Computer Security Symposium 2017
23 - 25 October 2017

多数のユーザの Web アクセスログから効率的に悪性サイトを抽出する手法

森島 周太[1]　中野 弘樹[1]　吉岡 克成[2]　松本 勉[2]　藤原 礼征[3]

概要：改ざんされた正規 Web サイトや不正な広告からクライアントを攻撃サイトに誘導し，クライアントの脆弱性を悪用してマルウェアをインストールさせるドライブバイダウンロード攻撃が猛威をふるっている．本報告では、数十万人規模のエンドユーザから得られる膨大な Web アクセスログを用いて悪性サイトを効率的に発見する方法を提案する．膨大な Web アクセスログに対して個別に詳細な分析を行うことは非効率的であるため，悪性サイトの URL に共通するパターン，悪性サイトにアクセスする頻度の高いユーザ群，既知のブラックリストとホワイトリストを用いて検査対象 URL の絞り込みを行うことで効率的に悪性サイトを発見できることを示す．

キーワード：Web アクセスログ，悪性サイト検知

An efficient method to extract malicious websites from massive end-user access log

Shuta Morishima[1]　Hiroki Nakano[1]
Katsunari Yoshioka[2]　Tsutomu Matsumoto[2]　Hiroyuki Hujiwara[3]

Abstract: Drive by download attack has been a huge threat in recent years. The attack redirects web clients from compromised websites and malicious advertisements to malicious websites that exploit vulnerability of clients and distribute malwares. In this report, we propose a method to efficiently extract malicious websites from web access log collected from over 100 thousand users. Since it is inefficient to analyze all access log in detail, we focus on URL patterns shared by malicious websites and users who frequently access malicious websites to extract candidates of malicious URLs for further detailed analysis.

Keywords: Web Access Log, Malicious Websites Detection

1. はじめに

近年ドライブバイダウンロード攻撃が猛威を奮っている．ドライブバイダウンロード攻撃は Web サイトを介してブラウザなどの Web クライアントに対して行われる攻撃であり，改ざんされた正規 Web サイトや不正な広告からクライアントを攻撃サイトに誘導し，クライアントの脆弱性を悪用してマルウェアをインストールさせる．

事例[1,2]では，多数のユーザのアクセスが予想される正規の Web サイトが不正アクセスにより改ざんされ，悪性サイトに誘導される状況となっていた．また，報告[3,4]では，エクスプロイトキットの一種である Rig エクスプロイトキットを利用したドライブバイダウンロード攻撃が近年多数観測されていることが報告されている．ドライブバイダウンロード攻撃への対策として，攻撃者に利用される悪性サイトを早期に発見，対策することが重要である．

本報告では，数十万人規模のエンドユーザから得られる膨大な Web アクセスログを用いて悪性サイトを効率的に発見する方法を提案する．膨大な Web アクセスログに対して個別に詳細な分析を行うことは非効率的であるため，提案手法は高速に検査対象 URL を絞り込む前段処理と，低速であるが精度の高い分析を行う後段処理で構成される．まず前段で悪性サイトの URL に共通するパターン，悪性サイトにアクセスする頻度の高いユーザ群，既知のブラックリストとホワイトリストを用いて膨大な Web アクセスログから悪性サイトの URL 候補を絞り込み，次に後段で検査対象 URL の詳細分析を行うことで悪性 URL を抽出する．

提案手法の評価のため，あるセキュリティベンダから提供をうけた Web アクセスログに提案手法を適用し，悪性サイトの URL の抽出を試みた．その結果，1 日分のアクセスログに含まれる約 1 億の URL から，効率的に悪性サイトを抽出することに成功した．

本報告の構成は次の通りである．まず第 2 章で関連研究について説明する．第 3 章で提案手法について説明し，第 4 章で提案手法に対する評価実験とその結果を説明する．第 5 章で考察を行い，最後に第 6 章でまとめと今後の課題を述べる．

1. 横浜国立大学
Yokohama National University
2. 横浜国立大学大学院環境情報研究院/先端科学高等研究院
Graduate School of Environment and Information Sciences, Yokohama National University / Institute of Advanced Sciences, Yokohama National University
3. トーテックアメニティ株式会社
Totec Amenity Corporation

