#### Traffic Analysis on a Mass Mailing Worm and DNS/SMTP

YASUO MUSASHI,<sup>†</sup> KENICHI SUGITANI,<sup>†</sup> and RYUICHI MATSUBA<sup>†</sup>

<sup>†</sup>Center for Multimedia and Information Technologies, Kumamoto University, Kumamoto 860-8555 Japan, E-mail: musashi@cc.kumamoto-u.ac.jp

#### **Domain Name System and Intrusion Detection System**

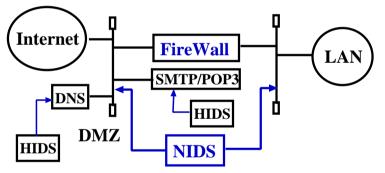
The most important network services on the Internet.

 $\mathbf{SMTP}/\mathbf{POP3}(\mathbf{Mail}), \mathbf{FTP}, \mathbf{HTTP}, \dots \ \Rightarrow \mathbf{gethostbyaddr}(), \mathbf{gethostbyname}(), \dots$ 

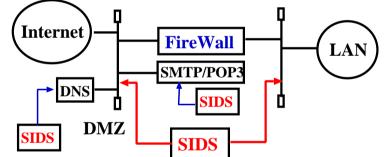
We need to protect the DNS server, firmly.



(B) Statistical Intrusion Detection System

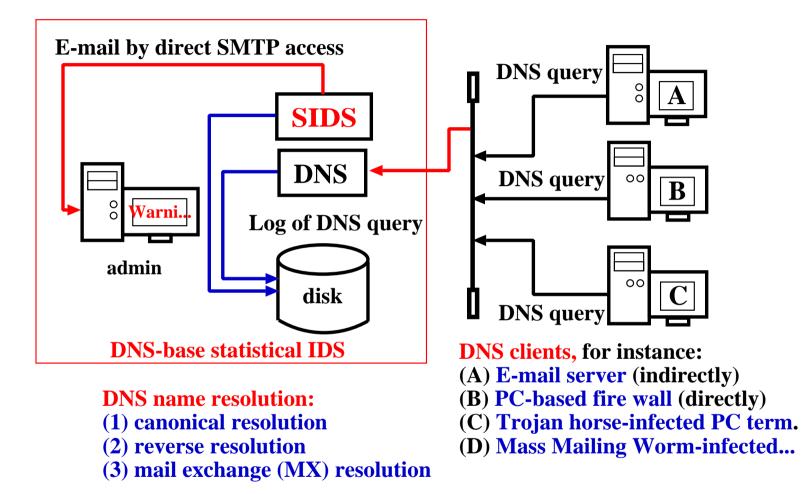


**Detection of a Signature with a Pattern File** 

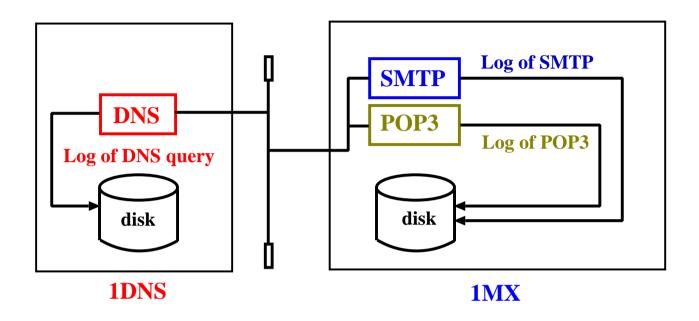


Detection of a Signature with a Statistical Method

#### Statistical Intrusion Detection by DNS query Access



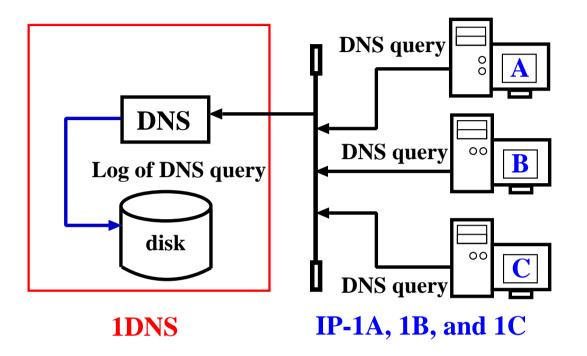
# This Work (1)



 Statistical investigation on traffic of the DNS query packets between the DNS server (1DNS) and the E-mail server (1MX).

