社団法人　情報処理学会　研究報告
IPSJ SIG Technical Report

2006－DPS－126（4）
2006－CSEC－32（4）
2006／3／16

メール発信元観測方式に関する検討

鬼頭 哲郎[†]　川崎 宏[†]　山田 知明[‡]　寺田 真敏[†]

[†] 株式会社 日立製作所 システム開発研究所　〒 212-8567 神奈川県川崎市幸区鹿島田 890

[‡] 株式会社 日立製作所 情報通信グループ　〒 212-8567 神奈川県川崎市幸区鹿島田 890

E-mail: [†]{tkito, hiroshi, terada}@sdl.hitachi.co.jp,　[‡]tomoaki.yamada.gu@hitachi.com

概要．近年セキュリティインシデントが多様化している中で電子メールが各種攻撃活動の基点になっている傾向がある．本研究では攻撃に用いられるような不正電子メールの発信元を広域で観測するためのシステムの構築が必要であると考え，フロー統計情報，スパムメール情報，ウイルスメール情報から不正電子メールの発信元の観測を行い，複数のサイト間での相関を把握することのできるシステムに関して検討を行った．このシステムを用いることにより電子メールを基点とした各種攻撃活動の傾向把握，および踏み台となっているコンピュータの検出を行うことができる．

キーワード．ウイルス対策，スパム対策，プロトコル推定

A Study on a Method for Observation of Mail Sender

Tetsuro Kito[†]　Hiroshi Kawasaki[†]　Tomoaki Yamada[‡]　Masato Terada[†]

[†]Systems Development Laboratory, Hitachi, Ltd. 890 Kashimada, Saiwai-ku, Kawasaki-shi, Kanagawa, 212-8567 Japan

[‡]Information & Telecommunication Group, Hitachi, Ltd. 890 Kashimada, Saiwai-ku, Kawasaki-shi, Kanagawa, 212-8567 Japan

E-mail: [†]{tkito, hiroshi, terada}@sdl.hitachi.co.jp　[‡]tomoaki.yamada.gu@hitachi.com

Summary. E-mails tend to be the base of various types of attacks such as phishing frauds and mass-mailing worms. Based on the idea that it is necessary to construct a system which we can observe senders of abuse mails over a wide area, in this paper, we studied on a system that we can observe senders of abuse mails using information such as flow statistics information, spam mail information and virus mail information, and we can look down upon correlation over several areas. By using this system, we can figure out the trend of various types of attacks and detect steppingstone computers.

Keywords. Anti-virus, Anti-spam, Protocol estimation

1　はじめに

近年，ネットワークワームや DDoS（Distributed Denial of Service）攻撃を初めとして，マスメール型のウイルスやフィッシング詐欺，大量のスパムメールなど，セキュリティインシデントが多様化している．また，インシデント発生に伴う影響の範囲も広範囲にわたるものから局所的なものへと変化してきており，インシデント発生の予兆や被害が見えにくくなっているという傾向が出てきている．このような現状においてわれわれは，フィッシング詐欺等のインシデントに見られるように，電子メールが各種攻撃活動の基点となっている点に着目し，不正電子メールを観測するためのシステムが必要であると考えた．現在，インターネット上のトラフィックを観測するシステムとして警視庁 @police の提供するインターネット定点観測 [1] や JPCERT/CC の提供する ISDAS(Internet Scan Data Acquisition System)[2] がある．これらの観測システムがトラフィックを広域において観測しているように，不正電子メールの発信元を広域において観測するためのシステムが必要であると考えられる．スパムメールの発信元や第三者中継を許可しているサーバのリストを提供するものとして DSBL(Distributed Sender Blackhole List)[3] や ORDB(Open Relay Database)[4] などがあり，これらを用いることで，ある IP アドレスがスパム

メールの発信元となっているかを調べることができる．広域で観測するためにはそのような情報に加えて，その IP アドレスからどのくらいの数のスパムメールがどのくらいの範囲に発信されているか，という情報を観測できる必要がある．

そこで，本研究では，複数サイトでフロー統計情報，スパムメール情報，ウイルスメール情報を収集し，不正電子メールの発信元を抽出して複数サイト間での相関性などを把握することのできるシステムに関して検討し，開発を行った．

