ネットワークアプリケーションのアクセス流量間相関の意味

武藏 泰雄，松葉 龍一，杉谷 賢一
ネットコミニュケーション研究部門
musashi@cc.kumamoto-u.ac.jp

概要

我々の研究室では，DNS および E-mail サーバ等のシスログを解析することにより DNS のアクセス流量や E-mail のアクセス流量等について部分的に相関が存在することを見出しております．またその結果を応用して，ウィルス対策やセキュリティインシデントの検知に結びつくシステムの開発研究を行っており今回は，DNS と SMTP との相関関係と Welchia の対策システムについてご紹介致します．

1 背景

我々の研究では，情報セキュリティ対策に素早く適用可能なノウハウ技術を開発することが責務と考えております．主な調査対象は IDS[1−14] のログや tcpdump, Snort などの syslog[15] 及び DNS サーバのログです．

ネットワークに流れているパケットを採取して，その中にウィルスやセキュリティ攻撃等のセキュリティインシデントを検出 (検知) するシステムを一般に侵入検知システム (IDS) と呼びます．

IDS は検知方式によって不正侵入検知 (MID) 型および異常性 (AID) 検知型の 2 つ大別できます．MID 型はいわゆる signature と呼ばれる攻撃パターン等を含むデータベースとパターン整合する方式で，[4,6] 採取したデータが signature のパターンと一致すれば，検知となります．長所はどのような攻撃かはっきり判ることで，短所は既知でないと見過ごしてしまうと点です．AID 型は signature 等のデータベースを使わず，アプリケーションプロトコルの使い方が変であるとか，例えばとあるアプリケーションプロトコロルの異常に流量が多いという点で検知します．[4−12] 長所は，異常性を検知しますので，未知のインシデントを検知することが可能です．短所は，単におかしい，ということしか判らないことです．

現在は AID/MID のハイブリッドが主流となっています．MID では，Dragon, RealSecure, Snort, CISCO secure IDS 等あり，AID/MID では，ManHunt/Decoy, Netdetector, Clear Sight 等が知られています．しかしどの IDS も大量のアラート[13]を吐くということ，設定が複雑でアラートの意味はなんとなく判るがどのようにこれを組織の情報セキュリティ対策に結びつけるかと点に関してはまだまだ暗中模索なのが現状です．

そこで，AID/MID のログやアプリケーションサーバプログラムが出力するログに基づいてセキュリティインシデントを検知する方法を研究することにしました．[16−21]

2 DNS 流量と SMTP 流量間相関

この研究の動機は，DNS サーバにどのようなパケットが送られてくるのか興味がわいたので，iplog と呼ばれる簡単なパケットログを採集可能なツールを DNS サーバにインストールしたことから始まります (Figure 1).[16]

この iplog[16] のメッセージには，パケットを送ってきたクライアント側の IP アドレスとポート番号にパケットサイズが含まれています．iplog は ICMP 及び UDP パケットはすべて捕獲しますが，TCP パケットはセッションのみです．

[13]検知したというメッセージで一種のログ

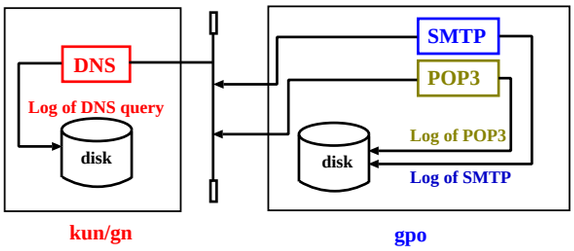


Figure 1. Investigated network system and network applications in 17th March, 2003.

DNS サーバにインストールしましたので，DNS クライアントからのアクセスは当然のように大量に得られました．そこでこれを IP アドレスごとに統計を取りますと，最もアクセス流量が多いのは E-mail サーバからのものであることが判ります．

Figure 2 は，2003 年 3 月 17 日おける DNS サーバ (gn) と E-mail サーバ (gpo) との間の DNS query アクセス量 (D_q) と E-mail サーバにおける SMTP クライアント接続アクセス流量との相関関係を示しています．[19] この Figure 2 により，E-mail サーバか

