多様な感染経路に対応したマルウェア収集・解析環境の構築

水谷 正慶 †　　武田 圭史 †　　村井 純 ††

† 慶應義塾大学大学院 政策・メディア研究科
†† 慶應義塾大学 環境情報学部
252-8520 神奈川県藤沢市遠藤 5322

{mizutani,keiji,jun}@sfc.wide.ad.jp

あらまし　マルウェアの活動形態の多様化に伴い，柔軟なマルウェアの収集・解析が求められている．これまでマルウェアの収集は感染元ホストからの通信を待ち受けるハニーポットが多く利用されていたが，Web ブラウザを経由して感染するマルウェアには対応できないなどの欠点がある．本稿では，多様な感染経路に対応したマルウェアの収集・動的解析の要件を定義し，必要に応じた環境の構成変更や通信制御が可能となる環境を実現した．また，CCC DATAset 2009 で提供されたハッシュ値に一致する検体を利用し，動的解析を実施した結果について示す．

The Implementation of an Environment to Collect and Analyze Malware Adapted to Various Infection Methods

Masayoshi Mizutani†　　Keiji Takeda †　　Jun Murai††

†Keio University - Graduate School of Media and Governance
††Keio University - Faculty of Environment and Information Studies
5322, Endo, Fujisawa-shi, Kanagawa 252-8520, Japan

Abstract　As malware (malicious software) activity patterns change, flexible collection and analysis of malware became necessary. Nowadays, most honeypots that are waiting for malware communications were used to collect their information, but these honeypots can't adapt to malware that infect computers over web browsers. This paper describes requirements of malware collection and analysis that can adapt to various infection methods, and this paper presents an implementation of malware collection and analysis environment that can change its structure and control its communications. Then, the results of dynamic analysis, which was done on malware that matched with hash values provided on CCC DATAset 2009, is also presented in the paper.

1　はじめに

マルウェアによる脅威が拡大する中，マルウェアの解析や対抗策の研究が急務となっている．マルウェアによる犯罪行為から経済的利益が得られるようになり [1, 2]，マルウェア作成者は多様な手段を用いて検知の回避や耐性の向上を図っている [3]．そのため，従来以上に感染が拡大し，新種のマルウェアの出現頻度や，傾向の変化が著しくなっている．マルウェアの対策を実施する際には，その活動形態の調査が必須である．急速に変化するマルウェアに対し，マルウェアの活動を知らなければ効果的な対策を提案するのは難しい．

マルウェアの解析には，感染対象となる OS 上でマルウェアを動作させて情報を収集する動的解析と，バイナリコードを解析して動作を把握する静的解析がある．静的解析はマルウェアに関する詳細な情報を入手できるが，近年のマルウェアは意図的に難読化されており，解析には多大な時間を要する．さら

に，新しいマルウェアの出現頻度が高いため [4]，短時間でマルウェアの動作に関する情報を得られる動的解析と併用するのが効果的である．

本稿では，マルウェアの収集・解析環境構築における要件を定義し，多様な感染経路に対応する設計・実装を実施した．これまで，収集と解析は攻撃を待ち受けるハニーポットを利用したものが多く，積極的に通信を開始し感染を拡大させるマルウェアのみに対応しているものがほとんどであった．これに対し，近年は Web サイトの閲覧などによって感染するマルウェアが増加しており，マルウェアの収集・解析環境の多様化が求められている．本稿では今後もマルウェアの活動形態が変化し続ける可能性を考慮し，複数の感染経路に対応できるマルウェア収集・解析環境を構築した．また，研究者が小規模な環境から構築する点を考慮することで，拡張性や柔軟性などの要件を定義し，これを満たすよう実装した．

2 マルウェア解析および収集環境構築における課題

従来，マルウェアの収集・解析は能動感染型のものが注目されてきた．能動感染型マルウェアとは，攻撃元ホストから通信を開始し，攻撃コードを送信することによって感染を拡大させるマルウェアである．能動感染型マルウェアは 2003 年に発生した Blaster[5] などを筆頭に大量に発生しており，擬似的なサービスを提供することで攻撃を待ち受けて攻撃コードやマルウェアを収集するハニーポットが多く利用されてきた．