本システムを用いることで，不正電子メールの発信状況を俯瞰的に観測することが可能になり，各種攻撃活動の傾向の把握，踏み台となっているシステムの検出とそれに伴う対策が可能になると考えられる．

2　システム概要

本研究で検討，開発したシステムのイメージは図 1 のようになっている．本システムはデータ収集部（図 1 (a)）および状況観測部（図 1 (b)）よりなる．データ収集部では複数の協力サイトで複数種類のデータを取得し，発信元に関する情報を抽出して観測センタへと送信する．状況観測部では観測センタに集められたデータを，サイト間や各情報間での相関を把握しやすい形で表示する．

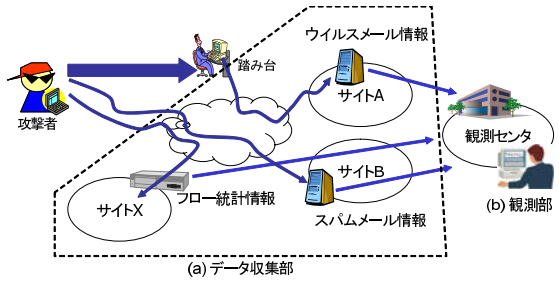

図 1: システムイメージ

2.1　データ収集部

収集するデータの種類としてはフロー統計情報，スパムメール情報，ウイルスメール情報の 3 種類とした．

各情報から発信元 IP アドレスを抽出する手法に関しては各情報の節で述べる．各サイトより提供された発信元 IP アドレス等の情報は観測センタに集められ，データベースへと格納される．なお，全てのサイトがフロー統計情報，スパムメール情報，ウイルスメール情報の全てを提供する必要はなく，フロー統計情報のみ提供，スパムメール情報とウイルスメール情報を提供といったように，収集対象としている 3 種類の情報のうち 1 種類または 2 種類の情報のみを扱う運用も想定している．

2.1.1　フロー統計情報

フロー統計情報としては RFC3176 で規定されている sFlow[5] を用いる．sFlow はネットワークを流れるパケットをルータやスイッチといったネットワーク機器でサンプリングしてフロー統計情報を得る技術である．sFlow ではサンプリングしたパケットからフローに関する情報を抽出し統計情報を作成するが，そのパケットの先頭から事前に指定されたバイト数分（ペイロード情報）を統計情報に含めることができる．悪質なユーザによるスパムメール送信の際には SMTP(Simple Mail Transfer Protocol) で用いられる標準ポートを使わずにそれ以外のポートを使用する場合があるという報告がある [6]．例えば，ボットネットを介したスパムメール送信の場合には Redirect を通じて送信が行われ，Redirect の待ち受けポートとして 25 番以外のポートを用いている．そのため，25 番ポート以外での SMTP トラフィックの捕捉が重要であると考え，本システムでは sFlow から得られるパケットのペイロード情報を用いてポート番号によらないプロトコル推定を行う．プロトコル推定の具体的な内容については 2.2 節で説明する．

フロー統計情報から抽出するデータは，サンプリングされたパケットの発信元 IP アドレス，発信元ポート番号，宛先 IP アドレス，宛先ポート番号といった基本的な情報に加え，各パケットに対して 2.2 節で説明するプロトコル推定を行うことで得られる SMTP らしさの度数（HIGH，MIDDLE，LOW のいずれか）である．これにパケットの到着時刻を加えて観測センタへと送信する．

2.1.2　スパムメール情報

サイトに設置されたスパムフィルタで検出された電子メールや，用意したダミーアカウントに到着した電子メールのヘッダの経路情報（Received ヘッダ）を解析することで発信元 IP アドレスを抽出し，スパムメール情報として収集する．

Received ヘッダからの発信元 IP アドレスの抽出は次のようにして行う．対象となる電子メールの Received ヘッダのうち，一番最後に追加されたもの（通常メールヘッダで一番上にある Received ヘッダ）を取り出し，そこからそのヘッダを追加したサーバの IP アドレスおよ

