1W-05

悪性 Web サイト解析に向けた通信データの収集のためのサイト探索手法

永井 達也† 神薗 雅紀‡ 白石 善明† 毛利 公美†† 高野 泰洋† 星澤 裕二‡ 森井 昌克†
神戸大学† PwC サイバーサービス‡ 岐阜大学††

1 はじめに

悪性 Web サイトによる被害が問題となっている．Drive-by Download 攻撃のような悪性 Web サイトによる攻撃を防御したり，その被害を早期に復旧させるためには，そのような攻撃のデータを収集し，技術の高度化を図ることになる．アンチウイルスベンダはユーザのクライアント端末に，ISP は網内の各所にセンサー機能を配置することで，不正なアクセスを検知したら直ちに通信データを収集し，解析することができる[1][2]．センサーインフラの所有者はこのときの通信データを研究等に利活用できる．しかしながら，アンチウイルスベンダや ISP のセンサーインフラはその加入者による通信とそれに関連するデータは，そのまま研究用データセットとして他者には供出しにくい．センサーインフラを持っていなければ，悪性 Web サイトを探索したうえで通信データを収集することになる．

公開されている情報をシードとして悪性 Web サイトの探索および通信データの収集ができれば，それを研究用データセットとして広く利活用できるようになる．本研究では悪性 Web サイトの研究に資する通信データの収集に向けて，センサーインフラを必要としない悪性 Web サイトの発見と，通信データを採取する方法の確立を目標としている．本稿ではその初期的段階として悪性 Web サイト探索のためのシードを検討している．アンチウイルスベンダの発行する解析レポートに掲載されている悪性 Web サイトを手がかりに，そのサイトが属する AS を探索するというアプローチをとる．その結果，悪性 Web サイト探索のシードとして解析レポートを用いると，探索した Web サイトのうち約 8%が悪性であったと判定された．

2 従来のアプローチ

Malware Domain List[3]のように悪性 Web サイトをブラックリスト化して共有する取り組みが行われている．

通信データは Web サイトをクロールして収集する．しかし，悪性 Web サイトの通信データが収集できると期待してブラックリストに掲載されているサイトをクロールしても，以下の事由により期待する通信データを容易に採取できない．

1. 悪性ではない Web サイトの情報が含まれる
2. ブラックリストに掲載された時点でその悪性 Web サイトが存在しない

これらはブラックリストサイトの多くは悪性 Web サイトを登録するための認証が無く，誰でも登録できるようになっていること，そして，攻撃者は悪性 Web サイトの URL を短時間で変更させることの 2 点が原因としてあげられる．

3 本研究のアプローチ

2 章に述べたようにブラックリストには悪性ではない Web サイトの情報が載せられていることがある．このような情報をシードとして使うと探索精度が下がってしまう．そこで，本研究ではアンチウイルスベンダの発行する解析レポートをシードとして，それに掲載されている悪性 Web サイトを手がかりとする．悪性 Web サイトが一部の AS に集中しているという報告[4]があることから，解析レポートが示す既知の悪性 Web サイトが属する AS を探索対象とする．

図 1 に悪性 Web サイト探索のアプローチを示す．アンチウイルスベンダは悪性 Web サイトの解析を十分に行ったのち，その脅威情報を解析レポートとして発行している．つまり，解析レポートに記載されている情報は悪性 Web サイトのものであることが確認されている．確かな情報源をシードとすることで探索精度の向上が期待できる．

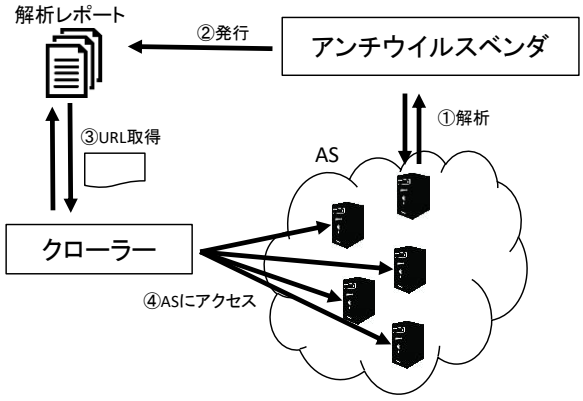

図 1:悪性 Web サイト探索のアプローチ

Searching websites for collecting communication data to analyze malicious websites

†Tatsuya NAGAI ‡Masaki KAMIZONO
†Yoshiaki SHIRAISHI ††Masami MOHRI
†Yasuhiro TAKANO ‡Yuji HOSHIZAWA †Masakatu MORII

