様式Ｃ－１９

# 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成２４年５月２５日現在

機関番号：12701
研究種目：若手研究(B)
研究期間：2010～2011
課題番号：22700065
研究課題名（和文）　マルウェア対策導出の自動化に関する研究

研究課題名（英文）　Research on Automation of Malware Response

研究代表者
吉岡　克成（YOSHIOKA　KATSUNARI）
横浜国立大学・環境情報研究院・准教授
研究者番号：60415841

研究成果の概要（和文）：本研究では，マルウェア動的解析技術の解析結果から検知・駆除といった対策を自動的に導出する方法について検討を行った．特に，ネットワークベース検知・ホストベース検知技術，駆除・無効化技術，遠隔検査技術について検討を行い，ネットワークベース検知手法，ホストベース検知手法，遠隔検査手法を提案・実装した．また，駆除・無効化に関する基礎検討を行った．

研究成果の概要（英文）：In this study, we address automated derivation of malware response in three approaches: namely, network-based and host-based detection, remote testing, and removal. We have proposed new methods of network-based detection, host-based detection and remote testing. Also, we have obtained basis for automated generation of removal tool.

交付決定額

（金額単位：円）

| | 直接経費 | 間接経費 | 合　計 |
|---|---|---|---|
| 2010 年度 | 1,000,000 | 300,000 | 1,300,000 |
| 2011 年度 | 600,000 | 180,000 | 780,000 |
| 総　計 | 1,600,000 | 480,000 | 2,080,000 |

研究分野：総合領域
科研費の分科・細目：情報学・計算機システム・ネットワーク
キーワード：マルウェア対策技術

## １．研究開始当初の背景

マルウェア対策は，マルウェアの活動観測・収集，解析，対策導出という一連のプロセスにより実施される．活動観測・収集技術としては，マルウェアからのネットワーク攻撃を観測するハニーポットや定点観測，Web サイト埋込型マルウェアを探索する Web クローリング，P2P ネットワークを流通するマルウェアの観測・収集，不正メールの収集などがある．次に，収集されたマルウェアを詳細に解析するための技術として動的解析や静的解析がある．動的解析は，解析対象のマルウェアを解析環境内で実際に実行し，その挙動を観測・分析する手法であり，コード自体の解析を行う静的解析に比べ自動化が容易であるため，大量化・多様化するマルウェアを解析する手段として注目を集めている．最近では，オンラインで検体の解析依頼を受付け，自動で動的解析を行った後，解析結果をユーザに送信する，マルウェア動的解析オンラインサービスがインターネット上に十以上も存在し(2009 年 10 月時点)，その中には総計百万検体以上の投稿を受けている人気サービスがあるなど，マルウェア動的解析技術の発展は著しい．一方で自動解析により得られる豊富な解析結果をいかに有効利用するかについては，まだ多くの検討がなされていない．マルウェア動的解析の解析結果は，検体のファイルアクセス，レジストリアクセ

ス，起動プロセス，通信，API・システムコール呼出等の詳細なログであるため，有効な対策を導出するためには，専門家による更なる検討や作業を必要とするのが現状であり，即時的かつ効果的な対応の妨げとなっている．

２．研究の目的

機密情報流出やサービス不能攻撃の要因とされるマルウェア(ウイルス・ボットなどの不正プログラム)は，ソースコードの流通や自動生成ツールの出現により近年大量化・多様化している．そのため，マルウェアの活動観測，収集，解析，対策導出技術の自動化を可能な限り進め，専門家の手動作業に強く依存した現状の体制から脱却することが必要である．特に対策導出技術は他の技術に比べて専門家への依存度が大きい．そこで本研究では，研究開発が進む動的解析技術の解析結果から検知・駆除といった対策を自動的に導出する方法について検討を行う．