2. 関連研究

悪性サイトを発見する方法は，URL のシグニチャを基に悪性判定を行う方法と，実際の悪性サイトにアクセスしてコンテンツを取得し分析する方法に大別される．

URL のシグニチャを基に悪性サイトの判定を行う研究としては，論文[5,6]が挙げられる．論文[5]では，一般に公開されているエクスプロイトキットの分析レポートを基に，Rig エクスプロイトキットによる攻撃に利用されるサイトの URL に共通する特徴的な文字列を抽出し，大規模な Web アクセスログからそれらの特徴と合致する URL を抽出することで，当該エクスプロイトキットによる攻撃に利用されるサイトのドメイン名の変化，及び活動期間の分析を行っている．また，論文[6]では，エクスプロイトキットを利用して作成された悪性サイトに表れる特徴的な文字列を文字列長や出現頻度といった要素に分解し，その要素を用いて悪性サイトを判別する手法の提案，評価を行っており，その結果，悪性サイトと良性サイトでは，URL 中の特定の要素の出現回数等に大きな差があることを確認している．これらの手法に対し，我々の提案手法は，まず Web アクセスログから既知のブラックリストを用いて悪性 URL を抽出し，それらの URL 群から悪性サイトに共通するパターンを生成し，次にそれらのパターンを用いて検査対象 URL を絞り込み，後段で詳細分析を行うことで悪性サイトを抽出するので，特定のエクスプロイトキットによる攻撃に利用されるサイトのデータセットや，サイトの URL に共通する特徴などの事前情報を必要としない点で有用である．

悪性サイトにアクセスし，コンテンツを用いて悪性サイトの判定を行う研究としては，論文[7,8]が挙げられる．論文[7]では，難読化の影響を受け難い HTTP ヘッダ情報に着目し，PHP のファイル情報やファイルタイプ，レスポンスの発行時刻情報等を悪性サイト判別の条件として分析を行っている．データセットに対して約 9 割の検知率を達成しており，特定の HTTP ヘッダ情報に着目することは悪性サイトを判別する上で重要な要素であることを示している．また，論文[8]では，アクセスした際にコンテンツ内に存在する JavaScript に着目し，抽象構文解析木から特徴的な木構造を保持する悪性な挙動を行う JavaScript を抽出する方法を提案しており，JavaScript の攻撃パターン別に特徴的な木構造が表れ，悪性サイト検知に有効なことが報告されている．論文[9]では，統計的な手法を用いて Web サイトの良悪性判定に有効な JavaScript の構文木の階層的な特徴を抽出し，それらの特徴を用いて単純ベイズ分類器の学習を行うことで，精度の高い悪性サイトの検知を行っている．論文[10]では膨大な Web ページのデータセットから良性のページを高速にフィルタリングする手法を提案している．当該手法では，検査対象の Web サイトの HTML，JavaScript，URL のそれぞれの特徴を基に機械学習を行い判別器を作成し，高速かつ精度の高いフィルタリングを行っている．これらの手法のように，Web サイトにアクセスし取得したコンテンツを用いる悪性サイトの判定手法は，検査対象の Web サイトが膨大である場合，すべての Web サイトのコンテンツを取得するための処理時間が課題となる．我々の提案手法では，Web サイトの詳細分析を行う前に，高速に検査対象 URL を絞り込む処理を行うことで，膨大な Web アクセスログから効率的に悪性サイトを抽出することができる．

3. 提案手法

3.1 Web アクセスログ

提案手法の入力となる Web アクセスログは，ユーザアクセスした URL，アクセス時間，ユーザ ID の情報を含んでいるものを想定する．なお，今回の検証実験では，あるセキュリティベンダから提供された，数十万規模のユーザに利用されるブラウザセンサから 2017 年 8 月 9 日～8 月 20 日の期間で収集されたアクセスログを利用する．2017 年 8 月 9～8 月 20 日の期間における，当該 Web アクセスログに含まれる全URL数とユニークユーザ数の推移を図1に示す．