- (2) How are the DNS query packets generated by the SMTP access which depends on the Mass Mailing Worm (MMW)-infected PC terminals?
- (3) To show methods to detect abnormality in E-mail server.

## This Work (2)



- (1) Statistical analysis on traffic of the DNS query  $(D_q)$  packets between the DNS server (1DNS) and the E-mail server (1A),
- (2) Statistical analysis on  $D_q$  traffic between the DNS server (1DNS) and the hijacked PCbased fire wall system (1B).
- (3) Statistical analysis on  $D_q$  traffic between the DNS server (1DNS) and the trojan horse virus (THV)-infected PC terminal (1C).

#### **Computations: Normal Equation 1**

$$D_{q} = R_{SMTP} + R_{POP3} + R_{FTP} + \cdots$$
<sup>(1)</sup>

$$\boldsymbol{R_i} = \boldsymbol{m_i} \boldsymbol{N_i} \tag{2}$$

 $egin{aligned} D_{ ext{q}} &= ext{the DNS query access between the 1DNS and 1MX.}\ R_i &= ext{the access numbers from the DNS clients.}\ i &= ext{a network protocol, such as SMTP, POP3, FTP, ...}\ N_i &= ext{the access counts of a network application,}\ m_i &= ext{a linear coefficient.}\ R_{ ext{smtp}} + R_{ ext{pop3}} \gg R_{ ext{ftp}} + \cdots ( ext{1MX}) \end{aligned}$ 

$$D_{\rm q} = m_{\rm SMTP} N_{\rm SMTP} + m_{\rm POP3} N_{\rm POP3}$$
(3)

# **Computations: Normal Equation 2**

$$egin{aligned} &A_{ ext{SMTP,POP3}} \; x_{ ext{SMTP,POP3}} = \; d_{ ext{SMTP,POP3}} \ & = \left[ egin{aligned} &\sum\limits_{j=1}^n N_{ ext{SMTP},j}^2 & \sum\limits_{j=1}^n N_{ ext{SMTP},j} N_{ ext{POP3},j} \ & \sum\limits_{j=1}^n N_{ ext{SMTP},j} N_{ ext{POP3},j} & \sum\limits_{j=1}^n N_{ ext{POP3},j}^2 \end{array} 
ight] \ & (j=1,2,3,\cdots,n; ext{days}) \ & x_{ ext{SMTP,POP3}} = (m_{ ext{SMTP},}m_{ ext{POP3}})^t \ & d_{ ext{SMTP,POP3}} = (\sum\limits_{j=1}^n N_{ ext{SMTP},j} D_{ ext{q},j}, \sum\limits_{j=1}^n N_{ ext{POP3},j} D_{ ext{q},j})^t \end{aligned}$$

(4)

#### **Used Server Daemon Programs and Estimation of Traffic**

Used server daemon programs

- 1DNS: The DNS server and the DNS packet recorder. BIND-9.1.3 and iplog-1.2
- 1MX:The SMTP and POP3 servers. ISC sendmail-8.9.3 and Qualcomm qpopper-4.0

**Estimation of Traffic** 

(1)  $D_{q}$ :

% grep domain /var/log/messages.1 | wc

(2)  $N_{
m SMTP}$ :

% grep "sendmail" /var/log/syslog.0 | wc

(3)  $N_{\text{POP3}}$ :

