情報処理学会論文誌 Vol. 52 No. 6 2009–2018 (June 2011)

推薦論文

TCP フィンガープリントによる悪意のある通信の分析

木佐森 幸太[†1,*1] 下田 晃弘[†1]
森 達哉[†2] 後藤 滋樹[†1]

カーネルマルウェアは独自のネットワークドライバを実装し，カーネルモードの通信を行うことで監視ツールからの隠匿を試みる．これらのネットワークドライバは独自の実装であるため，TCP ヘッダのパラメータを分析することでカーネルマルウェア発のトラヒックフローを検出することができる．本研究ではこの性質に基づき，カーネルマルウェアの可能性がある TCP フィンガープリントを整理した．そのフィンガープリントを複数の実運用ネットワークに適用し，フルカーネルマルウェアに感染した可能性が高いホストおよびその通信の特徴を分析する．

Analysis of Malicious Traffic Based on TCP Fingerprinting

Kouta Kisamori,[†1,*1] Akihiro Shimoda,[†1]
Tatsuya Mori[†2] and Shigeki Goto[†1]

Modern kernel malwares compose of their own network drivers and use them directly from kernel-mode to conceal their activities from anti-malware tools. Since these network drivers have specific characteristics, we can detect traffic flows originating from those drivers by analyzing some parameters recorded in TCP headers. On the basis of the above characteristics, we apply a fingerprinting technique to collect IP addresses of the hosts that are likely infected with kernel malwares. Using the method, we also aim to understand the characteristics of the hosts infected with kernel malware and their communications using network measurement data collected in several production networks.

1. はじめに

ボットネットは悪意のあるソフトウェア（マルウェア）に感染したホストによって構成されるネットワークであり，大規模のもので全世界に遍在する 100 万オーダの感染ホストから構成される[1)]．ボットネットは遠隔地の攻撃者によって独自の暗号化された通信チャネルを用いて操作され，スパム送信，DDoS 攻撃，フィッシング等，違法な目的で利用される．近年のマルウェアの高機能化にともない，ボットネットの構造や隠蔽のための手段はますます巧妙化している．そのため，攻撃者だけでなく，攻撃の踏み台である感染ホスト自体の検出が困難である．ボットネットによる加害元特定の困難さは社会的にも深刻な問題となっている．

近年のマルウェアの傾向の 1 つに，マルウェアのカーネル化が報告されている[2)]．カーネル・マルウェアとは最も権限の高いレベル（Ring0）で動作し，メモリ，CPU 命令，すべてのハードウェアデバイスへのフルアクセスが可能である．したがって，マルウェアへの感染検出はさらに困難なものとなる．カーネル・マルウェアのうち，すべてがカーネルモードドライバで実装され，コードのすべてが Ring0 で実行されるものをフル・カーネル・マルウェア（FKM）と呼ぶ．

FKM 自体の歴史は古く，1999 年には FKM として WinNT/Infis の存在が報告されているが，2006 年末頃までは数のうえでは少数にとどまっていた[2)]．2007 年中頃から 2008 年末までに猛威を振るい，全世界のスパムメールの約半分に貢献したとされる Srizbi.trojan は FKM の一種である Reactor Mailer を実装し，独自のネットワークドライバを用いて SMTP 通信を行う機能を有する[3)]．これらの独自ネットワークドライバは OS 由来の TCP/IP とは異なる実装であるため，TCP/IP ヘッダの組合せを注意深く観測することで（後述する TCP フィンガープリント技術）FKM のネットワークドライバ発のパケットと通常の Winsock 等を経由した OS 由来のパケットの識別が可能である[3)]．この性質を利用し，文献 3)–5) で

†1 早稲田大学理工学術院
Faculty of Science and Engineering, Waseda University
†2 NTT サービスインテグレーション基盤研究所
NTT Service Integration Laboratories
∗1 現在，NTT データ・セキュリティ株式会社
Presently with NTT DATA SECURITY CORPORATION
本論文の内容は 2009 年 10 月のマルウェア対策研究人材育成ワークショップ 2009（MWS2009）にて報告され，CSEC 研究会主査により情報処理学会論文誌ジャーナルへの掲載が推薦された論文である．

は TCP フィンガープリントを利用したスパムボットの検出およびスパムボットの全体像解明に向けて大域的な分析をしている．

本研究は FKM の検知手法として TCP フィンガープリントに注目し，その検知パラメータを工夫することで昨今の FKM 通信を検知できることを証明する．フィンガープリントを整理するための基礎データとして，ISP に設置したハニーポットの通信を計測したデータである CCC DATAset [6] を利用した．さらに，大学，企業，および学術ネットワークの複数計測ポイントで収集した TCP ヘッダデータを分析し，FKM に感染したホストの特徴を分析する．

本研究の特筆すべき貢献は，FKM の可能性が高いボットの存在を複数のデータセットで確認したこと，およびそれらの特徴を明らかにしたことである．特に CCC DATAset の分析においては FKM の可能性が高いホストが他の攻撃ホスト以上に観測され，これは我々の当初の予想を上回るものであった．また MAWI データセットを用いた分析では FKM の長期的なトレンドを明らかにした．このような時間軸での分析は FKM に対する抜本的な対策に必要な時間スケールを推し量るうえで重要な指標となり，有益である．

