研究用データセット「動的活動観測 2014」の検討

寺田真敏[*1]　青木　翔[*1]　楠美淳弥[*2]
重本倫宏[*1]　萩原健太[*3]

マルウェア検体の静的／動的解析では，指令サーバ接続，情報窃取，バックドアなどの機能の存在や挙動把握に重点が置かれ，これら機能のいずれを使ったのか，どの順番で使ったのかなど，攻撃者の行動という視点で把握や解析することはなかった．多くの場合，攻撃者の行動＝マルウェアの挙動という想定の下，静的／動的解析によって対応してきたというのが実情である．しかし，組織内ネットワークへの侵害活動においては，攻撃者の存在を意識する必要がある．本稿では，「攻撃者行動視点で脅威の特徴情報」を明らかにしていくために，攻撃者の行動を記録する研究用データセットの作成について報告する．

Feasibility Study of Research Data Set "Behavior Observable System 2014"

Masato Terada, Sho Aoki, Junya Kusumi, Tomohiro Shigemoto and Kenta Hagihara

Under the static / dynamic analysis of malware, mainly it focuses on the functions and behaviors of malware itself such as C&C server connection, information leak and backdoor. The analysis of malware does not include the viewpoint of actions of cyber attack actors. But under the targeted attack such as APT, we should focus on the actions of cyber attack actor, too. In this paper, firstly we will describe purpose of the research data set of the targeted attack age. Secondly, we will introduce our research data set "BOS_2014" for the countermeasures of targeted attack age.

1. はじめに

マルウェアを用いた攻撃手法の多様化と巧妙化は進んでおり，活動形態にも大きな変化がみられる．1999 年頃から電子メールを介したマルウェアの受動型感染が始まった．2001 年頃からはネットワーク型ワーム，2004 年頃からは遠隔操作可能なボットが流布した．その感染形態は，感染対象のホストに対してマルウェア自身が攻撃コードを送信する能動型感染が主流であった．2008 年頃からは，ブラウザが利用するプラグインやアプリケーションの脆弱性を利用して，マルウェアをダウンロードして実行する攻撃手法，ドライブバイダウンロード(drive-by-download)を用いた Web 感染型マルウェアが流布した．2011 年に入ると，電子メールと遠隔操作ツール(Remote Access Trojan/Remote Administration Tool)とを組合わせた組織内ネットワークへの侵害活動が台頭し始めた．2012 年からは，Web サイト群に仕掛けを蔵置し，組織内ネットワークへの侵害活動につなげる Web サイト待ち伏せ攻撃(Watering Hole Attack)が報告されるようになった．

このような大きな活動形態の変化と共に，攻撃活動の違いに視点を置いた CCC DATAset 2008～2011，IIJ MITF DATAset 2012，D3M 2010～2013，NICTER Darknet Dataset 2013，解析手法の違いに視点を置いた FFRI Dataset 2013，PRACTICE Dataset 2013 など，MWS で提供される研究用データセットも多様化してきている．

本研究の目的は，活動形態の変化を踏まえた研究用データセットを作成し，マルウェア対策の研究につなげることにある．そこで，本稿では，2011 年に入ってから台頭し始めた電子メールと遠隔操作ツールとを組合わせた組織内ネットワークへの侵害活動を想定した研究用データセットを作成したので報告する．

*1　(株)日立製作所, Hitachi Ltd.
*2　(株)日立システムズ, Hitachi Systems, Ltd.
*3　トレンドマイクロ(株), Trend Micro Incorporated.