しかし，2007 年の Provos らの調査 [6] から受動感染型マルウェアが注目されはじめた．受動感染型マルウェアは攻撃元ホストから一方的に攻撃コードが送信されるのではなく，ユーザの動作が起因となって攻撃が発生する．Provos らの調査では Web ブラウザであらかじめマルウェア本体をダウンロードさせ，Web ブラウザやプラグインの脆弱性を利用して感染する攻撃 (drive by download) に着目している．この攻撃手法はファイヤウォールや Network Address Translater(NAT) によって外部からの通信が遮断されているネットワークのホストでも，マルウェアに感染させることができ，マルウェア感染の危険性が高まっている．

このような感染経路をはじめとするマルウェアの活動形態の変化は，マルウェアの対策にも大きな影響を及ぼすため，新しいマルウェアに対応した解析環境が必要となる．筆者らの調査 [7] でも，能動感染型マルウェアと受動感染型マルウェアでは異なる傾向があることが明らかになっている．一部のマルウェア対策の研究では，通常の通信とマルウェアによる通信の傾向の違いを利用した検知手法が提案されているが，通信傾向が変化することで検知が困難になってしまう．そのため，継続的にマルウェアの活動を観測しつつ，多様な感染経路に対応できるマルウェアの収集・解析環境が求められている．

3 要件定義

- **並列性**: 動的解析は短時間で多くのマルウェアを解析できるのが利点であるため，複数のノードを同時に動作させる構成が望ましい．また，収集に関しても並列して実行することによって，より多くのマルウェアを収集できる可能性が高まる．
- **完全性**: マルウェアを実行した際に，意図しないデータや通信が改ざんされる事を防ぐ必要がある．また，同一の解析環境にある別のマルウェアによって，収集するデータが干渉されないように配慮しなければならない．
- **頑健性**: 一部の解析処理に問題が発生しても，全体の系は処理を継続できるように実装する必要がある．
- **拡張性**: 今後，マルウェアの増加速度はより加速するものと予想されるため，収集・解析環境を任意に拡張できる実装が必要である．特に，システム全体を停止させずに拡張・変更できる環境構成が望ましい．
- **柔軟性**: 近年，受動型攻撃が増加したように，今後も感染経路が変化していく可能性がある．感染経路や活動形態にあわせて，接続する機器やネットワークを選択できるような柔軟性が求められる．
- **通信の制御**: マルウェアの収集や一部の解析では，インターネットに接続しなければならない．しかし，マルウェアを実際に実行する動的解析では，他のネットワークに対する迷惑行為や攻撃が発生する可能性がある．そのため，必要に応じて通信を遮断する，外部ネッ

