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

石原 聖[1]　折戸 凜太郎[1]　佐藤 将也[1]　山内 利宏[1]

概要：モバイル端末の普及にともない，モバイル端末を対象とした攻撃が増加している．攻撃手法の 1 つとして，利用者の意図しない遷移により，利用者の意図しない Web サイトへ誘導する攻撃の手口がある．このような攻撃への対策の 1 つとして，URL やホスト名のブラックリストを構築し，悪性 Web サイトへのアクセスを未然に防止する手段がある．しかし，悪性 Web サイトの URL は短期間で変化するため，悪性 Web サイトを探索し，ブラックリストをタイムリーに更新する必要がある．そこで，本稿では，モバイル向けの悪性 Web サイトを探索し，ブラックリストを構築する手法を提案する．提案手法では，クローラを用いて Web 空間から収集した大量の HTML ファイルについて，既知の悪性 Web サイトから抽出したキーワードを用いて悪性である可能性が高い HTML ファイルを検索することで，効率的にモバイル向けの悪性 Web サイトを発見する．これにより，新たな悪性 Web サイトを早期にブラックリストに追加できる．本稿では，提案手法とその評価結果について報告する．

キーワード：Android、悪性 Web サイト，ブラックリスト，Web 媒介型攻撃

Method of Generating Blacklist by Searching Malicious Mobile Websites

Takashi Ishihara[1]　Rintaro Orito[1]　Masaya Sato[1]　Toshihiro Yamauchi[1]

Abstract: As mobile devices have become more popular, attacks to mobile devices increase. One of the methods to attack mobile devices induce a user to unwanted page by unwanted page-transition. One of the countermeasures against such an attack is to generate a blacklist of URLs and hostnames and prevent access to malicious websites using the blacklist. However, URLs of malicious websites change in a short time so it is necessary to search malicious websites and to timely update the blacklist. In this paper, we propose a method of generating blacklist by searching malicious mobile websites. The proposed method collects many HTML files from the web space using a crawler and searches for HTML files that are highly likely to be malicious using keywords extracted from known malicious websites to efficiently discover mobile malicious websites. Thus, new malicious websites can be timely added to the blacklist. This paper reports the design of the proposed method and evaluation results.

1. はじめに

スマートフォンやタブレットなどのモバイル端末が普及するにつれて，モバイル端末の利用者数が増加している．2019 年 1 月に公表された調査結果では，世界中のモバイル端末の利用者数は 2018 年から 1 億人増加し，世界人口の 67%に達したと報告されている [1]．2019 年 8 月に公表された日本国内のモバイル端末の利用実態調査では，調査対象者 1,200 人のモバイル端末所有率は 100%であったと報告されている [2]．文献 [2] におけるモバイル端末の機能ごとの利用率調査では，メール，ゲーム、および動画視聴などモバイル端末で利用できるサービスの選択肢が広がる中で，SNS の利用率は約 90%であり，SNS は日常的に使われる機能として定着している傾向があると報告されてい

[1] 岡山大学 大学院自然科学研究科
Graduate School of Natural Science and Technology, Okayama University

る．また，Google が開発しているスマートフォン向けのオペレーティングシステムである Android オペレーティングシステム（以降，Android）を搭載したスマートフォンが普及しており，2019 年の Android の市場占有率は約 70%以上となっている [3],[4].

一方，モバイル端末の利用者数が増加することで，モバイル端末がサイバー攻撃の標的にされやすくなっている．Android を標的とするモバイルマルウェアによる被害の調査結果では，モバイルマルウェアの被害を受けた利用者が増加しており，2018 年に検知した攻撃数は，2017 年に検知した攻撃数に比べて 43%増加したと報告されている [5]．また，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 [6].

モバイル端末における悪性 Web サイトによる攻撃の手口として，リダイレクトを用いて利用者の意図しないサイトへ誘導する手口が存在する．この手口ではまず，攻撃者はソーシャルメディア上の投稿やメッセージ，および不正広告により利用者を悪性 Web サイトへ誘導する．Android では，Drive–By Download 攻撃のように，利用者の同意なく自動的に不正アプリをインストールできる方法は確認されていないため，悪性 Web サイトではウイルス感染メッセージなどの偽警告を表示することで，ウイルス駆除の名目で利用者に不審なアプリをインストールさせようとする [7]．また，このような手口で誘導される悪性 Web サイトには，アプリをインストールさせる目的のもの以外にも，フィッシング詐欺サイトや不審なショッピングサイトなどが確認されている [6].