2. 関連研究

2.1 MWS 研究用データセット

MWS では，これまで，(1) サイバークリーンセンターのハニーポットで収集した研究用データセット CCC Dataset 2008～2013, (2) 独立行政法人 情報通信研究機構が所有する小規模攻撃再現テストベッドでのマルウェア検体の動作記録 MARS for MWS 2008～2010，(3) Web 感染型マルウェアの観測データ D3M 2010～2014，(4) ローインタラクション型ハニーポットで収集したマルウェア研究用データセット IIJ MITF DATAset 2012，(5) マルウェア動的解析データ FFRI Dataset 2013～2014，(6) 総務省「国際連携によるサイバー攻撃予知・即応に関する実証実験」プロジェクトで得られたマルウェア長期観測データ PRACTICE Dataset 2013，(7) 独立行政法人 情報通信研究機構が運用する NICTER にて観測したダークネットパケットデータ NICTER Darknet Dataset 2013～2014 が提供されてきた[1].

2.2 海外で提供されている研究用データセット

海外では，1999 年に米リンカーン研究所が開発した "1999 DARPA Intrusion Detection Evaluation Data Set"がある[2]．このデータは，侵入検知システムの有効性を確認するためのトラフィック評価データで，侵入検知技術の客観的な評価を行なうための評価データとしても活用されてきた．この他に，サイバー防御演習時のデータセット the 2009 Inter-Service Academy Cyber Defense Exercise datasets[3]，大規模セキュリティ関連データの収集と分析をもとに，より良いデータとナレッジの共有を図る BADGERS2011[4]，ネットワーク運用データをレポジトリとして蓄積し，インフラ防護と脅威評価に活用する PREDICT[5]，広域ネットワークの状況を分析し，幾つかのタイプのデータセットを提供する CAIDA[6]などが研究用データセットとして提供されている．

2.3 サイバー攻撃対策モデル

(1) 攻撃活動進行段階モデル

本稿で取り上げる電子メールと遠隔操作ツールとを組合わせた組織内ネットワークへの侵害活動は，攻撃対象となる組織に合う手法を選択し(標的型)，組織内ネットワークを活動基点とした(潜伏型)侵害活動と呼ばれている．その対策のために，進行段階がモデル化されている[7]. 文献[8]では，米国空軍の軍事コンセプトである Kill Chain(F2T2EA)をサイバーに応用し，対策視点でモデル化した Cyber Kill Chain を提案している．このモデルは，Reconnaissance(偵察)，Weaponization(武器化)，Delivery(配送)，Exploitation(攻撃)，Installation(インストール)，Command and Control(C2)(遠隔制御)，Actions on Objectives(実行)の 7 段階から成る．また，初期段階から対策として，配送段階での検知，武器化段階以前の分析と，攻撃者の意図，攻撃者のパターン，行動，TTP(Tactics, Techniques and Procedures：戦術，技術及び手順)を明らかにする攻撃活動分析(Campaign Analysis)の必要性を示している．

(2) 攻撃活動全般の構造化

進行段階のモデル化と共に，攻撃活動分析のための情報活用が検討されている．米 MITRE 社が開発した，脅威情報構造化記述形式 STIX(Structured Threat Information eXpression)[9]は，サイバー攻撃活動の攻撃から対策までを記録するための XML 仕様である．2010 年に，US-CERT と CERT/CC 間での脅威情報の交換から検討が始まり，2013 年 4 月に Ver1.0 がリリースされた．この STIX では，サイバー攻撃で狙っているソフトウェア，システムや設定の弱点，攻撃を検知するための事象だけではなく，攻撃者の行動や手口，サイバー攻撃に関与している人／組織などを関係付けていくためのサイバー攻撃活動の構造化が試みられている(図 1).

