P2P ファイル交換ネットワーク環境における マルウェアの検出手法の提案

四本木 正男† 堀合 啓一† 田中 英彦†
†情報セキュリティ大学院大学 情報セキュリティ研究科
Email: †{mgs084505, gs063101,tanaka}@iisec.ac.jp

あらまし P2P ファイル交換ネットワークで流通している実行ファイル形式のマルウェアは、同ネットワークで流通している他のコンテンツに偽装、もしくは混入されて流通している。一方、正常な実行ファイルは、通常、他のコンテンツに偽装、もしくは混入されて流通することは少ない。本稿では、この違いを利用し P2P ファイル交換ネットワーク上で流通している実行ファイルがマルウェアか否かを自動判別する検出手法を提案する。また、提案手法を実装したプロトタイプシステムの評価を通して提案手法の有効性を示す。
キーワード：P2P, マルウェア, ウイルス, 検知

Malware detection system on the P2P file exchange network environment

Masao Shihongi† Keiichi Horiai† Hidehiko Tanaka†
†Institute of Information Security University
Email: †{mgs084505, gs063101,tanaka}@iisec.ac.jp

Abstract. Executable malware on the P2P file exchange network environment is injected into other contents, or it is disguised as other contents. On the other hand, non-malicious software isn't injected into other contents, and it isn't disguised as other contents very much. We propose a malware detection system using the difference. Also we implemented a prototype system to show the validity of our approach.
Key words：P2P, malware, viruses, detection

1. はじめに

近年、WinnyやShareといった無管理なP2Pファイル交換ソフトウェアの利用者が、ウイルスに感染し機密情報や個人情報を流出させる事件が多発し社会的問題となっている。

このような流出事故が多発する一因として、P2Pファイル交換ネットワークで流通しているファイルのマルウェア含有率が非常に高いことがあげられる。P2Pファイル交換ネットワークのマルウェア含有率について 2006 年調査[1]によると Limewire では流通している実行ファイルの68%がマルウェアを含有していると報告されている。また、2008 年調査[2]ではWinnyに流通している圧縮ファイルの2割がマルウェアを含むと報告されている。

本稿では、このように非常に高い割合でマルウェアを含むネットワークにおいて、通常のウイルス対策ソフトによるウイルス検出だけでなく、P2P ファイル交換ネットワークに特化したマルウェアの検出手法を提案する。

2. 関連研究

P2P ファイル交換ネットワークに特化した研究として、P2P ファイル交換ネットワークを介して情報流出を引き起こす暴露ウイルスに対する研究が行われている。喜田らの P2P ファイル交換ソフトのアップロードフォルダを監視

し、事前に流通を許可したファイル以外がアップロードされていないかを監視する方式[3]や、鬼頭らの端末内の動作を監視し、暴露ウイルスの個人情報収集動作や情報公開動作を検知し暴露ウイルスを検出する方式[4]などである。しかし、いずれも情報流出対策であり、筆者らが調査した限りでは P2P ファイル交換ネットワークのマルウェア検出手法に関する研究は見あたらなかった。

3. 提案手法

P2P ファイル交換ネットワークで流通している圧縮ファイルに含まれる実行ファイルがマルウェアか否かを判定する手法として、P2P ファイル交換ネットワークに流通しているマルウェアと正常なファイルの流通状態の違いから、マルウェアを検出する手法を提案する。

マルウェアと正常なファイルの流通状態の違いについて、次の２つがあると推定できる。（１）マルウェアは P2P ファイル交換ネットワーク上で流通している圧縮ファイル等の様々なコンテンツに自身を混入して流通するが、正常なファイルは他のコンテンツに自身を混入することはない。（２）マルウェアは、自身を様々なファイル名に変更・偽装してコンテンツに混入するが、正常なファイルは、通常自身のファイル名は１つで複数のファイル名でコンテンツに混入することはない。

この２つの違いを元に以下の２つの検出手法を提案する。

（１）ハッシュ分類法 : 同じハッシュ値を持った別名の実行ファイルが複数存在する場合、それらのファイルをマルウェアと判定する。

（２）振る舞い分類法 : 同一の振る舞いを行う実行ファイルが、別名の実行ファイルで複数存在する場合、それらをマルウェアと判定する。

ハッシュ分類法は、P2P ファイル交換ネットワークで流通するコンテンツにマルウェアが自身を混入する際に、自身のバイナリを改変せずにそのまま混入するケースを検出するものである。一方、マルウェアではない実行ファイルでも、汎用性のあるものは複数のコンテンツに含まれて使い回される可能性があるが、その場合ファイル名は固定で変化しないと考えられる。