このような攻撃への対策の 1 つとして，URL やホスト名のブラックリストの利用がある．この対策により，リダイレクトの起点となる悪性 Web サイトへのアクセスを未然に防止する効果や，リダイレクト先の悪性 Web サイトへのアクセスを防止する効果が期待できる．しかし，攻撃者は悪性 Web サイトの IP アドレスやドメイン名を短期間のうちに変更することで，ブラックリストによる対策を困難にする場合がある [8] ため，悪性 Web サイトをタイムリーに探索し，ブラックリストを更新する必要がある．

そこで，本稿では，モバイル向けの悪性 Web サイトを探索し，ブラックリストを構築する手法を提案する．また，提案手法の実現課題，実現方式，および評価結果について述べる．

2. 利用者の意図しない Web サイトへ誘導する攻撃

モバイル端末において，利用者の意図しない遷移により，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．**図 1** に利用者の意図しない Web サイトへ誘導する攻撃の流れを示す．

利用者が遷移元サイトへアクセスすると，遷移元サイトは利用者を中継サイトにリダイレクトする．また，複数の中継サイトを経由して，利用者を攻撃サイトへリダイレクトする．このように，利用者の意図しない Web サイトへ誘導する攻撃は，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 サイトを利用する．

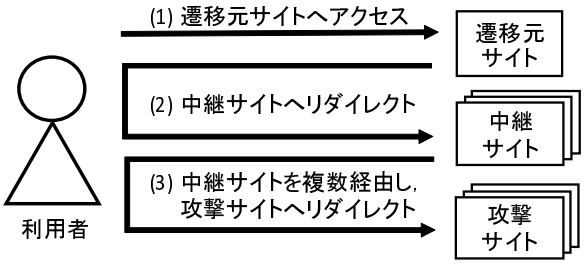

図 1 利用者の意図しない Web サイトへ誘導する攻撃の流れ

また，文献 [9] では，Cookie の保持や取得の有無により挙動が異なるという特徴や，遷移元サイトからのリダイレクト先や遷移元サイトからのリダイレクト数は，毎回同じであるとは限らないという特徴が報告されている．

攻撃サイトには，フィッシング詐欺サイト，不審なショッピングサイト，およびアプリをインストールさせようとするサイトなどがある．フィッシング詐欺サイトや不審なショッピングサイトでは，住所や口座番号といった情報を入力させることで，利用者の個人情報を収集する．アプリをインストールさせようするサイトでは，ウイルス感染メッセージなどの偽警告を表示することで，ウイルス駆除の名目で利用者に不審なアプリをインストールさせようとする．また，このような手口によりインストールさせようとするアプリには，アフェリエイトプログラムなどの利用により金銭的利益を得ることを目的とした正規アプリや，金銭要求を目的としたワンクリック詐欺ソフトや情報窃取型の不正アプリなどがあることが確認されている [6].

このようなモバイル端末を対象とした悪性 Web サイトについて，折戸らは Android 向けセキュリティアプリにおける悪性 Web サイト検知率の調査を行っている [10]．調査の結果，Google Safe Browsing，および Google Play ストアにおいてユーザ満足度が高い 2 つのセキュリティアプリにおける偽警告画面を表示する Web サイトの検知率は，最も高いアプリでも 65%であり，検知率は十分ではないことを示している．このため，モバイル向けの悪性 Web サイトにおける攻撃への対策が必要である．

3. ブラックリストの構築手法

3.1 目的と考え方

2 章で述べたモバイル向けの悪性 Web サイトにおける攻撃への対策の 1 つとして，URL やホスト名のブラックリストの利用が有効であると考える．URL やホスト名のブラックリストを利用することによって，リダイレクトの起

点となる悪性 Web サイトへのアクセスを未然に防止する効果や，リダイレクト先の悪性 Web サイトへのアクセスを防止する効果が期待できる．このため，本研究ではモバイル向けの悪性 Web サイトに対するブラックリストを構築することを目的とする．

