Computer Security Symposium 2017
23 - 25 October 2017

Drive-by-Download攻撃検知項目の評価

高田 真資[1,a)]　高橋 健一[2]　川村 尚生[2]　菅原 一孔[2]

概要：マルウェアの感染経路の一つとして Drive-by-Download 攻撃が存在する．Drive-by-Download 攻撃では，正規の Web サイトを，悪性な Web サイトへとリダイレクトするように改ざんすることで，正規の Web サイトにアクセスしたユーザーにマルウェアをダウンロードさせる．本稿ではリダイレクトを含んだ特徴的な通信をもとに攻撃の検知を行う既存研究について調査し，それぞれの研究で利用されている検知項目について良性データ，悪性データを使って評価した．検証結果をもとに各項目の有用性を考察する．

キーワード：Drive-by-Download 攻撃，悪性 Web サイト，評価

Evaluation of detection items against Drive-By-Download Attacks

Masashi Takada[1,a)]　Kenichi Takahashi[2]　Takao Kawamura[2]　Kazunori Sugahara[2]

Abstract: There is a Drive-By-Download attack as one of malware infection routes. In the Drive-By-Download Attack, a regular website is compromised and injected with codes to redirect users to malicious website. Then users who access the compromised website forced to download a malware. In the paper, we surveyed existing researches to detect the attacks based on characteristic communication including malicious redirects. We extracted the detection items from detection methods and verify them by benign data and malignant data and consider the usefulness of each method.

Keywords: Drive-By-Download Attack, malicious website, evaluation

1. はじめに

現在，Web におけるマルウェアの感染経路として Drive-by-Download 攻撃が存在する．Drive-by-Download 攻撃は，Web サイトにアクセスした Web ユーザーが意図せずにマルウェアをダウンロードさせられる攻撃であり，近年猛威を振るっている．Drive-by-Download 攻撃では，正規の Web サイトを，悪性な Web サイトへとリダイレクトするように改ざんする．改ざんされたサイトを経由し悪性サイトにリダイレクトされた Web ユーザーは，ユーザー PC に存在するソフトウェアの脆弱性を利用されてマルウェアをダウンロードされる．近年では，Drive-by-Download 攻撃を行うためのツールが Exploit Kit と呼ばれる形でパッケージ化されており，それらを用いることで経験や知識，技術力のない攻撃者でも容易に攻撃を行うことができる．

Drive-by-Download 攻撃の対策として，不正なドメイン名，IP アドレスをブラックリストに登録し，そのブラックリストを用いて悪性サイトへのアクセスを検知する方法がある [1]．しかし，この方法で登録されるドメイン名，IP アドレスは悪性であることが既知のものであり，ブラックリストに登録されていない未知の悪性サイトに対しては有効ではない．そこで，未知の悪性サイトへのリダイレクトを防ぐために，Drive-by-Download 攻撃が起きた際に発生する通信の特徴を利用して，攻撃を検知する手法が複数提案されている．それぞれの検知手法は通信中の異なる特徴を利用しており，発表された論文でも独自の評価基準に基

[1] 鳥取大学大学院持続性社会創生科学研究科
Graduate School of Sustinability Science, Tottori University
[2] 鳥取大学大学院工学研究科
Graduate School of Engineering, Tottori University
[a)] b132036@gmail.com

づいた評価が行われている．

本研究では，Drive-by-Download 攻撃により生じる特徴的な通信から攻撃の検知を行う手法を調査し，各手法が不正リダイレクト中のどのような特徴を利用しているかをまとめた．また，各手法で利用されている検知項目について良性データ，悪性データで評価を行った．

以降，2 章で Drive-by-Download 攻撃の概要について述べ，3 章で調査した検知手法について述べ，その中の数種類について，4 章で評価する．最後に 5 章でまとめる．

2. Drive-by-Download 攻撃

2.1 攻撃の概要

Drive-by Download 攻撃は，Web サイトにおいてユーザーが意図せずにマルウェアをダウンロードさせられる攻撃である．Drive-by Download 攻撃は主に，入り口サイト，踏み台サイト，攻撃サイト，マルウェア配布サイトからなる．攻撃の流れを**図 1** に示す．