振る舞い分類法は、多くの亜種を持つマルウェアや、自身のコードを変化させる自己変化型のマルウェアを検出するものである。ハッシュ分類法ではファイルの同一性をファイルハッシュで識別しているため、同じマルウェアが異なるファイルハッシュで流通すると、異なるファイルが個別に流通している正常な流通と区別できず、マルウェア検出できない可能性がある。そのため振る舞い分類法では、ファイルの同一性を、実行ファイルを実行した時にシステムにどのような変化を与えるかの「振る舞い」で判断する（振る舞い情報の内容については 4 章に記す）。これによって、マルウェアが異なるファイルハッシュで流通しても、振る舞いが同じであれば同一ファイルと識別し、流通状態の異常からマルウェア検出を行うことが可能となる。

4. 実装方法

提案手法を実装するシステムのシステム構成、機能一覧を図 1、表 1 に記す。本システムは次の二つの動作フェーズがある。

（１）振る舞い情報の収集フェーズ

本フェーズは、図 1 の①～④の動作を常時実施し、データベース（以下、ＤＢと記載）に振る舞い情報を蓄積する。①P2P ファイル交換ネットワークに流通している圧縮ファイルを無作為にダウンロード。②圧縮ファイルを展開し、実行ファイルが含まれる場合抽出する。③実行ファイルを検証環境で実行し、振る舞い情報を取得する。④得られた情報を「振る舞いＤＢ」に記録する。

（２）マルウェア判定フェーズ

本フェーズは、「振る舞いＤＢ」に記録された実行ファイルに対して、提案手法のハッシュ分類法、振る舞い分類法の二つでマルウェア判定を実施する。これまでダウンロードしたことのない実行ファイルのマルウェア判定を行う場合は、「振る舞い情報取得機能」にてその実行ファイルの振る舞い情報を取得し、得られた情報を「振る舞いＤＢ」に記録した後にマルウェア判定を実施する。

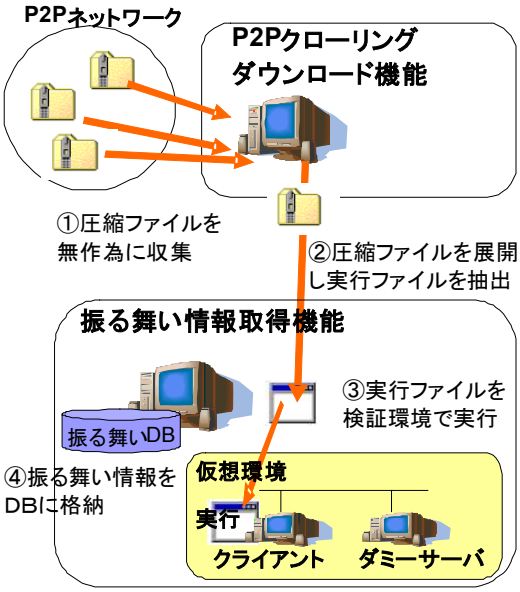


図 1 提案システムの構成

表 1 機能一覧

機能名	内容
P2Pクローリング/ダウンロード機能	P2Pファイル交換ネットワークに接続している端末に次々と接続、巡回し、流通しているファイルの情報を取得、未取得の圧縮ファイルのダウンロードを行う。なお、違法なファイルの流通を助長しないように、アップロードやキー情報の拡散等は行わない。
振る舞い情報取得機能	実行ファイルを仮想環境で実行し、実行前後のシステムの変化を記録することで、その実行ファイルの振る舞い情報を取得する。取得した振る舞い情報は「振る舞いDB」に格納する。
振る舞いDB	実行ファイルがマルウェアか否かの判定に利用する。振る舞いDBの蓄積フェーズでは、ファイル名、ファイルハッシュ、実行時の振る舞い情報を蓄積し、マルウェア判定フェーズでは、同一ファイルハッシュもしくは同一振る舞いの実行ファイルの抽出を行う。