びそのサーバに対する接続元 IP アドレスを抽出する．これらの情報がない場合，その Received ヘッダは無視し，解析対象を次の Received ヘッダに移す．サーバおよび接続元 IP アドレスを取得できた場合，それらの IP アドレスが信頼点であるかの判定を行う．判定はホワイトリスト方式によって行い，ホワイトリストに登録されている IP アドレスは信頼できるものとし，それ以外は信頼できないものとする．判定を行った結果，サーバの IP アドレスが信頼点で，接続元 IP アドレスが非信頼点であった場合，その接続元 IP アドレスをスパムメールの発信元 IP アドレスとする．サーバの IP アドレスと接続元 IP アドレスの両者がともに信頼点であった場合には次の Received ヘッダに対して同様の処理を行う．

図 2 に例を示す．図 2(a) のような Received ヘッダの場合，ヘッダの from 部と by 部から接続元の IP アドレスおよびメールサーバの IP アドレスを取得する（図 2(b)）．そしてホワイトリスト（図 2(c)）を参照し，抽出された aaa.aaa.aaa.aaa と bbb.bbb.bbb.bbb が信頼点であるか判定する．この場合両方ともホワイトリストに存在するので双方ともに信頼点であると判断し，解析対象を次の Received ヘッダ（2 行目）に移して同様の処理を行う．2 行目も同様にサーバ，接続元ともに信頼点であると判断される．3 行目ではサーバの IP アドレスである ccc.ccc.ccc.ccc はホワイトリストに存在するが，接続元 IP アドレスとして抽出された ddd.ddd.ddd.ddd はホワイトリストに存在しない．よって ddd.ddd.ddd.ddd は非信頼点となり，この IP アドレスを電子メール発信元 IP アドレスとする．

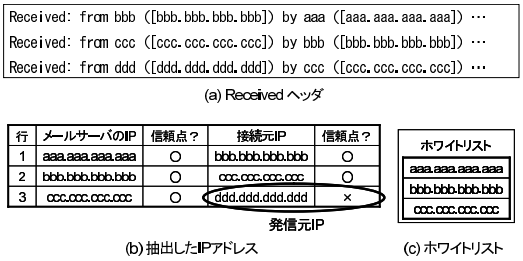

図 2: 発信元 IP アドレス抽出例

このようにして抽出された発信元 IP アドレスにスパムメールの到着時刻，発信元メールアドレスの情報を加え観測センタへと送信する．スパムメールの発信元メールアドレスは偽装されているものが多いと考えられるが，スパムメールの傾向を観測するためにこの情報を加えている．

2.1.3 ウイルス情報

本システムで対象とするウイルスは，自身を添付した電子メールを送信することで感染の拡大を計ろうとするウイルスである．従来のシステムで用いられているウイルス情報は，おもにサイトに設置されたウイルス検出システムなどの検出ログから得られる，どのウイルスが何件検出されたかという情報だけである．本システムにおいてはウイルス検出システムのログと，同地点に設置されたメールサーバのログとの対応をとることによってウイルスメールの情報を抽出し，検出されたウイルス名とともにウイルスメール情報として収集する．

ウイルスメールの発信元 IP アドレスの抽出法は，スパムメールにおける発信元 IP アドレスの抽出法と同様に対象電子メールの Received ヘッダの解析によって行う．これによって抽出された発信元 IP アドレス情報に検出時刻および検出されたウイルス名を加え，観測センタへと送信する．

2.2 sFlow データ分析手法

本研究では不正電子メールの発信元の観測を目的としているが，2.1.1 で述べたように，スパムメールの発信に際しては電子メール送信の標準ポートである 25 番ポート以外のポートを利用して送信が行われることがある．そこで，ルータやスイッチから得られる sFlow 統計情報に含まれるパケットのペイロード情報を分析し，SMTP トラフィックと思われるパケットの検出を行う．これによりポート番号によらないプロトコルの推定が可能になる．

ペイロード情報の分析は次のようにして行う（図 3）．まず，取得した sFlow 統計情報を，サンプリングされたパケットに関する情報であるフローサンプルに切り分け，フローサンプル内に格納された，サンプリングされたパケットのペイロードを取得する．つぎに，サンプリングされたパケットのうち TCP パケットを対象とし，データ部を抽出する．ここで得られたデータ部をプロトコル推定器にかけることでプロトコルの推定を行う．