まず，攻撃者は一般の Web サイトを改ざんし，入り口サイトを作成する．Web ユーザーが入り口サイトにアクセスすると，自動的にリダイレクトが発生し，踏み台サイトを経由して攻撃サイトへ誘導させられる．攻撃サイトではユーザー PC の環境に応じて，その脆弱性を利用した攻撃が行われ，その結果，ユーザー PC にマルウェア配布サイトからマルウェアがダウンロードされる．入り口サイトにアクセスしてからの一連の挙動はユーザに気づかれないように実行される．

以降では Drive-by Download 攻撃に関係するツール，技術について述べる．

2.2 Exploit Kit

Exploit Kit [2] は攻撃時に使用するシェルコードやマルウェアがパッケージ化されたものであり，Exploit Kit を用いることで攻撃者は容易に攻撃を行うことができる．Exploit Kit は複数の種類が存在し様々な脆弱性を利用しているが，中でも Java や Flash の脆弱性を利用することが多い．また，Exploit Kit はパッケージ化されているため，その攻撃には固有の特徴がみられ，それらを用いて攻撃の検知を行うことができる．しかし，Exploit Kit 自体が更新されるためその特徴も変化する．

2.3 不正リダイレクト

ユーザーは入り口サイトにアクセスすると，踏み台サイトへのリダイレクトを経て攻撃サイトへ誘導される．この時，攻撃サイトの特定を困難にするために複数回のリダイレクトが発生することがある．これにより，解析者は入り口サイトのソースコードを解析するだけでは攻撃サイトを特定できず，通信ログを解析する際も複数の余分なログを解析しなければならなくなる．また，通常の HTTP 通信では，HTTP 要求を送った際に Referrer ヘッダに参照元の Web ページの URL が設定されるが，悪性なリダイレクトでは Referrer ヘッダに値が設定されないことが多い．

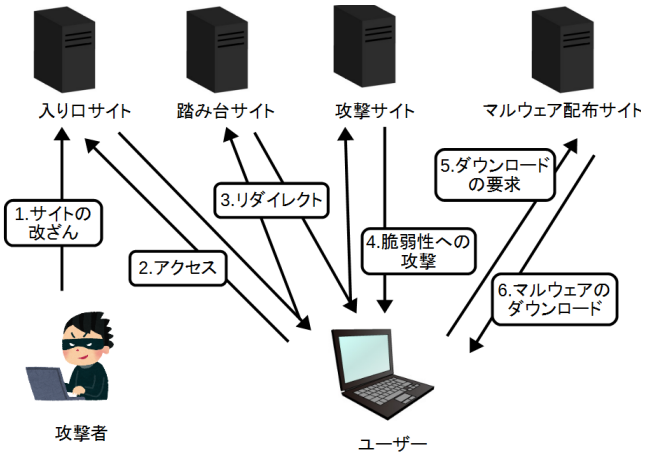

図 1 Drive-by-Download 攻撃の流れ
Fig. 1 Flow of Drive-by-Download Attack

2.4 プログラムの難読化

難読化とは，プログラムコードを人間が理解しづらいように記述する手法である．Drive-by Download 攻撃を行う悪性な Web サイトでは，コード中の変数名や関数名などの情報を隠し，シグネチャによる検知を回避する．この時，リダイレクト先のリンクが難読化され隠ぺいされる場合，HTTP 通信のレスポンスボディに記載のない URL への GET リクエストが発生するという特徴がある．

3. 調査

3.1 調査対象

Drive-by-Download 攻撃の特徴的な通信を利用した攻撃の検知手法を提案する論文を調査した．調査対象は CSS2000～2016 で発表されたものとした．この中から，Drive-By-Download 攻撃検知に関する 7 件を調査対象として選択した（表 1）．それぞれの論文で提案された各検知手法が利用する悪性通信中に含まれる情報を検知項目と定義して抽出した (**表 2**)．各検知手法では検知項目の組合わせにより不正リダイレクトの検知を行う．

