Vol.2015-DPS-162 No.48
Vol.2015-CSEC-68 No.48
2015/3/6

Drive-by Download攻撃対策フレームワーク実現に向けたリンク構造解析によるWebサイトの分析

松中 隆志[1,a)]　山田 明[1,b)]　窪田 歩[1,c)]

概要：著者らは，ユーザによるWebサイトの閲覧と同時に取得されるWebアクセスログをもとにWebを監視して，Drive-by Download攻撃サイトを検出・通報するフレームワークFCDBD(Framework for Countering Drive-By Download)を提案している．著者らは，FCDBD上でDrive-by Download攻撃におけるLandingサイトを検出する手法として，Webページから`<iframe>`タグなどで自動的に遷移する遷移先サイトの変化に着目して，改ざんによりLandingサイトとなったWebページを検出する手法を提案した．しかし，正規サイトにおいても主に広告サイト，トラヒック解析サイトなどでJavaScriptなどを用いて遷移先サイトを動的に変化させているサイトが多く存在しており，本来観測したい改ざんによる変化のみを抽出するのは困難である．本稿では，Webサイトのリンク構造の分析結果にもとづいて本手法のさらなる改良を提案する．

キーワード：Drive-by Download攻撃，Webリンク構造解析

A Feasibility Study for Enhancing the Framework for Countering Drive-by Download Attacks with Analysis of Web Link Structures of Websites

Abstract: The authors proposed the Framework for Countering Drive-By Download (FCDBD) which monitors the Web by utilizing web access logs from users and detects malicious websites related to the drive-by download attacks. Monitoring link-related behaviors is one of the approaches to detect the malicious websites in the framework. The authors proposed a detection method for the Landing site of Drive-by Download attacks. The method focused on the change of referred websites from a webpage. However, a legitimate webpage has many changes of referred websites caused by advertisement websites or traffic analysis websites. Therefore, it is hard to extract the change caused by the defacement and detect the Landing site correctly. In this paper, the authors propose the improved method for detecting the Landing site of drive-by download attacks.

Keywords: Drive-by download attack, Web link analysis

1. はじめに

Drive-by Download攻撃は，Web上における主要な脅威の一つである．この攻撃は，Webを利用してマルウェアを拡散する攻撃であり，ユーザは，攻撃が仕掛けられたWebページにアクセスするだけでマルウェアに感染させられてしまう．図1にDrive-by Downloadの典型的な攻撃フローを示す．攻撃者はユーザ環境(OS，ブラウザ，プラグインなど)の脆弱性を攻撃するサイト(Exploitサイト)，マルウェアを配布するサイト(Distributionサイト)，Exploitサイトまでユーザをリダイレクトさせるサイト(Intermediateサイト)を用意し，さらにIntermediateサイトへユーザをリダイレクトさせるサイト(Landingサイト)を準備する．Landingサイトは多くの場合正規のサイトが改ざんされて，Intermediateサイト，Exploitサイトへユーザをリダイレクトさせるスクリプト(JavaScript，PHPなど)を

[1] (株)KDDI研究所
KDDI R&D Laboratories Inc., 2–1–15 Ohara, Fujimino, Saitama 356-0003, Japan
[a)] ta-matsunaka@kddilabs.jp
[b)] ai-yamada@kddilabs.jp
[c)] kubota@kddilabs.jp

埋め込まれることにより Landing サイトとなる．ユーザが Landing サイトにアクセスすると，Intermediate サイトを経て Exploit サイトまでリダイレクトされる．Exploit サイトでは，ユーザの環境に合わせて攻撃コードを配信する．攻撃が成功すると，ユーザは Distribution サイトからマルウェアを自動的にダウンロードさせられ，最終的にマルウェアに感染する．Provos らの報告 [1] によると，Drive-by Download 攻撃に係る悪性サイト (Exploit サイト，Distribution サイト) は生存期間が短く，発見や解析が非常に困難である．

Drive-by Download 攻撃の早期発見, 検出および防御を目的として，著者らはフレームワーク (FCDBD: Framework for Countering Drive-By Download) を提案, 実装した [2], [6]．このフレームワークでは，ユーザが使用するブラウザおよび Web プロキシに観測センサを設置することで広域な観測網を構築し，ユーザの Web アクセスに関する情報を提供してもらうことで Web 上の Drive-by Download 攻撃に係る脅威をリアルタイムに把握する．そして検出された脅威の情報を観測センサに適宜フィードバックすることで，ユーザが攻撃の被害にあうのを未然に防ぐ．観測センサより収集される膨大な Web アクセスログなど情報をもとに Web ページのリンク構造の解析などを有効活用した悪性サイトの早期発見が期待できる．収集される Web アクセスログを用いた Drive-by Download 攻撃に係る悪性サイトの検出手法として，著者らはこれまでコンテンツのダウンロードに至るページ遷移の振る舞いにもとづく Drive-by Download 攻撃におけるマルウェアのダウンロードを検出する手法 [3]，Web ページの遷移元/遷移先のサイト数に着目して Exploit サイトを検出する手法 [5]，さらに Web ページからの遷移先サイトの変化に着目して正規のサイトが改ざんされた Landing サイトを検出する手法 [6] を提案した．しかし，遷移先サイトの変化に着目した Landing サイトの検出手法において，実際は，主に広告サイト，トラヒック解析サイトなどで JavaScript などを用いて遷移先サイトを動的に変化させているサイトが多く存在しており，本来観測したい改ざんによる変化のみを抽出するのは困難であった [4].