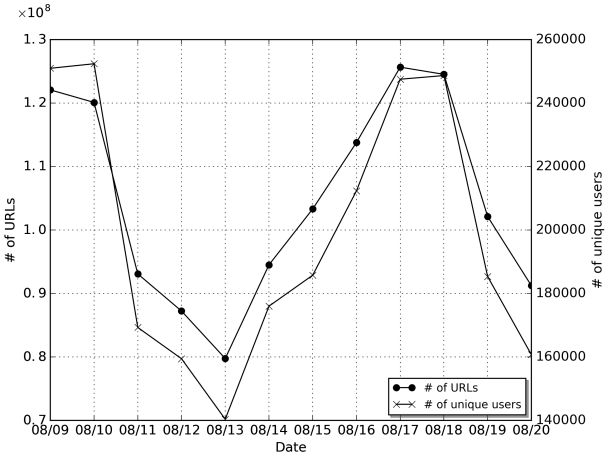

図 1 実験に使用したアクセスログにおける各日の全 URL 数とユニークユーザ数の推移

3.2 既知のブラックリスト，ホワイトリスト

後述する提案手法では既知のブラックリスト，ホワイトリストを利用する．今回の検証実験では，手法で用いる既知のブラックリストとして， Google Safe Browsing[11]（以下 GSB）を，ホワイトリストとして Alexa[12]をそれぞれ利用した．

GSB は Google 社が提供しているブラックリストであり，同社が継続的に悪性サイトのリストを更新している．GSB を利用することで，高速に悪性サイトを抽出することが可能である．

Alexa は世界中のドメインごとのアクセス数を元に作成されたものであり，Alexa 上位ドメインを利用することで，

良性サイトの高速なフィルタリングが可能である．

3.3 提案手法の概要

提案手法の流れを図 2 に示す．提案手法では Web アクセスログを入力とし，まず前段の高速な処理で悪性サイトの URL 候補を絞り込む．次に低速であるが精度の高い分析を行う後段処理で悪性 URL 群を抽出する．

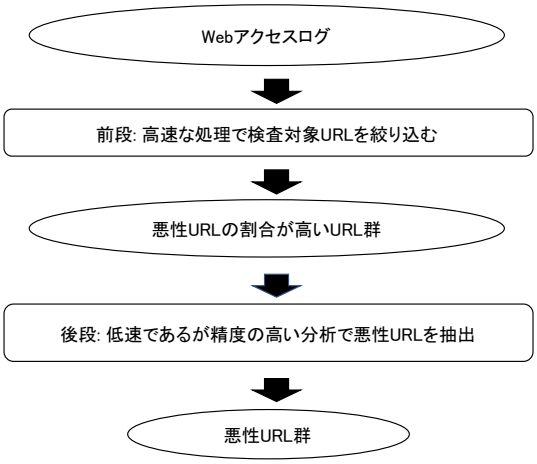

図 2 提案手法概要

本報告では，上記のうち，特に前段処理に関して以下の手法 A，手法 B を提案する．

A) 悪性サイトの URL に共通するパターンを用いた検査対象 URL の抽出

B) 悪性サイトにアクセスする頻度の高いユーザ群を用いた検査対象 URL の抽出

それぞれの方法についての詳細を 3.4, 3.5 節で説明する．悪性サイトの高精度な検知が期待される後段処理については，VirusTotal の URL 検査サービスを利用する．VirusTotal の詳細については 3.5 節で説明する．後段処理の実現方法としては他に Web クローラやクライアントハニーポットによる詳細分析などが考えられる．

3.4 手法 A

3.4.1 手法 A：概要

報告[3,4]では，Rig エクスプロイトキットによる攻撃に利用される悪性サイトの URL に共通する特徴として，URL に特定のパラメータの変数のパターンが用いられていることが報告されている．例として，URL ”hxxp://xxx.com/?a=xxx &b=yyy”のパラメータ変数のパターンは(a,b)となる．このように，同種の悪性サイトには特定のパラメータの変数のパターンが利用される場合があると想定される．手法 A では 1 日分の Web アクセスログを検査対象とし，その日付の当日及び過去の Web アクセスログから，GSB で悪性 URL を抽出し，それらの URL 群から悪性サイトの URL に共通するパラメータ変数のパターンを生成する．次に，それらのパターンを用いて入力の Web アクセスログから悪性サイトの候補を絞り込み，後段の処理で精度の高い分析を行うことで，GSB で未検知である悪性 URL 群を抽出する．

