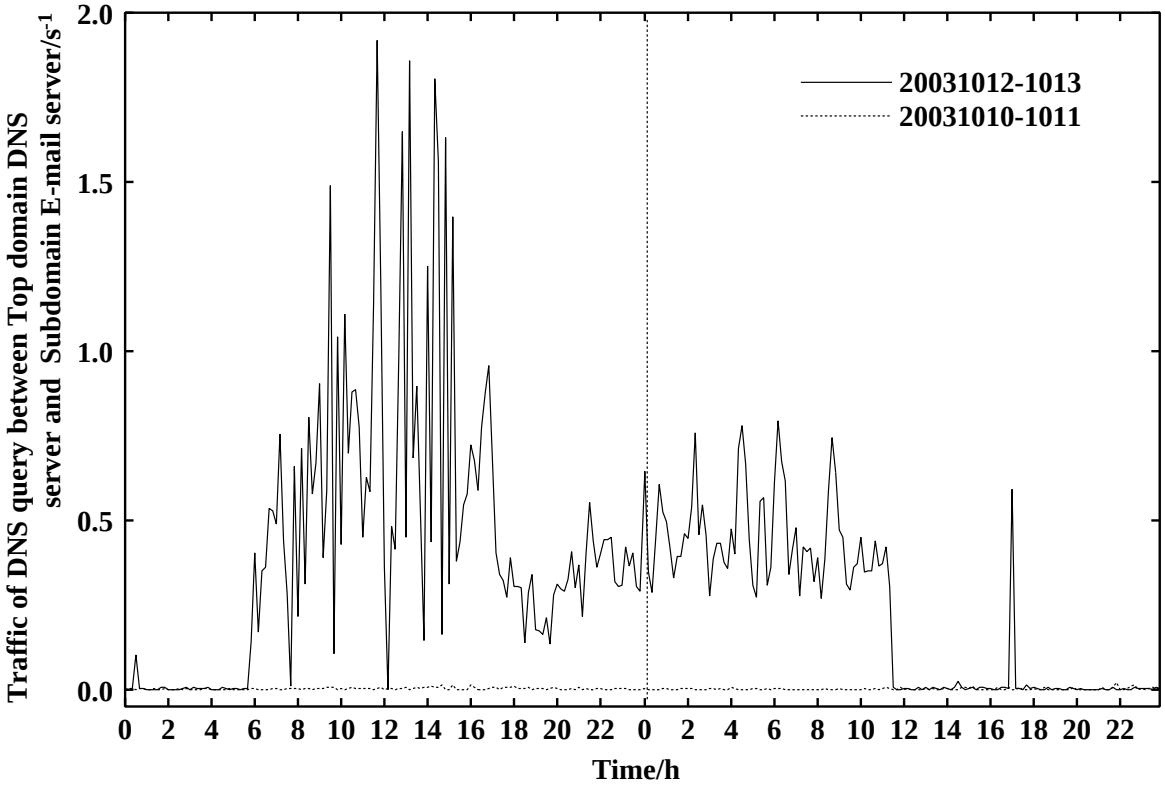

Figure 2. Traffic of the DNS query access between the top domain DNS server and the subdomain E-mail server through October 10th to 13th, 2003. The dotted line shows the DNS traffic through October 10th to 11th The solid line indicates the DNS traffic through October 12th to 13th. (s^{-1} unit).

Table 1. The total number of lines for MX, A, and PTR records per a day in the syslog file in **tDNS**, relating to the DNS client access from **sdMX**.

day	MX	A	PTR
Oct. 10th	27	114	36
Oct. 11th	36	58	16
Oct. 12th	35801	1544	22
Oct. 13th	17244	1269	21

2003. During the observation the traffic suddenly became a loud phase in the latter two days, though the faster two days was in quiescence.

We shows the observed DNS query access traffic in Figure 2. The abscissa is times in units of hour and the ordinate is access count rates from **sdMX** to **tDNS**. Since **sdMX** has a DNS cache system, **sdMX** generates only very small DNS query traffic in usual (see the dotted line in Figure 2). The subdomain DNS query traffic changed in a large scale manner after 05:30 in October 12th and the traffic stopped suddenly at 11:30 in October 13th. The large change in traffic was taken place by a

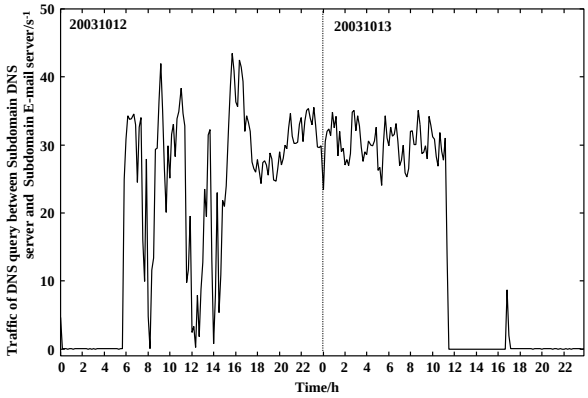

Figure 3. Traffic of the DNS query access between the subdomain DNS server and the subdomain E-mail server through October 12th to 13th, 2003 (s^{-1} unit).

spam relay in **sdMX**. How do we recognize the change as the spam attack ?