本稿では，本手法のさらなる改良を提案する．本稿では，上記の遷移サイトの変化に着目した Landing サイトを検出する手法において．各 Web ページの遷移先サイトの変化を監視し，新たな遷移先サイトが観測された際に，その新たな遷移先サイトの遷移元/遷移先のサイト数を調べ，その数が悪性サイトと同様の特徴を示していた場合に当該 Web ページを Landing サイトとする方法を提案する．これにより，より小さい誤検知率で Landing サイトと疑われるサイトを抽出できることが期待される．

本稿では，実際に FCDBD のブラウザ型の観測センサを用いて正規の Web ページにアクセスして取得した Web アクセスログをもとに，誤検知率 (False Positives) の評価を実施した．本評価によって，遷移先サイトの変化とその遷移先サイトのリンク構造をあわせて解析することで誤検知率を低減できる見込みを得たので，その内容を報告する．

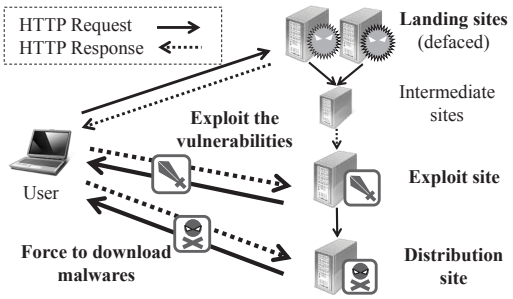

図 1 Drive-by Download 攻撃フロー例

具体的には，まず正規の Web ページを 2 週間にわたり観測し，遷移先サイトの変化を解析した．その結果，1 日あたり 6.0%のサイトにおいて遷移先サイトが変化している，新たな遷移先サイトが出現していることが確認された．次に，上記の観測にて確認された新たな遷移先サイトにおいて，さらに遷移元/遷移先のサイト数を解析したところ，新たな遷移先サイトにおいて [5] で提示した悪性サイト (Exploit サイト) の特徴を示すサイトに遷移していた正規サイトは 1.5%であった．このことから，遷移先サイトが変化している Web ページを抽出し，その後，新たに見られた遷移先サイトの遷移先／遷移元のサイト数を解析することで，誤検知率が低減できると考えられる．

2. 関連研究

Drive-by Download 攻撃の発見，検出

Drive-by Download 攻撃サイトを発見，検出する手法の一つとして，Web クローラ (honeyclient) を用いた Web サイトの巡回 (クローリング) がある [7]，[8]．honeyclient で効率的に悪性サイトを検知するためには，クローリングの起点となる seed を適切に与える必要がある．また，攻撃者が，自身の悪性サイトの検出を防ぐために，セキュリティ関連企業，研究機関によるクローリングと思われるアクセスに対して正常の Web サイトのようにふるまう (cloaking) ような対策を行うこともあり，クローリングによる悪性サイトの発見，検知は非常に困難である．さらに，悪性サイトの生存期間は数時間程度と短命なため，その実態をリアルタイムに把握することは困難である．

リンク構造に着目した Drive-by Download 攻撃の検出

悪性サイトを検出する手法として，Web ページ間の遷移関係の構造 (リンク構造) に着目して，未知の悪性サイトを検出する方法が提案されている．Zhang ら [10] の手法は，Drive-by Download 攻撃事例の HTTP トラヒック情報から悪性サイトのリンク遷移元元をたどり，URL の類似性などを考慮して複数の悪性サイトに共通のハブと

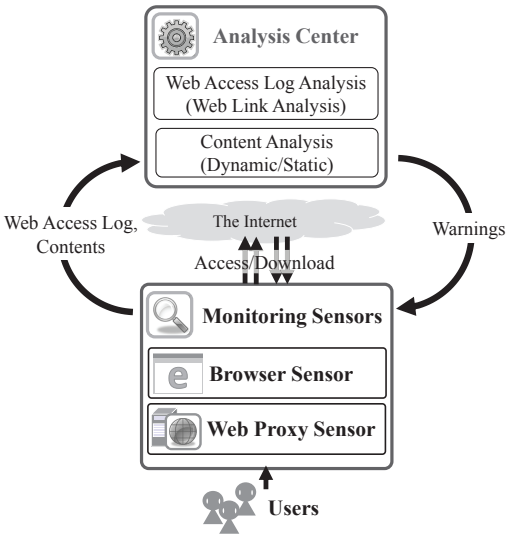

図 2 FCDBD 構成概要図

なるサイト (central server) を検出し，そのサイトをもとに MDN(Malware Distribution Network) を検出する．そして，その central server の URL の特徴をシグネチャとして，既知の MDN に属する未知の悪性サイトを検出するものである．Stringhini ら [11] の手法は，リンク構造上の特徴に加えてユーザのマシンの環境 (OS，ブラウザ，プラグインなど) も加味して，特定のマシン環境のユーザのみが到達するリンクのパスを抽出することで，膨大な Web アクセスログから Drive-by Download 攻撃の悪性サイトを検出するものである．Wand ら [12] の手法は，[10] による MDN の検出の後に，Landing サイトの HTML の内容，URL の特徴を抽出して，未知の悪性サイトを検出するものである．これらの手法は，新たな MDN の検出，URL，コンテンツなどの特徴の抽出のためには膨大な攻撃事例のデータが必要である．そのため，未知の MDN に属する Landing サイトなど悪性サイトの検出には即時的に対応できない．また，MDN の検出，特徴の抽出のための攻撃事例のデータ収集をいかに行うか，も重要な課題となる．