悪性 Web サイトは，新たに構築され続けているため，新たな悪性 Web サイトを早期にブラックリストに追加する必要がある．そこで，モバイル向けの悪性 Web サイトを探索し，発見した悪性 Web サイトからブラックリストを構築する手法を提案する．

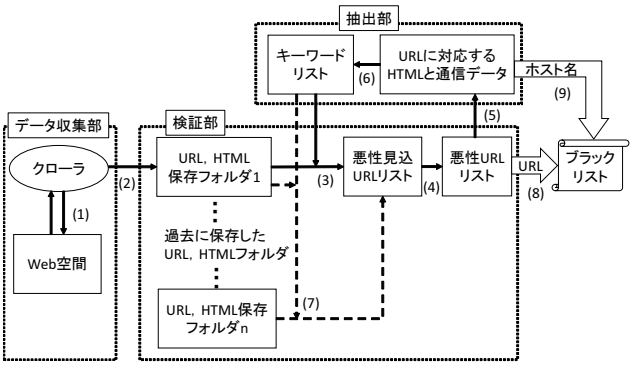

図 2　提案手法の処理流れ

3.2　課題

モバイル向けの悪性 Web サイトを探索し，発見した悪性 Web サイトを分析することによりブラックリストを構築する．本手法における課題を以下に示す．

(課題 1)　Web 上の大規模なデータへの対応

Web 空間には 16 億を超えるサイトが存在すると 2018 年の調査で報告されている [11]．悪性 Web サイトを探索するためには，広大な Web 空間における大規模なデータに対応する必要がある．

(課題 2)　悪性 Web サイトをタイムリーに発見

攻撃者は悪性 Web サイトの IP アドレスやドメイン名を短期間のうちに変化させることで，ブラックリストによる対策を困難にする手法を利用するものがある [8]．このため，悪性 Web サイトをタイムリーに発見し，ブラックリストを更新する必要がある．

(課題 3)　モバイル向けの悪性 Web サイトの攻撃の分析

ブラックリストの構築には，異常なプロセスの作成や繰り返しのリダイレクトなどの悪意のある攻撃における特徴を見つけるために，悪性 Web サイトの分析を行う [12]．このため，モバイル向けの悪性 Web サイトの攻撃を観測し，分析する必要がある．

3.3　対処

課題 1 と課題 2 の対処として，クローラを用いて Web 空間から大量の Web コンテンツを収集し，悪性 Web サイトを発見する．しかし，収集した大量の Web コンテンツ全てを検証し，悪性 Web サイトを発見することは困難である．このため，収集した Web コンテンツから悪意のある可能性のある Web コンテンツを抽出し，抽出した Web コンテンツを検証，および分析する．これにより，効率的に悪性 Web サイトを発見できる．

課題 3 の対処として，Android における WebView を介した Web アクセス観測機構（以降，WebView の観測機構）[13] を用いて，通信データの観測と分析を行う．多くの Android アプリケーションにおいて WebView が利用されており，Facebook や Twitter など多くの SNS の Android アプリケーションにおいても WebView が利用されている．また，SNS は日常的に使われる機能として定着している傾向があり，SNS 上の投稿により利用者を悪性 Web サイトに誘導する手法があることから，利用者は WebView を介してモバイル向けの悪性 Web サイトにアクセスする可能性が高い．このため，WebView の観測機構を用いてモバイル向けの悪性 Web サイトの攻撃にともなう通信を観測する．WebView の観測機構を用いることで，WebView を介した全ての Web アクセスの通信データを取得することができる．

3.4　提案手法の処理流れ

提案手法は，データ収集部，検証部，および抽出部の 3 つに分類される．データ収集部では，未知の URL と未知の URL に対応する Web コンテンツとして HTML ファイルを収集する．検証部では，既知の悪性 Web サイトから抽出したキーワードを用いて HTML ファイルを検索し，悪性である可能性が高い URL（以降，悪性見込 URL）を発見する．また，悪性見込 URL の悪性判定を行う．抽出部では，悪性と判定された Web サイトの通信データを分析し，キーワード検索に用いるキーワードを抽出する．提案手法におけるブラックリスト作成の処理流れを**図 2** 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 と収集した HTML ファイルを保存
(3) 既知の悪性 Web サイトから抽出したキーワードを用いて検索し，HTML ファイルが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 サイトか否か判定し，悪性 Web サイトであった場合，悪性 URL リストに URL を追加
(5) WebView の観測機構を用いて，悪性 Web サイトにアクセスした際の通信データを取得
(6) 取得した通信データから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，新しい悪性見込 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追加
(9) 攻撃にともなう通信を分析し，遷移において経由する中継サイトと攻撃サイトの URL からホスト名を抽出し，ブラックリストに追加