SMTP プロトコル推定器による推定結果は HIGH, MIDDLE，LOW のいずれかになる．それぞれ SMTP 通信である可能性が高い，やや高い，ほとんどない，という意味になっている．推定は次に示すパターンマッチングによって行われる．与えられたデータ部を見て，SMTP プロトコルでのみ用いられるコマンド文字列で始まっていた場合は，SMTP 通信である可能性が高いとして推定結果を HIGH とする．SMTP コマンド文字列ではあるが他のプロトコルでも用いられているコマンドで始まっていた場合は，SMTP 通信であるとは言い切れないが可

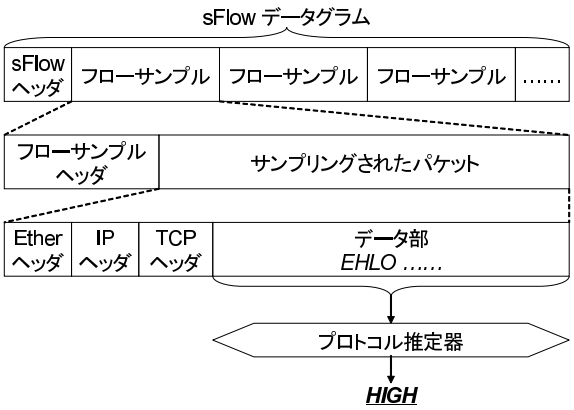

図 3: sFlow 統計情報の分析によるプロトコル推定

表 1: SMTP コマンド一覧

SMTP コマンド	SMTP 特有	特有でない
ARTN	○	
AUTH		○
BDAT	○	
DATA	○	
EHLO	○	
ETRN	○	
EXPN	○	
HELO	○	
HELP		○
MAIL	○	
NOOP		○
QUIT		○
RCPT	○	
RSET		○
SAML	○	
SEND	○	
SOML	○	
STARTTLS		○
TURN	○	
VRFY	○	

能性はあるとして推定結果を MIDDLE にする．最後に，HIGH でも MIDDLE でもない場合には，SMTP 通信である可能性が低いと考え，推定結果を LOW とする．表 1 に本システムのプロトコル推定で用いた SMTP コマンドの一覧と，それぞれ SMTP に特有のコマンドであるかどうかの分類を示す．

2.3 観測部

状況観測部として，前述したデータ収集部で収集したデータを用いて，複数サイト間での不正電子メールの観測および相関性の把握を行うためのインタフェースを開発した．

このインタフェースは概況画面（図 4）および詳細画面（図 6）で構成される．概況画面は，収集した 3 種類の情報に関して 3 サイト分を一画面で観測可能な構造になっている．各サイトごと，各情報ごとの観測が可能なように，各サイトを横軸に，各情報を縦軸にして格子状にグラフを配置している．ここで表示されるグラフは各情報の時系列変化であり，フロー情報の縦軸はパケット数，スパムメール情報とウイルスメール情報の縦軸は検出件数となっている．図において「No Data」となっているのは，そのサイトではその情報を収集していない，という意味である．グラフの下にはそれぞれの情報における発信元 IP アドレスを，フロー統計情報の場合はプロトコル推定により SMTP らしさの度数が MIDDLE 以上と推定されたパケット数の多い順，スパムメール情報やウイルスメール情報の場合は検出件数の多い順に 5 つ表示している．この IP アドレスリストに表示されている IP アドレスのうちの一つにポインタを合わせると，その IP アドレスが他のリストにも存在する場合にはその IP アドレスの背景色を変更しハイライトされる（図 5）．これによって，複数サイトや複数情報にまたがって発信元となっている IP アドレスを簡単に見つけ出すことができる．

概況画面でリストアップされている IP アドレスを選択すると，詳細画面へ遷移する．詳細画面では，全てのサイト，全ての種類の情報に対して，選択された IP アドレスを発信元としたデータをグラフにして表示する．さらに，フロー統計情報に関しては MIDDLE 以上と推定されたパケット数が多い順に宛先のポート番号を 5 つ，スパムメール情報では発信者メールアドレスのユーザ部，ドメイン部を件数の多い順に 5 つ，ウイルスメール情報に関しては検出されたウイルス名のトップ 5 を表示する．