A：Web 階層

Web 階層はリダイレクトによって生じる Web ページの階層的な構造である．ある Web ページ A にアクセスした際，A から別の Web ページ B へのリダイレクトが生じ，B からさらに別の Web ページ C へリダイレクトが生じる，というように階層的にリダイレクトが繰り返される場合がある．この際，始めにアクセスした Web ページ A の階層を 1，リダイレクト先の Web ページ B の階層を 2，さらに先の Web ページ C の階層を 3 というように Web 階層をカウントする．Drive-by-Download 攻撃では，リダイレ

表 1　調査対象
Table 1　Target Papers

論文	検知項目
通信の遷移に着目した不正リダイレクトの検出による悪性 Web サイト検知システムの提案 [3]	A, B, D
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 [4]	B, F
Drive-by-Download 攻撃における通信の定性的特徴とその遷移を捉えた検知方式 [5]	A, B, H, I
HTTP リクエストシーケンスに注目した不正リダイレクトの検出 [6]	A, D, I
ネットワーク通信の相関性に基づく Drive-by Download 攻撃検知手法 [7]	D, I
通信遷移と URL の属性情報を用いた悪性リダイレクト防止手法 [8]	A, B, D, F, G
難読化の特徴を用いたドライブバイダウンロード攻撃検知手法の設計と実装 [9]	C

表 2　検知項目
Table 2　Detection Items

項目名	番号
A	Web 階層
B	ドメイン名の遷移
C	レスポンスボディ内に未出現の URL へのリクエスト
D	ファイルの種類
E	Content-Type ヘッダとマジックナンバーの整合性
F	ドメイン年齢
G	X-Powered-By ヘッダの値
H	受信データ量の遷移
I	UserAgent の遷移

クト時に複数の踏み台サイトを経由する．そのため，深い Web 階層で読み込まれる Web ページの信頼度は低いと考えらる．そのため，読み込まれる Web ページの Web 階層が特定の深さに達したら悪性と判断する．

Web 階層をカウントするためには，リダイレクトが発生した際のリダイレクト元のパケットとリダイレクト先のパケットを対応付ける必要がある．リダイレクトの対応関係を推定するための方法としては以下の 2 つがある [10]．

(1) HTTP ヘッダの Referer ヘッダ，Location ヘッダから対応づける
(2) あるパケット P1 の HTTP レスポンスボディから抽出した URL（または，同一ホストの URL）に対し GET リクエストをおこなうパケット P2 を P1 より時間的に後に出現するパケットから探し対応づける（**図 2**）．この際，既に（1）で対応づけれたパケットは除外する．

B：ドメイン名の遷移

不正リダイレクトにより Web ユーザーが悪性ページにリダイレクトされる際にドメイン名が変化する．これは，入り口サイトはユーザーのアクセスを促すために他の Web サイトを改ざんして用意されるため，攻撃サイトとは別の

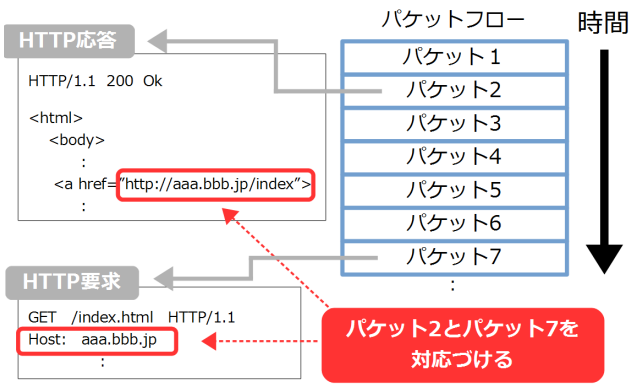

図 2　レスポンス内の URL によるリダイレクトを対応づけ
Fig. 2　Associating packets by using http response headers

ドメインとなるからである．このため，HTTP ヘッダの Host ヘッダからドメイン情報を取得し，リダイレクトの前後でドメイン名の変化があれば検知する．

C：HTML に出現していない URL へのリクエスト

改ざんコード，攻撃コードに対してのシグネチャによる検知を回避するため，悪性サイトでコードの難読化が施されることがある．この際，改ざんコードとともにリダイレクト先や攻撃コードのダウンロード元となる URL が難読化される．このため，読み込んだ Web ページに記載されていない URL への HTTP リクエストが発生する．このため，HTML から取得したリンク先情報を保存し，保存された URL のどれにも一致しないドメインへのリクエストがあった場合，難読化された URL へのアクセスとして検知する．