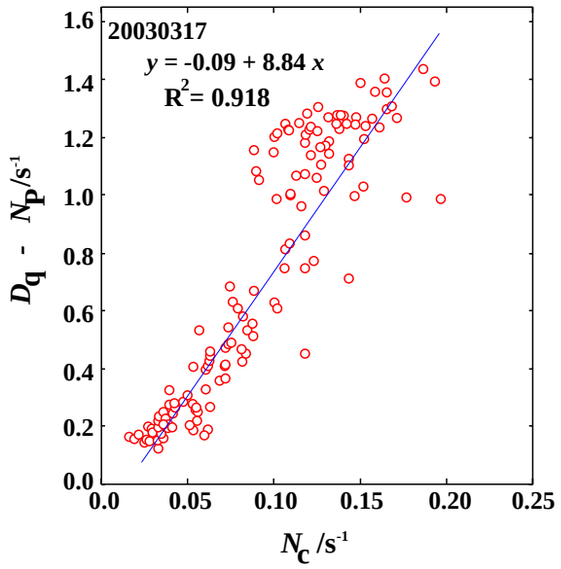


Figure 2. $D_q - N_{POP3}$ vs N_c plot (March 17th, 2003). The circle point shows a sampling data by ten minutes in the day (s^{-1} unit). Correlation coefficient (R^2) is 0.918

ら DNS サーバに対する DNS query アクセスとその SMTP アクセスは強い相関があることが示されています．

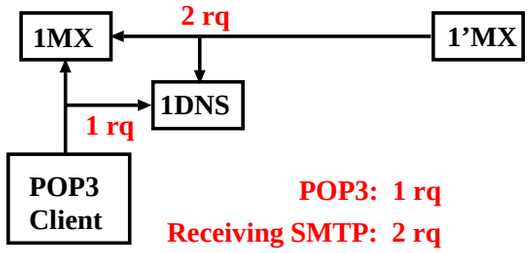
$$D_q = 8.8N_{SMTP} + N_{POP3} \quad (1)$$

なぜこのような強い相関が得られたでしょうか? そこで DNS サーバと私の研究実験用 E-mail サーバの DNS query を iplog や ethereal 等のパケット捕獲ツールで調査してみました．[16] その結果を Figure 3 に示しています．

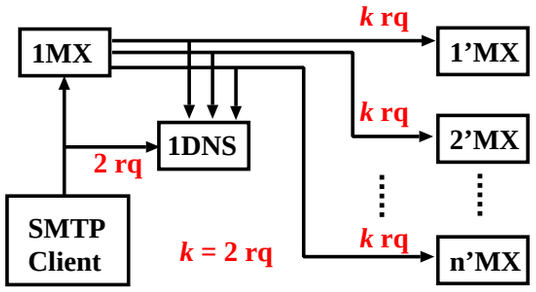
ここで E-mail の動作状態について簡単モデルを考えてみましょう．その簡単モデル化とは E-mail の動作状態を受信と送信との 2 つに分けることです．

さて受信時には，DNS サーバに対してどれくらいの DNS query が発生するのでしょうか．Figure 3A は E-mail サーバにおける受信時の動作状態を表しています．E-mail 受信時は，sendmail[22] や Postfix[23] 等の SMTP サーバプログラム (MTA) がまず E-mail を受信してサーバのディスクに全データを書き込みます．実際 E-mail を実験した MTA に送りつけると，2 つ DNS query を発生させます．ethereal で解析しますと，PTR record と A record を発生しています．後で判ったのですが，送信元の E-mail サーバがちゃんとした E-mail サーバなのかをチェックするためです．E-mail を受信する時，MTA が生成する DNS query パケットは 2 個となります．

(A) POP3 access and Receiving SMTP access



(B) Transmission SMTP access



1 rq = 1 request of DNS query packet

Figure 3. Investigated network system in a small scaled manner and how many DNS query packets are generated when receiving and/or transmitting E-mails.

$$D^{rec}_{SMTP} = 2N^{rec}_{SMTP} \quad (2)$$

ところで E-mail を PC に取り込む時は POP3 という方式を使います．この POP3 サーバプログラムも POP3 クライアント接続時に DNS サーバに登録されたかどうかチェックしています．ethereal 等で観察すると，PTR record のみチェックしているのが見られます．このことから POP3 方式は 1 回のアクセスにつき，1 個の DNS query パケットが生成されます．