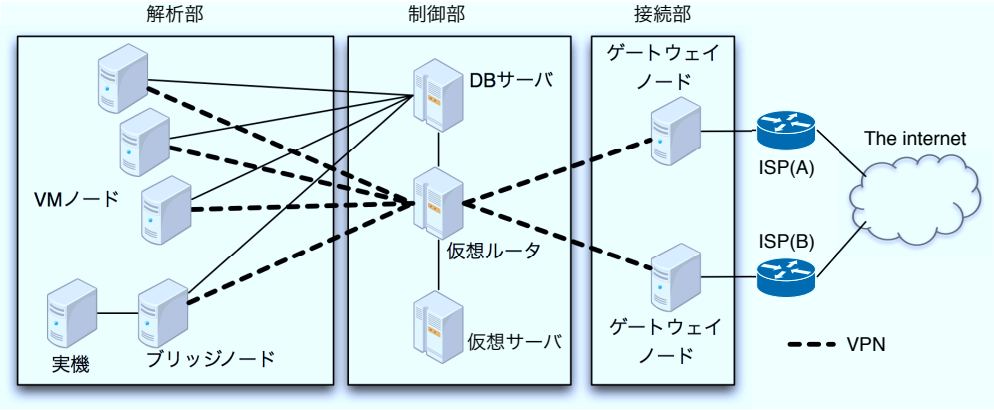

図 1: マルウェア収集・動的解析構成概要

トワークに見せかけた疑似サービスへ誘導する，通信量を制限するなどの措置が適宜必要となる．

4 多様な感染経路に対応した収集・解析環境の実装

第 3 節の要求に基づき，マルウェアの収集・解析システムを実装した．本実装ではマルウェアの収集において，攻撃元ホストから通信を開始する能動型攻撃だけではなく，悪意がある Web サイトの閲覧時に感染する受動型攻撃についても対応できるシステムを構築した．本システムでは，主に仮想マシンを利用してマルウェアの収集・解析を実施しており，仮想マシンの起動時に任意の命令を発行できるよう構成されている．外部ネットワークからの攻撃通信をシステム内のホストに転送すれば，従来の能動感染型マルウェアに対応したハニーポットとして動作し，起動時に攻撃コードが含まれると予想される Web サイトを自動的に閲覧する指示を発行すれば，受動感染型マルウェアに対応したハニーポットとなる．この構成を応用すれば，より新しい感染経路が発生しても柔軟に対応できると考えられる．

4.1 システム全体の構成

本システムは汎用的な大規模テストベッドではなく，必要に応じて部品を増設する専用のシステムとして設計されている．そのため他の用途に使うのは難しいが，比較的安価に設置・運用できるのが特徴の 1 つである．

図 1 は実装した収集・解析環境の概要を示している．本システムは VM ノードとブリッジノードおよび実機からなる解析部，仮想ルータ，仮想サーバからなる接続部，ゲートウェイノードからなる表層部の 3 つのパートで構成されている．

- **VM ノード**: 仮想マシンを用いて，マルウェアが感染対象とする OS を動作させ，マルウェアへ感染させる，あるいはマルウェアを動作させることで，マルウェア本体や活動に関する情報を収集するノード．各ノード上では複数の仮想マシンを動作させ，マルウェア収集・解析の並列処理を実現する．本システムではホスト OS として Ubuntu Linux，仮想マシンに VirtualBox[8] を用い，VirtualBox 上で Windows XP(R) Professional(パッチ未適用) を基本として利用した．

- **ブリッジノード・実機**: 実機上にインストールしたマルウェア感染対称の OS を動作させ，収集・解析を実施する．自身の解析を防ぐために，仮想マシン上での動作を検知して動作を変更させる機能を有するマルウェア [9] が存在するため，同一の OS やソフトウェアを使っていても，仮想マシン上での動作と実機上での動作が異なる可能性があり，実機での収集・解析環境も必要となる．実機で動作させた場合，特殊な制御を当該ホスト上で実施するのは難しいため，実機上で発生した通信をブリッジノードを介してネットワークに接続させる．

- **仮想ルータ**: VM ノードとブリッジノードの解析部とゲートウェイノードを仲介する本シ

ステムの基幹ルータ．ただし，各 VM ノード，ブリッジノード，ゲートウェイノードとは全て VPN を利用して接続している．仮想ルータは VM ノード，ブリッジノードからきた通信を収集・解析内容に応じてルーティングし，必要に応じて特定のサービスを仮想サーバに誘導する．さらに，ファイアウォールを用いることで外部へ送信してはならない通信を遮断し，IDS を用いることで通信の概要を把握できる．本実装では OS に Ubuntu Linux，VPN に OpenVPN[10]，ファイアウォールに iptables，IDS に Snort[11] を利用した．

- **仮想サーバ**: マルウェアが外部のホストに対して送信するリクエストの宛先を仮想サーバに改ざんし，あたかも外部のホストが応答しているかのように見せかけるためのサーバ．通信の誘導には宛先 Network Address Translation (NAT) を利用している．例えば DNS の場合，偽のアドレスを応答したり，DNS サーバを経由して実際の問い合わせ結果を応答するのも可能である．また，SMTP の場合は大量の迷惑メールが外部に送信されないように，受信したメールを捨て続けるブラックホール SMTP サーバとして動作させる事も可能である．また，TCP の SYN パケットに対して偽の SYN-ACK パケットを送信する dumnet[12] を利用することで，擬似的に TCP のセッションを確立させ，マルウェアにパケットを送信させるようにも設定できる．本実装では OS に Ubuntu Linux，DNS サーバに bind9，SMTP サーバに Postfix を利用した．