3. FCDBD: Framework for Countering Drive-By Download

3.1 FCDBD 概要

FCDBD(Framework for Countering Drive-By Download) は，Drive-by Download に係る悪性サイトの早期発見，検出および防御を目的としたフレームワークである [2]，[6]．図 2 に FCDBD フレームワークの構成を示す．FCDBD フレームワークはユーザ側に配置される観測センサと，観測センサから提供された情報を解析する解析センタからなる．

観測センサは，ユーザの Web ブラウザ (ブラウザセンサ) および Web プロキシサーバ (Web プロキシセンサ) に設置され，ユーザの Web アクセスに関する情報を観測し，得られた情報を解析センタに送信する．ブラウザセンサは Web ブラウザのプラグインソフトウェアとして実装され，表 1 に記載した内容を含む Web アクセスログを解析センタに送信する．その際，個々のブラウザセンサは ID で識別されるが，この ID は Web ブラウザが起動されるごとにランダムに変更されるため，解析センタ側で同一ユーザの Web アクセスログを追跡できない．ブラウザセンサはまた取得した Web サイト上のコンテンツをセンタに送信する．その際，解析センタ側で悪性が疑われると判断されたサイトのコンテンツのみをセンタからの要求に応じて送信する．

表 1 Web アクセスログの主な内容

・観測センサの ID
・ユーザがアクセスした URL
・ダウンロードしたコンテンツのハッシュ値
・Web ページ遷移時のマウスイベントの有無
・HTTP Request/Response ヘッダ

解析センタは，観測センサから送信された情報を解析し，悪性と思われるサイトを検出する．そして，観測センサに検出された悪性サイトの情報を送信し，観測センサを利用するユーザが悪性サイトにアクセスするのを防ぐ．解析センタで行う解析として，Web アクセスログを用いた各 Web ページのリンク構造の解析，コンテンツ情報を用いたコンテンツ解析がある．Web ページのリンク構造の解析事例として，著者らはこれまで各 Web ページの遷移先ホストの変化を観測し，不明なホストへの遷移が観測されるような変化がみられた場合に改ざんなどによって Landing サイトとなったとする手法 [6]，各 Web ページの遷移先/遷移元サイトの数から Exploit サイトとみられるサイトを抽出する手法 [5]，および観測センサからの Web アクセスログの送信に対してリアルタイムに解析を行い，即時的に判定結果を観測センサに返答するための処理として，Web ページのアクセスから引き起こされる一連のページ遷移の挙動を監視し，Drive-by Download 攻撃におけるマルウェアのダウンロードに至るまでの Web ページの遷移に関する挙動と似た特徴を示した場合に，ダウンロードされた実行形式ファイルを悪性とみなす手法 [3] を提案した．前者 2 つについての詳細は後述する．

また，観測センサから収集されたコンテンツは，動的解析/静的解析 [13] により詳細に解析される．その結果，悪性と判定されたコンテンツの情報 (URL，ハッシュ値) はブラックリストとして解析センタ内に格納される．以降，観測センサのユーザがブラックリスト内のコンテンツにアクセスしようとした際には，解析センタから観測センサに警告を送信する．また，良性と判定されたコンテンツで，当該コンテンツの情報がブラックリストに記載されていた場合は，当該コンテンツの情報を即座にブラックリストから削除することで以降の誤検知を防ぐ．

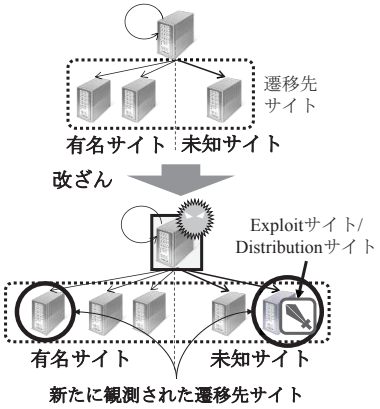

図 3 遷移先サイトの変化に着目した Landing サイト検出手法

改ざん時

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01
Transitional//EN">
<HTML>
...
</DIV><!--c47665--><script type="text/javascript"
src="hxxp://[redacted].co.za/.../nfyvrkb8.php?id=2
544721"></script><!--/c47665-->
...
</HTML>
```

修復後

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01
Transitional//EN">
<HTML>
...
</DIV>
...
</HTML>
```

図 4 Landing サイトのコンテンツ事例

3.2 Web アクセスログを利用した Drive-by Download 攻撃検出手法

FCDBD の観測センサから収集される Web アクセスログを利用した Drive-by Download 攻撃の検出手法として，著者らはこれまで Web ページ間のリンク構造に着目した手法を提案した [6]，[5]，[3]．以下にこれらの手法について説明する．