Table 1 gives the total number of lines described MX, A, and PTR records on **sdMX** for the observed days. We confirm that the DNS query is dominated by MX records, and that the query

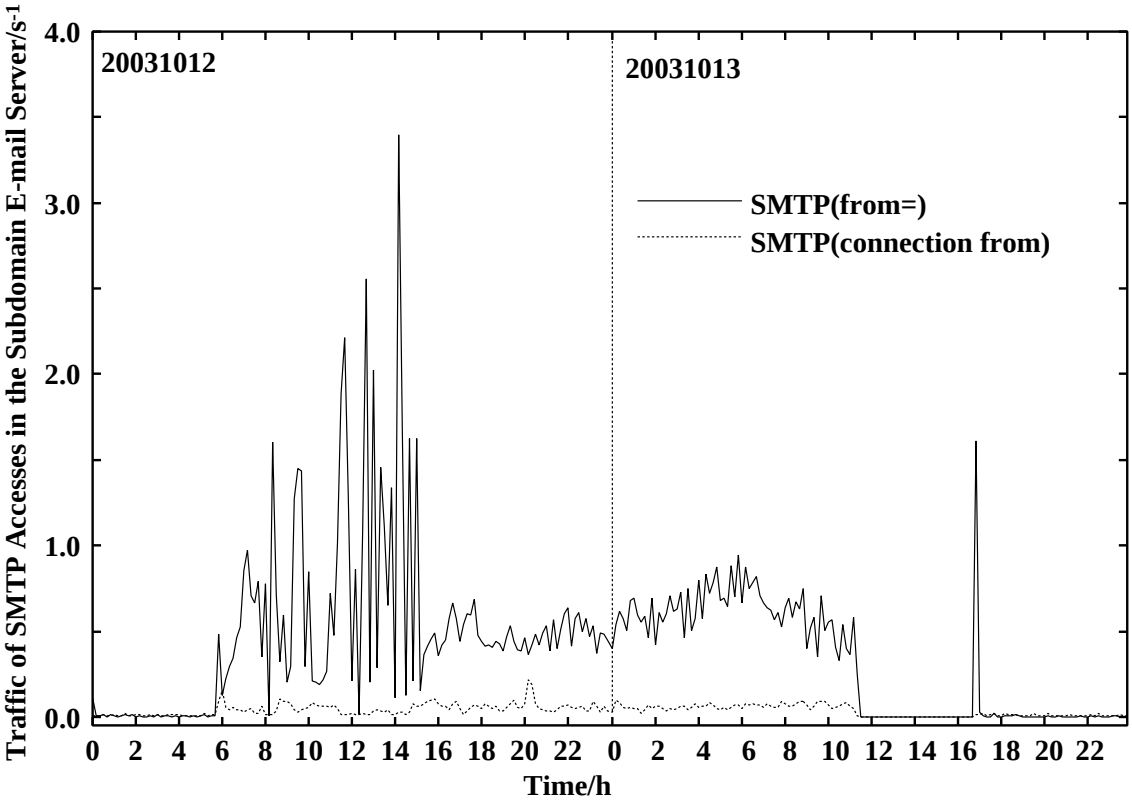

Figure 4. Traffic of SMTP accesses of the subdomain E-mail server (**sdMX**) in October 12th and 13th, 2003. The solid line indicate the access number of "from=" line ($N_{\rm f}$) and the dotted line show the access number of "connect from" lines ($N_{\rm c}$) ($\rm s^{-1}$ unit).

drastically increases in the latter two days. When we see the syslog file we encounter many lines in which "reject: RCPT" is written. The lines are distinguished as "recipient address rejected" or "user unknown".[18] Although, in usual, just one SMTP connection leads one exchange of E-mail, i.e., $\alpha = N_{\rm c}/N_{\rm f} \sim 1$, in loud phase the values of $N_{\rm c}$ become much larger or less than that of $N_{\rm f}$ values. We have already obtained α in usual to be 0.35-0.70.[16] Therefore, we can make a clear distinction between an attack and a spontaneous increase in traffic by comparing the α values in observed days with that usual one.

It is noted that we killed the process of the SMTP server daemon in **sdSMTP** at 11:30 in October 13th, so that the sudden stop in the query traffic appeared. When restarting and stopping the SMTP daemon at 17:00 in October 13th, 2003, a peak emerged again. It is clear that **sdMX** is still under the spamming SMTP relay at the time.