- **DB サーバ**: マルウェアに関する情報を一元的に管理するサーバ．収集したマルウェアの検体や，悪意のある URL とドメイン名，解析結果であるマルウェア通信などの情報が蓄積されている．VM ノード上の仮想マシンの多くは自動で実行され，その際にどのような解析をするか，どの検体を利用するか，どの URL に接続するかなど，動作に関する指示を発行する役割も DB サーバが担う．さらに解析結果に関する情報も自動的に DB サーバに送信される．

- **ゲートウェイノード**: インターネットへと接続するため，外部接続性を提供するノード．マルウェアの振る舞いは感染対象のネットワークによって異なるという仮説があるため，これらの検証のために複数の外部ネットワークと接続性を用意した．また，冗長性などの点からも複数の回線を持つ構造が望ましいと考えられる．

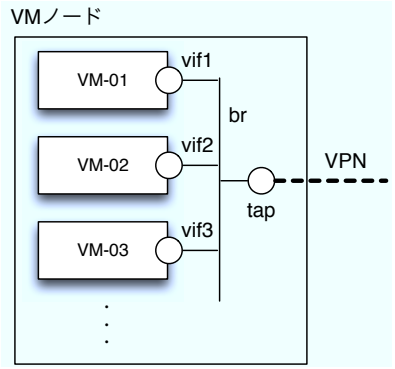

図 2: VM ノード構成

一定規模のテストベッドを構築する場合，専用のホスト設置場所やネットワークを用意し運用しなければならないが，本システムは規模に関わらず容易な構築が可能である．インターネット上の保護された通信上に環境を構築しているため，インターネットが利用できれば物理的な設置場所やネットワーク構成などに依存しない．また，システム構成を変更する場合も，比較的容易に VM ノードやブリッジノードを追加できる．

4.2 VM ノードの構成

本システムの収集・動的解析は VM ノードが主体となっている．村上らの調査 [9] で述べられている通り，仮想マシン上では動作が制限されるマルウェアも存在するが，本システムでは効率的な収集・動的解析を優先している．図 2 に VM ノードの構成を示す．VM ノードは複数の仮想マシンを有し，それぞれを並行して動的解析を実行できる．仮想マシン数は VM ノードの性能とゲスト OS の推奨性能を考慮して設定するのが望ましい．各仮想マシンには仮想ネットワークインターフェース (vif_n) を有し，そのインターフェースと VPN によって作成される仮想ネットワークインターフェース (tap) をブリッジすることで，直接インターネットに接続するのを防いでいる．各仮想マシンのネットワークインターフェース vif_n では通信内容を記録すると同時に，iptables

と ebtables[13] を用いて同一セグメント内の別の仮想マシンに送信される通信を遮断し，複数種類のマルウェアが混在しないよう配慮している．

各 VM ノードは仮想マシン上のゲスト OS を起動する時に DB サーバから収集・解析に関する指示，収集・解析の継続時間を取得する．現在，1) 取得済みのマルウェアを実行する，2) 悪意のあると考えられる Web サイトを Internet Exproler で閲覧する，の 2 種類の動作が実装されており，各 VM は HTTP 経由で DB サーバから指示を得る．各 VM はホスト OS との共有ディレクトリが設定されていり，DB サーバから指示された内容を元に共有ディレクトリにバッチファイルを生成する．バッチファイルには Internet Explorer で参照する URL や，実行するマルウェア検体のファイル名が記述されており，ゲスト OS は起動時に共有ディレクトリに設置されているバッチファイルを自動的に起動する．また，とくにゲスト OS に指示を与えず，任意のゲートウェイノードに送信される攻撃通信を転送することによって，従来の能動型攻撃を受け付けるハニーポットとしても動作する．指定された収集・解析の継続時間を経過した後はゲスト OS を終了させ，感染前の状態に復元する．