4.1. 振る舞い情報の取得方法

実行ファイルを仮想環境で実行し、実行前後のシステム変化を記録することで振る舞い情報を収集する。

4.1.1. 振る舞い取得環境

取得環境は、外部ネットワークから隔離された環境で、実行ファイルを実行するクライアント PC とダミーサーバをネットワーク接続した構成とする。両マシンは仮想 OS で構築し、クライアント PC は Windows XP Pro SP2、ダミーサーバは Linux(Fedora9)とする。ダミーサーバは一部 Truman[5]の機能を利用し、クライアント PC からの通信「DNS, TELNET, FTP, SMTP, POP, Web, IRC, SMB」に疑似応答する。クライアント PC のデフォルトゲートウェイと DNS はダミーサーバに指定し、ネットワーク外への通信は全てダミーサーバが応答する。

4.1.2. 取得する振る舞い情報

クライアント PC 上で実行ファイルを実行し、実行前後のシステムの変化を取得した後、クライアント PC を実行前の元の環境に戻す。この作業を複数回(n 回)繰り返して得たシステムの変化を整理し、ファイル名、ファイルハッシュとともに振る舞いＤＢに記録する。実行ファイルによっては実行する度に挙動が変わるものもあるため、複数回動作を取得する。なお今回は n=4 回とした。記録する変化情報は次に記す５つである。(1)ファイルの変化情報（２）プロセスの変化情報（３）レジストリの変化情報（４）Listen Port の変化情報（５）通信先ドメイン情報。以下にそれぞれの詳細を記す。

4.1.2.1. ファイルの変化情報

クライアント PC の全ファイルのファイル名とそれぞれの全ファイルハッシュをあらかじめ取得しておき、実行ファイルを実行後、再度全ファイルのハッシュを取得し、変化(作成、変更、削除)した全てのファイル名を記録する。

ただし、プログラムによっては、テンポラリファイルなど、実行する度に異なるファイル名を変化させる場合がある。このようなファイルをそのまま振る舞い情報として記録すると、同一の実行ファイルでも振る舞い情報が一致せず、実行ファイルの同一性の判定に利用できない。そのため、ファイル名が不定の場合、次の２つの決まりに従い記録を行う。（１）複数回(n 回)実行し、変化するファイルのディレクトリの階層の深さが共通であれば変化するディレクトリとファイル名にワイルドカードを用いて振る舞い情報を記録する（２）変化するディレクトリの階層が一致しない場合や、変化したりしなかったりする場合は、振る舞い情報として記録しない。

例を図 2 に記す。「ファイルを実行し、変化したファイル名を取得する」という作業を４回行った結果、左の４つの枠内のファイルが変化した例である。「C:¥WINDOWS/win.ini」は４回とも変化しているので、そのまま DB に記録する。「C:/Winny2/・・」は、４回ともに変化しているが、ファイル名とディレクトリの２階層目が一致しない。そのため、ファイル名と二階

層目を「*」にして DB に記録する。「c:/temp/yyy.tmp」「c:/tmp/xxx.tmp」は 2,3 回目に変化が記録されているが、1,4 回目で対応するファイルの変化が取得できていないため、DB に記録しない。

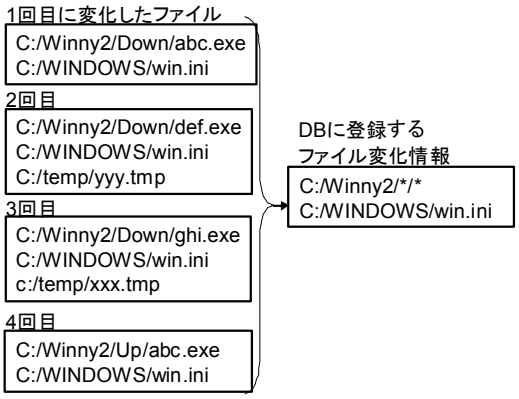


図 2 変化するファイル名の記録方法

4.1.2.2. プロセス名変化情報

実行ファイルを実行する前後でプロセス名一覧を取得し、増減したプロセス名とその増減数を取得する。実行ファイル１つに複数回(n 回)この情報取得作業を行い、必ず増減するプロセス名をＤＢに記録する。ただし、実行ファイル名と同一のプロセス名が増加する場合は、プロセス名ではなく「*ME*」という名前で記録する。実行する度に異なるプロセス名が増減する場合、もしくは、増減したりしなかったりする振る舞いはＤＢに記録しない。

