Statistical Study of Unusual DNS Query Traffic

Dennis Arturo Ludeña Romaña,* Yasuo Musashi,† Hirofumi Nagatomi,* and Kenichi Sugitani†
*Graduate School of Science and Technology
Kumamoto University, Kumamoto, 860-8555, Japan
Tel: +81-96-342-3824, Fax: +81-96-3829
E-mail: {dennis,nagatomi}@st.cs.kumamoto-u.ac.jp
†Center for Multimedia and Information Technology
Kumamoto University, Kumamoto, 860-8555, Japan
Tel: +81-96-342-3915, Fax: +81-96-3829
E-mail: {musashi,sugitani}@cc.kumamoto-u.ac.jp

Abstract— **We statistically investigated on the unusual big DNS resolution traffic toward the top domain DNS server from a university local campus network in April 11th, 2006. The following results are obtained: (1) In April 11th, the DNS query traffic includes a lot of fully qualified domain names (FQDNs) of several specific web sites as name resolution keywords. (2) Also, the DNS query traffic includes a plenty of source IP addresses of PC clients. Usually, we can observe the source IP addresses of E-mail and/or Web servers in the usual DNS query traffic, mainly. From this point, it can be concluded that the PC clients are probably infected with bot worms (BWs) and they have tried to crash the top domain DNS server.**

I. Introduction

It is of considerable importance to raise up a detection rate of bot worms (BWs), since they infect with the PC clients as well as hijacks the compromised PC clients [1-4]. After the infection or hijacking, the BW-infected PC clients becomes usually a component of the bot network (a bot) that is used to send a lot of unsolicited E-mails like spam, phishing, and mass mailing (a SMTP proxy; spam bot), to carry out a distributed denial of service (DDoS) attack (a base for cyber attack; a DDoS bot), to launch new upgraded internet worms that infect with the next victim PC clients (bot propagation), to spy out or disclosure private information (information leakage), and so on [1]. From these points, it is required to develop a countermeasure method to detect the bot worm activity.

Conventionally, we can detect spam bots by observing the clients based MX (Mail Exchange) resource record DNS query access when supposing that the client based MX RR based DNS resolution access is suspicious because the usual PC clients send only Address (A) RR based DNS query packets [7-10]. This spam bot detection model is very useful to detected a mass mailing worm (MMW) like W32/Netsky and W32/Mydoom MMWs [12,13] as well as the BW-infected PC clients when transmitting spam mails like W32/Mytob and W32/Zotob BWs [14,15]. However, it is generally difficult to detect distributed denial of service (DDoS) BW-infected/compromised PC clients.

In April 11th, 2006, we observed unusual but a large scale DNS query traffic in a campus top domain DNS (**tDNS**) server like a denial of service (DoS) attack from the campus network (Figure 1).

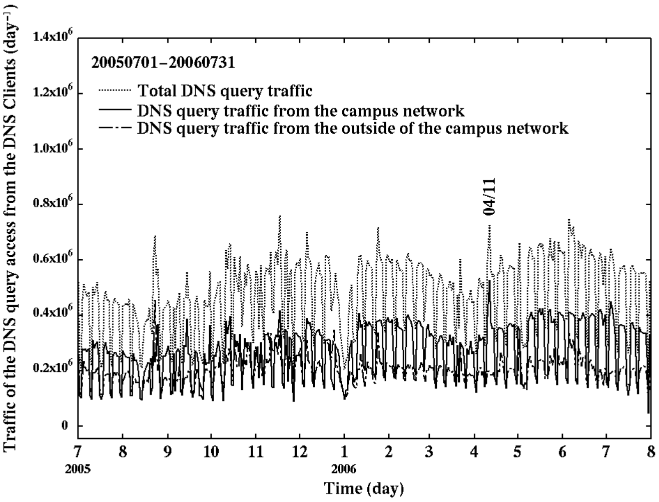


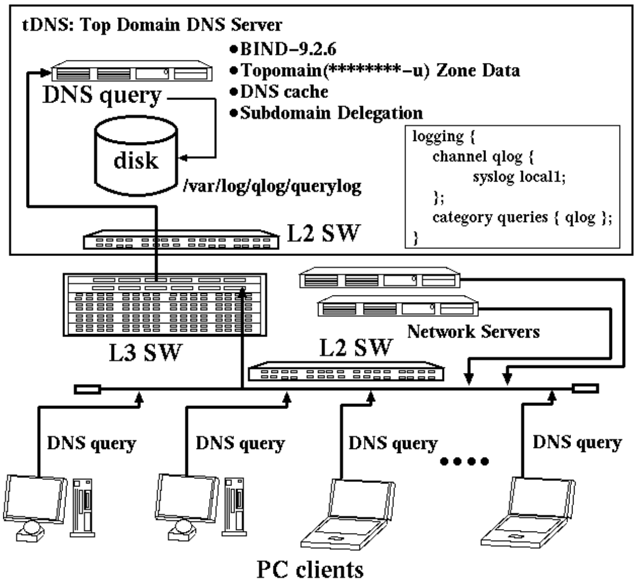
Fig. 1. Total traffic of the DNS query packets to the top domain DNS server (**tDNS**) and the traffic from the inside- and the outside-DNS clients in a university through July 1st, 2005 to July 31st, 2006 (day^{-1} unit).