各 VM ノードは自立分散的に動作しており，システムへの追加や離脱がシステム全体に及ぼす影響は小さい．また，何らかのトラブルで 1 部の VM ノードが停止しても他の VM ノードは継続して動作し続ける．これによって，拡張性や頑健性が実現されている．

5 実験

5.1 受動感染型マルウェアの収集

2009 年 7 月 2 日から同年 9 月 3 日まで，1,515 件の悪意があると考えられる Web サイトを調査した．調査に利用した URL は Malware Domain List[14] で提供されている URL と，hpHosts[15] で提供されているドメイン名が含まれる Web ページの URL の 2 種類である．ゲスト OS の起動後，自動的に Web ページを閲覧し，20 分間通信データを取得した．

通信データにおける HTTP の通信を解析したところ，174 種類の Windows の実行形式ファイルを取得した．これらを AVG Free Edition 8.5.409[16] (ウィルス DB ヴァージョン 270.13.76/2345) を用いて検査したところ，131 件をマルウェアとして検出した．さらに検知できなかった 43 件についても，ハッシュ値を Malware Hash Registry[17] で検索したところ，マルウェアとして登録されている事が明らかとなった．この結果から，受動感染型マルウェアの収集システムとして有効に機能していると言える．

5.2 CCC DATAset 2009 を用いた動的解析

CCC DATAset 2009 で提供されているハッシュ値と同様の検体を収集し，本環境において解析を実施した．今回はインターネットには接続せず，隔離されたネットワーク構成で解析を試みた．基本的に全ての通信を仮想サーバに誘導し，DNS と SMTP 以外のサービスは dumnet によって擬似的な TCP セッションの確立のみとした．2009 年 9 月 3 日 20 時から同日 24 時までに 1 つの検体につき 9 回から 10 回動的解析を実施し，解析結果として通信データを取得した．

通信開始のパターン　あるマルウェアは DNS による名前解決をせずに，10 秒あたり 20 から 30 の IP アドレスに対して UDP パケットを送信していた．本解析ではインターネットへの接続を許可していないため，外部から接続先に関する情報は取得できない．一つのマルウェアに対して複数回の動的解析を実施していたため，宛先 IP アドレスと宛先ポート番号を比較したところ，各解析結果で宛先 IP アドレスと宛先ポート番号の出現順序が同じだった．今回の解析では解析開始時刻がそれぞれ異なるため，宛先 IP アドレスとポート番号の生成関数が固定であるか，あらかじめプログラムに宛先 IP アドレスとポート番号のリストが指定されているかの，いずれかであると考えられる．この情報は静的解析における手がかりとして利用できるほか，シグネチャ型 IDS のルール作成などにも応用できる．

マルウェアの使用文字列およびプロトコルの特定　dumnet を利用し擬似的に TCP セッションを確立させることで，マルウェアに TCP のセグメントデータを送信させることができる．このセグメントデータを解析することで，マルウェアが利用する文字列やプロトコルを特定 [18] できる．解析の結果，あるマルウェアは必ず特定の文字列を含む URL に対し，HTTP によってアクセスしている事が明らかになった．出現する文字列が毎回同じ結果であれば，マル

ウェアのコード内に該当する文字列が含まれる可能性が高い．近年の多くのマルウェアは難読化されているため直接の解析は困難だが，動的解析時にメモリ上に展開するなどの手法を用いた後の解析には有用であると考えられる．また，この解析結果も IDS のルール作成などに活用できる．

6　今後の課題

VM ノードやブリッジノードのセキュリティ管理について考慮する必要がある．現在，VM ノードやブリッジノードは同一の管理ネットワーク内で運用されているが，今後の拡張性を考慮すると複数のネットワークに散在できる分散性が求められる．しかし，VM ノードの VM 管理ソフトやブリッジノードの OS，VPN ソフトに対して攻撃を実施し，各ノードを攻略する未知のマルウェアも存在すると見られる．このようなマルウェアを実行した場合，収集・解析システム全体に障害が発生したり，内外を問わずネットワークシステムが脅威にさらされる可能性がある．そのため，Secure OS の利用やファイアウォール，IDS の導入を視野にいれ，収集・解析システムのセキュリティを高めるため，監視と制御に関する実装を考慮していく必要がある．