4.1.2.3. レジストリ変化情報

実行ファイルを実行する前後で、自動実行(Autorun)に関わるレジストリ情報を取得し、変化するレジストリのキーと値の名前を取得する。複数回(n 回)この情報取得作業を行い、必ず変化するキーと値の名前を記録する。

4.1.2.4. ListenPort 変化情報

実行ファイルを実行する前後で、Listen している TCP/UDP ポート番号を取得し、Open/Close したポートを取得する。複数回(n 回)実施し必ず変化するポートを記録する。

4.1.2.5. 通信先ドメイン情報

ダミーサーバでクライアント PC の通信パケットをキャプチャし、クライアント PC が DNS の名前解決を行った場合、そのドメイン名を通信先ドメイン情報として記録する。複数回(n 回)実施し、必ず通信するドメインを記録する。

4.2. マルウェア判定方法

4.2.1. ハッシュ分類法

ファイルのファイルハッシュが同一で、複数個の別名のファイルが存在する場合に、マルウェアと判定する。

4.2.2. 振る舞い分類法

取得した変化情報５つ(4.1.2)が全て一致する実行ファイルを抽出し、それらのうち、ひとつでも異なるファイル名が存在する場合、全てマルウェアと判定する。ただし、次の２つの振る舞いについては、マルウェアと判定するには情報が少なすぎるため、複数のファイル名があってもマルウェアと判定しない。（１）プロセス変化情報に実行したファイルのファイル名が１つ増え、それ以外は変化がない振る舞い（２）一切変化のない振る舞い。

5. 検証

5.1. 検証方法

まず、Winny ネットワークに提案システムを接続し、2009/5/1～6/29 の約 2 ヶ月間、実行ファイルの収集と、振る舞い DB の蓄積と行う。その後、収集した実行ファイルを、ハッシュ分類法、振る舞い分類法でマルウェア判定を行う。

5.2. 検証結果

5.2.1. 収集した実行ファイル

収集期間中にダウンロードした圧縮ファイル(zip,lzh)は 11476 個であり、その圧縮ファイルに含まれる実行ファイルは 2885 個であった。これらのファイルをファイルハッシュで分類すると、876 個であった。これをウイルスバスター2009(2009/8/6 最新パターン)でスキャンすると 876 個中 692 個を 28 種類のウイルスとして検出した。検出したウイルス名の Top10 を表 2 に記す。

表 2 ウイルスバスター2009 検出結果

検出したウイルス名	検出数
PE_PARITE.A	341
WORM_ANTINNY.JB	145
WORM_ANTINNY.AE	85
WORM_ANTINNY.GEN	54
WORM_ANTINNY.J	10
WORM_ANTINNY.BU	9
WORM_ANTINNY.JA	8
TROJ_DELF.CTJ	6
Cryp_Xed-17	4
TROJ_Generic.DIF	4

5.3. マルウェア判定結果

876 個の実行ファイルを、ハッシュ分類法、振る舞い分類法によってマルウェア判定した結果、前者は 217 個、後者は 649 個をマルウェア判定した。これらの検出結果とウイルスバスター2009 でのウイルス検出結果との比較を表 3 に記す。ハッシュ分類法で検出した 217 個(表 3-a の XT+X)は全てウイルスバスター2009 でウイルス検出されている(表 3-a の X)。振る舞い分類法で検出した 649 個(表 3-b の XT+X)は、ウイルスバスター2009 で 642 個をウイルス検出し(表 3-b XT)、残り 7 個は未検出のファイルであった(表 3-b X)。

表 3 検出数比較

表3-a ハッシュ分類法

XT	217
T	475
X	0
--	184
合計	876

表3-b 振る舞い分類法

XT	642
T	50
X	7
--	177
合計	876

■記号の意味
XT ..ウイルスバスター2009と提案手法が検出したファイル数
T ..ウイルスバスター2009のみ検出したファイル数
X ..提案手法のみ検出したファイル数
-- ..ウイルスバスター2009と提案手法が未検出のファイル数

■参考
ウイルスバスターの検出数合計:(表3-a,b) XT+T=692
ハッシュ分類法の検出数合計 :(表3-a) XT+X=217
振る舞い分類法の検出数合計 :(表3-b) XT+X=649