今回の分析において我々が前提とした仮説は，「**一般的な OS では利用されない TCP フィンガープリントを有する攻撃パケットは FKM 感染ホストが発信した**」と解釈することである．上記の仮説に基づく我々の推論をより確固とした根拠に基いて検証するためには，たとえば文献 7) のようなカーネル・ルートキットのプロファイリング等，得られたマルウェア検体の詳細な分析に基づく裏付けが必要である．大規模なマルウェアの収集と動的なマルウェア解析による通信の分析も有効な手段と考えられる．これらは我々の近い将来の課題である．

2. TCP フィンガープリントを用いた FKM の検出方法

本章では TCP フィンガープリントを用い，FKM に感染した可能性が高いホストおよび通信を検出する方法について述べる．はじめに TCP フィンガープリントの原理を述べる．次に FKM の可能性が高いフィンガープリントの抽出方法を述べる．

2.1 TCP フィンガープリントの原理

TCP/IP の仕様は RFC で定義されているが[8]，OS ごとにその実装は異なることが知られている[9]．TCP ヘッダに記載された各種オプションの設定値や応答の挙動の特徴を注意深く観測することによって対象システムの OS を推定することができる．このような技術を TCP フィンガープリント（または OS フィンガープリント）と呼び，主に攻撃元の素性を非侵襲的に推察するための手段として利用されている[9]．前章で述べたように，FKM は既存の OS とは異なる独自の TCP/IP 実装を持つため，特徴的な TCP フィンガープリントを有すると期待される．

表 1　p0f シグネチャの構成要素
Table 1　Composition of p0f signature.

W	ウィンドウサイズ
T	TTL の初期値
D	Don't Fragment ビット
S	SYN パケット全体のサイズ
O	TCP オプション（NOP，最大セグメントサイズ等）
Q	その他特徴的な点等
OS	OS の種類
V	OS のバージョン等

p0f [10] は代表的な受動的 TCP フィンガープリントのツールとして知られており，広く利用されている．p0f は入力として tcpdump [11] で計測した pcap 形式のデータを用いることができる．tcpdump は標準的に利用されるネットワーク計測ツールであり，パケットヘッダデータをキャプチャすることができる．p0f にはいくつかのモードがあるが，本研究では SYN パケットを用いるモードを用いる．SYN パケットに対する p0f シグネチャは [W:T:D:S:O:Q:OS:V] のようにコロンで区切られたフォーマットとなっている．各フィールド値の説明を**表 1** に示す．なお複数の TCP オプション（O）が設定されている場合は [W:T:D:S:O,O,O:Q:OS:V] のようにコンマで区切ってオプションを並べて書く．

本研究では p0f を用いて TCP フィンガープリントの分析を行う．ただし p0f のシグネチャは 2006 年以降更新されていないため，本研究ではマニュアルで収集した Windows Vista, Linux 2.6+，FreeBSD 7+，Mac OS 10.5+ 等のシグネチャを追加することによって既知の OS のフィンガープリントを最新の情報に更新している．

2.2 FKM の可能性が高いフィンガープリント

ネットワーク上を流れるパケットを監視し，TCP ヘッダの分析を行うことによって FKM の可能性が高いフィンガープリントを抽出する方法を検討する．ここでは「一般的な OS では利用されない TCP フィンガープリントを有する攻撃パケットは FKM 感染ホストが発信した」という前提を置く．この前提が正しいことを CCC DATAset を用いて確認する．なお文献 3)–5) で指摘されているように，FKM の一種である Srizbi が有するフィンガープリントは既知の OS と異なる特徴を有することが知られている．

後述する例で示すように現在のインターネットにおける通信の大多数は p0f によって識別可能な**既知 OS** が送信している．したがって，既知の OS による通信が大多数を占める一般のインターネット回線の計測データから FKM の可能性が高いフィンガープリントの情報を収集するアプローチは効率的ではない．一方，ハニーポットに対する攻撃通信では悪意のある通信が集中するため，FKM の通信を効率的に収集できることが期待できる．

本研究ではハニーポットで収集した通信データから FKM の可能性が高いフィンガープリントを抽出し，その通信を検証するアプローチをとる．ハニーポットに対する通信としては CCC DATAset2008，2009 [6] の攻撃通信データ（インバウンド）を用いる．CCC DATAset における攻撃通信はハニーポットの通信を tcpdump でパケットキャプチャしたデータである．ハニーポットの台数は 2 台であり，同一 ISP に設置されている．CCC DATAset 2008 では Windows XP SP1 と Windows 2000，CCC DATAset2009 では Windows XP SP1 と Windows 2000 がハニーポットのゲスト OS として用いられている．それぞれのデータの計測時期は 2008 年 4 月 28・29 日（CCC DATAset 2008），および 2009 年 3 月 13・14 日（CCC DATAset 2009）である．

以下では FKM の可能性が高いフィンガープリントの抽出手順を述べる．はじめにハニーポットに対する攻撃通信を記録した pcap データに対して最新の OS に対応させた p0f を適用し，既存の OS ではない UNKNOWN と判定された未知のフィンガープリントを抽出する．