3.5 期待される効果

(1) 提案手法では，キーワードを抽出するごとに，キーワードリストを拡張する．また，拡張したキーワードリストを用いて，過去に保存した HTML ファイルに対して検索を行う．これにより，拡張する前のキーワードリストでは見逃した悪性 Web サイトを新たに発見できる可能性がある．
(2) 悪性 Web サイトを発見する手法として，Google 検索エンジンで高度な演算子を使用して検索を行う Google Dorks と呼ばれる手法がある [14]．そこで，キーワード抽出で得られたキーワードを検証部における HTML ファイルの検索に用いるだけでなく，検索エンジンでの検索キーワードとして利用する．これにより，悪性 Web サイトを検索エンジンから発見できる可能性がある．
(3) 提案手法では，利用者の意図しない Web サイトへ誘導する攻撃において経由する中継サイトと攻撃サイトからホスト名を抽出する．これにより，遷移元サイトは異なっていても，同じ中継サイトや攻撃サイトを利用する場合，未知の悪性 Web サイトからの遷移を防止できる可能性がある．

4. 実現方式

4.1 実現における課題

提案手法を実現するために，以下の実現課題に対処する必要がある．

(1) クロールする URL の選定方法の検討
　データ収集部では，クローラを用いて大量の未知 URL に対応する Web コンテンツを収集する．しかし，Web 空間には膨大な数の Web サイトが存在し，悪性 Web サイトはその URL を短期間で変更するため，Web 空間を手当たり次第にクロールする方法では，悪性 Web サイトの出現に対応できない．ここで，Web 空間から悪性 Web サイトを効率的に発見するために，クロールする URL を選定する必要がある．
(2) 悪性 Web サイトであるか否かを判定する方法の検討
　ブラックリストを構築するためには，収集した Web サイトの中から悪性 Web サイトを検出し，攻撃にともなう通信データを分析する必要がある．このため，収集した Web サイトについて，悪性 Web サイトであるか否かを判定する必要がある．
(3) キーワードの抽出方法の検討
　悪性と判定された Web サイトの HTML ファイル，および WebView の Web アクセス観測機構により取得した通信データにおいて，悪性ではない Web サイトにも含まれる多くのキーワードを抽出した場合，検証部での HTML ファイル検索の際に悪性ではない Web サイトの HTML ファイルが多くヒットしてしまい，探索の効率が落ちてしまう．このため，悪性 Web サイトである可能性が高い HTML ファイルがヒットするようなキーワードを抽出する必要がある．

4.2 課題への対処

4.2.1 クロールする URL の選定方法の検討

クロールする URL として，Twitter の Streaming API[15] の statuses/filter を利用して URL を収集する．これは，攻撃者が Twitter や Facebook，YouTube などの SNS に，悪性 Web サイトの URL を投稿するという手口があるためである [16]．また，Twitter を対象とした悪性 URL の分析結果では，悪性 URL を含む Twitter の投稿は 1 千個に 1 個程度の割合で存在すると報告されている [17]．このため，Twitter 上から URL を収集することで，Web 空間を手当たり次第にクロールする方法よりも効率的に悪性 Web サイトを発見できると考える．また，Twitter の Streaming API は，一定量が間引かれて提供される Tweet をほぼリアルタイムで取得することができるため，攻撃者により投稿された未知の悪性 Web サイトの URL をタイムリーに発見できる可能性がある．

4.2.2 悪性 Web サイトであるか否かを判定する方法の検討

セキュリティアプリは，悪性 Web サイトへの Web アクセスを検知する．このため，悪性 Web サイトであるか否かの判定に利用できる可能性がある．しかし，2 章で述べたように，セキュリティアプリにおける悪性 Web サイトの検知率は十分ではない．このため，悪性 Web サイトであるか否かの判定に，セキュリティアプリを利用するのは適切ではないと考える．