6. 考察

6.1. 検出率とマルウェアの流通状態

ハッシュ分類法はウイルスバスター2009 がウイルスを検出したファイルの 31%、振る舞い分類法では 93%をマルウェアと判定しており、振る舞い分類法の検出精度が高い。これは、検体のマルウェアが多くの亜種を持つため、もしくは自己変化型であるためと推定される。マルウェアの流通状態を示す一例として、ウイルスバスター2009 が WORM_ANTINNY.GEN として検出したファイルについて、ファイルハッシュとそのファイルハッシュが持つファイル名の種類数を図 3(a)、振る舞いパターンとその振る舞いパターンが持つファイル名の種類数を図 3(b)に記す。なお、図のファイルハッシュは先頭 4 文字のみを省略表示し、上位 40 件まで表示している。図 3(a)より WORM_ANTINNY.GEN は複数のファイルハッシュを持ち、一種類のファイル名しかないファイルハッシュも多く存在することが分かる。ハッシュ分類法では一種類のファイル名しか持たないファイルハッシュは正常と判定するため、これらの実行ファイルはフォールス・ネガティブとなる。一方、振る舞い分類では、図 3(b)より WORM_ANTINNY.GEN は 5 種類に分類されている。振る舞い分類法では、このように同一マルウェアが集約されるため、一種類のファイル名しか持たない振る舞いが存在しにくくなり検出率が高くなっている。

6.2. 誤検出率と正常ファイルの流通状態

振る舞い分類法でマルウェア検出した 649 ファイル中、7 個はウイルスバスター2009 で検出していない。この 7 個は別のウイルス対策ソフト NOD32 Antivirus でスキャンしたところ、3 個が Win32/Antinny.J,Antinny.A として検出された。残り 4 個は、実行するとエラーダイアログが表示されるものが 2 つ、動画変換ソフトと思われるアプリケーションのインストーラが起動するものが 2 つであった。この 4 個の振る舞い情報を調査した結果、誤検出と思われる。前者の 2 個はファイルが壊れており、ファイル名の異なる 2 つの実行ファイルが「不正終了しエラーダイアログを出す」という同一の振る舞いを行ったためマルウェアと誤判定し、後者の 2 個はインストーラを起動する振る舞いが同一であったためマルウェアと誤判定していた。結果、振る舞い分類法の誤検出数は 649 個中 4 個であり、誤検出率は 0.6%である。

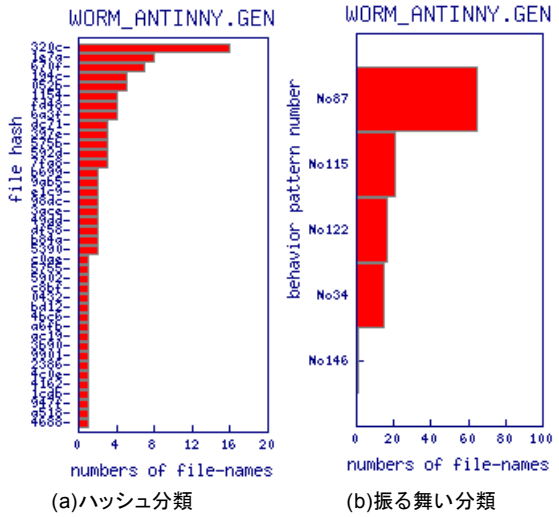
ウイルスバスター2009 がウイルス検出しなかったファイルの流通状態を図 4 (a)(b)に記す。なお、図 4 (b)の最も多い振る舞い No11 は「プロセス変化情報に実行ファイルのファイル名が増え、それ以外は変化がない振る舞い」、次に多い No55 は「変化情報が一切ない振る舞い」である。図 4 (a)より正常なファイルはファイル名を 1 つしか持たないという仮説どおりの流通であることが分かる。図 4 (b)はマルウェア判定の例外パターンである No11,55 をのぞくと、基本的に 1 つの振る舞いは 1 つのファイル名で流通していることが分かる。ただし、振る舞いパターン No103,56,109 はファイル名種類数がそれぞれ 3,2,2 個と複数あるため、この計 7 個はウイルスバスター2009 ではウイルス未検出であるが、振る舞い分類法ではマルウェアと判定される。この 7 個は先に挙げた、NOD32 でウイルスとして検出される 3 個と、振る舞い分類法が誤検出した 4 個である。