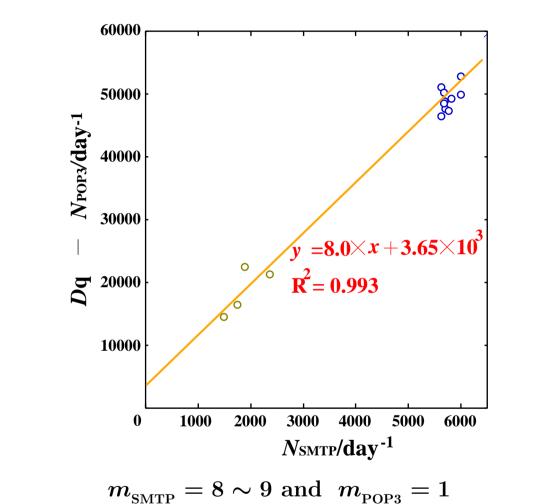
```
% grep "poppe\[" syslog.0 | wc
```

# Observed data of $N_{\text{SMTP}}$ , $N_{\text{POP3}}$ , and $D_{\text{q}}$ (day<sup>-1</sup>).

j	$N_{\scriptscriptstyle \mathrm{SMTP}}$	$N_{ m pop3}$	$D_{ m q}$
2002/02/11	1878	4480	26845
02/13	6010	17701	70327
02/14	5647	17663	68574
02/15	5744	16469	65849
02/17	1487	4004	18370
02/18	5973	16959	67262
02/19	5594	16118	62489
02/20	5666	17178	66718
02/21	5701	15851	63614
02/23	2363	6451	27540
02/24	1749	3814	20199
02/25	5731	16020	63626
02/26	5675	17688	68612

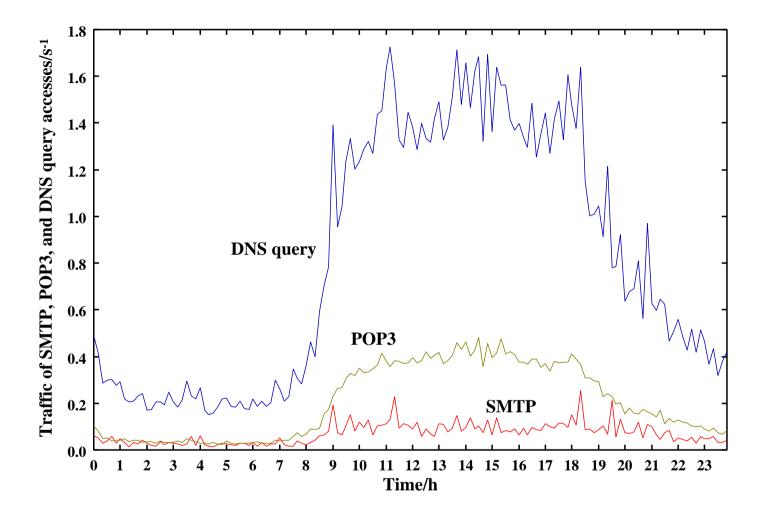
$$egin{aligned} A_{_{ ext{SMTP,POP3}}} &= egin{bmatrix} 3.120 imes 10^8 & 9.084 imes 10^8 \ 9.084 imes 10^8 & 2.652 imes 10^9 \end{bmatrix}, \ d_{_{ ext{SMTP,POP3}}} &= (3.612 imes 10^9, 1.052 imes 10^{10})^t, \ x_{_{ ext{SMTP,POP3}}} &= (8.6, 1.0)^t \ D_{_{ ext{q}}} &= 8.6 N_{_{ ext{SMTP}}} + N_{_{ ext{POP3}}} \end{aligned}$$

# $D_{\rm q} - N_{\rm POP3}$ versus $N_{\rm SMTP}$ plot



The SMTP access generates the DNS query, rather than that of the POP3 access.

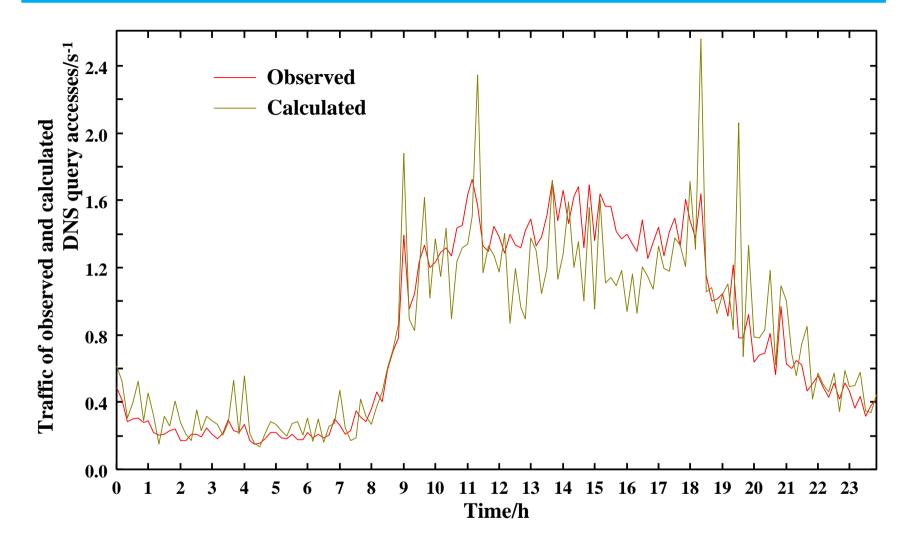
#### Traffic of SMTP, POP3, and DNS query in 2002/02/13



(1) There are three peaks.