そこで，本研究では悪性 Web サイトに手動でアクセスし，利用者の意図しない遷移により利用者の意図しないサイトへ誘導されるか否かを確認することで，Web サイトの悪性判定を行う．手動でのアクセスは手間がかかるものの，本手法では悪性見込 URL リストだけをチェックするため，チェックする Web サイト数は多くない．

4.2.3 キーワードの抽出方法の検討

提案手法におけるキーワードの抽出方法を**表 1** に示し，以下で説明する．悪性と判定された Web サイトの HTML ファイルは，攻撃の起点となる遷移元サイトの HTML ファイルである．遷移元サイトには，利用者を攻撃サイトまで遷移させる起点となるファイルが存在すると考える．また，

表 1　キーワードの抽出方法

対象	抽出するキーワード
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ホスト名
中継サイトの URL	ホスト名
攻撃サイトの URL	ホスト名

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 [14]．このため，HTML ファイルからは，遷移の起点となるファイル名とこのファイルを提供するホスト名をキーワードとして抽出する．通信データからは，利用者の意図しないサイトへ遷移する途中に経由する中継サイト，および攻撃サイトの URL からホスト名をキーワードとして抽出する．これは，遷移元サイトから中継サイトや攻撃サイトに直接遷移する可能性が考えられるためである．また，検索エンジンを用いて独自に発見した悪性 Web サイトからも同様にキーワードを抽出し，検証部における検索に利用する．

5. 評価

5.1 評価内容と評価環境

表 2 に示す環境において，提案手法の有用性を明らかにするために，以下の評価を行った．

(1) 探索によるモバイル向け悪性 Web サイトの発見数
　提案手法により，モバイル向けの悪性 Web サイトをどの程度発見できるかを示す．

(2) ブラックリストを用いた悪性 Web サイトの検知実験
　提案手法により構築したブラックリストを用いて，モバイル向けの悪性 Web サイトへのアクセスを検知できることを示す．

5.2 探索によるモバイル向け悪性 Web サイトの発見数

探索により発見できるモバイル向け悪性 Web サイト数を評価した．探索は 2019 年 7 月 23 日から 2019 年 8 月 5 日の間で行い，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．

表 3 に結果を示す．表 3 において，収集日は URL と HTML ファイルを収集した日である．HTML ファイル数は，クロールにより収集した HTML ファイルの数である．検証日は，検証部において HTML ファイルの検索と手動による悪性判定を行った日である．1 つの収集日に検証日が複数あるのは，見逃した悪性 Web サイトを新たに発見することを目的として，発見した悪性 Web サイトから抽出したキーワードを用いてキーワードリストを拡張し，過去に収集した HTML ファイルを検索したためである．具体的には，7 月 25 日に収集した 2,036 個の HTML ファイルについて，7 月 25 日の検証により悪性 Web サイトを 2 件発見した．8 月 2 日にキーワードリストを拡張し，再検証することにより，7 月 25 日の検証で発見した悪性 Web サイトとは異なる悪性 Web サイトを 3 件発見した．7 月 29 日と 8 月 6 にキーワードリストを拡張して行った再検証では，過去に発見した悪性 Web サイトと異なる悪性 Web サイトは発見できなかった．

収集した HTML ファイルの合計数は 25,093 個であり，発見した悪性 Web サイトの合計数は 20 件であった．発見した 20 件の悪性 Web サイトのうち，ユニークなホスト名は 8 件であった．なお，発見した 20 件の悪性 Web サイトのうち，遷移元サイトの HTML ファイルから抽出したキーワードにより発見したサイト数は 18 件であり，中継サイ

表 2　評価環境

OS	Windows 8.1 Pro（64bit）
CPU	Intel Core i5-4460
メモリ	12 GB
Android エミュレータ	Android 6.0