3.4.2 手法 A：処理の手順

Web アクセスログ 1 日分を入力とし，以下の処理で悪性サイトの URL に共通するパターンの生成，及び悪性 URL の抽出を行う．

1. 入力のアクセスログの日付を基準とし，過去 k 日以内のアクセスログから GSB で悪性と検知された URL を抽出する．（k は 1 以上の整数）
2. 処理 1 で抽出された URL をパラメータ変数のパターンでグループ化する．
3. 全てのグループの中，パラメータ変数の数が 2 以上であり，かつグループに含まれる URL のドメインの種類が 2 以上であるものを抽出し，そのパラメータ変数のパターンを悪性サイトの URL に共通するパターンとする．
4. 処理 3 で抽出したパターンと合致する URL を入力のアクセスログから抽出する．その際，GSB で既に検知されている URL は除外する．
5. 処理 4 で抽出した URL を Alexa 上位 10000 ドメインでフィルタリングし，フィルタ後の URL 群を検査対象 URL とする．
6. 検査対象 URL を VirusTotal で詳細分析し，悪性 URL を抽出する．

3.5 手法 B

3.5.1 手法 B：概要

一般的にドライブバイダウンロード攻撃は，改ざんされた正規 Web サイトや不正な広告からクライアントを攻撃サイトに誘導し，攻撃サイトがクライアントの脆弱性を悪用してマルウェアをインストールさせるというように，役割の異なる複数の悪性サイトが攻撃に利用される．しかし，ブラックリストではそれらのサイトの一部しか検知できない可能性がある．また，悪性サイトに頻繁にアクセスするユーザは，通常のユーザと比較して，ブラックリストで未検知の悪性サイトにアクセスする可能性が高いと考えられる．手法 B では，まず前段で Web アクセスログから GSB により検知された URL にアクセスする頻度の高いユーザを抽出し，それらのユーザがアクセスした URL について後段で精度の高い分析を行うことで，GSB で未検知である悪性 URL 群を抽出する．

3.5.2 手法 B：処理の手順

Web アクセスログ 1 日分を入力とし，以下の処理で悪性サイトにアクセスする頻度の高いユーザ群の抽出，及び悪性 URL の抽出を行う．

1. 入力のアクセスログの中，GSB で検知された URL とそれらの URL にアクセスしたユーザを抽出する．
2. 処理 1 で抽出した各ユーザについて，全アクセス数

に占める GSB で検知された URL へのアクセス数の割合を求める．

3. 入力のアクセスログから，処理 2 で求めた割合が閾値以上のユーザがアクセスした URL を抽出する．この際既に GSB で検知されているものは除く．なお，後述する評価実験においては閾値を 1%に設定した．
4. 処理 3 で抽出した URL を Alexa 上位 10000 ドメインでフィルタリングし，残った URL を後段の検査対象とする．
5. 検査対象 URL を VirusTotal で詳細分析し，悪性 URL を抽出する．

3.6 VirusTotal

VirusTotal はユーザから投稿された URL やファイルをウィルス対策エンジンや動的解析システムなどによって解析し，その結果を提示するサービスである．一部の機能は有料であるが，URL スキャン機能や過去の URL のスキャン結果の取得機能は無料で利用することができる．ただし，機能の使用可能回数は一定時間ごとに上限が設定されている．提案手法では検査対象 URL に対し，過去のスキャン結果が存在すればそれを取得し，存在しなかった場合はスキャンを申請し結果を取得することで良悪性を判定した．なお，後述する評価実験においては， VirusTotal のウィルス対策エンジンの中 1 件でも悪性と判定した場合その URL を悪性と判断するものとした．

4. 評価実験

4.1 GSB で検知された URL

2017 年 8 月 9〜8 月 20 日の期間における，Web アクセスログに含まれる全 URL を対象とし，Alexa 上位 10000 ドメインでフィルタリング後，GSB で悪性 URL の抽出を行った．GSB で検知された URL 数の推移を図 3 に示す．