(2) The DNS traffic resembles well the SMTP one.

#### Observed and calculated DNS traffic in 2002/02/13



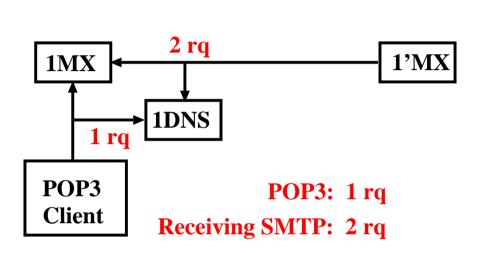
The calculated curve resembles well the observed one.

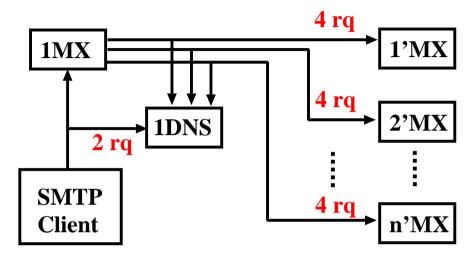
#### Why does SMTP generate larger DNS traffic than POP3?

(A) POP3 access and

**Receiving SMTP access** 

(B) Transmission SMTP access





**Transmission SMTP :** 2 + 4n rq

**1** rq = **1** request of DNS query packet

#### DNS query accesses by a SMTP access

$$R_{_{
m POP3}}=N_{_{
m POP3}}$$

(5)

(6)

(7)

# **Receiving SMTP access**

$$R^{
m rec}_{_{
m SMTP}}=2N^{
m rec}_{_{
m SMTP}}$$

#### Transmitting SMTP access

$$R^{
m tr}_{
m SMTP} = (2+4n) N^{
m tr}_{
m SMTP}$$

#### DNS vs SMTP/POP3

$$R_{\rm SMTP} = R_{\rm SMTP}^{\rm rec} + R_{\rm SMTP}^{\rm tr}$$
(8)

$$q = \frac{N_{\rm SMTP}^{\rm rec}}{N_{\rm SMTP}^{\rm rec} + N_{\rm SMTP}^{\rm tr}} \tag{9}$$

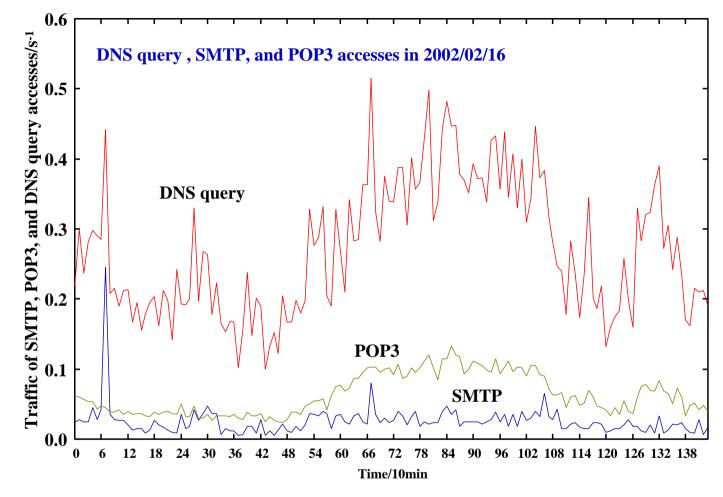
$$\begin{split} m_{\rm SMTP} N_{\rm SMTP} &= 2q N_{\rm SMTP} + (1-q)(2+4n) N_{\rm SMTP} \ (N_{\rm SMTP} > 0) \\ m_{\rm SMTP} &= 2q + (1-q)(2+4n) \\ &= 2+4n(1-q) \end{split}$$
 (10)

$$D_{q} = (2 + 4n(1 - q))N_{SMTP} + N_{POP3}$$
 (11)

If  $q = 0.50 \sim 0.75$  and  $m_{_{\rm SMTP}} = 8.6; n = 3.3 \sim 6.6.$ 

The user of 1MX sends to at least 3  $\sim$  7 persons by one E-mailing.