$$D_{POP3} = N_{POP3} \quad (3)$$

Figure 3B では，E-mail 送信時の MTA の動作とその時に発生する DNS query パケットの個数を表しています．E-mail 送信時に MTA は，SMTP クライアントから受信したメッセージを送信先 E-mail サーバ (MTA) に中継する，SMTP 中継という役割を担当します．MTA はまず DNS query パケットの発生について 2 つのフレーズに分けられます．最初は，SMTP クライアントがちゃんとしたクラアイアントであるどうか調査します．この動作は MTA が E-mail 受信時と同じ状態です．この時 DNS query パケットは，PTR record と A record の 2 個が生成されます．次に SMTP 中継先，つまり送り先，具体的に言えば宛先アドレス (To:) の処理を行います．E-mail アドレスを@をデリミッタとすると，アカウント部分とドメイン名部分に分割できます．ドメイン部分は必ず FQDN であるとは限らないので，つまりホスト・ドメイン名ではない可能性が多いので，ドメイン名を FQDN に変換する作業が必要となります．その時 DNS サーバについてドメイン名の E-mail サーバの FQDN を得るために MX record と呼ばれる DNS query パケットを DNS サーバへ送ります．それで，FQDN が返されましたら，それをあらためて IP アドレスへ変換するため，A record を DNS query パケットとして 1 個生成します．一回の E-mail 送信時に複数ドメイン名の宛先が，あればその分だけ MX record と A record の 2 個の DNS query が必要ですので，解決すべきドメイン名が n 個であれば，下記の様な式になります．

$$D^{\rm tr}_{\rm SMTP} = (2+2n)N^{\rm tr}_{\rm SMTP} \tag{4}$$

次にネットワークサーバから DNS サーバへ送られる DNS query パケットはネットワークアプリケーションが生成する DNS query パケットを用いて下記の様に表現できます．

$$D_{\rm q} = D_{\rm SMTP} + D_{\rm POP3} + D_{\rm FTP} + \cdots \tag{5}$$

ここで E-mail サーバであるという条件を考慮すれば，上記の式は下記様な条件式が成立し，

$$D_{\rm SMTP} + D_{\rm POP3} >> D_{\rm FTP} + \cdots \tag{6}$$

結局，

$$D_{\rm q} = m_{\rm SMTP}N_{\rm SMTP} + m_{\rm POP3}N_{\rm POP3} \tag{7}$$

が得られます．ここで $m_{\rm SMTP}$ や $m_{\rm POP3}$ は線形係数であり，$N_{\rm SMTP}$ や $N_{\rm POP3}$ は SMTP および POP3 のアクセス流量を表します．

Figure 3A より，$m_{\rm POP3}$ は 1 だから式 (7) は，となり，

$$D_{\rm q} = m_{\rm SMTP}N_{\rm SMTP} + N_{\rm POP3} \tag{8}$$

となります．次に受信率を下記の式で定義します．

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

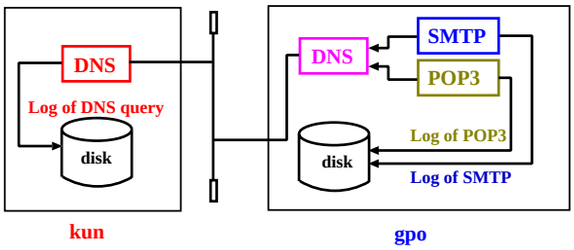


Figure 4. Investigated network system and network applications through 11th to 16th March, 2002.

この式を使って，式 (2) および (4) を下記の様に書き直すします．

$$D^{\rm rec}_{\rm SMTP} = 2qN_{\rm SMTP} \tag{10}$$

および

$$D^{\rm tr}_{\rm SMTP} = (2+2n)(1-q)N_{\rm SMTP} \tag{11}$$

となります．更に，下記の式が成立するとします．

$$D_{\rm SMTP} = D^{\rm rec}_{\rm SMTP} + D^{\rm tr}_{\rm SMTP} \tag{12}$$

上記の式は下記の様に書き直します．

$$m_{\rm SMTP}N_{\rm SMTP} = 2qN_{\rm SMTP} + (1-q)(2+2n)N_{\rm SMTP}$$

ここで $N_{\rm SMTP}$ は観測値なので

$$\begin{aligned} m_{\rm SMTP} &= 2q + (1-q)(2+2n) \\ &= 2 + 2n(1-q) \end{aligned} \tag{13}$$

となります．結論として DNS query パケット流量と SMTP アクセス流量の関する式が得られます．

$$D_{\rm q} = (2+2n(1-q))N_{\rm SMTP} + N_{\rm POP3} \tag{14}$$

評価環境の $m_{\rm SMTP}$ は 8.8 であり，仮に受信率が 0.5 とすれば，n は 6.8 となり，その E-mail サーバは一回の E-mail 送信あたり，少なくとも 6〜7 箇所の異なる E-mail サーバに同時にメッセージを送信していることを意味しています．

さてせっかく求められた式 (1) を早速使ってみましょう．[16] 筆者が咄嗟に思いついたのは，DNS cache というDNS サーバに実装されている機能です．DNS cacheってどれくらい効くのだろうかと思い，評価環境の DNS サーバと E-mail サーバの間に DNS cache 入れる前と入れた後の環境で DNS アクセス流量を測定してみました (Figure 4)．