It is expected that the inside DNS traffic between the subdomain DNS server (**sdDNS**) and the SMTP server daemon (**sdSMTP**) in **sdMX** is almost the same or very similar to each other. We find, surprisingly, that the DNS traffic curve in Figure 3 does not resemble well that in Figure 2. This feature indicates that the DNS query traffic from the **sdMX** to **tDNS** is not including all the contents of the DNS cache server **sdDNS**, though **sdDNS** is run inside **sdMX**.

We illustrate the SMTP traffic of **sdMX** in Figure 4. Interestingly, the number of the "from=" line ($N_{\rm f}$) curve is quite similar to the DNS query access traffic curve in Figure 2. This specifically shows that the traffic of DNS query access from **sdMX** mainly consists of the MX record. If accesses flock to the DNS server to resolve FQDN/IP address, the DNS cache system processing the MX records may break down. Therefore, we consider that very large amounts of SMTP accesses, i.e.,

the mass SMTP spamming relay accesses can easily destroy the DNS cache and the DNS server.

4. Concluding Remarks

We statistically investigated system log (syslog) files in the top domain DNS server (**tDNS**), the subdomain DNS server (**sdDNS**) and the E-mail server (**sdMX**). By monitoring the DNS query accesses on **tDNS**, we have found information about detection of abnormality in **sdMX**: (1) Usually, the DNS client traffic from **sdMX** to **tDNS**, is very small, but it increases when **sdMX** is receiving the spamming SMTP relay accesses. (2) Large number of DNS query accesses make the DNS cache system break down easily. This is because the query accesses are chiefly driven by MX records (3) The broken DNS cache, as undesirable feedback, generates a mass of the DNS query traffic to **tDNS**.

We continue further investigation in order to get more information to develop an automated system detecting the subdomain E-mail server attacked by a spam relay.

Acknowledgement. All the calculations and investigations were carried out in Center for Multimedia and Information Technologies, Kumamoto University. We specially thank to technical officers, K. Tsuji, M. Shimamoto and T. Kida, and K. Makino who is a system engineer of MQS (Kumamoto) for daily supports and constructive cooperations.

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) Cisco Systems: The Science of Intrusion Detection System Attack Identification, http://www.cisco.com/warp/public/cc/pd/-sqsw/sqidsz/prodlit/idssa_wp.htm, 2002.

6) Laing, B.: How To Guide-Implementing a Network Based Intrusion Detection System, http://www.snort.org/docs/iss-placement.pdf, ISS, 2000.

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

8) Barbará, D., Wu, S., and Jajodia, S.: Experience with EMERALD to DATE", Proceedings 1st USENIX Workshop on Intrusion Detection and Network Monitoring, Santa Clara, California, April 1999, pp.73-80, http://www.csl.sri.com/neumann/det99.html

9) Neumann, P. and Porras, P.: Detecting Novel Network Intrusions using Bayes Estimators", First SIAM International Conference on Data Mining, 2001, http://www.siam.org/meetings/sdm01/pdf/-sdm01_29.pdf

10) Warrender, C., Forrest, S., and Pearlmutter, B.: Detecting Intrusions Using System Calls: Alternative Data Models, *Proc. IEEE Symposium on Security and Privacy*, No.1, pp.133-145 (1999).

11) 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).

12) Ptacek, T. H. and Newsham, T. N.: Insertion, Evasion, and Denial os Service: Eluding Network Detection, January, 1998,

http://www.robertgraham.com/mirror/Ptacek-Newsham-Evasion-98.html

13) Anderson, D., Lunt, T. F., Javitz, H., Tamaru, A., and Valdes, A.: Detecting unusual program behavior using statistical component of the Next-generation Intrusion Detection Expert System (NIDES), *Computer Science Laboratory SRI-CSL-95-06*,1995.

14) Symantec: ManHunt, http://enterprisesecurity.symantec.com/products/products.cfm?ProductID=156&EID=0

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

16) 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).

17) Musashi, Y., Sugitani, K., and Matsuba, R.: Traffic Analysis on Mass Mailing Worm and DNS/SMTP, *IPSJ SIG Notes, Computer Security 19th*, Vol. 2002, No.122, pp.19-24 (2002).

18) Musashi, Y., Matsuba, R., and Sugitani, K.: Statistical Analysis in Logs of DNS Traffic and E-mail Server, *IPSJ SIG Notes, Computer Security 20th*, Vol. 2003, No.18, pp.185-189 (2003).

19) Musashi, Y., Matsuba, R., and Sugitani, K.: Statistical Analysis in Log Files of Electronic-Mail Server and Domain Name System Server. SPAM Mail Generates Many DNS Query Packets Traffic Analysis on a Domain Name System Server, *Journal for Academic Computing and Networking*, No.7, pp.5-11 (2003).

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

21) http://www.postfix.org/

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

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