#### Indirect Detection of Mass Mailing Worm-infected PC



- (1) The DNS traffic resembles well the SMTP one.
- (2) Several peaks are found in the curve.

(3) The rippled part in the DNS traffic emerges when the rippled one takes place in the SMTP curve.

#### Total SMTP and User SMTP accesses

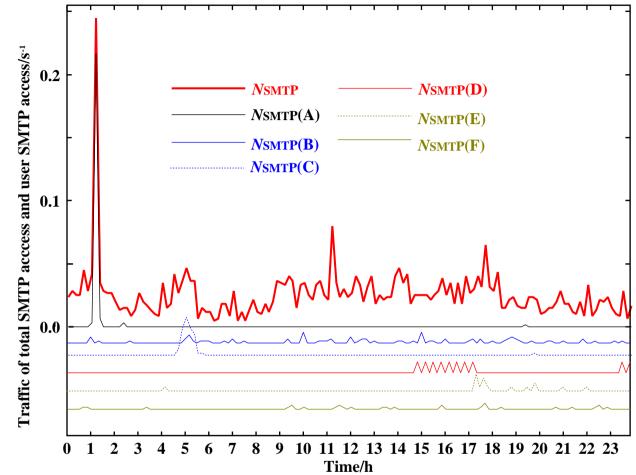
$$N_{\text{SMTP}} = \sum_{i} N_{\text{SMTP}}(i)$$
 (12)

 $N_{\rm SMTP}$  = the total number of the SMTP access in 1MX.

 $N_{\text{SMTP}}(i)$  = the number of the SMTP access by User i,

where i represents User  $A \sim F$  is the first  $\sim$  the sixth top SMTP users of 1MX.

#### Traffic of total and user SMTP accesses in 2002/02/16





- (2) Users B, C, E, and F are mailing-list (ML) accounts.
- (3) The saw tooth shaped (rippled) part can be observed the  $N_{\rm SMTP}(D)$  curve.

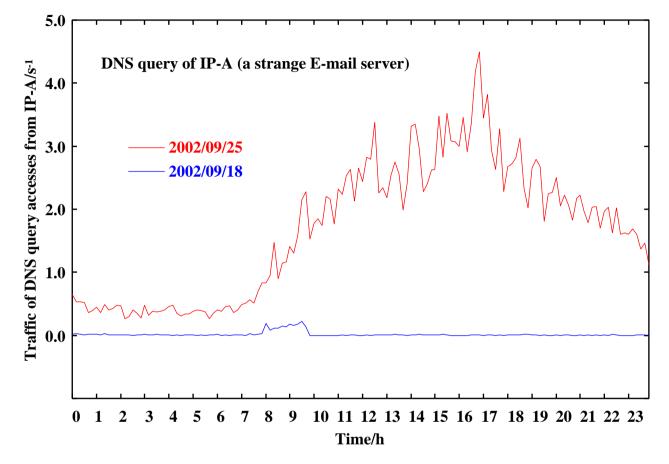
 $\Rightarrow$  Is the PC terminal of user D infected with a mass mailing worm?

#### Total DNS query and IP-terminal DNS client accesses

$$\boldsymbol{D}_{\mathbf{q}} = \sum_{i} \boldsymbol{D}_{\mathbf{q}}(i) \tag{13}$$

 $D_{q}$  = the total number of the DNS query access to 1DNS.  $D_{q}(i)$  = the number of the DNS query access by IP terminal *i*, where *i* represents IP terminals A~C is the top DNS clients of 1DNS.

#### $D_{\rm q}$ traffic curves of IP-A in normal and abnormal days



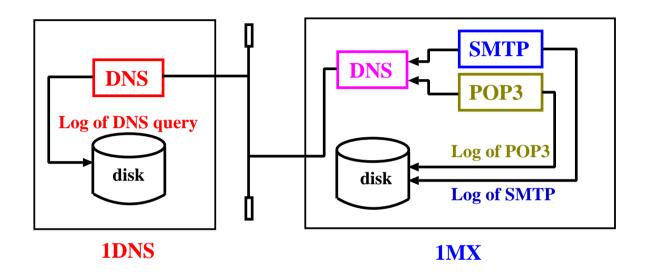
(1) In a normal day (15th), the  $D_q$  curve exhibits nearly zero.

(2) The  $D_q$  curve shows a normal curve of the E-mail server.