現在，本システムでは VM ノードが 3 台，ブリッジノードが 1 台という小規模な運用となっている．本システムの構成では仮想ルータにトラフィックが集中するため，単一障害点になる恐れがある．したがって，今後規模を拡大させる過程で仮想ルータの性能を吟味しつつ，仮想ルータの冗長化や負荷分散を可能とする構成に変更していく必要がある．

7　まとめ

本稿では，従来着目されてきた能動感染型マルウェアだけではなく，Web サイトの閲覧によって感染する受動感染型マルウェアの研究を促すため，多様な感染経路に対応したマルウェア収集・解析環境を構築した．感染対象となる OS を複数用意し，それぞれに任意の指示を発行できるシステムを構築することで，悪意があると考えられる Web サイトを自動的に閲覧し，意図的に感染させることで受動感染型マルウェアを収集した．また，本システムはマルウェアを実際に動作させて情報を収集する動的解析にも利用できるため，CCC DATAset2009 で示されたハッシュ値を持つマルウェア検体を動的解析し，その結果を示した．

参考文献

[1] Nathan Friess and John Aycock. Black market botnets. MIT Spam Conference, 2008.

[2] Brett Stone-Gross, Marco Cova, Lorenzo Cavallaro, Bob Gilber t, Mar tin Szydlowski, Richard Kemmerer, Chris Kruegel, and Giovanni Vigna. Your botnet is my botnet: Analysis of a botnet takeover, Apr 2009.

[3] 竹森敬祐, 磯原隆将, 三宅優, 西垣正勝. ボットネットおよびボットコードセットの耐性解析. In **マルウェア対策研究人材育成ワークショップ** *2008*, Oct 2008.

[4] Cyber Clean Center. 2008 年 08 月度　サイバークリーンセンター活動実績, Aug 2008. `https://www.ccc.go.jp/report/200808/0808monthly.html`.

[5] CERT Advisory CA-2003-20 W32/Blaster worm. *`http://www.cert.org/advisories/CA-2003-20.html`*, Aug 2003.

[6] Niels Provos, Dean McNamee, Panayiotis Mavrommatis, Ke Wang, and Nagendra Modadugu. The ghost in the browser analysis of web-based malware. In *HotBots'07: Proceedings of the first conference on First Workshop on Hot Topics in Understanding Botnets*, pages 4–4, Berkeley, CA, USA, 2007. USENIX Association.

[7] 水谷 正慶, 武田 圭史, 村井 純. Web 感染型悪性プログラムの分析と検知手法の提案. **セキュアでサステイナブルなインターネットアーキテクチャ論文特集**, J92-B(10):1–12, Sep 2009.

[8] Sun Microsystems, Inc. VirtualBox. `http://www.virtualbox.org/`.

[9] 村上 純一 松木 隆宏. 悪性プログラムの耐解析技術を逆用した活動抑制手法の提案. *Computer Security Symposium 2006*, 2006.

[10] OpenVPN Technologies, Inc. OpenVPN Open Source Project. `http://www.openvpn.net/index.php/open-source.html`.

[11] Martin Roesch. Snort, 12 1998. `http://www.snort.org`.

[12] Junichi Murakami. dumnet.

[13] Bart De Schuymer. ebtables. `http://ebtables.sourceforge.net/`.

[14] MDL. Malware Domain List, 2007. `http://www.malwaredomainlist.com/`.

[15] Ur I.T. Mate Group. hpHosts, 2009. `http://hosts-file.net/`.

[16] Grisoft. AVG Free Edition. `http://free.grisoft.com/`.

[17] Team Cymru Research NFP. Malware Hash Registry. `http://www.team-cymru.org/`.

[18] *Lightweight, Payload-Based Traffic Classification: An Experimental Evaluation.* IEEE, 2008.