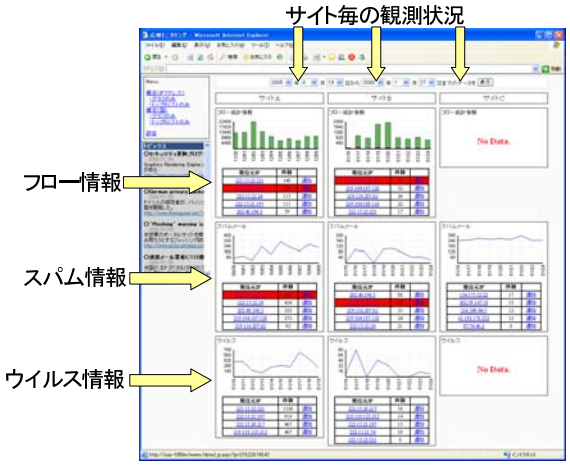

図 4: 概況画面

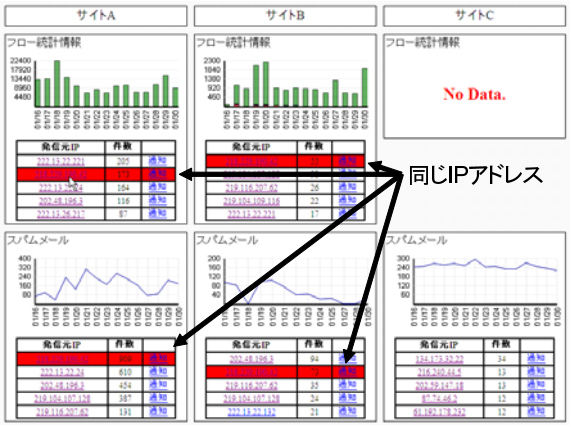

図 5: 発信元 IP アドレスのハイライト表示

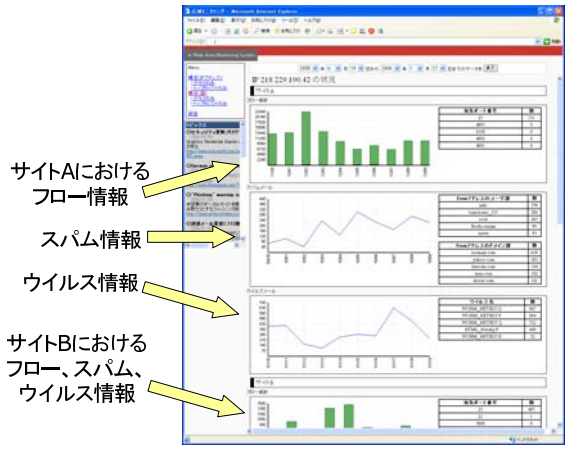

図 6: 詳細画面

3 考察

現在，データ収集部として稼動しているサイトは 1 サイトで，そこで収集しているデータはスパムメール情報である．このサイトで収集されたデータをもとに考察・検討を行う．

まず，観測センタに送信されたデータは 3738 件であり，これに対して分析を行った結果を図 7 に示す．図 7(a) はスパムメール数の時系列推移（上）と発信元 IP アドレス数の時系列推移（下）のグラフである．スパムメール数は日によって大きな違いはなく，メール数の 9 割程度の数の発信元 IP アドレスがある．さらにそのうちのほとんどが，検出されたスパムメールの件数が 1 件のみの発信元 IP アドレスであることがわかる．観測期間全体での発信元 IP アドレスの数を，スパムメール検出件数ごとに示したものが図 7(b) である．発信されたスパムメールが 2 件までの IP アドレスで全体の IP アドレス数の 95% ほど，メール数で言うと全体の約 9 割を占めていることがわかる．