D：ファイルの種類

Exploit Kit ではソフトウェアの脆弱性を用いて攻撃を行う．中でも Java，Flash の脆弱性が利用されることが多く.jar や.swf を読み込む通信は悪性の可能性が高いといえる．また，実行形式のファイルはマルウェア本体である可能性があり，危険度が高い．上記を踏まえ，通信中に現れた特定の種類のファイルを検知する．ファイルの種類は URL 中の拡張子，HTTP の Content-Type ヘッダ，ファイルのマジックナンバーから確認する．

E：Content-Type ヘッダとマジックナンバーの整合性

正常な HTTP レスポンスでは Content-Type ヘッダで指定されたファイルタイプと同様のファイルが受信される．しかし，マルウェア本体のダウンロード時には偽装のために，実際のファイルとは異なる Content-Type が設定されることがある．このため，Content-Type ヘッダに記されたファイルタイプとレスポンスボディにマジックナンバーを比較し，それらが一致するかどうかを確認することで検知を行う．

F：ドメイン年齢

ドメイン年齢は，Web サイトのドメインが登録されてか

ら現在までの期間を表す．攻撃サイトでは，ブラックリストによるフィルタリングを回避するために短期間でドメインが変更される．関連研究では悪性サイトのドメイン生存期間は最長でも 6 か月と述べている．そのため，ドメイン年齢が 6 か月以下のもの，または取得できないものを危険とみなして検知する．

G：X-Powered-By ヘッダの値

X-Powered-By ヘッダは HTTP ヘッダの 1 つであり，PHP のバージョン情報を表す．Exploit Kit を用いた Drive-by-Download 攻撃では，X-Powered-By ヘッダに古い PHP バージョンが記載されることが多い．そのため，古い PHP バージョンが記された X-Powered-By ヘッダを検知する．

H：受信データ量の遷移

Drive-by-Download 攻撃では，リダイレクト時，攻撃時などの段階ごとにユーザーが受信するデータの大きさが変化する．入り口サイトへのアクセス時，リダイレクト時には数百〜数千バイトほどの大きさだが，攻撃ファイルのダウンロード時には数十〜数百キロバイトに上昇する．そのため，このような受信データ量の変化を検知する．

I：UserAgent の遷移

JRE の脆弱性を利用した Exploit Kit では攻撃サイトにアクセスする際に，Java アプリケーションにより通信が行われるため，UserAgent が Java となる．また，マルウェアが C & C サーバと通信する際には，ユーザーのブラウザ環境とは異なる UserAgent が出現する．そのため，通信中の UserAgent の変化を検知する．

4. 評価

3 章で述べた検知項目のうち A：Web 階層，B：ドメインの遷移，D：ファイルの種類，G：X-Powered-By ヘッダの値について悪性データ，良性データ中にどのくらい出現するかの検証をおこなった．

4.1 評価データ

4.1.1 悪性データ

悪性データとして D3M データセット [11] に含まれる攻撃通信データを使用した．D3M(Drive-by-Download Data by Marionette) データセットは NTT セキュアプラットフォーム研究所の Web クライアント型ハニーポット (Maronette) により収集された Web 感染型マルウェアの観測データ群である．D3M データセットには，ブラックリストに登録された URL をハニーポットが巡回した際にキャプチャした攻撃通信データ，巡回した URL が収録されており，その攻撃通信データを利用した．攻撃通信データは pcap 形式のファイルであり，1 つのファイル内に複数種類の URL へアクセスした際の通信データが収録されている．

4.1.2 良性データ

アクセス数の多い Web サイトは悪性である危険性が低いと仮定し，以下の Web ページからアクセス数の多い Web サイトのランキング情報を得た．

- Alexa [12]
- 日本の人気サイトランキング 500 [13]

Alexa からは上位 50 サイトを日本の人気サイトランキング 500 からは上位 500 サイトを抽出し，それらを良性サイトとした．抽出したサイトの中にはアクセスできないサイトが存在したため，それらはデータから省いた．また，2 つのサイトのランキングで重複した Web サイトが存在したため良性サイトの合計は 446 サイトとなった．それぞれの良性サイトに対してアクセスした際の通信データを Wireshark によりキャプチャした．良性データはすべて pcap 形式のファイルである．