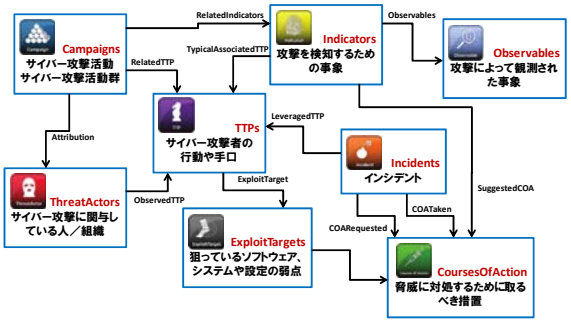

図 1：STIX による脅威情報構造化

3. 研究用データセット「動的活動観測 2014」

本章では，研究用データセット「動的活動観測 2014(BOS_2014)」の概要について述べる．

3.1 動的活動観測

(1) 目的

これまで，マルウェア検体の静的／動的解析では，マルウェアの挙動に着目したものであった．例えば，指令サーバ接続，情報窃取，バックドアなどの機能の存在や挙動把握に重点が置かれ，これら機能のいずれを使ったのか，どの順番で使ったのかなど，攻撃者の行動という視点で把握や解析することはなかった．多くの場合，攻撃者の行動＝マルウェアの挙動という想定の下，静的／動的解析によって対応してきた．

しかし，組織内ネットワークへの侵害活動においては，攻撃者の存在を意識する必要がある．そこで，動的活動観測では，マルウェアの挙動に加えて，どのような操作をしたのか，どのようなファイルにアクセスしたのかなど攻撃者の行動と組合わせていくことで，攻撃者行動視点で脅威を特徴付けできる研究用データセットの作成を試みる．

(2) 観測環境

動的活動観測環境は，実インターネット上の攻撃者が試みる組織内ネットワークへの侵害活動を観測するシステムで，システムそのものが組織内ネットワークを模擬している(図 2)．クライアントは，電子メールに添付された検体を実行する PC であり，プロキシ経由／プロキシ経由なしのいずれかの形態で，インターネットとの接続性を持つことができる．

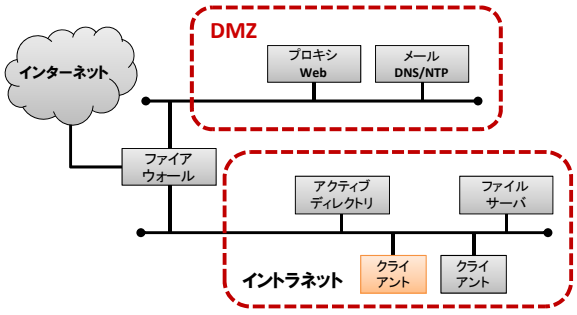

図 2：動的活動観測環境の概要図

(3) データセット構成

動的活動観測 2014 で作成したデータセットは，マルウェア検体，通信観測データ，プロセス観測データの 3 つである．

(a) マルウェア検体

動的活動観測に使用したマルウェア検体のハッシュ値をテキスト形式で記載したファイルである．

(b) 通信観測データ

マルウェア検体を実行した際の通信のフルキャプチャデータであり，攻撃者の行動に関する解析が可能である．

(c) プロセス観測データ

マルウェア検体を実行したクライアントでのプロセスの稼働状況を記録したデータであり，攻撃者の行動に関する解析が可能である．

3.2 研究用データセットの特徴

本節では，動的活動観測 2014(BOS_2014)で作成した研究用データセットの特徴ついて述べる．なお，攻撃者の行動視点に注目するために，操作者を併記した時系列イベントの形で観測事象を記載する．

(1) Case11

マルウェア検体 Case11 は，exe ファイルであり，Microsoft Word ファイルのアイコンで偽装されていた．実行後の観測事象を表 1 に示す．

(2) Case21

マルウェア検体 Case21 は，exe ファイルであり，フォルダアイコンで偽装されていた．実行後の観測事象を表 2 に示す．

表 1：マルウェア検体 Case11 での観測事象

#	時刻	操作者	観測事象
1	11:06	O	C:¥data 直下にて，マルウェア検体(exe ファイル)をダブルクリック実行
2	11:06	M	C:¥windows¥FlashHelpx64.exe をドロップ
3	11:06	M	自動起動を目的としたレジストリ改変
4	11:06	M	**.**.160.125 との接続を確立
5	11:15	A	スクリーンキャプチャの取得
6	11:15	A	端末基本情報の取得
7	11:15	A	スクリーンキャプチャの取得
8	11:40	A	スクリーンキャプチャの取得
9	11:40	A	C:¥RECYCLER¥に a.exe をアップロード
10	11:41	A	cmd.exe からコマンド「a /stext aaa.txt」経由で a.exe を実行
11	11:41	A	コマンド「del a.*」で a.exe を削除
12	11:41	M	**.**.160.125 との接続を解除
13	11:41	M	プロセスが終了