３．研究の方法

本研究では，マルウェア動的解析の結果を利用し，(1)検知 (2)駆除・無効化(3)遠隔検査，の 3 つの観点でマルウェア対策導出の自動化レベルの向上を目指した．

(1)ネットワークベース検知およびホストベース検知技術の検討
まずネットワークベースの検知手法として，マルウェア動的解析の通信ログから攻撃通信を検出する方法を検討した．自己暗号化が施された攻撃コードを検知するため CPU エミュレータ上で攻撃コードを実行し，その振る舞いに基づき攻撃を検知する手法について検討した．
　次に，ホストベースマルウェア検知については，マルウェアを複数回実行した際に生成されるファイル名，変更・追加されるレジストリ情報，通信先ホストが毎回異なることに着目し，このような挙動のランダム性に基づいてマルウェアを検知する手法について検討した．また，ホストベースのマルウェア検知で用いるシグネチャを自動生成する方法を検討した．近年のマルウェアはパッカーと呼ばれる圧縮・暗号化ツールにより自己暗号化されており，多様なバイナリ表現を持ち得るため，まず，暗号化を解除(復号)し，実際のコード(オリジナルコード)を取得した上で特徴的なバイナリパターンの抽出を試みた．

(2) 駆除・無効化技術の検討
感染中のマルウェアを活動できない状態にするための駆除ツールの自動生成方法を検討した．駆除を行うには，マルウェアが感染状態を維持するために必要な要素(本体ファイル，再起動用レジストリキー，起動プロセス)を動的解析により特定し，これを削除・修正する必要があるが，毎回の動的解析において全ての要素が観測できるわけではない点に注意が必要である．そのため，マルウェアがもつ環境依存性やランダム性の影響を調査した．

(3) 遠隔検査技術の検討
ボットのように外部からの命令通信を受信しそれに応じて活動するマルウェアは，感染時に特定またはランダムに決定されるポートで待ち受け状態となるものが多い．これらのポートに対して試験的に接続を行い，検査用データを送信し，検査対象ホストからの返答を調べることで，マルウェア感染状態のマシンを識別する方法を検討した．

４．研究成果

本研究の成果について(1)検知 (2)駆除・無効化(3)遠隔検査の順に説明する．

(1)ネットワークベース検知およびホストベース検知技術の検討
まずネットワークベースの検知手法として CPU エミュレータを用いた動的検知手法を提案した．通信トラヒックを CPU エミュレータ上で機械語命令として実行することで攻撃コードを検知する手法は既に提案されているが，既存手法では，全てのトラヒックに対して動的検知を適用するために検査コストが高かった．そこで，我々は機械語と思われる箇所を静的に走査し，攻撃コードの候補を検知した場所に限り，動的検知を行う手法を提案した(図 1)．提案手法は，既存手法と同等の精度を保ちつつ，検査速度をおよそ 5 倍にすることに成功した(表 1)．
次にホストベースのマルウェア検知技術として，マルウェアを複数回実行した際に生成されるファイル名，変更・追加されるレジストリ情報，通信先ホストが毎回異なることに着目し，このような挙動のランダム性に基づいてマルウェアを検知する手法を提案した．
　具体的には，表 2 に示すような API の呼び出しをフックすることで観測し，引数を確認することで，挙動のランダム性を判定することとした．
　表 3 と表 4 にマルウェアと正規プログラムにおける，挙動のランダム性に基づく検知結果を示す．表 3 にあるように実験に用いた 1,157 個のマルウェア検体のうち平均して 65.2%において挙動のランダム性が見られた．一方，正規プログラム 1,582 個のうち約 1.6%

において挙動のランダム性が見られた．このことから，挙動のランダム性はプログラムの良悪性を判断する 1 つの基準と成り得る可能性があることが分かった．

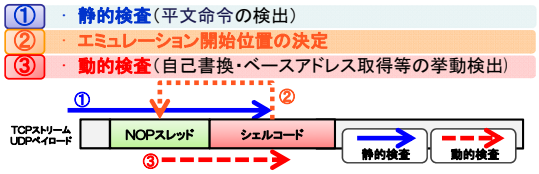


図 1. 静的検査と動的検査の組み合わせによる効率的な攻撃コード検知

表 1．既存手法と提案手法の比較

(a) CCC DATAset 2011 での検知結果

| | 提案手法 | 既存手法 |
|---|---|---|
| 通信データ(PCAP)サイズ[MB] | 298 | |
| 検査対象の合計サイズ[MB] | 185 | |
| 検知数 | 499 | 495 |
| 静的検査回数 | 859847 | |
| 動的検査回数 | 10798341 | 92783230 |
| (割合) | 1 | 8.59 |
| 実行時間[sec] | 36142 | 180323 |
| (割合) | 1 | 4.99 |