4.2 結果・考察

A，B：Web 階層とドメインの遷移

リダイレクト元とリダイレクト先のパケットを関連づけ，URL へのアクセスからリダイレクトをたどり，リダイレクト先がなくなるまでの一連の通信を通信セッションとした．各通信セッションにおける Web 階層をカウントした結果を**表 3** に示す．

悪性データでは 18 層や 20 層といった深い Web 階層がみられた．悪性，良性とも 2 層，3 層の通信セッションが多く，階層が深くなればなるほど検知される数は減少した．2 層，3 層の通信セッションについて，悪性データでは 3 割以上でドメインの遷移が起きたが，良性データではドメインの遷移が起きる割合は 2 割以下となっていた．

また，HTTP リクエストのうち Referer ヘッダが存在しないものをカウントした．**表 4** に結果を示す．

C：ファイルの種類

Content-Type ヘッダが HTTP レスポンスに含まれるパケットは悪性データで 23337 個，良性データで 27369 個存在した．これらのパケットの中で特定の Content- Type をものをカウントした．表 3 に結果を示す．

良性データでは application/x-msdownload，application/x-download，application/x-msdos-program，application/java-archive，application/pdf を持つものがなかった．application/x-msdownload，application/x-download，application/x-msdos-program は exe のような実行ファイルを受信する際にあらわれる Content-Type である．これらが出現しなかった理由は，良性のデータ（ランキング情報から得た Web サイト）がトップページであるものが多く，トップページに pdf や実行形式のファイルを置くことはほとんどないためだと考えられる．一方，application/x-shockwave-ash，application/javascript，application/x-javascript，text/javascript は一定数出現していた．また，具体的なファイルの種類が明示されていない（application/octet-stream である）場合も出現してい

表 3　Web 階層とドメイン遷移
Table 3　Layers of Web page and Domain transition

Web 階層	悪性	ドメインの遷移	良性	ドメインの遷移
0	326		144	
1	711	390	8131	1241
2	453	175	9291	1369
3	239	85	3231	551
4	164	29	805	213
5	132	23	131	45
6	59	9	17	10
7	50	15	7	5
8	18	5	2	1
0	15	5	0	
10	5	2	1	
11	4	1		
12	5			
13	2			
14	9			
15	13			
16	1			
17	0			
18	2			
19	0			
20	1			
合計	2209	739	21760	3435

表 4　Referer ヘッダの有無
Table 4　Existence of Referer header

	悪性（割合）	良性（割合）
Referer のない HTTP リクエスト	4545（18.7）	2987（10.2）
全ての HTTP リクエスト	24250	29182

た．これは HTTP リダイレクトの際の空のテキストファイルや jpeg や png のファイルで使われていることが多かった．

一方，悪性データでは全てのファイルが出現していた．特に，良性データと比べて，実行ファイルが指定されている割合や，ファイルの種類が明示されていない場合，Flash や JAR が指定されている割合が高くなっていた．一方，javascript が指定されている割合は，良性データと比較して少ない結果となった．

G：X-Powered-By ヘッダの値

全ての HTTP レスポンスの中で，X-Powered-By ヘッダを含むもの，X-Powered-By ヘッダが PHP のバージョン情報を持つもの，さらにその中で古い PHP のバージョンのものを調べた．今回は PHP 5.4.0 以前のバージョンのものを古い PHP として扱った．結果を表 **6** に示す．

X-Powered-By ヘッダは全体的に良性データよりも悪性データで多く出現していた．特に PHP のバージョン情報を含むものが多く，脆弱性の攻撃に用いられているものだと推測できる．ただし，良性データにおいても古い PHP バージョンを設定されている場合が存在していることが確認できる．

