5ZC-06

BOS2 データセットにおける接続先関連情報取得に関する検討

建部大聖[†1]　小林尚生[†1]　藤井翔太[†2]　青木翔[†2]　佐藤隆行[†2]　寺田真敏[†1†2]
東京電機大学[†1]　日立製作所[†2]

1. はじめに

マルウェアやサイバー攻撃を動的に観測したデータセットは，脅威の実態を踏まえた対策を進める上で有用である．本研究は，MWS 活動目標のひとつである研究用標準データセットを策定するという考えの下，サイバー攻撃の動的観測に係る BOS データセットの後継としての BOS2 データセットを作成し，提供することを目的としている．本稿では，BOS2 データセットにおいて接続先のスナップショット取得，具体的には，DNS や WHOIS 等の接続先に係る周辺情報について，取得する項目を整理し，作成した取得ツールについて報告する．さらに，取得ツールの予備実験から接続先に係る周辺情報の活用について考察する．

2. 関連研究

本章では研究用データセットである MWS データセット，BOS データセットについて述べる．

2.1 MWS データセット

MWS では，2008 年以降，マルウェア分析技術を研究／評価するための適切な素材として，(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，(2)Web 感染型マルウェアの観測データ D3M，(3)マルウェア動的解析データ FFRI Dataset，(4)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，(5)国立研究開発法人 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 Dataset，(6)マルウェア解析環境で得られたマルウェアの挙動ログ Soliton Dataset などを提供してきた[1].

2.2 BOS データセット

動的活動観測 BOS(Behavior Observable System)データセットは，2014 年から 2019 年の MWS データセットのひとつである．目的は，攻撃者のアトリビューションの一部として，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せていくことで，攻撃者行動視点で脅威の特徴付けを試みることにある．本稿で取り上げる BOS2 は，BOS の後継であり，その仕様について検討状況にある[2]

3. BOS2 データセットの構成

本章では BOS2 の構成と，BOS2 で新たに追加する接続先の間接的な周辺情報である DNS, WHOIS についてのデータ項目に述べる．

3.1 構成

BOS2 データセットの構成を図 1 に示す．動的活動観測では，マルウェアのハッシュ値，プロセス，イベントログ，通信記録など，マルウェアならびに攻撃者の行動を観測した記録をデータ項目とする．観測周辺情報は，マルウェアならびに攻撃者の行動に関連するデータ項目であり，不正接続先などのサーバから直接取得する必要のある情報グループ(以下，直接関連情報)と，DNS，WHOIS などから得られる不正接続先自身に関する情報グループ(以下，間接関連情報)に分かれる．直接関連情報としては，ページのスクリーンショット，ソースファイル，SSL 証明書等が該当する．間接関連情報は，いわゆる公開情報源に該当するもので DNS，WHOIS などから得られる不正接続先自身に関する情報と，VirusTotal などのレピュテーションサイトに掲載される評価情報とがある．

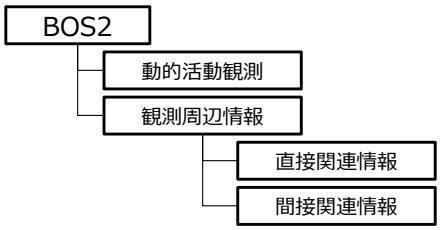

図 1：BOS2 データセットの構成

3.2 間接関連情報

BOS2 データセットでは，間接関連情報として，動的活動観測の事後分析に必要となる DNS, WHOIS から得られる不正接続先自身に関する情報を同梱することを想定している．本節では，BOS2 データセットに同梱する間接関連情報である DNS と WHOIS のデータ項目について述べる．

(1)DNS データ項目

BOS2 データセットに同梱する DNS データ項目を表 1 に示す．三原[3]らは C2 サーバの検知に MX レコードを活用しており，表 1 に示すデータ項目は，C2 サーバの稼働状況確認など，動的活動観測の事後分析に活用できるものと考えている．

(2)WHOIS データ項目

BOS2 データセットに同梱する WHOIS データ項目を表 2 に示す．久山[4]らは，C2 サーバの検知にドメインの登録期間及び登録された連絡先メールアドレスを活用している．表 2 に示すデータ項目も C2 サーバの状況把握など，動的活動観測の事後分析に活用できるものと考えている．

Feasibility Study for Obtaining Connect Related Information in the BOS2 Dataset
†1 TAISEI KEMPE, NAOKI KOBAYASHI and MASATO TERADA, Tokyo Denki University
†2 SHOTA FUJII, SHO AOKI and TAKAYUKI SATO, Hitachi Ltd.