次に UNKNOWN と判定されたフィンガープリントについて，観測した TTL（Time to live）値を 2 のベキ乗の値に切り上げて正規化する．TTL は IP ヘッダの 8 ビットのフィールドで 0 から 255 の値をとる．TTL は OS ごとに特徴的な初期値をとることが知られている．TTL は通信経路上のルータを通過するたびに 1 ずつ減少するため，ハニーポットで観測している TTL の値は初期値から通信経路上のルータの数を減算したものとなる．つまり TTL の観測値は OS の特徴的なパラメータよりも少しだけ小さな値となる．本研究では，TTL の観測値をそのまま用いるのではなく，一般的に TTL の初期値として知られている 32，64，128，255 に切り上げてフィンガープリントを整理した．これを式で表すと次のようになる．TTL の観測値 t に対して，正規化した値は $t' = 2^{\lceil \log_2 t \rceil}$ となる．ただし例外として正規化後の値が $t' = 256$ となる場合は，TTL の最大値が 255 であることから $t' = 255$ に補正する．また，$t < 32$ の場合には $t' = 32$ と正規化する．正規化の適用例を**表 2** に示す．たとえば観測した TTL が $t = 53$ であれば $t' = 64$ に，$t = 249$ であれば $t' = 255$ と正規化する．

表 2　TTL の正規化
Table 2　Normalization of observed TTL.

観測値	正規化後
0〜32	32
33〜64	64
65〜128	128
129〜255	255

表 3　MWS シグネチャの例
Table 3　Examples of MWS signatures.

番号	シグネチャ
MWS 60352_1	[60352:64:0:52:M1240,N,W2,N,N,S:.:MWS:60352_1]
MWS 60352_6	[60352:64:0:52:M1414,N,W2,N,N,S:.:MWS:60352_6]
MWS 53760_4	[53760:64:0:64:M1414,N,W3,N,N,T0,N,N,S:.:MWS:53760_4]
MWS 16384_1	[16384:128:0:40:.:.:MWS:16384_1]
MWS 65535_2	[65535:64:0:48:M1414,N,N,S:.:MWS:65535_2]

上記の処理を CCC DATAset 2008 および 2009 に適用することにより 43 種類のユニークなフィンガープリントを得ることができた．本研究ではこれらのフィンガープリントを有するホストが FKM に感染している可能性が高いと仮定し，その通信を分析することで有効性の検証を行う．なお今回収集したフィンガープリントの集合には Srizbi のフィンガープリントは存在しなかった．Srizbi の主な活動はスパム送信であるため，観測を行ったハニーポットで電子メールを受信していないことが原因であると考えられる．

以降，本研究では上記で抽出した 43 種類のフィンガープリントを MWS シグネチャと呼び，個々の MWS シグネチャの最初の要素（W: TCP のウィンドウサイズ）によって名称をつける．**表 3** に例を掲げる．たとえば，TCP のウィンドウサイズ（W）が 65535 バイトのシグネチャであれば，MWS65535_1，MWS 65535_2…となる．ここで TCP オプション（O），その他の特徴的な点（Q）の値がない場合にはピリオドで表す．たとえば表 3 の MWS 16384_1 シグネチャ[16384:128:0:40:.:.:MWS:16384_1] は TCP オプション（O）がなく，その他の特徴的な点（Q）もないシグネチャであり，該当する要素がピリオドとなっている．ここで抽出した MWS シグネチャすべてに共通な特徴は DF ビットが 0 に設定されていることである．CCC DATAset においては MWS シグネチャ以外のインバウンドの通信についても DF ビットが 0 であったため，計測した環境に依存する現象である可能性が高いと考えられる．後に示す他のネットワークにおける検証ではこの点を考慮した分析を行っている．

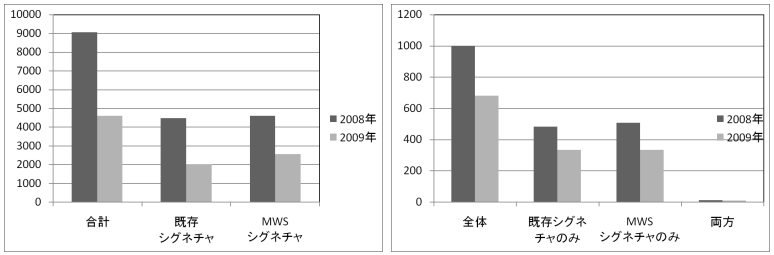

図 1　MWS シグネチャの統計：(左) 出現回数, (右) シグネチャごとの送信元 IP アドレス数
Fig. 1　Statistics of MWS signatures: frequencies (left) and the number of unique source IP addresses (right).

3. MWS シグネチャの分析

本章では CCC DATAset を用い，MWS シグネチャを有するホストの通信を分析する．分析の結果，通信の大半が悪意のある可能性が高い通信であること，およびハニーポットに対する攻撃が高い確率で成立していることを示す．

3.1 MWS シグネチャの出現頻度

はじめに MWS シグネチャの出現頻度を分析する．**図 1** は CCC DATAset 2008 および 2009 における MWS シグネチャの出現回数，ならびにシグネチャごとのユニークな送信元 IP アドレスの数を示す．2008 年，2009 年ともに，MWS シグネチャによる通信の方が既存シグネチャと比較して出現回数が多いことが分かる．2009 年は全体として通信回数が減少しているものの，MWS シグネチャによる通信が占める割合は高くなっている．一方 MWS シグネチャで通信してくるホストの割合は 2008 年から 2009 年にかけて減少しているものの，依然として観測したホストの半数以上が MWS シグネチャを有することが分かる．