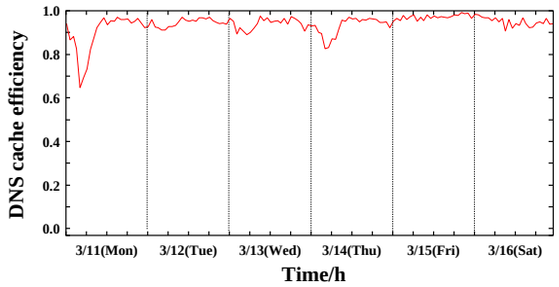


Figure 5. Changes in DNS cache efficiency upon going from March 11th-16th (2002).

下記の様な DNS cache の効率を求める式を定義します. [16]

$$\mathrm{DCE} = 1 - \frac{D_{\mathrm{q}}^{\mathrm{obs}}}{D_{\mathrm{q}}^{\mathrm{calc}}} \tag{15}$$

2002 年 3 月 11 日〜16 日までの DNS 実測値と計算値から求めた DNS cache 効率を Figure 5 にプロットしてみました．Figure 5 より，DNS cache は非常に cache 効率が高い事が判明致しました．

ここまで判ったことは，DNS 流量のある部分は SMTP 流量と強い相関があることです．つまりネットワーク上を流れるパケット流量は部分的な相関が存在することを示しています．この部分的相関を解析して行くことによって，セキュリティインシデントを検出する方法がどんどん明らかになると考えられます．[17−20] 総合情報基盤センターでは，この様なネットワークアプリケーション流量の相関分析を元にしたネットワークインシデント検知システムを構築中です．[21]

次の節では W32/Welchia.A の検出についての調査と報告です．

3 W32/Welchia.A 対策と結果

2003 年 7 月 16 日に Microsoft Security Bulletin MS03-026 が発表されましたので，この時点で Windows Update を行えば Blaster/Welchia 等の感染回避は可能であったことは良く知られている事実でした．本センターと致しましても，学内 LAN と上位 WAN との接続点でフィルタリングをただちに行いましたが，持ち込み PC による感染被害が発生致しまして，またたく間に感染が拡大し，一部の部局ではネットワーク接続が遅延や全くの不通が発生し，ほぼネットワーク全体が麻痺したかの様に思える時期が，少なくとも 2 週間程度は続きました．その後，Welchia の検出方法が判明致しましたので感染 PC の IP アドレス表を提示して事態の収拾を計りましたが，新たに購入したばかりの PC が感染拡大の大きな原因となりはじめました．その後自動検出及び自動管理担当者への即時通報システムを開発しかなり感染を減少させることに成功致しましたのでご報告したいと思います．

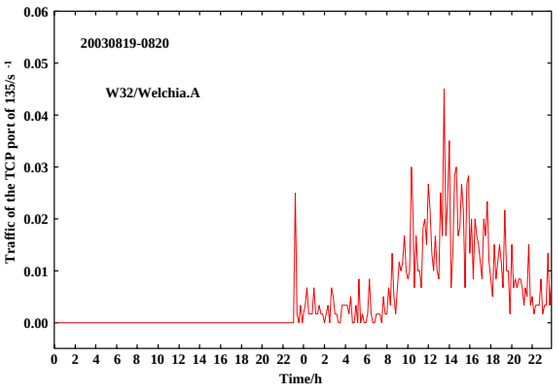


Figure 6. Traffic of the TCP port 135 trial access to the IP address of 133.95.10.3 through August 19th to 20th, 2003 (s^{-1} unit).

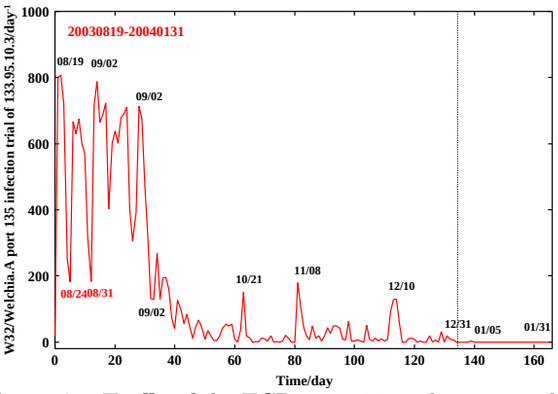


Figure 7. Traffic of the TCP port 135 trial access to the IP address of 133.95.10.3 through August 21st to January 31st, 2004 (day^{-1} unit).