手法 1: 遷移先サイトの変化に着目した Landing サイト検出手法

図 3 に概要を示す．本手法では，Web ページの遷移先サイトを監視し，遷移先サイトに新たなサイトが追加されていることが観測された場合に，当該 Web ページが改ざんされ Landing サイトなった疑いがあるとして検出するものである．しかし，Web ページが管理者によって正規に内容を変更される場合にも，遷移先サイトが変化する可能性がある．そのような正規な事例と区別するために，本手法では遷移先サイトを google，twitter などの有名なサイトとそれ以外の未知のサイトにわけ，それ以外の未知のサイトが新たに遷移先サイトとして追加された際にのみ悪性化を疑うこととした．なお，有名なサイトのリストとして Alexa[14] がサイト内で提示しているアクセスランキング (全世界でのトップ 500，日本でのトップ 500) を参照した．さらに，遷移先サイトの変化を引き起こす要因の一つとして，広告系サイトによる動的な遷移先サイトの変更がある．そこで本手法では，Firefox などのプラグインソフトウェアである Adblock Plus[15] で用いられているフィルタを利用して，広告系サイトと思われるサイトをログから取り除いて判定を行う．[6]

図 4 は，セキュリティ関連企業から当該企業が提供する検査機構により悪性と判定された URL の提供を受け，実際に当該 URL へアクセスして取得した Landing サイトのコンテンツの一例である．図 4 の例 (改ざん時) では，単純に悪性スクリプトへ誘導する<script>タグが挿入されていた．後日，再度同じ Web ページにアクセスすると，該当する箇所は削除されていた (修復後)．改ざん時と修復後の Web アクセスログを比較すると，改ざん時には<script>タグで指定されたサイトへのアクセスが確認できたが，修復後は当該サイトへのアクセスは確認されなかった．このように，正規 Web ページが改ざんされて Landing サイトとなる際には，遷移先サイトに違いがみられると考えられる．

手法 2: 遷移先/遷移元のサイト数にもとづく Exploit サイト検出手法

図 5 に Drive-by Download 攻撃に係る悪性サイトのページ遷移の事例を示す．本事例はマルウェア対策研究人材育成ワークショップ [16] でマルウェア対策技術の技術者の育成および研究促進を目的として配布されているデータセット，D3M(Drive-by Download Dataset by Marionette) データセット [17] に収録されている攻撃事例のうちの 1 つである．攻撃者はマルウェアを広域に拡散させるために，複数の Web ページを改ざんし，自身の Exploit サイトへ転送させるスクリプトを埋め込む．そのため，図 5 のように，Exploit サイトは複数の遷移元サイトを有するような形になる．そして Exploit サイトではユーザに対して攻撃を仕掛け，最終的にマルウェアをダウンロードさせるため，Exploit サイトからの遷移先サイトは単一サイト (Exploit サイト/Distribution サイト) のみとなると考えられる．一方，正規の Web ページにおいては，例えば広告ネットワークは，Web ページにアクセスしたユーザを複数の広告主のコンテンツへ誘導するため，Exploit サイトと同様に複数のサイトから参照される一方で，複数の遷移先サイトへ遷移する構成になると考えられる．

以上より，ある Web ページの遷移元サイトの数を *#fan-in*，遷移先サイトの数を *#fan-out* とすると，$\#fan\text{-}in > 1 \wedge \#fan\text{-}out = 1$ となる Web ページは Exploit サイトの疑いがあると判定する．