後述するように大学のネットワークにおいて観測された MWS シグネチャを有するホストの割合はわずかに 0.03%（=280/954,100）であり，CCC DATAset で観測された MWS シグネチャホストの比率はきわめて高いことが分かる．すなわち，MWS シグネチャは悪意のある通信が集中するハニーポットにおいて多く観測されることから，FKM である可能性がより高いことが推察できる．

3.2 MWS シグネチャを有するホストの通信

次に個々の MWS シグネチャを有するホストが発生した通信を分析する．**図 2**，**図 3** はシグネチャごとの出現回数およびユニークな送信元 IP アドレス数を集計したグラフである

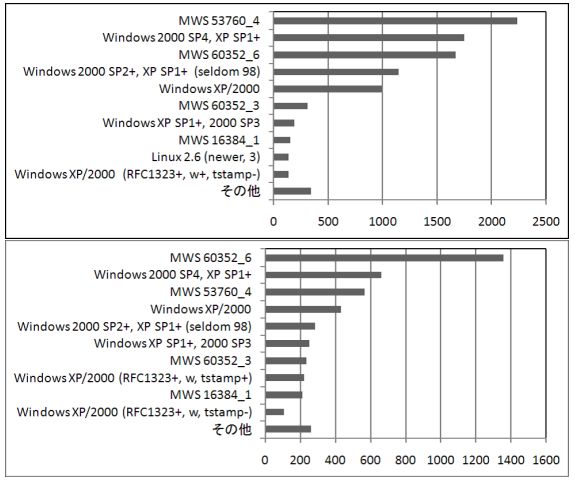

図 2　シグネチャごとの出現回数：(上) 2008 年, (下) 2009 年
Fig. 2　Frequencies of each signature: Year 2008 (top) and year 2009 (bottom).

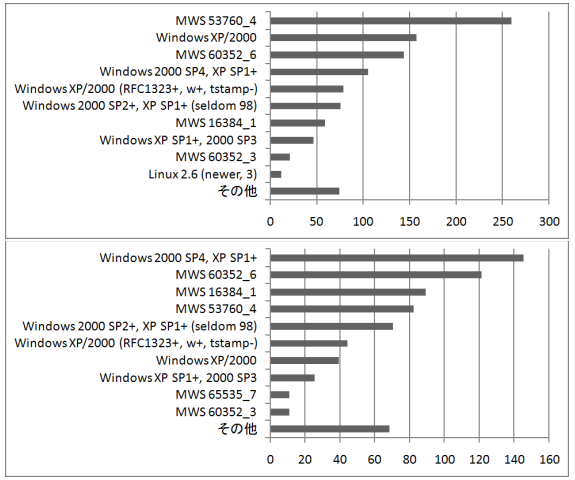

図 3　シグネチャごとの送信 IP アドレス数：(上) 2008 年, (下) 2009 年
Fig. 3　Number of source IP addresses for each signature: Year 2008 (top) and year 2009 (bottom).

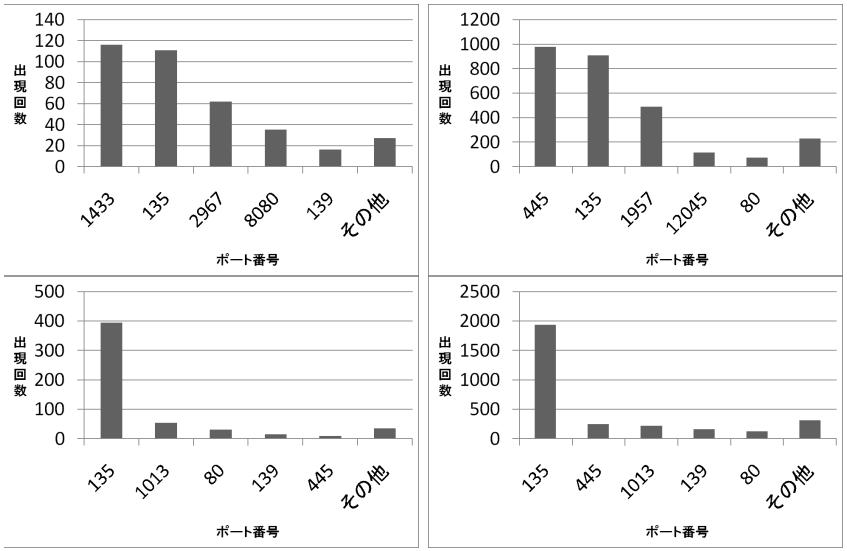

図 4　送信先ポート番号分布：(左上) MWS 16384_1, (右上) MWS 53760_4, (左下) MWS 60352_3, (右下) MWS 60352_6

Fig. 4　Distribution of destination port numbers: MWS 16384_1 (upper left), MWS 53760_4 (upper right), MWS 60352_3 (lower left), and MWS 60352_6 (lower right).