表 1：DNS データ項目

データ項目
● A：ホスト名に対応する IPv4 のアドレス ● AAAA：ホスト名に対応する IPv6 のアドレス ● MX：ドメイン宛てのメールアドレスの配送先サーバ ● PTR：IP アドレスに対応するホスト名 ● SOA：ドメインのゾーン管理のための情報や設定 ● NS：権威サーバの情報 ● TXT：ホスト名に関連付けるテキスト情報 ● CNAME：ドメインやホスト名の別名 など

表 2：WHOIS データ項目

データ項目
● ドメイン名の有効期限 ● ドメイン名の登録者に関する情報 ● 登録に関する問題が発生した際の連絡先 ● AS に関する情報 ● IP の属しているネットワーク範囲 など

4. 間接関連情報取得ツール

本章では，BOS2 データセットの間接関連情報を取得するためのツールについて，その仕様概要と取得ツールの予備実験について述べる．

4.1 概要

間接関連情報取得ツールは，動的活動観測開始時，観測中，観測後のいずれのタイミングでも利用できるようコマンドライン型のインタフェースという形で作成している．入力の引数は IP アドレス，ドメイン，URL のいずれかである．出力は，図 2 に示すフォルダ構成をとる．間接関連情報を格納するフォルダには，次のように構成した DNS 及び WHOIS データ項目格納ファイルを格納する．

- 同一対象の繰り返し調査を考慮し各調査対象フォルダおよびファイルには一意となる名称を付与する．
- メタ情報とデータの 2 つのフィールド構成とする
- メタ情報フィールドは，データを記録した年月日を示す date フィールド，入力データの種別(IP アドレス，ドメイン)を記録する type フィールドを持つ
- データフィールドにはメタ情報の type フィールドに応じて取得した情報を記録する

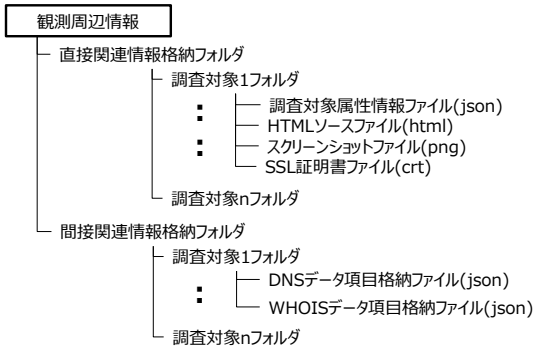

図 2：BOS2 データセットのフォルダ・ファイル構成

4.2 ツールを用いた予備実験

ツールを用いた予備実験では，間接関連情報取得ツールの動作確認に加えて，新たに追加したデータ項目の活用について検討した．予備実験で注目したデータ項目は，WHOIS の「ドメイン名の登録者に関する情報」にある組織名で，正規に運営しているドメイン(以下，正規ドメイン) 32 件と，過去に悪性と判定されたドメイン(以下，悪性ドメイン) 12 件が調査対象である．なお，調査対象となる正規ドメインは業種の異なる分野から選択し，悪性ドメインは，公的機関から配布組織を限定して配信された不正接続先を対象とした．調査結果である表 3 の#1 は登録者情報のあるドメインを分母とし，表 3 の#2，#3 については組織名の公開数を分母としている．

表 3：正規ドメインと悪性ドメインの登録者情報の調査

#	項目	正規ドメイン	悪性ドメイン
1	組織名の公開数	30 / 32 (93%)	7 / 8 (87%)
2	組織名が企業名	28 / 30 (93%)	1 / 7 (14%)
3	組織名が代行公開	2 / 30 (6%)	6 / 7 (85%)

4.3 考察

調査結果から，正規ドメインでは，一部で組織名の非公開，代行公開があるものの，企業名を公開している割合が高い．特に差異がある項目は，#2：組織名が企業名，#3：組織名が代行公開で，これらの項目については，不正接続先の判定に利活用できるのではないかと考える．

5. おわりに

本稿では，BOS2 データセットに新たに追加する観測周辺情報のうち，DNS，WHOIS などから得られる間接関連情報の概要ならびに，その取得ツールについて報告した．さらに，ツールの予備実験を通して，新たに追加したデータ項目の活用の可能性を示した．今後は，BOS2 データセットのリリースに向け，観測周辺情報取得ツールの作成改善ならびに，観測周辺情報を活用した分析を進めていく予定である．

参考文献

[1] 寺田真敏，他: マルウェア対策のための研究用データセット MWS Datasets ～コミュニティへの貢献とその課題～，情報処理学会, Vol.2020-IFAT-139 No.8, pp.1-6(2020).
[2] 藤井翔太ほか：BOS2 データセットに向けた検討，コンピュータセキュリティシンポジウム 2023 論文集, pp.878-885(2023).
[3] 三原元ほか：数量化理論と攻撃データ(CCCDataset2009)を利用したボットネットの C&C サーバ特定手法の提案と評価，情報処理学会論文誌, Vol51, No.9, pp.1579-1590(2010).
[4] 久山真宏ほか：攻撃者に察知されにくい情報を用いた C&C サーバの検知手法の提案と評価，情報処理学会論文誌, Vol58, No9, pp.1410-1418(2017).