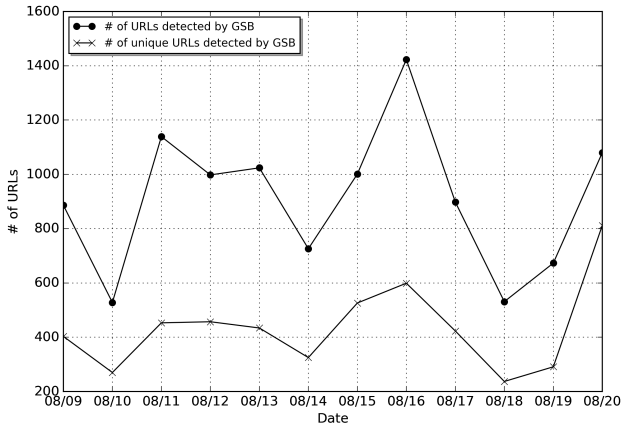

図 3 GSB で検知された URL 数の推移

日付	k の値	パターン数	検査対象 URL 数	悪性 URL 数	割合 (%)
8/9	1	5	19	16	84.2
	7	21	109	20	18.3
	30	80	1337	28	2.1
8/14	1	5	21	3	14.3
	7	14	103	7	6.8
	30	68	867	17	2.0
8/16	1	9	26	6	23.1
	7	16	65	8	12.3
	30	70	1826	27	14.8

表 4 手法 A：実験 1 結果

4.2 評価実験: 手法 A

手法 A の評価のため，以下の２つの実験を行った．

4.2.1 実験 1 概要

2017 年 8 月 9 日, 14 日, 16 日の 3 日間 Web アクセスログをそれぞれ入力とし，手法 A を適用して悪性 URL を試みた．その際，処理 1 の k の値を 1，7，30 の 3 パターンに設定して実験を行い，k の値と抽出される悪性サイトに共通するパターンの数，手法の前段で抽出される検査対象 URL の数，後段の詳細分析で悪性と判定される URL の数，検査対象 URL から抽出された悪性 URL の割合を調査した．

4.2.2 実験 1 結果

実験 1 の結果を表 4 に示す．

実験結果から，k の値が増加するほど後段で抽出できる悪性 URL の数が増加する一方，悪性 URL が抽出されないパターンと，後段で悪性と判定されない検査対象 URL が増加する傾向が確認できた．

4.2.3 実験 2 概要

2017 年 8 月 13 日〜8 月 20 日の期間において，各日の Web アクセスログに対し手法 A を適用し，悪性 URL の抽出を試みた．なお，処理 1 の k の値は k=7 に設定して実験を行った．

4.2.4 実験 2 結果

実験 2 で抽出された悪性パターン数，前段の処理で抽出された検査対象 URL 数，後段で悪性と判定された URL 数，それらの URL に含まれていたユニークなドメイン数，検査対象 URL から抽出された悪性 URL の割合を表 5 に示す．

実験結果から，実験対象期間において，手法 A により GSB で未検知であった悪性 URL を 1 日あたり平均で約 23 件抽出した．また，1 日分のアクセスログから手法の前段で抽出された検査対象 URL に占める，後段で検知された悪性 URL の割合の平均は約 11.4%であった．

日付	パターン数	検査対象 URL 数	悪性 URL 数	悪性ドメイン数	割合 (%)
8/13	13	126	34	12	27.0
8/14	14	103	7	6	6.8
8/15	14	58	4	4	6.9
8/16	16	65	8	7	12.3
8/17	17	131	20	18	15.3
8/18	18	395	29	28	7.3
8/19	15	524	62	55	11.8
8/20	14	540	22	18	4.1

表 5　手法 A：実験 2 結果

実験結果から, 悪性 URL の抽出に特に有効であったパターンの例について，以下で説明する．

(1) mano, pano, {shop または work または gift}

実験 2 の対象期間において，これらのパターンに該当するユニークな URL が，GSB で 478 件，提案手法の後段で 17 件検知された．これらのパターンは実験の実施と同時期に報告された Rig EK の URL のパターンと一致していることから，抽出された URL は Rig EK による攻撃に関連するサイトのものであると推測される．