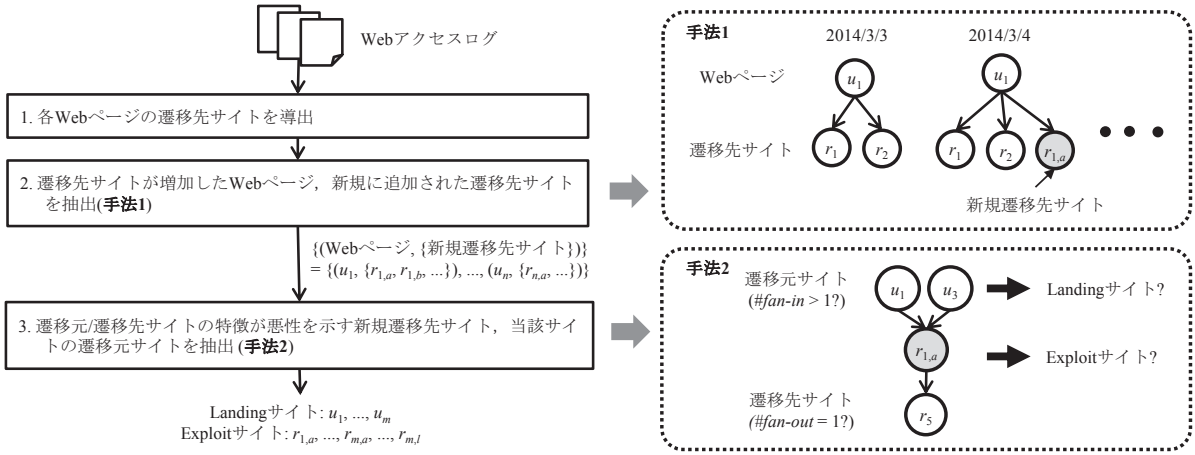

図 7 改良された Landing サイト検出方法

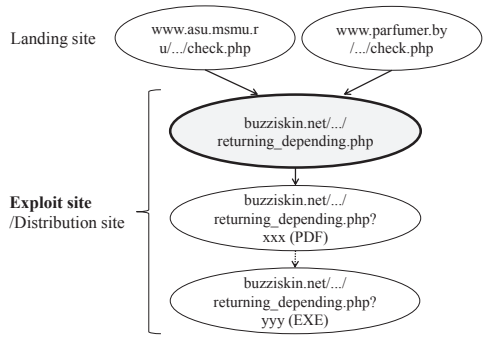

図 5 Drive-by Download 攻撃におけるページ遷移の事例

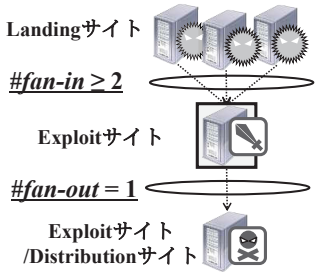

図 6 遷移先/遷移元のサイト数にもとづく Exploit サイト検出手法

4. 手法 1 と手法 2 を組み合わせた Landing サイト検出手法の改良

図 7 に本稿で提案する Landing サイト検出手法の改良方法を示す．観測センサから送られてくる Web アクセスログをもとに，まず各 Web ページの遷移先サイトを導出する．その際，元に Web ページから 1 ホップ先の遷移先サイトのみを対象とする．次にその Web ページが過去にアクセスされたものである場合に，過去の遷移先サイトと現在の遷移先サイトを比較し，新たに遷移先サイトとして追加されたサイト (新規遷移先サイト) があるかどうか調べる．そして，新たな遷移先サイトが存在する場合は，その新たな遷移先サイトと元の Web ページを抽出する．最後に，その抽出された新規遷移先サイトに対して，過去の解析結果から遷移元/遷移先サイトの数を調べ，その数が悪性サイトの特徴と同様の特徴を示す場合，すなわち遷移元サイトの数を *#fan-in*，遷移先サイトの数を *#fan-out* とすると，$\#fan\text{-}in > 1 \wedge \#fan\text{-}out = 1$ を満たす場合に，該当する新規遷移先サイトを Exploit サイト，その新規遷移先サイトの遷移元である Web ページを Landing サイトとして抽出する．

表 2 Web アクセスログの諸元

・アクセスした Web ページ数: 1,684
・ログ取得期間: 2014/3/3 – 2014/3/14(土日は除く)
・遷移先サイトを含めた総 URL 数: 7,899

5. Landing サイトの検出手法の評価

3.2 節で述べた遷移先サイトの変化に着目した Landing サイトの検出手法について，FCDBD のブラウザセンサで実際の正規 Web ページにアクセスして取得した Web アクセスログを用いて誤検知率の評価を実施した．

評価に用いた Web アクセスログについて

利用した Web アクセスログの諸元を図 2 にまとめる．利用した Web アクセスログは，正規の Web ページ 1,684URL に 1 日 1 回，2014 年 3 月 3 日から 2014 年 3 月 14 日までアクセスして取得された．アクセスの対象とした Web ページは，セキュリティ製品のログをもとに事前に解析し，遷移先サイト数が比較的多いと推測される URL を抽出したものである．ログの取得には，FCDBD のブラウザセンサを用いた．Web ブラウザは Internet Explorer 8 を利用した．また，今回アクセスした Web ページはいずれもアクセスした期間内に改ざんされた報告が確認されなかったものである．よって，検出手法の評価によって検出されてしまった事象はすべて偽陽性 (false positives) の事象となる．

実施内容

評価の実施内容について図 8 に概要を示す．はじめに

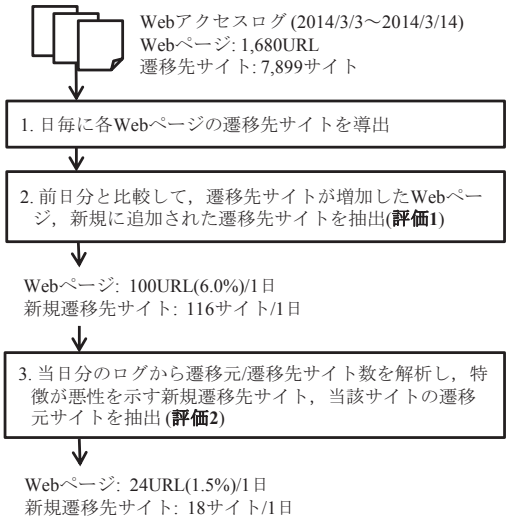

図 8 評価の実施内容

3.2 節の手法 1 の評価を実施した．上述の Web アクセスログを用いて正規 Web ページ 1,680URL について遷移先サイトを導出し，遷移先サイトの日ごとの変化を解析した．そして，前日と比較して遷移先サイトの増加がみられた正規 Web ページ，および増加した遷移先サイト (新規遷移先サイト) を抽出した (**評価 1**)．次に，遷移先サイトの増加がみられた正規 Web ページにおいて，各正規 Web ページの新規遷移先サイトの遷移元/遷移先サイトの数を，同日のログを用いて解析し，3.2 節の手法 2 で示したような悪性 (Exploit サイト) の特徴を示すかどうかを調べた．そして，悪性の特徴を示す新規遷移先サイト，および当該新規遷移先サイトの遷移元である正規 Web ページを抽出した (**評価 2**)．以上の実施内容の結果を集計して，手法 1 のみ，および手法 1 にさらに手法 2 を適用したときのそれぞれの誤検知 (false positives) 率を評価した．