7. 課題

本稿で実装した振る舞い分類法では、同一マルウェアとして集約できないマルウェアが存在する。その例として PE_PARITE.A の流通状態を図 5(a)(b)に記す。同一マルウェアにも関わらず振る舞いパターンが多数記録され、結果的に１つしかファイル名を持たない振る舞いが存在してしまい、フォールス・ネガティブに繋がっている。振る舞いが一致しない理由の一つとして、PE_PARITE.A が正常な実行ファイルに寄生して感染するウイルスであることがあげられる。PE_PARITE.A に感染した実行ファイルの振る舞いは、「マルウェアの動作＋元のファイルの動作」を行うため、同一のマルウェアであっても、宿主のプログラムの振る舞いが異なれば振る舞い情報が一致しない場合がある。マルウェア動作の振る舞いのみを取得する手法などの検討が必要である。

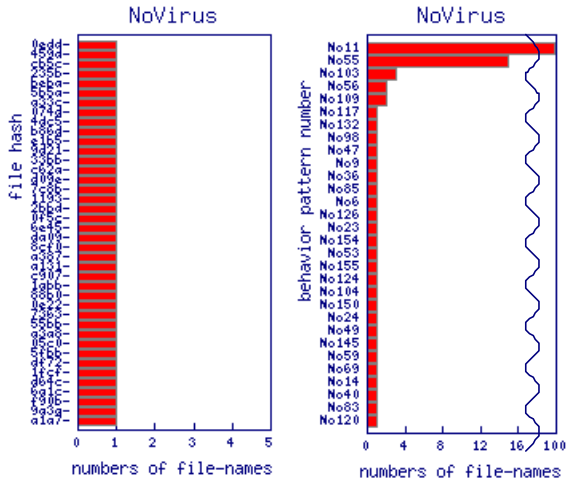
8. まとめ

本稿では、P2P ファイル交換ネットワークに流通しているコンテンツに含まれる実行ファイルがマルウェアか否かを、流通状態から判定する検出手法を提案し、実装、評価した。その結果、振る舞い分類法で検出率 90%以上、誤検出率 1%未満の結果を得られた。しかし、現状の分類法では検出が難しいマルウェアも存在するため、今後の課題である。



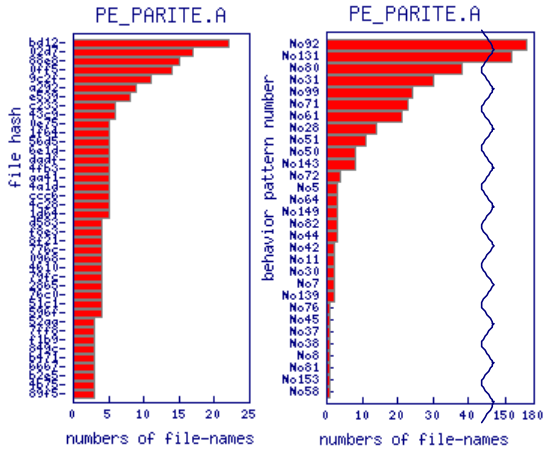
(a)ハッシュ分類　(b)振る舞い分類

図 3 分類方法とファイル名数



(a)ハッシュ分類　(b)振る舞い分類

図 4 正常ファイルのファイル名数



(a)ハッシュ分類　(b)振る舞い分類

図 5 振る舞い分類できていない例

文献

[1] Andrew Kalafut, Abhinav Acharya, Minaxi Gupta, "A Study of Malware in PeertoPeer Networks" Proceedings of the 6th ACM SIGCOMM conference on Internet measurement pp. 327-332, 2006.

[2] "P2P の現状～Winny、Share ネットワーク状況調査報告～" 安心・安全インターネット推進協議会 P2P 研究会 第１回 セミナー 2008/9/10

[3] 喜田弘司, 坂本久, 島津秀雄, 垂水浩幸 "ファイルアクセス制御エージェントの提案" 情報処理学会論文誌 Vol.48, No.1(20070115) pp. 200-212

[4] 鬼頭哲郎, 松木隆宏,松岡正明, 仲小路博史, 寺田真敏, "端末内の動作監視に基づく情報漏えいウイルスの検知手法に関する検討" IPSJ SIG Technical Reports 2008-CSEC-42(45) 2008/7/25

[5]http://www.lurhq.com/truman/