[操作者] O：観測者，M：マルウェア検体，A：攻撃者

3.3 MWS Dataset 2014 との関係性

本節では，MWS Dataset 2014 のデータセット D3M と，動的活動観測 2014 との関係性について，検体ハッシュ値と接続先サイトを基点に，トレンドマイクロのビックデータとを組合せて調査した結果について述べる．

- D3M 2010～2014 と動的活動観測 2014 との間には，検体ハッシュ値，検体ハッシュ値から導かれる接続先サイトに関して直接的な関係性を見出すことはできなかった．
- Case11 観測時点では，D3M との関係性を見出すことはできなかったが，2014 年 8 月中旬に再調査した時点では，D3M 2014 の検体 SHA1:B76A94A81A*(TROJ_VILSEL. BK)の接続先サイトの IP アドレスと，マルウェア検体 Case11(BKDR_POISON. BWB)の接続先サイトの IP アドレスは同一の管理組織の配下で管理されていた(図 3)．脅威情報等の関係性を可視化するために Maltego[10]を用いて作成した図 3 の赤丸が検体ハッシュ値，緑丸が接続先サイト，青四角が IP アドレス管理組織である．

表 2：マルウェア検体 Case21 での観測事象

#	時刻	操作者	観測事象
1	15:06	O	デスクトップ上でマルウェア検体(exe ファイル)をダブルクリック実行
2	15:06	M	www.google**.com/windowsxp/Snews.asp に対して HTTP POST 要求を送信
3	15:06	M	HTTP POST 応答「HTTP/1.0 200 OK」を受信，以降継続
4	15:07	M	検体のプロセス lplus.exe が起動した cmd.exe で，コマンド「net start」，「tasklist」，「systeminfo」，「netstat -an」などを実行
5	16:29	A	検体のプロセス lplus.exe が起動した cmd.exe で，「arp -a」を実行
6	16:38	A	検体のプロセス lplus.exe が起動した cmd.exe で，「at ¥¥I160V01」を実行
7	17:06	A	検体のプロセス lplus.exe が起動した cmd.exe で，「net group "domain computers" /domain」を実行
8	17:19	A	検体のプロセス lplus.exe が起動した cmd.exe で，「ping 10.2.149.1」を実行(10.2.149.1 は共有フォルダを提供するファイルサーバ)
9	17:49	A	C:¥WINDOWS¥Debug¥Rar.exe で¥¥10.2.149.1¥public¥mail ¥testMail にアクセス
10	17:49	A	C:¥WINDOWS¥Debug ¥Rar.exe で¥¥10.2.149.1 ¥public¥012 営業本部¥顧客先アドレス**.zip にアクセス
11	17:55	A	検体のプロセス lplus.exe が起動した cmd.exe で，「ftp -s:c¥windows¥debug¥ftpo.txt」を実行