評価結果

図 9，図 10 に評価の結果を示す．図 9 は，評価 1，評価 2 それぞれの 1 日あたりに抽出された Web ページ，遷移先サイトの数を示す．図 10 は，評価 1，評価 2 それぞれの 1 日ごとの抽出 Web ページ数，遷移先サイト数を重複なくカウントした累積総数の変化を示す．また表 3 に収集したログに含まれる 1 日あたりのアクセスした Web ページ，遷移先サイトの総数を示す．表 3 の「累積」は，全日のログに含まれる Web ページ，遷移先サイトを重複なくカウントした総数を示す．

評価 1. 図 9 より，1 日あたり 78〜125 ページ (4.6%〜7.5%) の Web ページにおいて遷移先サイトが変化したことがわかる．また 1 日あたりに抽出された遷移先サイトの数は 83〜171 サイト (1.6%〜3.2%) であった．また図 10 より，2 週間の累積では 408 ページ (総ページ数の 24.2%)，492 サイト (総サイト数の 2.3%) が抽出された．また図 10 より，1 日あたりに新たに抽出される Web ページの数は，観測初期はおおよそ 70 ページ (4.1%) であったが，観測後期にはおおよそ 20 ページ程度 (1.2%) になっており，観測期間が長くなるに従い新規の抽出数が収束していくことがわかる．

評価 2. 図 9 より，1 日あたりに抽出される Web ページの数が 12〜55 ページ (0.7%〜3.3%) に減少しており，評価 1 に比べて 1 日あたり 3.5%〜5.8%の改善が確認された．また 1 日あたりに抽出される遷移先サイト数も 10〜40 サイト (0.2%〜0.7%) に減少していた．また図 10 より，2 週間の累積では 150 ページ (8.9%)，87 サイト (0.5%) が抽出された．

表 4 遷移先サイト (87 件) のカテゴリ

カテゴリ名	サイト数
Web Advertisement	29
Computers/Internet	21
Internet/Infrastructure /Search Engines/Portals	11
Business/Economy/Finantial	10
Blogs/Web Communications	3
Pornography/Adult/Mature	3
Arts/Entertainment/Games	3
Auctions/Shopping	2
Brokerages/Trading	2
Health/Society/Life Style	2
News/Media	1

6. 考察

誤検知の事例

評価 2 で誤検知された遷移先サイト 87 件について，サイトのカテゴリを調べた結果を表 4 に示す．なお，カテゴリの調査に Trend Micro 社の Site Safety Center [18] の Web ページ内で提供されている Web レピュテーションサービスを利用した．結果，広告 (Web Advertisement) 系がもっとも多く，続いてトラヒック解析系 (Computer/Internet/Infrastructure) のサイトが多く存在することがわかった．広告系のサイトは，比較的サイズの小さい画像，テキストへ遷移させるケースが多くみられる．また，トラヒック解析系のサイトの多くは，例えば webbug のようなサイズの非常に小さい画像ファイルをトラヒック解析対象の Web ページに挿入して，当該ファイルへのアクセス履歴を解析することで，Web ページ間のユーザの行動のトラッキングなどを行う．そのため，遷移先サイトのコンテンツ情報を利用して，例えばサイズがある値より小さい遷移は遷移先としてカウントしない，特定のコンテンツタイプへの遷移はカウントしない，などの対策が考えられる．

遷移元/遷移先サイト数の解析について

評価 2 の遷移元/遷移先サイト数の解析において，上記の評価では 1 日分のログを用いて解析を行い評価した．利用するログの日数に応じた評価への影響を調査するために，利用するログの日数を 2014 年 3 月 3 日から 3 月 14 日の 1 日〜10 日分までに設定し，それぞれの日数分のログを用い

表 3 1 日あたりの総アクセス数

日付 (YYMMDD)	140303	140304	140305	140306	140307	140310	140311	140312	140313	140314	累積
Web ページ数	1,684	1,684	1,684	1,684	1,684	1,684	1,684	1,684	1,684	1,684	1,684
遷移先サイト数	5,397	5,315	5,240	5,375	5,192	5,318	5,158	5,516	5,285	5,232	7,889

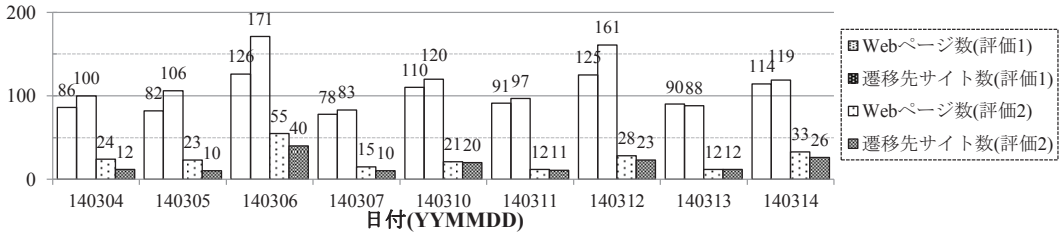

図 9 評価結果 (1 日あたりの平均: (評価 1)Web ページ:100.2(6.0%), 遷移先サイト:116.1(2.2%), (評価 2)Web ページ:24.8(1.5%), 遷移先サイト:18.2(0.3%))