†Kobe University
‡PwC Cyber Services
††Gifu University

4 提案手法：解析レポートをシードとした悪性 Web サイトの探索

3 章のアプローチを具体的に以下の手順で実行する．

Step1) 解析レポートから既知の悪性 Web サイトの URL を抜き出す．

Step2) 悪性 Web サイトの URL から，そのサイトが属する AS 番号を抜き出す．

Step3) AS に対応する IP アドレス群を取得し，アクセスする．

Step4) アクセスしたサイトに対して悪性 Web サイトであるかどうかの判定を行う．

5 評価実験

5.1 実験概要

解析レポートをシードとしたときの悪性 Web サイトの探索精度を確認するために，4 章で述べた探索手法を表 1 の環境で実装した．

Step1 の解析レポートには MalwarebytesLab が 2016 年 12 月 2 日に発行したもの[5]と IIJ-SECT が 2016 年 10 月 17 日に発行したもの[4]を利用した．この二つの解析レポートから探索を行い，悪性 Web サイトの数を集計する．Step4 の悪性 Web サイトかどうかの判定に VirusTotal[6]を利用した．Web サイトの IP アドレスを VirusTotal に送ると，過去のホスト名や関連する URL の検査情報が得られる．これらの情報からアクセスした Web サイトが過去に悪性だと判定されていたかどうかを確認する．

5.2 実験結果

2016 年 12 月 14 日～2017 年 1 月 10 日の間で，MalwarebytesLab の解析レポートに記載されていた AS 番号（AS X とする）の IP アドレス 3700 個，IIJ-SECT の解析レポートに記載されていた AS 番号（AS Y とする）の IP アドレス 415 個にアクセスを行い，Web サイトであった IP アドレスに対して悪性であるかどうかの判定を行った．

悪性 Web サイト探索の集計結果を表 1 に示す．AS X，AS Y の両方とも，Web サイトであった IP アドレスのうち約 8%が悪性であった．そのうちの，12 月 15 日にアクセスした AS X に含まれるある IP アドレスはホスト名を変えながら頻繁に悪性 Web サイトになっていることが確認できた．この現象は解析レポート発行日である 12 月 2 日以降も続いていた．

5.3 考察

解析レポートから得られた AS に存在する Web サイトのうち約 8%が悪性 Web サイトであった．GoogleSafeBrowsing のレポートでは『1 日に数十億件の URL を調査して，安全ではないウェブサイトが新たに数千件も見つかり』と報告している[7]．

表 1: Web クローラーの実装環境

OS	Windows7 Home Premium
ブラウザ	HtmlUnit 2.23
悪性判定	VirusTotal
収集期間	2016/12/14～2017/1/10

表 2: 解析レポートをシードとした Web サイト探索の結果

	AS X	AS Y
良性だった Web サイト	1176	156
悪性だった Web サイト	100	12
Web サイト合計	1276	168

単純な比較はできないものの，本研究のアプローチで発見した悪性 Web サイトの割合は低いものではないと考えられる．

6 まとめ

本稿では悪性 Web サイトの通信データの収集に向けて，その探索のためのシードとしてアンチウイルスベンダが公開している解析レポートを利用する手法を検討した．実験により，ある 2 つの AS に対して，Web サイトのうち約 8%が悪性であったことが確認できた．本論文での探索手法をもとにして，出現してから間もない悪性 Web サイトを見つける方法を考えていく予定である．

参考文献

[1] 八木毅, 谷本直人, 浜田雅樹, 伊藤光恭. (2009). プロバイダによる Web サイトへのマルウェア配布防御方式. 電子情報通信学会技術研究報告. IN, 情報ネットワーク, 109(119), 55-60.

[2] 針生剛男, 秋山満昭, 青木一史, 八木毅, 岩村誠, 倉上弘. (2012). 進化するマルウェア等によるサイバー攻撃の検知・解析・対策技術. NTT 技術ジャーナル, 13-17.

[3] Malware Domain List, https://www.malwaredomainlist.com/mdl.php

[4] IIJ-SECT, "Rig Exploit Kit 観測数の拡大に関する注意喚起", https://sect.iij.ad.jp/d/2016/10/178746.html

[5] Malwarebytes Lab, "Tor Browser zero-day strikes again", https://blog.malwarebytes.com/threat-analysis/2016/11/tor-browser-zero-day-strikes-again/

[6] VirusTotal, https://www.virustotal.com/ ja/

[7] Google Safe Browsing, "セーフ ブラウジング - 透明性レポート", https://www.google.com/transparencyreport/safebrowsing/?hl=ja