1ZA-06

悪性Webサイトの探索による モバイル向けブラックリスト構築手法の評価

石原 聖† 佐藤 将也† 山内 利宏†
†岡山大学大学院自然科学研究科

1 はじめに

悪性 Web サイトへアクセスしたモバイル端末の利用者数が増加傾向にある．モバイル端末に対する悪性 Web サイトによる攻撃の 1 つに，利用者の意図しない Web サイトへ誘導する攻撃が存在する．この攻撃では，遷移元サイト，中継サイト，および攻撃サイトといった複数の悪性 Web サイトが利用される．

そこで，我々は，悪性 Web サイトを探索し，モバイル向けのブラックリストを構築する手法を提案した [1]．本稿では，文献 [1] の手法について，悪性 Web サイトの発見数と構築したブラックリストを用いた悪性 Web サイトの検知率について評価した結果を述べる．

2 モバイル向けブラックリストの構築手法

2.1 考え方

提案手法は，探索により発見した悪性 Web サイトの URL，および表 1 に示すキーワードをブラックリストに登録する．既知の悪性 Web サイトについて，類似するドメインには共通の悪性 Web コンテンツが配置される可能性が高いという特徴がある [2]．このため，遷移元サイトの HTML ファイルから遷移の起点となるファイル名とそのファイルを提供する FQDN を抽出する．また，文献 [3] より，中継サイトの URL は，指定された URL とランダムに作成された文字列から作成される場合がある．さらに，攻撃サイトの URL は，利用者の端末情報を含む場合がある．このように URL が変化するため，URL のみでは悪性 Web サイトへのアクセスを検知できない可能性がある．このため，中継サイトと攻撃サイトの URL から FQDN を抽出する．

2.2 提案手法の処理流れ

提案手法におけるブラックリスト作成の処理流れを図 1 に示し，以下で説明する．

(1) クローラを用いて Web 空間から Web コンテンツとして HTML ファイルを収集
(2) クロール先の URL と HTML ファイルを保存

表 1 抽出するキーワード

対象	抽出するキーワード
遷移元サイトの HTML ファイル	遷移の起点となるファイル名
	遷移の起点となるファイルを提供する FQDN
中継サイトの URL	FQDN
攻撃サイトの URL	FQDN

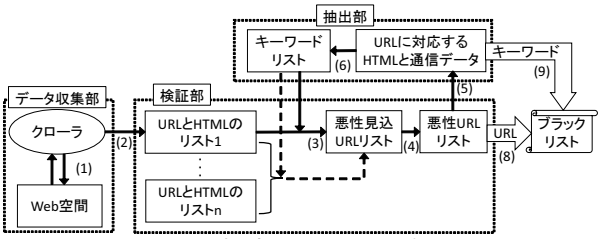

図 1 提案手法の処理流れ

(3) キーワードを用いて HTML ファイルを検索し，ファイル内にキーワードを含む場合，HTML ファイルに対応する URL を悪性見込 URL リストに追加
(4) 悪性見込 URL リストに追加された URL が悪性 Web サイトであれば，悪性 URL リストに URL を追加
(5) Google Chrome を利用して Web サイトへアクセスした際のアクセス先 URL などを収集するアプリを用いて，悪性 Web サイトにアクセスした際の通信データを取得
(6) 検証部における検索に用いるキーワードを抽出し，キーワードリストを拡張
(7) 拡張したキーワードリストを用いて，悪性見込 URL が見つからなくなるまで (3)〜(6) を繰り返し
(8) 悪性 URL リストの URL をブラックリストに登録
(9) 抽出したキーワードをブラックリストに登録

3 評価

3.1 評価内容と評価環境

以下の 2 つの評価を行った．

(1) 探索による悪性 Web サイトの発見数
(2) ブラックリストによる悪性 Web サイトの検知率

(1) について，2019 年 7 月 23 日から 2019 年 12 月 16 日の間で 2.2 節の (1)〜(9) まで手順を行い，悪性 Web サイトをどの程度発見できるのかを評価した．なお，初期のキーワードリストには，独自に発見した悪性 Web サイトから抽出したキーワードを設定した．

次に，(2) について，上記の評価で得られたブラックリストを用いて，悪性 Web サイトへのアクセスを検知

Evaluation of Method of Generating Blacklist for Mobile Devices by Searching Malicious Websites
Takashi Ishihara†, Masaya Sato†, Toshihiro Yamauchi†
†Graduate School of Natural Science and Technology, Okayama University

表 2 ブラックリストによる検知結果

	異なる中継/攻撃サイトの利用	検知率 (検知数/アクセス数)	検知時にマッチしたキーワード
サイト A	無	100% (10/10)	中継サイトの FQDN1，攻撃サイトの FQDN1
サイト B	無	100% (10/10)	中継サイトの FQDN1，攻撃サイトの FQDN1
サイト C	無	100% (10/10)	中継サイトの FQDN2，攻撃サイトの FQDN2
サイト D	無	100% (10/10)	中継サイトの FQDN3
サイト E	有	100% (10/10)	(遷移 a) 中継サイトの FQDN4〜6，攻撃サイトの FQDN3
		60% (6/10)	(遷移 b) 中継サイトの FQDN7