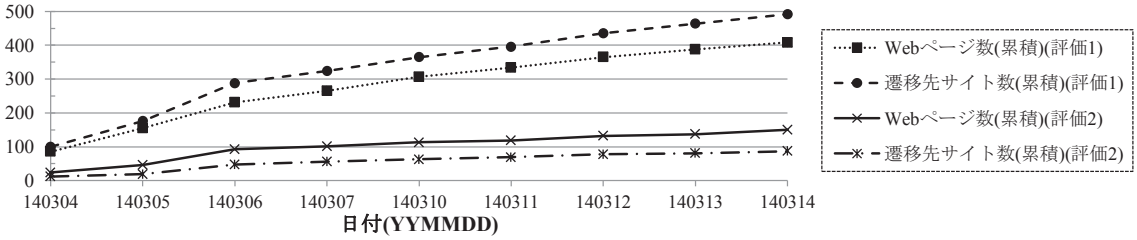

図 10 評価結果 (累積)

て各サイトの遷移元/遷移先サイトを解析して，悪性サイトの特徴 ($\#fan\text{-}in > 1 \wedge \#fan\text{-}out = 1$) を示すサイトを集計した．その結果を表 5 に示す．表 5 より，利用するログの日数が増加することで誤検知率も増加していることがわかる．これは，ログの増加にともなう解析対象となる遷移先サイト数の増加よりも，ログの増加によってあるサイトにおける遷移元サイトの数が増加し，結果として遷移元サイトが複数になり誤検知された事例が起因していると考えられる．また，Drive-by Download 攻撃の悪性サイトの存在時間は数時間程度と短命である [1] ため，悪性サイトの遷移元/遷移先サイト数を解析し，検出するためには比較的短期間のログで十分と考えられる．以上より，遷移元/遷移先サイト数の解析には，誤検知率の低減という観点から 1 日分程度のログで十分であると考えられる．

FCDBD への Landing サイト検出手法の適用

本手法を FCDBD に適用することによって, 未知の Landing サイトによる Drive-by Download 攻撃を防ぐ効果が期待できる．解析センタは，観測センサから Web ページにアクセスしたときの一連のログを受信する度に，当該ログから当該 Web ページの遷移先サイトを解析する．その際，解析した結果は逐次センタ内に保存しておく．そして，当該 Web ページの遷移先サイトに，以前のアクセス時にはみられなかったサイトが追加されていた時に，過去の解析により得られた当該遷移先サイトの遷移先/遷移元サイトの数を参照する．そして，当該遷移先サイトが悪性サイトのような特徴を示した場合は，当該 Web ページが改ざんにより Landing サイトとなった疑いがあるとして観測センサに警告する．

しかし，FCDBD の実際の運用を考慮すると誤検知率をさらに低減させる必要がある．1.5%の誤検知率だと，単純におおよそ 100 回の Web アクセスに対して 1〜2 回誤った警告が発生することとなる．これは，フレームワークを使用しているユーザに不快感を与えかねない．また, FCDBD では，悪性が疑われるコンテンツを実際に観測センサより収集してコンテンツ解析を行う．さらに，システムへの影響を考慮すると，悪性が疑われると判定された Web ページのコンテンツは，実際に収集されコンテンツ解析を行うこととなる．例えば，100 万人規模で観測センサを配布すると仮定すると，100 万人規模のユーザ数を有するセキュリティソフトウェアのログデータ [*1] を利用した試算結果より，1 日あたりおよそ 150 万 URL の Web ページへアクセスされることとなる．そのため，毎日その 1.5%であるおよそ 23,000 サイトのコンテンツが誤検知にも関わらず解析されることとなる．ユーザへの影響，システムへの影響を鑑みるとさらなる改善が必要である．

さらなる手法の改良

正規 Web ページが改ざんされ Landing サイトとなる時には，ユーザを Exploit サイトへ転送させるスクリプトを挿入されるなど，改ざん前後で当該 Web ページのコンテンツが変わっていると容易に推測される．そのため，コンテンツの変化を判定条件に加味することでさらなる手法の高精度化が期待できる．ブラウザセンサはコンテンツのハッ

[*1] ソフトウェアのユーザの同意のもと，利用許諾約款の範囲内で当該ログデータを使用した

表 5 ログの日数分にともなう手法 2 の誤検知率の変化

日付 (YYMMDD)	140303	140304	140305	140306	140307	140310	140311	140312	140313	140314
遷移先サイト数	5,397	5,836	6,152	6,507	6,718	6,947	7,152	7,495	7,720	7,899
誤検知数	414	509	582	693	754	820	883	1,007	1,094	1,158
誤検知率	7.7%	8.7%	9.5%	10.7%	11.2%	11.8%	12.3%	13.4%	14.2%	14.7%

シュ値を計算し，その値を Web アクセスログとしてセンタに送信する．そのため，例えばセンタ側で当該情報を過去に送信されたハッシュ値と照合し，ハッシュ値が，過去のものと異なる場合に手法 1，手法 2 を用いて判定を行うことが考えられる．

7. まとめ

本稿では，Drive-by Download 攻撃における Landing サイトを検出する手法として以前著者らが提案した Web ページからの遷移先サイトの変化に着目して，改ざんにより Landing サイトとなった Web ページを検出する手法の改良案として，新たな遷移先サイトがみつかった際に当該サイトの遷移元/遷移先サイトの数を調べて，その数が悪性サイトの特徴と同様の特徴を示した際に，元の Web ページを Landing サイトとして検出する方法を提案した．本稿での評価では，本改良により誤検知率が 1 日あたりの平均で 6.0%から 1.5%に低減されることが確認された．今後，本手法のさらなる改良を行い，FCDBD フレームワークに当該手法を実装してその有効性検証を検証する．