表 5　ファイルの種類
Table 5　File type

Content-Type	悪性（割合）	良性（割合）
全ての Content-Type	23337	27369
application/octet-stream	926（3.97）	101（0.0036）
application/x-msdownload	112（0.48）	0（0）
application/x-download	37（0.16）	0（0）
application/x-msdos-program	3（0.013）	0（0）
application/x-shockwave-flash	399（1.71）	68（0.0024）
application/java-archive	345（1.48）	0（0）
application/pdf	466（2.0）	0（0）
application/javascript	893（3.83）	1276（0.046）
application/x-javascript	660（2.83）	1145（0.0418）
text/javascript	427（1.83）	697（0.025）

表 6　X-Powered-By ヘッダの値
Table 6　Value of X-Powered-By header

	悪性（割合）	良性（割合）
X-Powered-By ヘッダを含む	5611（23.1）	1489（4.9）
PHP のバージョン情報を含む	3647（15.0）	317（1.0）
古い PHP のバージョン情報を含む	1446（5.9）	84（0.28）
全ての HTTP レスポンス	24329	30107

5. おわりに

本稿では Drive-by-Download 攻撃の不正リダイレクトの検知を行う手法について調査し，各検知手法で使用される特徴を検知項目として抽出した．抽出した検知項目について良性データ，悪性データ中での出現率を調べた．

今後の課題として他の項目について調査するとともに，それらの包含関係を調査することがあげられる．

参考文献

[1] , 福島祥郎, 堀良彰, 櫻井幸一. ”悪性 Web サイト間の関連性に着目した信頼性評価によるブラックリスト方式の検討.”, 研究報告コンピュータセキュリティ（CSEC）, Vol.2011 No.38, pp.1-8, 2011.

[2] エクスプロイトキット最新動向分析：Web サイト改ざんと不正広告を経由し、Flash 脆弱性を攻撃, http://blog.trendmicro.co.jp/archives/13017.

[3] , 安藤槙悟, 寺田真敏, 菊池浩明, 趙晋輝ほか. 通信の遷移に着目した不正リダイレクトの検出による悪性 Web サイト検知システムの提案. 研究報告コンピュータセキュリティ (CSEC), Vol.2011, No.32, pp.1–6, 2011.

[4] 酒井裕亮, 佐々木良一ほか, 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, 研究報告コンピュータセキュリティ (CSEC), Vol.2013, No.29, pp.1–6, 2013.

[5] 北野美紗, 大谷尚通, 宮本久仁男ほか, Drive-by-Download 攻撃における通信の定性的特徴とその遷移を捉えた検知方式, コンピュータセキュリティシンポジウム 2013 論文集, Vol.2013, No.4, pp.595–602, 2013.

[6] 工藤聖, トラン・コン・マン, 中村康弘ほか, HTTP リクエ

ストシーケンスに注目した不正リダイレクトの検出, コンピュータセキュリティシンポジウム 2015 論文集, Vol.2015, No.3, pp.221–225, 2015.
[7] 寺田成吾, 小林峻, 小出和弘, 羽藤逸文, 瀬戸口武研, 道根慶治, 山下康一ほか, ネットワーク通信の相関性に基づく Drive-by Download 攻撃検知手法, コンピュータセキュリティシンポジウム 2015 論文集, Vol.2015, No.3, pp.1–7, 2015.
[8] 佐藤祐磨, 中村嘉隆, 高橋修ほか, 通信遷移と URL の属性情報を用いた悪性リダイレクト防止手法, コンピュータセキュリティシンポジウム 2015 論文集, Vol.2015, No.3, pp.8–15, 2015.
[9] 藤原寛高, 難読化の特徴を用いたドライブバイダウンロード攻撃検知手法の設計と実装, 2015.
[10] 高田雄太, 森達哉, 後藤滋樹. ”Web 感染型マルウェアのリダイレクト解析.” 第 73 回全国大会講演論文集 Vol..2011, No.1, pp.497-498. 2011
[11] 高田雄太, 寺田真敏, 村上純一, 笠間貴弘, 吉岡克成, 畑田充弘ほか, ”マルウェア対策のための研究用データセット MWS Datasets 2016”, 研究報告セキュリティ心理学とトラスト (SPT), Vol.2016, No.17, pp.1–8, 2016.
[12] Alexa Top 500 Global Sites, https://www.alexa.com/topsites.
[13] 日本の人気サイトランキング 500, http://akimoto.jp/japan/.