表 2. 挙動のランダム性判定に用いる API とその引数リスト

| API | 比較に用いた引数 |
|---|---|
| RegSetValueEx | レジストリエントリ名 |
| RegSetValue | レジストリエントリ名 |
| CreateFile | ファイル名 |
| LZOpenFile | ファイル名 |
| _lcreat | ファイル名 |
| CreateDirectoryEx | ディレクトリ名 |
| CreateDirectory | ディレクトリ名 |
| CopyFileEx | コピー先ファイル名 |
| CopyFile | コピー先ファイル名 |
| LZCopy | コピー先ファイル名 |
| MoveFileEx | 移動先ファイル名 |
| MoveFile | 移動先ファイル名 |
| DnsQuery | 名前解決するドメイン名 |
| gethostbyname | 名前解決するドメイン名 |
| InternetOpenUrl | アクセス先URL |
| HttpOpenRequest | リクエストパス |
| InternetConnect | アクセス先ドメイン名（or IPアドレス） |
| URLDownloadToFile | ダウンロードファイルのURL |

表 3. 挙動のランダム性に基づくマルウェアの検知結果

| | 検知 | 未検知 | 検知率 |
|---|---|---|---|
| 全体 | 754 | 403 | 65.2% |
| レジストリ系API | 314 | 843 | 27.1% |
| ファイル系API | 725 | 432 | 62.7% |
| ネットワーク系API | 541 | 616 | 46.8% |

表 4. 挙動のランダム性に基づく正規プログラム誤検知

| | 検知 | 未検知 | 誤検知率 |
|---|---|---|---|
| 全体 | 9 | 543 | 1.6% |
| レジストリ系API | 2 | 550 | 0.4% |
| ファイル系API | 8 | 544 | 1.4% |
| ネットワーク系API | 3 | 549 | 0.5% |

(2) 駆除・無効化技術の検討

感染中のマルウェアを活動できない状態にするための駆除ツールの自動生成方法を検討した．駆除を行うには，マルウェアが感染状態を維持するために必要な要素（本体ファイル，再起動用レジストリキー，起動プロセス）を動的解析により特定し，これを削除・修正する必要があるが，マルウェアは駆除を防ぐために実行時に毎回異なるファイル名で自身をコピーすることがある．そこで，同一の検体を複数回，動的解析することにより，生成されるファイル名やレジストリを調査した．結果は表 3 に示す通りであり，検体のうち 1,157 検体のうち，62.7%が実行のたびに異なる名前のファイルを生成しており，この点を考慮した自動駆除ツールの生成が必須であることがわかった．

(3) 遠隔検査技術の検討

ボットのように外部からの命令通信を受信しそれに応じて活動するマルウェアは，感染時に特定またはランダムに決定されるポートで待ち受け状態となるものが多い．これらのポートに対して試験的に接続を行い，検査用データを送信し，検査対象ホストからの返答を調べることで，マルウェア感染状態のマシンを識別する方法を検討した．提案手法では，動的解析環境内でマルウェアに感染したホストにポートスキャンを行い，さらにテスト入力データを送信することで得られる返信データに基づき，検知用シグネチャを自動生成する．提案システムの全体図を図 2 に示す．評価実験の結果，実験に用いた全検体 434 検体のうち，約 1/3 にあたる 132 検体についてポート待ち受けが確認され，それらは全て提案手法により正しく検知できることが確認された．また，33 種類の正規サービスに対して検知シグネチャを適用したところ，誤検知がないことを確認した．