[操作者] O：観測者，M：マルウェア検体，A：攻撃者

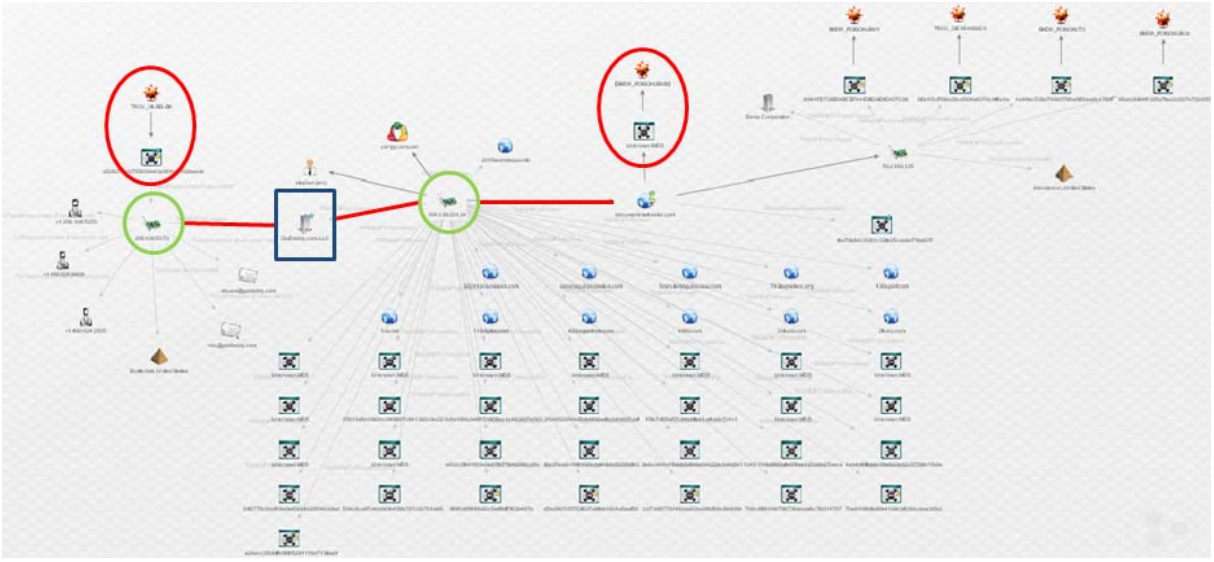

図 3：Maltego によるマルウェア検体 Case11(BKDR_POISON.BWB)[右赤丸]と
D3M 検体(SHA1:B76A94A81A*,TROJ_VILSEL.BK)[左赤丸]との関係図(2014 年 8 月時点)

4. おわりに

本稿では，組織内ネットワークへの侵害活動を想定した研究用データセット「動的活動観測 2014(BOS_2014)」について報告した．

研究用データセット「動的活動観測 2014(BOS_2014)」は，どのような操作をしたのか，どのようなファイルにアクセスしたのかなど，攻撃者行動視点で脅威を特徴付けできるようにするため，マルウェア検体(ハッシュ値)，通信観測データ，プロセス観測データから構成している．

今後の課題は，研究用データセット「動的活動観測」の拡充，サイバー攻撃対策への活用，活動形態の変化を踏まえた新たな研究用データセットの作成などを検討していきたいと考えている．

謝辞

本研究は総務省実証事業「サイバー攻撃解析・防御モデル実践演習の実証実験の請負」で実施したものです．本研究を進めるにあたって有益な助言と協力を頂いた北陸 StarBED 技術センターならびに関係者各位に深く感謝致します．

参考文献

1) MWS2014 実行委員会，研究用データセット MWS 2014 Datasets について，
http://www.iwsec.org/mws/2014/about.html
2) MIT Lincoln Laboratory, DARPA Intrusion D etection Evaluation Data Sets,
http://www.ll.mit.edu/mission/communications/ist/corpora/ideval/index.html
3) B. Sangster, et al.: Toward Instrumenting N etwork Warfare Competitions to Generate Label ed Datasets, 18th USENIX Security Symposium CSET'09 (2009 年 8 月)
4) BADGERS2011: Building Analysis Datasets and Gathering Experience Returns for Security,
http://iseclab.org/badgers2011/ (2011 年 4 月)
5) PREDICT: the Protected Repository for the Defense of Infrastructure Against Cyber Threats,
https://www.predict.org/
6) CAIDA: The Cooperative Association for Internet Data Analysis, http://www.caida.org/home/
7) IPA：『新しいタイプの攻撃』の対策に向けた設計・運用ガイド(2011 年 11 月),
http://www.ipa.go.jp/security/vuln/newattack.html
8) Eric M. Hutchins, et.al.：Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains (2011 年 3 月)
9) STIX, http://stix.mitre.org/
10) Maltego,
https://www.paterva.com/web6/products/maltego.php