(2) sessionid, {sslchannel または securessl}

実験 2 の対象期間において，これらのパターンに該当するユニークな URL が，GSB で 128 件，提案手法の後段で 28 件検知された．これらの URL はドメイン名，パス名に類似性がみられることから，同種の悪性サイトであると推測される．現時点でこれらのサイトが利用される攻撃に関する情報は得られていないので，調査を継続する．

4.3 評価実験: 手法 B

4.3.1 実験概要

2017 年 8 月 15 日〜8 月 20 日の期間の Web アクセスログに対し手法 B を適用し，悪性 URL の抽出を試みた．

4.3.2 実験結果

実験 2 で検査対象としたユーザ数，前段の処理で抽出された検査対象 URL 数, 後段で悪性と判定された URL 数, それらの URL に含まれていたユニークなドメイン数, 検査対象 URL から抽出された悪性 URL の割合を表 6 に示す.

実験結果から，実験対象期間において，手法 B により GSB で未検知であった悪性 URL を 1 日あたり平均で約 318 件抽出することができた．また，1 日分のアクセスログから手法の前段で抽出された検査対象 URL に占める, 後段で検知された悪性 URL の割合の平均は約 9.2%であった.

日付	検査対象ユーザ数	検査対象 URL 数	悪性 URL 数	悪性ドメイン数	割合 (%)
8/15	35	4690	254	56	5.4
8/16	38	4997	358	45	7.2
8/17	32	3725	267	38	7.2
8/18	28	3075	277	36	9.0
8/19	20	1736	195	47	11.2
8/20	33	3575	554	43	15.5

表 6 手法 B：実験結果

5. 考察

本章では，実験結果を元に提案手法の有効性について考察を行う．

効率の観点では, 実験に用いた Web アクセスログ 1 日分に含まれる全 URL 数が約 1 億件，その中 GSB で検知されるものが数百件であることを考慮すると，評価実験において，1 日分のアクセスログから提案手法の前段で抽出された検査対象 URL に占める，後段で検知された悪性 URL の割合が手法 A, B でそれぞれ約 11.4%, 9.2%であったことから, 提案手法は当該 Web アクセスログから効率的に悪性サイトを抽出できたと言える．

手法 A については，悪性サイトに関する事前情報や詳細な分析を必要とせず，Rig エクスプロイトキットに利用される悪性サイトなど, 悪性サイトの URL 共通するパターンを抽出することができることが示された．今後手法 A による悪性サイトの抽出を継続することで，新規のエクスプロイトキットによる攻撃が行われた場合も，攻撃に利用される悪性サイトを抽出できる可能性がある．

手法 B により GSB で未検知であった悪性サイトの URL を多数抽出することができたことから，ユーザの傾向に着目した悪性サイトの抽出は有効であると考えられる. なお, 今回の評価実験では 1 日分の Web アクセスログを基に悪性サイトへのアクセス頻度が高いユーザを抽出したが，より長期間のアクセスログからユーザの傾向を分析し，継続的に悪性サイトにアクセスするユーザを抽出する手法も検討する価値があると考えられる．当該手法の提案については今後の課題とする．

今回の評価実験においては，提案手法の後段の処理は VirusTotal の URL 検査サービスを利用し, VirusTotal のウィルス対策エンジンの中 1 件でも悪性と判定した場合検査対象の URL を悪性と判断した．しかし，これらのウィルス対策エンジンのいずれかが誤検知をする可能性も考えられるため，より精度の高い判定を行うために，後段の処理を Web クローラやクライアントハニーポットで行うことも有効であると考えられる．