 $\Rightarrow The DNS query cache system virtually crashes with the increase of the mass mailing worm(MMW).$ 

#### Cache Effects on DNS traffic from E-mail servers

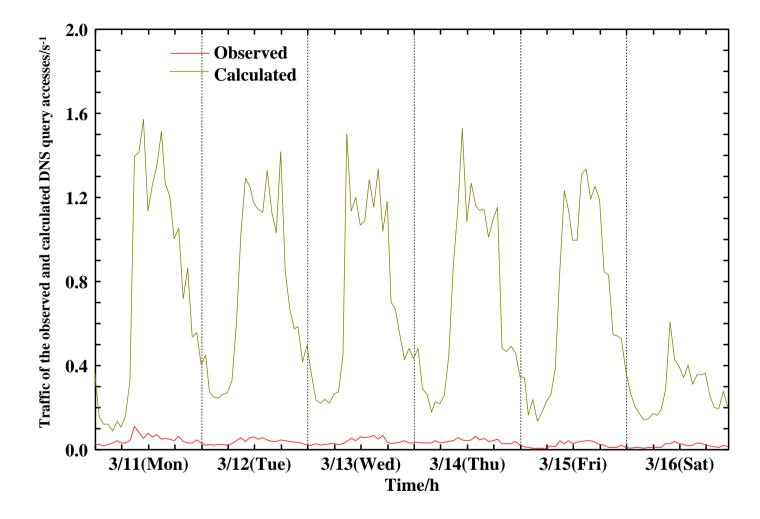
We present the DNS cache effects of the DNS query access between 1DNS and 1MX with the equation (  $D_{\rm q}=8.6N_{\rm SMTP}+N_{\rm POP3}).$ 



Used server daemon programs

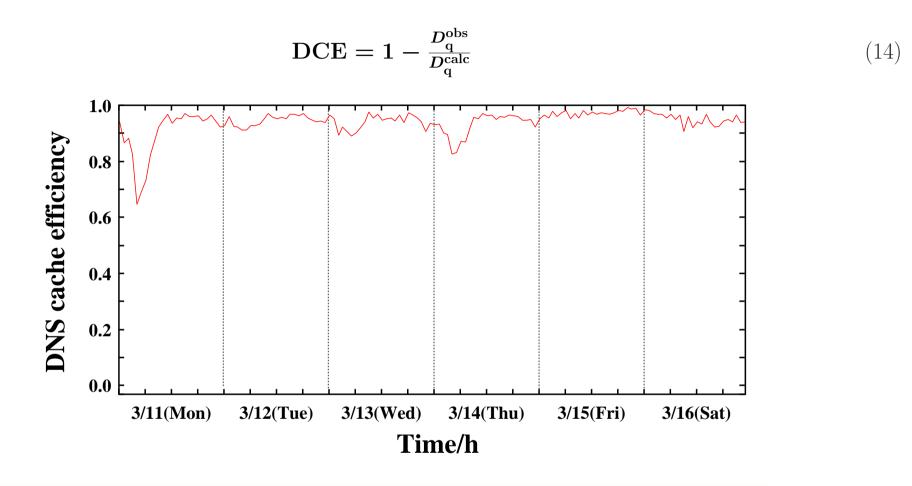
- 1DNS: The DNS server and the DNS packet recorder. BIND-9.1.3 and iplog-1.2
- 1MX:The SMTP and POP3 servers. ISC sendmail-8.9.3 and Qualcomm qpopper-4.0

#### Observed and calculated DNS traffics in 20020311-0316



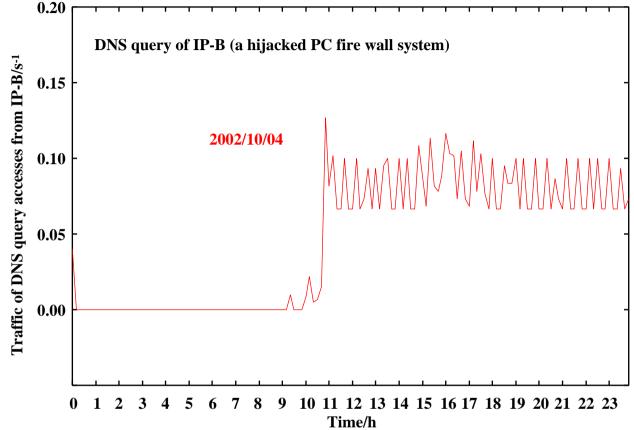
The observed traffic is considerably much smaller than the calculated one.