(上位 10 位まで)．2009 年の送信元 IP 数以外は MWS シグネチャがトップを占めていることが分かる．ボットネットの大部分を占めるとされる Windows ファミリのシグネチャが比較的多く混在するが，MWS シグネチャによる通信の割合が出現回数，送信元 IP アドレスともに高いことが読み取れる．

続いて出現回数，送信元 IP ともに上位に位置している 4 種の MWS シグネチャについて，上記 5 位までの送信先ポート番号を集計した結果を**図 4** に示す．135 番，139 番，445 番，1433 番，2967 番といった，広く知られた脆弱性に関連のあるポート番号の比率がきわめて高いことから，悪意のある通信である可能性が高いことが分かる．また，これらの傾向はシグネチャごとに異なる特徴があることも読み取れる．

3.3　MWS シグネチャと典型的な攻撃パターン

最後に MWS シグネチャを有するホストの典型的な攻撃パターンを示す．CCC DATAset 2009 を対象とし，MWS シグネチャを有する 345 の外部ホストのうち，3-way ハンドシェイクによりハニーポットと TCP コネクションを確立した 122 ホストを対象として通信を分析した結果，代表的な 2 種類の攻撃パターンを得たので以下に示す．なお攻撃パターンシーケンスにおいて H_X はインターネット上のホスト X，P_X はハニーポット X を示す．

```
echo open xxx.xxx.xxx.xxx 2766 >\\
       i&echo user yyyyy zzzzz\\
    >> i&echo get wmsoft05006.exe\\
    >> i&echo quit
    >> i&ftp -n -s:i&startwmsoft05006.exe\\
```

図 5　ケース I におけるシェルコードの例

Fig. 5　An example of shell code for case I.

3.3.1　ケース I（MWS 60352_6）

```
H_A:9109 -> P_A:135 (scan)
H_A:9110 -> P_A:135 (rpc)
H_A:9197 -> P_A:135 (rpc)
H_A:9203 -> P_A:1013 (シェルコード送信)
P_A:1028 -> H_A:3450 (malware 要求)
P_A:1028 -> H_A:3450 (malware 要求)
```

最初の 135 番ポートに対する通信はスリーウェイハンドシェイク成立後に何もせず終了するが，2 度目・3 度目の 135 番ポートに対する通信では RPC プロトコルによる通信が行われていた．さらに，1013 番ポートに対する通信では，**図 5** に示すようなシェルコードによって ftp でファイルをダウンロードしてそれを実行するよう命令を送っていた．ハニーポット P_A からホスト H_A に対する通信では ftp コマンドのやりとりによって実行ファイルが転送されるため，当該ファイルのダウンロードであると見られる．以上の観測結果より当該シグネチャの攻撃パターンは RPC のバッファ・オーバフロー脆弱性に起因することが推察される．

3.3.2　ケース II（MWS 53760_4）

```
H_B:56101 -> P_B:135 (rpc)
P_B:1027 -> H_B:47602 (malware 要求)
P_B:1027 -> H_B:47602 (malware 要求)
```

ケース I とは異なり，最初の 135 番ポートに対する通信で RPC プロトコルによる通信を行っている．P_B から H_B に対する通信では，“This program cannot be run in DOS mode.” 等の文字列が観測されることから，Windows のポータブル実行可能ファイルをダウンロードしていることが分かる．また，CCC DATAset 2009 の攻撃元データにおいて，時刻・IP・ポート番号が合致する記録が存在したため，実際にマルウェアを送信していることが分かる．

4. MWS 通信の広域分析

本章では，前章の分析の結果得られた MWS シグネチャを他のネットワーク計測データに適用し，ボット感染の疑いホストを検出・分析した結果を示す．はじめにハニーポットが設置されたネットワーク環境に特有であった可能性のある条件を一般化するために MWS シグネチャの拡張を行う．次にキャンパスネットワーク，企業メール網，学術ネットワークの 3 種類のデータセットを対象とし，MWS シグネチャを有するホストの通信を分析する．特に学術ネットワークの分析においては長期間にわたる MWS シグネチャの時間的変化の分析を行うことを狙いとしている．

4.1 MWS シグネチャの拡張

前章で示したように CCC DATAset では，MWS シグネチャに限らず，すべての通信の DF ビットが 0 であった．DF ビットはルータやファイアウォールによって削除されることがあるため，CCC DATAset の収集環境でも経路上で削除された可能性がある．したがって，観測した DF ビット値と本来の値が異なる可能性がある．そのため，MWS シグネチャの DF ビットを 1 とした MWS+DF シグネチャを作成した．この拡張によって MWS シグネチャが既存のシグネチャと一致することはなかった．なお，今日広く利用されている大多数の OS は DF ビットを 1 に設定している．

さらに，通信環境を考慮するために最大セグメントサイズ（MSS）オプションの値に関する拡張を行う．MWS シグネチャの中には MSS のみが異なるグループが 6 群存在した．これらのグループについては DF ビットを 1 としたうえで MSS の値をワイルドカードとし，6 種の MWS_Gen シグネチャを作成した．

以下の分析では MWS_Gen シグネチャについては MWS シグネチャ，MWS+DF シグネチャよりも検出の優先度を下げる．これにより，MWS シグネチャ，MWS+DF シグネチャとは MSS の値だけが異なる通信を MWS_Gen シグネチャとして検出できる．

表 4　キャンパスネットワークにおける TCP フィンガープリントの観測数
Table 4　Cumulative number of TCP fingerprints observed in the campus Network.

	観測数	全体に占める割合
MWS	12,132,095	5.140%
MWS+DF	2,132,886	0.904%
MWS_Gen	6,062,306	2.569%
UNKNOWN	47,627,301	20.180%
既存 OS	168,056,675	71.207%
合計	236,011,263	100.000%

表 5　キャンパスネットワークにおけるフィンガープリント別ユニーク送信 IP アドレス数
Table 5　Number of unique IP addresses for each TCP fingerprint in the campus network.