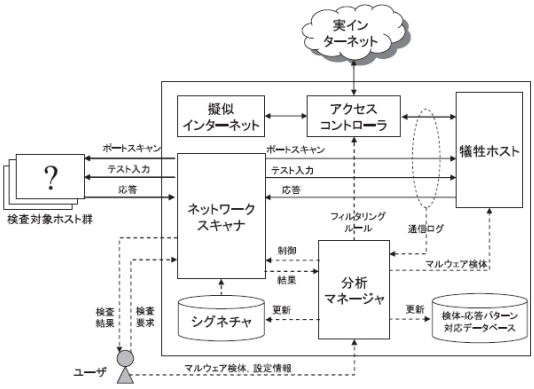


図 2. マルウェア感染ホスト検知シグネチャ生成システム

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計５件）

1.笠間貴弘，織井達憲，吉岡克成，松本勉，″公開型マルウェア動的解析システムに対するデコイ挿入攻撃の脅威，″ Journal of Information Processing, Vol. 52, No. 9, pp. 2761 - 2774, 2011.
2. K. Yoshioka, Y. Hosobuchi, T. Orii, and T. Matsumoto, ″Your Sandbox is Blinded: Impact of Decoy Injection to Public Malware Analysis Systems,″ Journal of Information Processing, Vol. 52, No.3, 1144-1159, 2011.

3. H. C. Kim, T. Orii, K. Yoshioka, D. Inoue, J. Song, M. Eto, J. Shikata, T. Matsumoto, and K. Nakao, ″An Empirical Evaluation of an Unpacking Method Implemented with Dynamic Binary Instrumentation,″ IEICE Trans. Vol. E94D, No. 9, pp. 1778 - 1791, 2011.

4. 吉岡克成，村上洸介，松本勉，″マルウェア感染ホスト検出のためのネットワークスキャン手法と検出用シグネチャの自動生成″ 情報処理学会論文誌 Vol.51, No.9, pp. 1633 - 1644, 2010.

〔学会発表〕（計 16 件）

1. 織井達憲，吉岡克成，四方順司，松本勉，″マルウェア解析の効率化を目指した自己書換え動作の可視化方法，″ 電子情報通信学会暗号と情報セキュリティシンポジウム 2011 CD-ROM 論文集，セッション 3B3-5，2011.

2. 村上洸介，藤井孝好，吉岡克成，松本勉，″リモートエクスプロイト攻撃を効率的に観測可能なマルウェア動的解析手法の提案，″ 情報処理学会コンピュータセキュリティシンポジウム(CSS2011) 論文集 CD-ROM，セッション 3B3-3，2011.

3. 笠間貴弘，吉岡克成，井上大介，松本勉，″実行毎の挙動の差異に基づくマルウェア検知手法の提案，″ 情報処理学会コンピュータセキュリティシンポジウム(CSS2011) 論文集 CD-ROM，セッション 3B3-4，2011.

4. K. Yoshioka, Y. Hosobuchi, T. Orii, and T. Matsumoto, ″Vulnerability in Public Malware Sandbox Analysis Systems,″ IEEE 10th Annual International Symposium on Applications and the Internet, SAINT2010, pp 265 - 268, 2010.

5. T. Kasama, K. Yoshioka, T. Matsumoto, M. Yamagata, M. Eto, D. Inoue, and K. Nakao, ″Malware Sandbox Analysis with Automatic Collection of Server Responses using Dummy Client,″ Proc. 5th Joint Workshop Workshop on Information Security, JWIS 2010, pp. 67 - 81, 2010.

6. 藤井孝好，吉岡克成，四方順司，松本勉，″エミュレーションに基づくシェルコード検知手法の改善，″ マルウェア対策研究人材育成ワークショップ 2010 (MWS2010) 論文集 CD-ROM，セッション 2A-3，2010.

7. 神保千晶，吉岡克成，四方順司，松本勉，衛藤将史，井上大介，中尾康二，″CPU エミュレータと Dynamic Binary Instrumentation の併用によるシェルコード動的分析手法の提案，″ 電子情報通信学会技術報告 ICSS2010-54，pp.59-64，2010.

６．研究組織
(1)研究代表者
吉岡　克成　（YOSHIOKA KATSUNARI）
横浜国立大学・環境情報研究院・准教授
研究者番号：60415841

(2)研究分担者
なし

(3)連携研究者
なし