できるか否かを検証した．評価対象の悪性 Web サイトは，2019 年 12 月 20 日から 2019 年 12 月 30 日の間で提案手法により収集した 5 件の悪性 Web サイト（サイト A〜サイト E）を利用した．悪性 Web サイトは，アクセスごとに遷移先が同じであるとは限らないため，10 回アクセスした．なお，サイト E は，画面をタップするタイミングによって，2 つの異なる中継サイトや攻撃サイトへの遷移が発生することを確認した．このため，サイト E ではそれぞれ 10 回アクセスした．また，検証には Google Chrome の URL バーの文字列を取得し，取得した文字列にブラックリストへ登録した文字列が含まれる場合，利用者に悪性 Web サイトへのアクセスを通知するアプリを用いた．

悪性 Web サイトへのアクセスは，OS が Android 6.0 の実機端末上で，Google Chrome を用いて実施した．

3.2 評価結果と考察

3.2.1 探索による悪性 Web サイトの発見数

探索の結果，クロールした 122,350 件の Web サイトから 200 件の遷移元サイトを発見した．発見した 200 件の遷移元サイトのうち，ユニークな FQDN は 54 件であった．

ユニークな FQDN を持つ 54 件の遷移元サイトにアクセスした際の通信データから 111 個のキーワードを抽出した．抽出したキーワードのうち，ファイル名は 3 個であり，FQDN は 108 個であった．キーワードとして抽出した FQDN のうち，遷移の起点となるファイルを提供する FQDN は，ファイル名の抽出数と同様に 3 個と少ない．これは，JavaScript コードの難読化により，遷移元サイトにおいて遷移の起点となるファイルを特定できない場合があったためである．

なお，発見した 200 件の遷移元サイトのうち，2020 年 1 月 7 日時点で遷移が発生した 182 件の遷移元サイトについて，URL ブラックリスト方式である Google Safe Browsing により検知できるものはなかった．

3.2.2 ブラックリストによる悪性 Web サイトの検知率

2019 年 7 月 23 日から 2019 年 12 月 16 日の間の悪性 Web サイトの探索により構築したブラックリストを用いた悪性 Web サイトの検知結果を表 2 に示す．サイト A〜サイト D は，10 回のアクセスすべてにおいて，それぞれ表 2 に示す検知時にマッチしたキーワードにより，悪性 Web サイトへのアクセスを検知した．

サイト E では，(遷移 a) の場合，中継サイトの FQDN4〜6，および攻撃サイトの FQDN3 により，10 回のアクセスすべてにおいて，悪性 Web サイトへのアクセスを検知した．一方，(遷移 b) の場合，中継サイトの FQDN7 により，10 回のうち 6 回の悪性 Web サイトへのアクセスを検知した．(遷移 b) において，中継サイトの FQDN7 と 1 文字だけ異なるブラックリストに登録されていない FQDN が中継サイトに利用された場合に，見逃しが発生した．

検知時にマッチしたキーワードは中継サイトの FQDN と攻撃サイトの FQDN のみである．遷移の起点となるファイル名とそのファイルを提供する FQDN は抽出数が少ないため，検知時にマッチしなかったと推察する．

4 おわりに

本評価結果から，キーワードを利用したブラックリストは，特定のサイトでは検知率が充分に高いことが分かる．ただし，検知率が高い原因として，提案手法により収集した悪性 Web サイトを評価に利用したことが挙げられる．残された課題として，提案手法を用いずに収集した悪性 Web サイトにおける検知率の評価がある．

謝辞 本研究成果は，国立研究開発法人情報通信研究機構 (NICT) の委託研究「Web 媒介型攻撃対策技術の実用化に向けた研究開発」により得られたものです．

参考文献

[1] 石原 聖，折戸 凜太郎，佐藤 将也，山内 利宏：モバイル向け悪性 Web サイトの探索によるブラックリスト構築手法，コンピュータセキュリティシンポジウム 2019（CSS2019）論文集，pp.1025–1032 (2019).

[2] Invernizzi, L., Comparetti, P. M.: Evilseed: A Guided Approach to Finding Malicious Web Pages, Proc. IEEE Symposium on Securityand Privacy, pp.428–442 (2012).

[3] Imamura, Y., Orito, R., Chaikaew, K., Manardo, C., Leelaprute, P., Sato, M., Yamauchi, T.: Threat Analysis of Fake Virus Alerts Using WebView Monitor, Proc. 2019 Seventh International Symposium on Computing and Networking (CANDAR), pp.28–36 (2019).