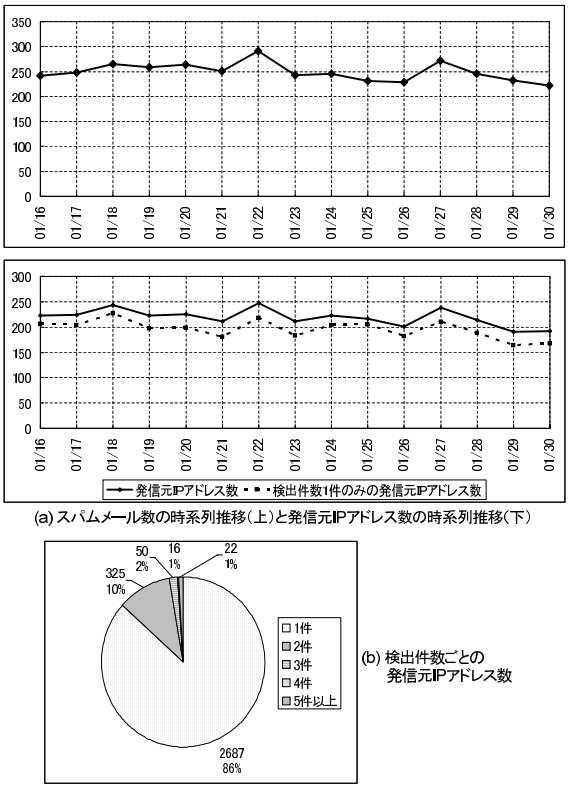

図 7: スパムメール情報の分析

上記のような傾向が当該サイトではみられるが，1 サイトだけの情報では，この傾向がこのサイト特有のものであるか，それとも他のサイトと同じ傾向であるかとい

うことはわからない．複数のサイトにおける情報を集めることで，発生している攻撃がそのサイトのみを狙って行われているものなのかそうでないか，といった他サイトとの相関が求められると考えられる．そのために，不正電子メールの発生状況がサイトによって異なるか等のサイト間の相関を把握できるようなインタフェースに作りこんでいく必要がある．

同じサイト内でも他の情報と組み合わせることで，スパムメール情報だけでは見えない各情報間の相関を見ることが可能になると考えられる．例えば，ある発信元からスパムメールだけでなくウイルスメールが送られている場合，その発信元がボットネットに組み込まれて感染の拡大やスパムメールの送信のための踏み台になっている可能性があると判断できる．

また，本システムを，ISP が個別に実施しているスパムメールフィルタリングサービス [7, 8, 9, 10] などと連携させ，情報共有をしていくことで，より広域での状況把握が可能になると考えられる．

4　まとめと今後の課題

本研究ではルータからのフロー統計情報，およびメールサーバからのスパムメール・ウイルスメール情報を収集し，これらの情報を用いて不正電子メールの発信元の観測を行うためのシステムについて検討し，開発を行った．このシステムを用いることで不正電子メールの発生状況を俯瞰的に観測でき，サイトにまたがった発信元の状況，種別にまたがった発信元の状況を把握することができると考えられる．また，本システムを，他の観測システム，例えばボットネットを対象とした観測システムなどと連携させることで，電子メールを対象とした不正な活動を俯瞰的に観測することができ，より広い範囲での現状把握が可能になると考えられる．

今後の課題としては，データ収集部でのサイト数，収集情報数を増やして観測を行い，どのような情報を見せるとよりよい観測が行えるようになるか，という点に関して検討を行う必要がある．また，本システムを用いて抽出された発信元 IP アドレスに対して，集中的により詳しい観測を行うための方式を検討していきたいと考えている．さらに，本システムはアプリケーションレイヤを対象とした情報を扱っているが，他のレイヤを観測対象にした観測システムとの連携による情報分析方法に関して検討していくことも考えている．

謝辞

本研究は独立行政法人情報通信研究機構から委託を受け実施している「広域モニタリングシステムに関する基盤技術の研究開発」の成果の一部である．

参考文献

[1] 警視庁 @police, インターネット定点観測, http://www.cyberpolice.go.jp/detect/observation.html

[2] JPCERT/CC, ISDAS(Internet Scan Data Acquisition System), http://www.jpcert.or.jp/isdas/

[3] Distributed Sender Blackhole List, http://dsbl.org/main/

[4] Open Relay Database, http://www.ordb.org/

[5] RFC3176: InMon Corporation's sFlow: A Method for Monitoring Traffic in Switched and Routed Networks, http://www.ietf.org/rfc/rfc3176.txt

[6] 小山覚, ボットネット実態調査結果 “Our security depends on your security.”, BlackHat Japan 2005

[7] BIGLOBE 迷惑メールブロックサービス, http://email.biglobe.ne.jp/spam/

[8] DION 迷惑メールフィルター, http://www.dion.ne.jp/service/mail/spam/filter.html

[9] IIJ4U 迷惑メール処理機能, http://www.iij4u.or.jp/guide/option/antiv/spam/

[10] @nifty スパムメールブロック, http://www.nifty.com/mail/reject/index.htm