表 3　悪性 Web サイト探索の結果

収集日	HTML ファイル数	検証日	悪性 Web サイトの発見数
7 月 23 日	1,982	7 月 23 日	1
		7 月 24 日	0
		7 月 29 日	0
		8 月 2 日	0
		8 月 6 日	0
7 月 24 日	2,545	7 月 24 日	0
		7 月 29 日	0
		8 月 2 日	1
		8 月 6 日	0
7 月 25 日	2,036	7 月 25 日	2
		7 月 29 日	0
		8 月 2 日	3
		8 月 6 日	0
7 月 26 日	3,299	7 月 26 日	2
		7 月 29 日	0
		8 月 2 日	2
		8 月 6 日	0
7 月 29 日	2,672	7 月 29 日	0
		8 月 2 日	0
		8 月 6 日	0
7 月 31 日	3,614	7 月 31 日	1
		8 月 2 日	1
		8 月 6 日	0
8 月 1 日	2,326	8 月 1 日	1
		8 月 2 日	0
		8 月 6 日	0
8 月 2 日	3,601	8 月 2 日	2
		8 月 6 日	0
8 月 5 日	3,018	8 月 5 日	4
		8 月 6 日	0
合計	25,093		20

表 4　キーワードとして抽出したファイル名とホスト名の数

	抽出したキーワード数
ファイル名	3
ホスト名	53

トと攻撃サイトから抽出したキーワードにより発見したサイト数は 2 件であった．

また，**表 4** にキーワードとして抽出したファイル名の数とホスト名の数を示す．遷移元サイトにおいて，動画再生ボタンなどのタップにより中継サイトへ直接遷移し，遷移の起点となるファイルが存在しない場合があったため，ファイル名の抽出数は少ない．また，JavaScipt 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コードを含むファイルを特定できず，キーワードとして抽出できない場合があった．難読化 JavaScript コードへの対処は今後の課題とする．

5.3 構築したブラックリストを用いた悪性 Web サイトの検知実験

モバイル向けの悪性 Web サイトの探索により構築したブラックリストによる検知実験を行った．実験には，自作したテストアプリを使用した．このアプリは，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として登録され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である．

評価では，提案手法を用いて 2019 年 7 月 23 日より前に収集した悪性 Web サイトと文献 [9] で分析に用いられた悪性 Web サイトを利用した．URL とホスト名のブラックリストにより悪性 Web サイトへのアクセスを検知できるか否かを検証した．悪性 Web サイトはアクセスごとに異なる中継サイトと攻撃サイトに遷移する可能性があるため，サイトごとに 5 回ずつアクセスし，検知実験を行った．また，悪性 Web サイトでは，遷移元サイトの Cookie を保持しておらず，遷移元サイトから Cookie を取得する場合において意図しない遷移が起きる [9]．このため，Cookie を消去してアクセスを行った．

2019 年 8 月 8 日に実施した検知実験の結果を**表 5** に示す．なお，表 5 中の「○」は，検知アプリが悪性 Web サイトへのアクセスを検知したことを示し，「–」は悪性 Web サイトへのアクセスを検知できなかったこと示す．以下では，結果の詳細を述べる．

(1) サイト A
5 回のアクセスにおいて，悪性 Web サイトへのアクセスを一度も検知しなかった．これは，抽出したキーワード数が充分ではなかったためであると考える．

(2) サイト B
ブラックリストに登録した「ads–app–online–new.pw」により悪性 Web サイトへのアクセスを 2 回検知した．また，「news-easy.com」により悪性 Web サイトへのアクセスを 1 回検知した．検知できなかった 2 回のアクセス時には，URL バーに「ads–app–online–new.pw」から「–」を抜いた「adsapponlinenew.pw」のホスト名が表示されていた．

(3) サイト C
5 回のアクセスにおいて，「smart–update.info」により悪性 Web サイトへのアクセスを検知した．

(4) サイト D
「ads–app–online–new.pw」により悪性 Web サイトへのアクセスを 2 回検知した．しかし，サイト B と同様に，URL バーに「adsapponlinenew.pw」と表示される 3 回の場合については，検知できなかった．

(5) サイト E
悪性 Web サイトへの 5 回のアクセスすべてを検知した．5 回のアクセスにおいて「allowplay.pro」，「adaranth.com」，「smart–update.info」により，複数の中継サイトへのアクセスを検知した．また，アクセスごとに「bnsjb1able.com」，「65vk1fba34.com」，および「0byv9mgbn0.com」と異なる中継サイトへのアクセスを検知した．