	送信元 IP 数	全体に占める割合
MWS	401	0.007%
MWS+DF	44,520	0.770%
MWS_Gen	269,373	4.656%
UNKNOWN	372,833	6.444%
既存 OS	5,222,683	90.272%
合計	5,785,478	100.000%

4.2 キャンパスネットワーク

早稲田大学の対外接続回線において TCP ヘッダデータを収集し，p0f を用いて拡張 MWS シグネチャの通信を分析した．当該回線は学術（帯域 10 Gbps）および商用網（帯域 300 Mbps）を収容しており，収集データには両者の回線を総合したトラフィック情報が含まれる．データ収集期間は 2009 年 12 月 25 日から 12 月 31 日の 1 週間であり，TCP SYN パケットのみを収集の対象とした．

早稲田大学の通信データに p0f を適用した結果を，MWS シグネチャ，MWS+DF シグネチャ，MWS_Gen シグネチャ，各種 MWS シグネチャ以外の UNKNOWN，既存 OS，の 5 種に分類した．**表 4**，**表 5** に各シグネチャごとの観測 SYN パケット数とユニークな送信元 IP アドレス数を示す．複数種別のシグネチャで通信している送信元 IP は重複してカウントしているため，各種別の値を合計しても全体の送信元 IP 数とは一致しない．MWS シグネチャを有するホストおよび通信量はたかだか数パーセントであり，CCC DATAset で観測された割合と比較すると非常に少ないことが分かる．また DF の拡張の影響はそれほど多くないが，MSS の拡張によりさらに多くのホストが検出されることが分かる．

拡張した MWS シグネチャにおいて，送信元 IP 数の割合は少ないにもかかわらず SYN

表 6　MWS 16384_1 の送信先ポート別通信回数
Table 6　The number of connections destined to each port number for packets originated from MWS 16384_1.

送信先ポート番号	回数
2967	5,827,791
1433	2,968,309
135	1,460,904
3306	344,411
1521	223,939
8088	201,510
8080	196,786
445	84,127
その他合計	750,668

表 7　送信ポート別出現数：MWS_Gen 65535_1（左上），MWS_Gen 53760（右上），MWS_Gen 60352_1（左下），MWS_Gen 65535_2（右下）
Table 7　Frequencies of connections for each destination port: MWS_Gen 65535_1 (upper left), MWS_Gen 53760 (upper right), MWS_Gen 60352_1 (lower left), and MWS_Gen 65535_2 (lower right).

ポート番号	出現回数
445	1,827,882
80	120,384
6889	48,207
21053	11,828
8080	10,566

ポート番号	出現回数
445	1,391,316
139	2,301
25	879
80	871
135	736

ポート番号	出現回数
445	1,322,036
1433	7,660
80	1,843
25	925
139	368

ポート番号	出現回数
445	533,677
80	378,072
6889	75,996
21053	20,479
6649	12,757

パケット数が多いのは，少数の送信元 IP から MWS 16384_1 シグネチャによる大量の SYN パケットが送信されているためである．1 つの送信元 IP あたりの SYN パケット数は，多いもので 100 万以上にもなる．MWS 16384_1 シグネチャ発通信の送信先ポートの分布を**表 6** に示す．この結果，特定のポート番号 2967 に通信が集中していることが分かる．この事実から，シグネチャと攻撃パターンに相関があることが推察される．

MWS 16384_1 シグネチャについで通信量が多かった 4 シグネチャについて，送信先ポート番号を集計したのが**表 7** である（上位 5 位まで）．4 シグネチャのすべてに共通してポート 445 番が大部分を占めていることが分かる．前章でも述べたとおりポート 445 番はポートスキャンの対象として著名なものの 1 つである．その他観測される 135 番，139 番，2967 番，1433 番といったポートも同様である．その他 HTTP アクセスや SMTP に対するアクセスがあることが分かる．

表 8　FKM の可能性が高いシグネチャを有するホストが発信したメールの統計（上位 10 ホストのみ）．表中の S はスパム数，H は通常メール数，I は IP アドレス数
Table 8　Statistics of e-mail messages originating from hosts that use MWS signatures (top 10). S, H, and I are the number of spam messages, the number of legitimate messages, and the number of unique source IP addresses, respectively.

シグネチャ名	スパム	通常メール	送信元 IP 数
MWS 65535_8	290	0	9
MWS 65535_5	252	0	8
MWS 65535_3	90	0	4
MWS 65535_7	64	0	6
MWS 16384_3	25	0	3
MWS 65535_4	16	0	7
MWS 53760_4	16	0	2
MWS 65535_12	9	0	1

4.3　企業メール網

ある企業の電子メールサーバ（MTA）に接続したネットワークセグメントで TCP ヘッダデータを収集した．この回線で観測可能な通信は SMTP のみであるため，ボットネットのスパム活動が主な分析対象である．キャンパスネットワークと同様に，TCP SYN パケットのみを収集した．データの収集時期は 2009 年 3 月 1 日から 3 月 31 日の 1 カ月間である．上記の MTA ではスパムアプライアンスが動作しているため，ある IP アドレスから送信されたメールがスパムであったか否かの判定が可能である．

本データにおいて観測された IP アドレス数は 1,230,830 であり，そのうちわずかに 53 アドレスが今回発見したシグネチャを有するホストであった．今回の検討外であるが，FKM の一種である Srizbi のシグネチャを有する IP アドレスの数は 40,322 であった．