現在，著者らは FCDBD フレームワークの実験を実施しており，観測センサを実際に利用して当実験にご協力いただける方を募集している．実験に参加されたい方，実験にご興味がある方は，著者らまでご連絡いただきたい．また，実験の実施にさきがけ，参加者が安心して実験に参加できるように取得するデータの内容，利用用途，管理を実験の参加者に明示した参加規約など文書を整備し，公開している [*2]．参照されたい方は著者らまでご連絡いただきたい．

謝辞 本研究成果は，独立行政法人情報通信研究機構 ((以下，NICT) 理事長: 坂内正夫，本部: 東京都小金井市) の委託研究「ドライブ・バイ・ダウンロード攻撃対策フレームワークの研究開発」により得られたものである．ここに深謝する．

参考文献

[1] N. Provos, P. Mavrommatis, M. A. Rajab and F. Monrose, *All Your iFRAMEs Point to Us*, Proc. the 17th USENIX Security Symposium, 2008.

[2] 笠間貴弘, 井上大介, 衛藤将史, 中里純二, 中尾康二,「ドライブ・バイ・ダウンロード攻撃対策フレームワークの提案」, コンピュータセキュリティシンポジウム 2011(CSS2011), 2011.

[3] T. Matsunaka, A. Kubota and T. Kasama, *An Approach to Detect Drive-by Download by Observing the Web Page Transition Behaviors*, Proc. of 9th Asia Joint Conference on Information Security (AsiaJCIS2014), 2014.

[4] 松中隆志. 半井明大, 浦川順平, 窪田歩,「ドライブ・バイ・ダウンロード攻撃対策フレームワークにおけるリンク構造解析による改ざんサイト検出手法の一検討」, 第 31 回暗号と情報セキュリティシンポジウム (SCIS2014), 2014.

[5] 松中隆志, 窪田歩, 星澤裕二,「Drive-by Download 攻撃対策フレームワークにおける Web アクセスログを用いた Web リンク構造の解析による悪性サイト検出手法の提案」, コンピュータセキュリティシンポジウム 2014(CSS2014), 2014.

[6] T. Matsunaka, J. Urakawa and A. Kubota, *Detecting and Preventing Drive-by Download Attack via Participative Monitoring of the Web*, Proc. of 8th Asia Joint Conference on Information Security (AsiaJCIS2013), 2013.

[7] M. Akiyama, M. Iwamura, Y. Kawakoya, K. Aoki and M. Itoh, *Design and Implementation of High Interaction Client Honeypot for Drive-by-Download Attack*, IEEE Trans. of Communication, Vol. E93–B, No. 5, pp. 1131–1139, May. 2010.

[8] Y–M. Wang, D. Beck, X. Jiang, C. Verbowski, S. Chen and S. King, *Automated Web Patrol with Strider HoneyMonkeys: Finding Web Sites That Exploit Browser Vulnerabilities*, Proc. 13th Annual Network & Distributed System Security Symposium (NDSS2006), 2006.

[9] J. W. Stokes, R. Andersen, C. Seifert and K. Chellapilla, *WebCop: Locating Neighborhoods of Malware on the Web*, Proc. 3rd USENIX Workshop on Large-scale Exploits and Emergent Threats (LEET2010), 2010.

[10] J. Zhang, C. Seifert, J. W. Stokes and W. Lee, *ARROW: GenerAting SignatuRes to Detect DRive-By DOWnloads*, Proc. 20th International World Wide Web Conference (WWW2011), 2011.

[11] G. Stringhini, C. Kruegel and G. Vigna, *Shady Paths: Leveraging Surfing Crowds to Detect Malicious Web Pages*, Proc. 20th ACM Conference on Computer and Communications Security (CCS2013), 2013.

[12] G. Wand, J. W. Stokes, C. Herley and D. Felstead, *Detecting Malicious Landing Pages in Malware Distribution Networks*, Proc. 43rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN2013), 2013.

[13] 西田雅太, 星澤裕二, 笠間貴弘, 衛藤将史, 井上大介, 中尾康二,「文字出現頻度をパラメータとした機械学習による悪質な難読化 JavaScript の検出」, 情報処理学会研究報告コンピュータセキュリティ (CSEC), Vol. 2014-CSEC-64, No. 21, 2014.

[14] *Alexa – Actionable Analytics of the Web*, `http://www.alexa.com`.

[15] *Adblock Plus*, `https://adblockplus.org/`.

[16]「マルウェア対策研究人材育成ワークショップ 2014(MWS2014)」, `http://www.iwsec.org/mws/2014/`.

[17] 秋山満昭, 神薗雅紀, 松木隆宏, 畑田充弘,「マルウェア対策のための研究用データセット〜MWS Datasets 2014〜」, 情報処理学会研究報告コンピュータセキュリティ (CSEC) Vol. 2014-CSEC-66, No. 19, 2014.

[18] Trend Micro, *Site Safety Center*, `http://global.sitesafety.trendmicro.com/index.php`.

[*2] `http://www.fcdbd.jp/`．アクセスを希望される方は著者らまでご連絡いただきたい．