表 5　ブラックリストによる検知結果

	1 回目	2 回目	3 回目	4 回目	5 回目	結果
サイト A	–	–	–	–	–	0/5
サイト B	○	○	–	–	○	3/5
サイト C	○	○	○	○	○	5/5
サイト D	○	–	○	–	–	2/5
サイト E	○	○	○	○	○	5/5

本実験結果から，提案手法により構築したブラックリストにより悪性 Web サイトへのアクセスを検知できる場合があることを示した．収集期間が 9 日間と充分でなかったため，キーワードが充分に集められなかった可能性がある．このため，検知率が充分高いとはいえない．ただし，特定のサイトでは充分に高い検知率であることから，キーワードを利用したブラックリストの有効なサイトとそうでないサイトがあると考えられる．誤検知 (false positive) は起こっていない．このため，利便性を損なわないブラックリストを構築できている．ただし，誤検知 (false negative) による見逃しが多い．この問題にはキーワードの拡充により対処できる可能性があり，今後の課題とする．

6. 関連研究

悪性 Web サイトへの対策にブラックリストを利用する研究がある [18]．文献 [18] では，Web 空間から新しい悪意

のある URL を自動的に検出し，自動でブラックリストを生成する AutoBLG を提案している．AutoBLG は，悪意のある URL の IP アドレスを利用して未知の URL を収集し，フィルター処理により分析する URL の量を減らすことで，効率的に悪意のある URL を発見する．上記の論文は，Web サイトを介した攻撃として Drive–By Download 攻撃に着目している．一方で，提案手法では，Web サイトを介した攻撃として利用者の意図しない Web サイトへ誘導する攻撃に着目している．

モバイル向けの悪性 Web サイトについて，リアルタイムで検出することを目的とした研究に，文献 [19] がある．文献 [19] は HTML コンテンツ，JavaScript コンテンツ，URL，およびモバイル固有の機能などの静的な特徴から，教師あり機械学習を利用してモバイル向けの悪性 Web サイトを検出する kAYO を提案している．kAYO は，教師あり機械学習を利用しているため，学習に用いるラベル付きの教師データが事前に必要になる．しかし，このようなデータの作成はコストが高い問題がある [18]．一方で，提案手法は，悪性 Web サイトから抽出した比較的少量のキーワードを用いることで，モバイル向けの悪性 Web サイトを発見できる．

7. おわりに

利用者の意図しない Web サイトへ誘導する攻撃を検知するために，モバイル向けの悪性 Web サイトを探索し，ブラックリストを構築する手法を提案した．提案手法は，クローラを用いて Web 空間から収集した大量の HTML ファイルについて，既知の悪性 Web サイトから抽出したキーワードを用いて悪性である可能性が高い HTML ファイルを検索することで，効率的にモバイル向けの悪性 Web サイトを発見できる．さらに，キーワードを抽出するごとにキーワードリストを拡張し，拡張したキーワードリストを用いて過去に保存した HTML ファイルから悪性である可能性が高い HTML ファイルを検索することで，拡張する前のキーワードリストでは見逃した悪性 Web サイトを新たに発見できる．また，提案手法は発見した悪性 Web サイトを分析し，利用者の意図しない Web サイトへ誘導する攻撃で利用される中継サイトと攻撃サイトのホスト名をブラックリストに登録する．これにより，遷移元サイトへのアクセスだけでなく，中継サイトと攻撃サイトへのアクセスを検知できる．

提案手法を用いて探索を行い，モバイル向けの悪性 Web サイトを発見できることを示した．また，提案手法により構築したブラックリストを用いてモバイル向けの悪性 Web サイトへのアクセスを検知する実験を実施し，ホスト名とファイル名をブラックリストと照合することで，モバイル向けの悪性 Web サイトへのアクセスを検知できる場合があることを示した．

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] DataReportal: Digital 2019: Global Digital Overview，DataReportal (online)，available from〈https://datareportal.com/reports/digital-2019-global-digital-overview〉(accessed 2019–08–08).

[2] 情報通信ネットワーク産業協会：「2019 年度 モバイル通信端末の利用実態調査」 ～端末買替えニーズは長期化、格安通信契約は転換期を迎えたか？５Ｇへの期待高まる！～，情報通信ネットワーク産業協会 (オンライン)，入手先〈https://www.ciaj.or.jp/news/pressreleases/press_releases_past_issue/pressrelease2019/4794.html〉(参照 2019–08–08).