In this paper, we discuss on (1) the unusual DNS query traffic from the inside of the campus network through April 10th to 11th, 2006, (2) the source IP addresses- and query keywords-based statistical analysis on the DNS query traffic, and (3) how to detect the unusual DNS query traffic.

II. Observations

A. Network System

We investigated traffic of the DNS query packets access between the top domain DNS server (**tDNS**) and the PC clients. Figure 2 shows an observed network system in the present study, an optional configuration of BIND-9.2.6 server program daemon in **tDNS**, and the three typical DNS query types. The DNS server, **tDNS**, is one of the top level DNS (kumamoto-u) servers and plays an important role of domain name resolution and subdomain delegation services for many PC clients and the subdomain network servers in the university, respectively, and the operating system is CentOS



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

Fig. 2. A schematic diagram of the network observed in the present study.

4.3Final and is currently employed kernel-2.6.9 with the Intel Xeon 3.20GHz Quadruple SMP system, the 2GB core memory, and Intel 1000Mbps Ethernet Pro Network Interface Card.

B. *Capture of DNS Query Packets*

In **tDNS**, BIND-9.2.6 program package has been employed as a DNS server daemon [16]. The DNS query packets and their query keywords (query contents) have been captured and decoded by a query logging option (Figure 2, see % man named.conf in more detail). The log of DNS query access has been recorded in the syslog files which are daily updated/rotated by the crond system. The line of syslog messages mainly consists of a source IP address and query keywords (payloads) in the DNS query packets like a fully qualified domain name (an A resource record (RR) type: standard name resolution), an IP address (a PTR RR type: reverse name resolution), and a mail exchange (an MX RR type).

III. RESULTS AND DISCUSSION

A. *Statistics of DNS Clients*

Firstly, we can show statistics of the resource record (RR) based analysis on the observed DNS query traffic from the DNS clients to the top domain DNS server (**tDNS**) in the April 10th and 11th, 2006, as shown in Table I.

Interestingly, in Table I, the DNS query traffic is mainly dominated by the A (Address) RR based DNS query one in April 10th, and 11th, 2006. Also, we can find a large difference between the both A RR DNS query traffics. This result shows that we should investigate on the A RR based DNS query traffic.

We carried out statistical analysis on the source IP addresses and query keywords in the A RR based DNS query traffic, as shown in Table II. Unexpectedly, we cannot find any suspicious source IP addresses in Table II, while we can clearly obtain significant several specific DNS query keywords *i.e.* fully qualified domain names (FQDNs) including a common keyword as "y****".

TABLE I

Statistics of the observed the A, PTR, MX and the other resource records (RRs) based DNS query traffic in April 10th and 11th, 2006 (day^{-1} unit).

	April 10th, 2006	April 11th, 2006
Total	598,425	724,887
From the campus network	398,494	527,751
A RR based	264,281	406,722
PTR RR based	68,942	61,423
MX RR based	9,588	11,396
The others RR based	55,683	48,210

TABLE II

The source IP addresses- and query keywords-based statistics of the A RR based DNS query traffic in April 11th, 2006 (day^{-1} unit).

Source IP addresses	frequency		13,991
133.95.1*.1(E-mail server)	33,195	www.y****.co.jp	13,991
133.95.1**.60	7,160	k***.***.com	6,279
133.95.1*.170	6,606	mail.c****.net	6,169
133.95.1**.41	6,543	gateway.****.com	5,604
13.95.12*.4*	5,376	bc.y****.co.jp	5,226
133.95.1**.1	5,121	ai.****.jp	4,759
133.95.12*.42	4,872	i.****.jp	4,340
133.95.1*2.133	4,720	img.y****.co.jp	4,012
133.95.1*2.73	4,376	pa.y****.co.jp	3,963
133.95.16*.58	3,850	dns**.y****.co.jp	3,749

As a result, it is possible that the unusual A RR based DNS query traffic is mainly driven by the A RR based DNS query one which consist of several specific query keywords including the common keyword as "y****".

B. *Analysis of A RR based DNS Traffic*

We illustrate the observed total A resource record (RR) based DNS query traffic and the A RR based DNS query traffic including only the specific query keywords through April 10th to 11th, 2006, as shown in Figure 3.