W32/Welchia.A の前にまず，2003 年 8 月 11 日前後に CERT より Windows の RPC 脆弱性を狙った Blaster Worm の拡大感染の通知がありました．RPC 関連の TCP ポートを学内外から制限を掛けましたので，W32/Blaster.A をネットワーク経由で阻止することにはなんとか成功致しました．しかしながら 2003 年 8 月 19 日午後 10 時頃から W32/Welchia.A の感染が確認されました (Figure 6)．いずれも学内の PC の IP アドレスからでした．

W32/Welchia.A の検知方法，Windows XP/2000 における対処方法についてある程度技術が確立して来た 2003 年 9 月 17 日より，学内担当者を中心に，感染 PC の IP アドレスの通知を開始し，更に同年 10 月 1 日までに，LAN 切り離し等の準備および部局長などへ通知依頼を行いました．これらの処置により，9 月中旬より，W32/Welchia.A に感染した PC の検出がかなり劇的に

減少しはじめました (Figure 7).

同年 10 月 2 日より，L2 スイッチの port disable/L3 スイッチのフィルタリングを併用して LAN 切り離し等の対策を開始しました．しかしこの LAN 切り離しはあまり効果はなく，LAN 切り離しの時期を計るの非常に困難であることが判り，自動検知および自動通報システムの構築を行いました．

自動検知は非常に簡単です．前節で述べました iplog を使います．W32/Welchia.A は TCP port 135 番を介して感染します．そのため port 135 番への TCP アクセスを監視することによって検知することが可能です．

まず 10 秒間 port 135 番に対する TCP アクセスをチェックします．検知したデータが新しければ，設定した E-mail アドレスへ即時自動通報します．最初は全部局の LAN 管理担当者に通知を行っていましたが，クレームが寄せられましたので，感染 PC の IP アドレスに関係のある LAN 管理担当者に絞って通知する通報システムとしました．

このシステムによりまして，多量の W32/Welchia.A 感染 PC の IP アドレスをできるだけリアルタイムで自動検知することができ，かつ当該部局担当者は E-mail による通報よって迅速に適切な対応を取ることが可能となりました．

同年 8 月下旬で一日にのべ 800 IP の感染端末の検知されるという事態に陥っておりましたが，2004 年 1 月 5 日以降はまったく W32/Welchia.A の検知が観測されませんでした．この自動検知・自動通報システムは 2004 年 2 月 1 日に停止致しました．

そして対策中非常に驚いたことは，各部局担当者の大変貴重な時間を割いて，この自動通報 E-mail の情報を頼りに対処していただいたということです．こればかりにはひたすら頭が下がるばかりで，大変恐縮致しているところです．また，自動検知・自動通報システムは通常の IDS には良くある機能の一つですが，私の知る範囲では IDS の管理者に通報するもの意外はまだ見たことがありません．ただこのシステムは検知したその都度 E-mail するので，言い替えれば，使い方を誤ればただの DoS 攻撃システムになります．IDS の研究を行うにあたって，このシステムの適用や実装の方法に関する研究は，まだまったく行われておりません．今後我々の研究室では，IDS の実装と運用技術について，汎用的なインシデント検知技術の開発研究とともに，ログ解析が可能な人材の育成等を念頭に置きながら，進めて行きたいと思います．

4 今後の展開

総合情報基盤センターでは，更にネットワークインシデント検知システムと通報システム，それにインシデントデータベースの構築，ログの更なる解析方法を提案し，即実装して本大学の将来の情報セキュリティを確保する技術を探求して行きたいと思います．

謝辞. 我々のすべて研究は総合情報基盤センターの設備を使って行われています．これらの研究が行えるのも本センターの教職員のおかげです．また MQS の SE の方々にも大変お世話になっております．そして，熊本大学の教職員および学生の皆様のご理解ご協力があってこそ成立する研究分野でもあります．この場を借りて厚く感謝申し上げます．

参考文献

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

[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] Matsuba, R., Musashi, Y., and Sugitani, K.: Statistical Analysis in Syslog Log Files ins DNS and Spam SMTP Relay Servers, *IPSJ Symposium Series, DSM 2004*, No.2004, pp.31-36 (2004).

[21] Matsuba, R., Musashi, Y., and Sugitani, K.: Statistical Analysis in Syslog Log Files ins DNS and Spam SMTP Relay Servers, *IPSJ SIG Technical Reports, Distributed System and Management 32nd*, No.2004, pp.37 (67-72)2004.

[22] http://www.sendmail.org/

[23] http://www.postfix.org/

[24] http://www.isc.org/products/BIND/