上記の 53 アドレスによる通信をシグネチャごとにまとめたものが**表 8** である．本データセットにおいて観測されたスパムメールの総数は約 1,500 万であり，FKM と識別されたホスト発のスパムは非常に少ない．しかしながら，これらの識別されたホストが発信したメッセージの大多数がスパムであり，マルウェアの構成によってはスパム送信モジュールを搭載するものも存在することがうかがえる．

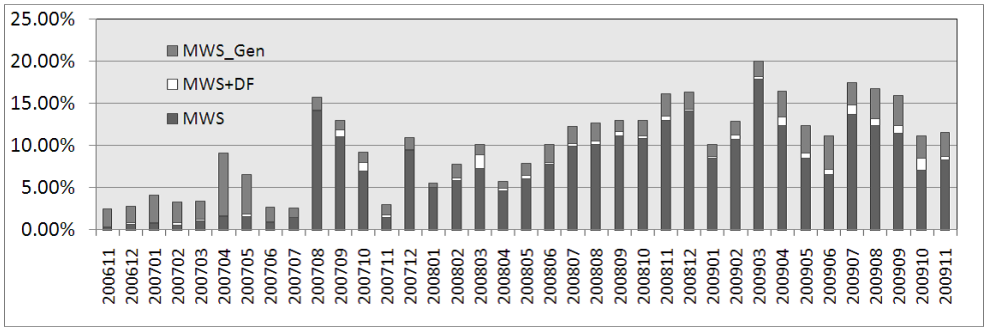

図 6　MWS シグネチャの正規化された出現回数の推移
Fig. 6　Time series of normalized frequencies for each MWS signature.

4.4　学術ネットワーク

本節の狙いは TCP フィンガープリントの時間的な変化を分析することである．このために学術ネットワークを定点観測した MAWI データセット[12] を用いる．MAWI データセットは WIDE プロジェクトによって研究用に公開されているデータであり，同プロジェクトの WEB ページ[12] より取得することができる．本研究では太平洋を横断する国際回線を計測した samplepoint-F を用いる．samplepoint-F のデータは毎日 14:00〜14:15 の 15 分間取得されているパケットキャプチャデータであり，統計データとともに公開されている．2006 年 11 月から 2009 年 11 月の 37 カ月間に計測されたデータを対象とし，分析を行った．取得したデータを 1 カ月ごとに集計をし，分析を行った．月ごとの日数の差やデータの欠損を考慮し，各々の月における全観測パケット数で正規化した統計を用いる．

MWS，MWS+DF，および MWS_Gen のそれぞれのシグネチャについて，正規化した出現回数の推移を示したものが**図 6** である．2006 年 11 月時点ですでに MWS シグネチャを有するホストが存在している可能性があること，および 2009 年 11 月時点においても感染ホストが存在している可能性があることが示唆される．したがって，FKM の活動期間は比較的長いタイムスケールであることが推察される．

観測した MWS シグネチャにおいて，MWS+DF および MWS_Gen による拡張の寄与は少ない．また，すべての月において観測した MWS シグネチャによる通信のおよそ 99%以上が MWS16384_1 シグネチャによるものであった．この傾向はキャンパスネットワークで観測された傾向と同様であり，同シグネチャの活動が局所的なものではないことが示唆される．

さらに MWS 16384_1 シグネチャは 2007 年 7 月以降，急激に増えていることが見て取れる．文献 3)，5) に示されているようにこの時期はちょうど Srizbi の活動が顕著になり始めた時期と一致している．マルウェアの開発・配布に関して両者に相関が存在する可能性があると考えられる．より詳細なマルウェアの分析は今後の課題である．

5.　まとめと今後の課題

フルカーネル・マルウェアの可能性が高いホストの詳細分析，および実ネットワークにおける実態調査を行った．分析における鍵となるアイディアは TCP フィンガープリントを利用することである．一般の OS では利用されていないにもかかわらず，出現頻度および送信 IP アドレス数の高い TCP フィンガープリントを抽出し，それらの抽出した TCP フィンガープリントを有する IP アドレス発の通信について，送信先ポートの調査と典型的な攻撃パターンを分析することで FKM の可能性が高い通信およびホストを検出することができる．

本研究では CCC DATAset におけるハニーポットへの攻撃通信を分析し，FKM の可能性が高い TCP フィンガープリントを MWS シグネチャとして抽出した．この結果，ハニーポットへの通信には MWS シグネチャを有するホストによる通信が当初の予想以上に存在することが明らかになった．また，MWS シグネチャにはいくつかの種類が存在し，いずれのシグネチャも通信の大多数が攻撃に紐づいていることを示した．さらに各々のシグネチャは特有の攻撃パターンを有すること，およびネットワークによって観測されるシグネチャが異なることを明らかにした．これらの事実は我々が前提とした仮説「**一般的な OS では利**

用されない TCP フィンガープリントを有する攻撃パケットは FKM 感染ホストが発信した」を支持するものである．この仮説をより確実な根拠に基づいて検証するためには，感染ホストおよびマルウェアの動的および静的解析に基づいたプロファイリングが必要不可欠である．これらは我々の今後の課題である．

MWS シグネチャは CCC Dataset だけでなく，広く様々なネットワークで観察されていることから，マルウェアにおける FKM の割合は今後さらに増えていく可能性がある．FKM は Ring0 で動作するため，通常のアンチウィルスソフトウェア等の監視から隠匿する動作が可能である．新規に開発する OS だけでなく，現存する大多数の OS に対して FKM の動作・実行を防止するための防衛メカニズムを構築することが重要である．