In Figure 3, we can easily find that the both DNS query traffic curves considerably resemble well each other through 12:00-15:00 in April 11th, 2006. This feature shows that the unusual DNS query traffic is mainly driven by the A RR based DNS query traffic only including several specific query keywords.

Furthermore, we statistically investigated on the source IP addresses in the A RR based traffic that includes only several keywords, as shown in Table III. Expectedly, in Table III, the A resource record (RR) based DNS query traffic including the specific keywords is mainly driven by the PC clients based DNS query traffic and the PC clients based DNS query traffic drastically changes through April 10th to 11th, 2006.

In order to confirm these result in more detail, we performed regression analysis on the total A RR based DNS query traffic including the specific keywords versus the traffic from the PC clients. The data are April 10th to 11th, 2006. As shown in Figure 4, the correlation coefficient (R^2) is calculated to be 0.999. This also means that the total A RR based DNS query traffic including the specific keywords

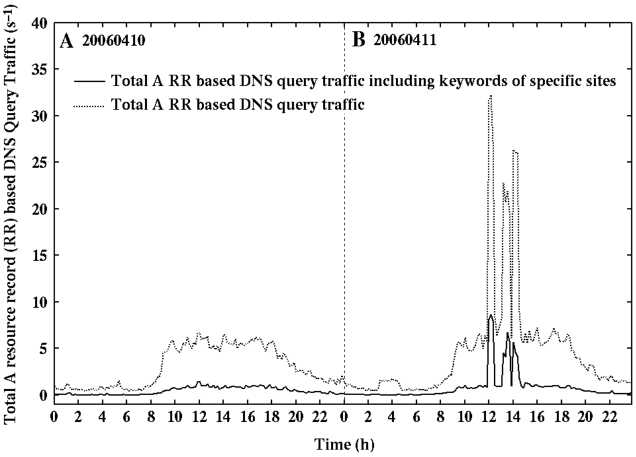


Fig. 3. Total A resource record (RR) based traffic from the campus network to the top domain DNS (tDNS) server and the traffic including query keywords as specific sites through April 10th to 11th, 2006 (s^{-1} unit).

TABLE III

Statistics of the observed the A resource record (RR) based DNS query traffic including only the specific keywords in April 10th and 11th, 2006 (day^{-1} unit).

	April 10th, 2006	April 11th, 2006
Total	39,359	72,594
PC Clients	35,789	67,391
Servers	3,570	5,197

strongly correlates with the PC clients based DNS query traffic.

Fortunately, we have already found a bot worm (BW) compromised PC client in April 11th, 2006, after investigation employing an entropy-based analysis [17].

As a result, it can be clearly concluded that the unusual DNS query traffic is mainly driven by the A RR based DNS query traffic including only the specific keywords from the compromised PC clients.

IV. Conclusions

We carried out investigation on the unusual DNS query traffic from the campus network in April 11th, 2006. Interestingly, the unusual DNS query traffic is clearly driven by the A resource record (RR) based DNS query traffic including several fully qualified domain names (FQDNs) as query keywords of a specific large-scale Web site. Also, since the DNS query traffic mainly includes the PC clients based source IP addresses and the numbers of the IP addreses are drastically increased, it can be said that the DNS query traffic is mainly generated from the PC clients on the campus network and it is possible that the PC clients are bots for distributed denial of service (DDoS) attacks or prescanning before DDoS attacks toward the top domain name server (**tDNS**) or the other specific sites.

We continue to develop detection and prevention systems based on the results of the present paper and to evaluate detection rate for the bot worm (BW) –infected PC clients as DDoS bots in the university campus and/or enterprise network because results show that the newly developed countermeasure technology is expected to detect the outbound DDoS attack or prescanning in a high precise manner.

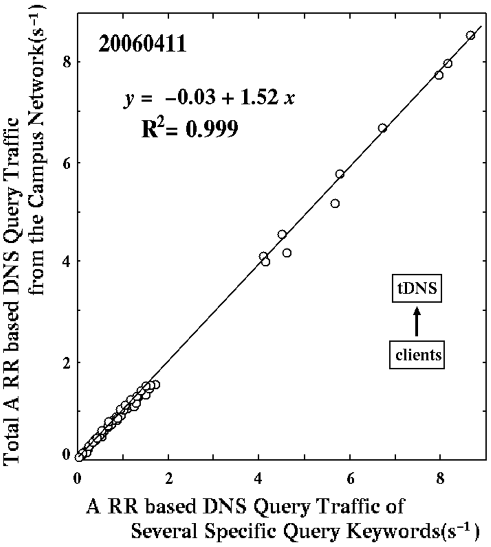


Fig. 4. Total A resource record (RR) based traffic including only the specific keywords versus the PC clients based DNS query traffic including only the specific keywords through April 10th to 11th, 2006 (s^{-1} unit).