[3] Net Market Share: Net Market Share (online)，available from〈https://netmarketshare.com〉(accessed 2019–08–08).

[4] StatCounter: StatCounter Global Stats (online), available from〈https://gs.statcounter.com/os-market-share/mobile/worldwide〉(accessed 2019–08–08).

[5] Chebyshev, V.: Mobile malware evolution 2018，Securelist (online), available from〈https://securelist.com/mobile-malware-evolution-2018/89689/〉(accessed 2019–08–08).

[6] 岡本勝之：最新モバイル脅威事情：モバイルこそ「Web 経由」に注意，トレンドマイクロセキュリティブログ (オンライン)，入手先〈https://blog.trendmicro.co.jp/archives/13524〉(参照 2019–08–08).

[7] 岡本勝之：実例で学ぶネットの危険：スマホで突然の「ウイルスに感染」表示、開くとどうなる？, トレンドマイクロセキュリティブログ (オンライン)，入手先〈https://blog.trendmicro.co.jp/archives/16382〉(参照 2019–08–08).

[8] 吉田豊，中村嘉隆，稲村浩，高橋修：ドライブ・バイ・ダウンロード攻撃検知のための悪性サイト情報収集手法の改善，マルチメディア，分散協調とモバイルシンポジウム 2016 論文集，pp.813–818 (2016).

[9] 今村祐太，折戸凜太郎，Kritsana Chaikaew，Célia Manardo，Pattara Leelaprute，佐藤将也，山内利宏：Android における WebView の Web アクセス観測機構を利用した悪性 Web サイトの脅威分析と対策の提案，コンピュータセキュリティシンポジウム 2018 (CSS2018) 論文集，pp.137–144 (2018).

[10] 折戸凜太郎，佐藤将也，山内利宏：Android 向けセキュリティアプリにおける悪性 Web サイト検知率の調査，第 18 回情報科学技術フォーラム (FIT2019) 講演論文集，(2019).

[11] Internet Live Stats: Total number of Websites, Internet Live Stats (online)，available from〈https://www.internetlivestats.com/total-number-of-websites/〉(accessed 2019–08–08).

[12] Sahoo, D., Liu, C., and Hoi, S.C.: Malicious URL detection using machine learning: A survey，arXiv.org e-Print archive (online), available from〈https://arxiv.org/pdf/1701.07179.pdf〉(2017).

[13] Imamura, Y., Uekawa, H., Ishihara, Y., Sato, M. and Yamauchi, T.: Web Access Monitoring Mechanism for Android WebView，*Proceedings of the Australasian In-*

formation Security Conference (AISC 2018), pp.1–8 (2018).

[14] Invernizzi, L., Comparetti, P. M.: Evilseed: A guided approach tofinding malicious web pages, *Proc. IEEE Symposium on Securityand Privacy*, pp.428–442 (2012).

[15] Twitter Streaming APIs, available from 〈`https://dev.twitter.com/streaming/overview`〉(accessed 2019–06–25).

[16] Trend Micro:「サポート詐欺」の手口が変化、SNS の投稿を検索結果に表示させ詐欺ページに誘導，トレンドマイクロセキュリティブログ (オンライン)，入手先〈`https://blog.trendmicro.co.jp/archives/20549`〉(参照 2019–08–09).

[17] 國分佑太朗，中村章人：SNS における悪性 URL の分析と防御，社会情報学会 (SSI)2017(オンライン), 入手先〈`http://gmshattori.komazawa-u.ac.jp/ssi2017/wp-content/uploads/2017/03/30.pdf`〉(参照 2019–08–08).

[18] Bo, S., Akiyama, M., Yagi, T., Hatada, M. and Mori, T.: Automating URL Blacklist Generation with Similarity Search Approach, *IEICE Transactions on Information and Systems*, vol.99, no.4, pp.873–882 (2016).

[19] Amrutkar, C., Kim, Y. S. and Traynor, P.: Detecting Mobile Malicious WebPages in Real Time, *IEEE Transactions on Mobile Computing*, vol.16, no. 8, pp.2184–2197 (2017).