#### **Estimated Cache Efficiency of DNS traffic**



The DNS cache for SMTP/POP3 services is considerably effective.

#### The $D_{q}$ traffic curve of the Hijacked Fire Wall System

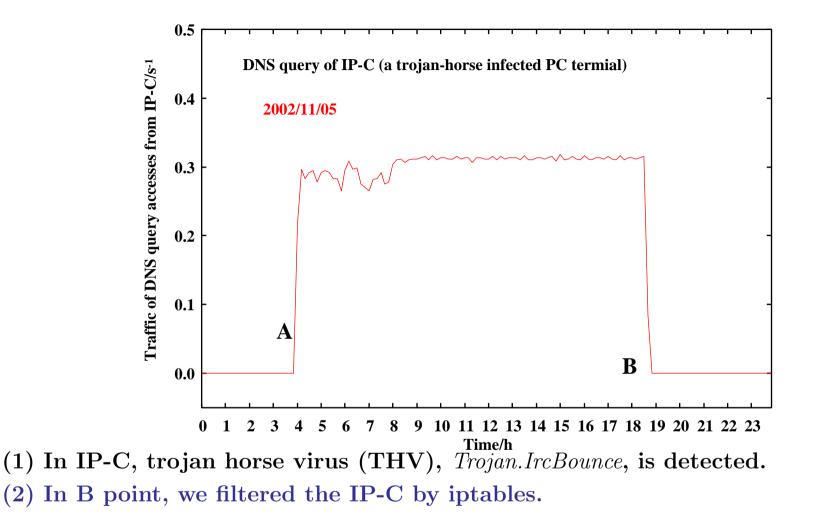


(1) The  $D_{\rm q}$  curve shows zero in the early morning.

- (2) It rises straightly upon going from 10:30 to 11:00.
- (3) The rippled part can be observed after 11:00 and the system was hijacked.

 $\Rightarrow$  The rippled curve means an indication of remotely hijacked system.

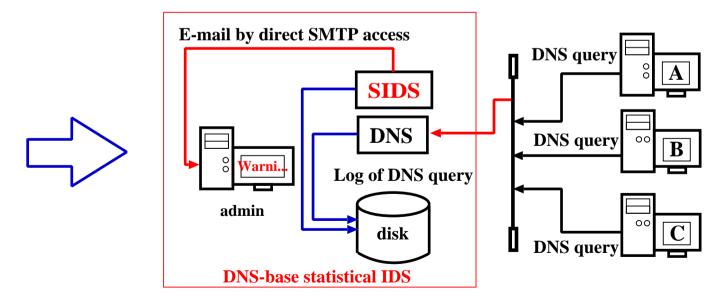
#### The $D_q$ traffic curve of the Trojan Horse Virus-infected PC



We can detected THV by only observing DNS query access from the DNS client.

### Conclusions (1)

- (1) The total number of DNS packets,  $D_q$ , are represented as,  $D_q = m_{\text{SMTP}} N_{\text{SMTP}} + m_{\text{POP3}} N_{\text{POP3}}$ , where  $N_{\text{SMTP}}$  and  $N_{\text{POP3}}$  represent the number of the SMTP access and that of the POP3 access, respectively. The linear coefficients  $m_{\text{SMTP}}$  and  $m_{\text{POP3}}$  are calculated to be 8.0-8.6 and 1.0.  $m_{\text{SMTP}} = 2 + 4n(1-q)$ , where q is a mail-receiving rate and n is a number of different domain hosts.  $\Rightarrow$  Useful information for estimation and design of an E-mail server.
- (2) In the DNS query and SMTP access curves, a rippled curve emerges when a PC terminal is infected with mass mailing worm.  $\Rightarrow$  Mass mailing worm can be detected by only observing DNS query access from E-mail server/PC terminal.



# Conclusions (2)

- (1) The DNS query  $(D_q)$  cache system virtually crashes with increase in mass mailing worm infection of PC terminals.
- (2) The plateau is observed in the curve of  $D_q$  traffic from hijacked and PC-based fire wall system.
- (3) The rippled curve emerges in the  $D_q$  traffic from the trojan horse virus-infected PC terminals.