Acknowledgment

All the studies were carried out in CMIT of Kumamoto University. We gratefully thank to all the CMIT staffs.

References

[1] P. Barford and V. Yegneswaran, “An Inside Look at Botnets, Special Workshop on Malware Detection,” Advances in Information Security, Springer Verlag, 2006.

[2] J. Nazario, “Defense and Detection Strategies against Internet Worms,” I Edition; Computer Security Series, Artech House, 2004.

[3] (a) J. Kristoff, “Botnets, detection and mitigation: DNS-based techniques,” Northwestern University, 2005, http://www.it.northwestern.edu/bin/docs/bots_kristoff_jul05.ppt. (b) J. Kristoff, “Botnets”, North American Network Operators Group (NANOG32), Reston, Virginia (2004), http://www.nanog.org/mtg-0410/kristoff.html.

[4] D. David, C. Zou, and W. Lee, “Model Botnet Propagation Using Time Zones,” *Proceeding of the Network and Distributed System Security (NDSS) Symposium 2006*, http://www.isoc.org/isoc/conferences/ndss/06/proceedings/html/2006/.

[5] A. Schonewille and D. –J. v. Helmond, “The Domain Name Service as an IDS. How DNS can be used for detecting and monitoring badware in a network,” 2006, http://staff.science.uva.nl/~delaat/snb-2005-2006/p12/report.pdf

[6] B. McCarty, “Botnets: Big and Bigger,” *IEEE Security and Privacy*, No. 1, pp.87-90, 2003.

[7] (a) Y. Musashi, R.Matsuba, and K. Sugitani, “Detection, Prevention, and Managements of Security Incidents in a DNS Server,” *Proceeding of the 4th International Conference on Emerging e-learning Technologies and Applications (ICETA2005)*, Košice, Slovakia, pp.207-211, 2005. (b) Y.

Musashi, R. Matsuba, and K. Sugitani, "Indirect Detection of Mass Mailing Worm-Infected PC terminals for Learners," *Proceeding for the 3rd International Conference on Emerging Telecommunications Technologies and Applications (ICETA2004)*, Košice, Slovakia, pp.233-237, 2004.

[8] (a) Y. Musashi, R. Matsuba, and K. Sugitani, "Prevention of A-record based DNS Query Packets Distributed Denial of Service Attack by Protocol Anomaly Detection," *IPSJ SIG Technical Reports, Distributed System and Management 38th (DSM38)*, Vol. 2005, No.83, pp.23-28, 2005. (b) R. Matsuba, Y. Musashi, and K. Sugitani, "Detection of Mass Mailing Worm-infected IP address by Analysis of Syslog for DNS server," *IPSJ SIG Technical Reports, Distributed System and Management 32nd (DSM32)*, Vol. 2004, No.37, pp.67-72, 2004.

[9] D.Whyte, P. C. van Ororschot, and E. Kranakis, "Addressing Malicious SMTP-based Mass-Mailing Activity Within an Enterprise Network," Carleton University, School of Computer Science, Technical Report TR-05-06, May, 2005, http://www.scs.carleton.ca/research/tech_reports/2005/download/TR-05-06.pdf.

[10] K. Ishibashi, T. Toyono, K. Toyoma, M. Ishino, H. Ohshima, and I. Mizukoshi, "Detecting Mass-Mailing Worm infected Hosts by Mining DNS Traffic Data," *Proceeding of the 2005 ACM SIGCOMM workshop on Mining network data*, Philadelphia, Pennsylvania, USA, pp.159-164, 2005.

[11] Y. Musashi, S. Hayashida, R. Matsuba, K. Sugitani, and K. Rannenberg, "Detection- and Prevention-System of DNS query-based Distributed Denial-of-Service Attack," *Proceeding for the 8th Asia-Pacific Network Operations and Management Symposium Toward Managed Ubiquitous Information Society (APNOMS2005)*, Okinawa, Japan, pp.574-585, 2005.

[12] http://www.trendmicro.com/vinfo/virusencyclo/default5.asp?-VName=WORM_NETSKY.Q

[13] http://www.trendmicro.com/vinfo/virusencyclo/default5.asp?-VName=WORM_MYDOOM.A

[14] http://www.trendmicro.com/vinfo/virusencyclo/default5.asp?-VName=WORM_MYTOB.A

[15] http://www.trendmicro.com/vinfo/virusencyclo/default5.asp?-VName=WORM_ZOTOB.A

[16] BIND-9.2.6: http://www.isc.org/products/BIND/

[17] A. Wagner and B. Plattner, "Entropy Based Worm and Anomaly Detection in Fast IP Networks," Proceeding of 14th IEEE Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE 2006), Liköping, Sweden, pp.172-177, 2005.