謝辞　本論文はマルウェア対策研究人材育成ワークショップ 2009 における筆者らの発表[13]が推薦論文として採択された際に，査読者からいただいたアドバイスに基づき，ワークショップの発表論文に加筆修正を施したものである．研究運営委員会の先生方に感謝いたします．

参　考　文　献

1) Weblog, F.-S.: Calculating the Size of the Downadup Outbreak (2009). http://www.f-secure.com/weblog/archives/00001584.html
2) Kasslin, K.: Kernel Malware: The Attack from Within (2006). http://www.f-secure.com/weblog/archives/kasslin_AVAR2006_KernelMalware_paper.pdf
3) Stern, H.: The Rise and Fall of Reactor Mailer, *Proc. MIT Spam Conference 2009* (2009).
4) Esquivel, H., Mori, T. and Akella, A.: Router-Level Spam Filtering Using TCP Fingerprints: Architecture and Measurement-Based Evaluation, *CEAS* (2009).
5) Mori, T., Esquivel, H., Akella, A., Shimoda, A. and Goto, S.: Understanding the World's Worst Spamming Botnet, 技術報告 TR1660，University of Wisconsin Madison Technical Report (2009).
6) 畑田充弘：マルウェア対策のための研究用データセットとワークショップを通じた研究成果の共有，マルウェア対策研究人材育成ワークショップ 2009（MWS2009），pp.1–8 (2009).
7) Riley, R., Jiang, X. and Xu, D.: Multi-aspect profiling of kernel rootkit behavior, *EuroSys '09: Proc. 4th ACM European Conference on Computer Systems*, New York, NY, USA, ACM, pp.47–60 (2009).
8) Stevens, W.R.: *TCP/IP Illustrated, Volume 3: TCP for Transactions, HTTP, NNTP, and the UNIX Domain Protocols*, Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA (1996).
9) Honeynet Project: Know your enemy: Passive fingerprinting (2002). http://project.honeynet.org/papers/finger
10) Zalewski, M.: the new p0f: 2.0.8 (2006). http://lcamtuf.coredump.cx/p0f.shtml
11) TCPDUMP/LIBPCAP public repository. http://www.tcpdump.org
12) MAWI Working Group Traffic Archive. http://mawi.wide.ad.jp/mawi/
13) 木佐森幸太，下田晃弘，森　達哉，後藤滋樹：TCP フィンガープリントによる悪意のある通信の分析，マルウェア対策研究人材育成ワークショップ 2009（MWS2009），pp.553–558 (2009).

(平成 22 年 4 月 24 日受付)
(平成 23 年 3 月 7 日採録)

推　薦　文

本論文は，一般的なホスト型パケットモニタで検知できないフルカーネルマルウェアをモニタする手法を提案している．そのために，プロトコルスタックの実装による細かな違いを利用している．本手法の着眼点が素晴らしい．また，実ネットワーク上で評価を行うことによって，有効性を証明している．さらに，検知のためのシグネチャを網羅的に作成し，MWS データセットだけでなく他のデータセットもあわせて分析しており，今後の対策のための参照データとしてもきわめて有益である．よって，研究会推薦論文として推薦する．

（コンピュータセキュリティ研究会主査　菊池浩明）

木佐森幸太

昭和 54 年生．平成 16 年東京大学経済学部経済学科卒業．平成 20 年早稲田大学基幹理工学部コンピュータ・ネットワーク工学科卒業．平成 22 年早稲田大学大学院基幹理工学研究科情報理工学専攻修士課程修了．同年 NTT データ・セキュリティ（株）入社．

下田　晃弘

昭和 60 年生．平成 19 年早稲田大学大学院基幹理工学研究科情報理工学専攻修士課程修了．平成 23 年同大学院基幹理工学研究科博士後期課程修了．現在，同大学部基幹理工学科助手．主としてネットワーク・フローに注目したセキュリティに関する研究に従事．博士（工学）．

森　　達哉

昭和 48 年生．平成 9 年早稲田大学理工学部応用物理科卒業，平成 11 年同大学大学院修士課程修了．同年日本電信電話（株）入社．平成 19 年から 20 年にかけて米国ウィスコンシン州立大学マディソン校客員研究員．現在，NTT サービスインテグレーション基盤研究所主任研究員．インターネットの計測・分析およびネットワークセキュリティに関する研究に従事．情報科学博士．平成 21 年電子情報通信学会英文 B 誌論文賞，平成 22 年電気通信普及財団テレコムシステム技術賞受賞．電子情報通信学会，ACM 各会員．

後藤　滋樹（フェロー）

昭和 23 年生．昭和 48 年東京大学大学院理学系研究科修士課程修了．同年電電公社武蔵野電気通信研究所に入所．昭和 59 年から 60 年にかけて米国スタンフォード大学客員研究員．現在，早稲田大学理工学術院教授．コンピュータ・アーキテクチャ，自然言語処理，プログラム理論，演繹的プログラム合成，コンピュータ・ネットワークの研究に従事．工学博士（平成 3 年，東京大学）．平成 8 年情報処理学会ベストオーサ賞，平成 15 年情報通信月間総務大臣表彰，平成 15 年情報処理学会フェロー．電子情報通信学会，ソフトウェア科学会，人工知能学会，応用数理学会，IEEE，ACM 各会員．