6. まとめと今後の課題

本報告では，数十万人規模のエンドユーザから得られる膨大な Web アクセスログを用いて悪性サイトを効率的に発見する方法を提案した．提案手法は高速に検査対象 URL を絞り込む前段処理と，低速であるが精度の高い分析を行う後段処理で構成され，本報告では特に前段処理について，悪性サイトの URL に共通するパターンを用いた検査対象 URL の抽出(手法 A)と，悪性サイトにアクセスする頻度の高いユーザ群を用いた検査対象 URL の抽出(手法 B)を提案し，後段処理については，VirusTotal の URL 検査サービスを利用した．また，提案手法の評価のため，実際の Web アクセスログを用いて悪性 URL の抽出を行った．その結果，手法 A では 2017 年 8 月 13 日～2017 年 8 月 20 日の実験期間において， 各日の Web アクセスログから GSB で未検知であった悪性 URL を平均で約 23 件抽出した．手法 B では 2017 年 8 月 15 日～2017 年 8 月 20 日の実験期間において，各日の Web アクセスログから GSB で未検知であった悪性 URL を平均で約 318 件抽出した．手法 A,B の効率については，前段で抽出された検査対象 URL に占める，後段で検知された悪性 URL の割合の平均は手法 A で約 11.4%，手法 B で約 9.2%であり，これらの手法により膨大な Web アクセスログから効率的に悪性サイトを抽出できたと考えられる．

今後の課題としては，手法 A による悪性サイトの抽出を継続し，新規のエクスプロイトキットによる攻撃に利用される悪性サイトを検知できるかを検証することに取り組みたいと考えている．また，手法 B については，本報告では 1 日分の Web アクセスログから悪性サイトへのアクセス頻度が高いユーザを抽出し，それらのユーザがアクセスする URL を詳細分析の対象としたが，今後はより長期間のアクセスログからユーザの傾向を分析し，継続的に悪性サイトにアクセスするユーザを抽出する手法の提案や，それらのユーザがアクセスするサイトのコンテンツに着目することで，悪性サイトにアクセスするユーザ群とユーザがアクセスする Web コンテンツの嗜好性の関係の分析を行う手法の提案に取り組む価値があると考えている．

謝辞

本研究の一部は文部科学省国立大学改革強化推進事業の支援を受けて行われた．

本研究成果の一部は、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」の支援により得られた．

参考文献

[1] “弊社ホームページ改ざんに関するお詫びと復旧のご報告”. http://www.narita-airport.jp/jp/news/150305/, (参照 2017-08-26).

[2] “ローソンホームページ「採用」サイト改ざんについて”. http://www.lawson.co.jp/company/news/010798/, (参照 2017-08-26).

[3] “RIG エクスプロイトキット解析レポート”. https://www.nttsecurity.com/-/media/nttsecurity/files/resource-center/what-we-think/rigek-analysis-report.pdf, (参照 2017-08-26)

[4] “Rig Exploit Kit によるドライブ・バイ・ダウンロード攻撃の検知状況”. https://www.ibm.com/blogs/tokyo-soc/rig-exploit-kit/, (参照 2017-08-26)

[5] 嶌田一郎，太田敏史，岡田晃市郎，山田 明， “ユーザ環境における RIG Exploit Kit の実態調査方法の提案”, 電子情報通信学会信学技報 ICSS2017-15.

[6] 佐藤祐磨，中村嘉隆，高橋修，“エクスプロイトキットで利用される文字列特徴を用いた悪性 URL 検出手法の提案”，情報処理学会研究報告 Vol.2016-CSEC-72 No.25.

[7] 酒井裕亮，佐々木良一，“Drive By Download 攻撃に対する HTTP ヘッダ情報に基づく検知手法の提案”，情報処理学会研究報告 Vol.2013-CSEC-60 No.29.

[8] 神園雅樹，西田雅太，小島恵美，星澤裕二，“抽象構文解析木による不正な JavaScript の特異点抽出手法の提案”，情報処理学会論文誌 Vol.54 No.1 349-456.

[9] C. Curtsinger， B. Livshits， B. Zorn, C. Seifert, ” ZOZZLE: Fast and Precise In-Browser JavaScript Malware Detection”， USENIX Security, 2011

[10] D. Canali, M. Cova, G.Vigna, C. Kruegel, ” Prophiler: a fast filter for the large-scale detection of malicious web pages”, International conference on World wide web (WWW), 2011

[11] ”Google Safe Browsing”. https://developers.google.com/safe-browsing/, (参照 2017-08-26)

[12] “The top 500 sites on the web”. https://www.alexa.com/topsites, (参照 2017-08-26)