「マルチメディア，分散，協調とモバイル
(DICOMO2015)シンポジウム」 平成27年7月

# 通信遷移と PageRank を用いた悪性リダイレクト防止手法の評価

佐藤祐磨[†1]　中村嘉隆[†2]　高橋修[†2]

**概要：** Web の普及に伴い，Web を通じたサイバー攻撃が発生している．Web 上のサイバー攻撃であるドライブバイダウンロード攻撃による被害が多く見られる．難読化スクリプトコードの解析，悪性 URL の収集，HTTP 通信の情報解析など，ドライブバイダウンロード攻撃に関連する様々な対策手法が提案されている．しかし，現在の攻撃対策手法では，ドライブバイダウンロード攻撃を正確に検出できず，ドライブバイダウンロード攻撃によるウイルスの感染を防止することができていないという問題がある．そこで，本論文では，HTTP 通信遷移と PageRank を利用してドライブバイダウンロード攻撃において利用される悪性リダイレクトを防止し，既存手法よりも正確に攻撃を検出する手法を提案し，手法の評価を行う．

# Evaluation of the preventing method for malicious Web redirections based on HTTP transitions and PageRank status

YUMA SATO[†1]　YOSHITAKA NAKAMURA[†2]　OSAMU TAKAHASHI[†2]

## 1. はじめに

近年，Web の普及にともなって，ドライブバイダウンロード攻撃による被害が増加している．ドライブバイダウンロード攻撃は Web 上を介して行われるサイバー攻撃であり，Web を利用するユーザの PC にウイルスをダウンロードさせる攻撃である．マカフィー株式会社のマンスリーウイルスレポートによると 2013 年検知会社数年間ランキングにおいて，トップ 10 の内 7 つがドライブバイダウンロード攻撃であったと示している[1]．表 1 に 2013 年検知会社数年間ランキングを示す．灰色のウイルスがドライブバイダウンロード攻撃によってダウンロードされるウイルスである．マカフィー株式会社は，今後もドライブバイダウンロード攻撃がサイバー攻撃の脅威となると予想している．ドライブバイダウンロード攻撃は，攻撃者が正規の Web サイトを改ざんして，その Web サイトを閲覧したユーザを攻撃サイトに誘導し，ウイルスに感染させる．ウイルスは，Web を利用するユーザの使用するソフトウェアの脆弱性を利用し，ダウンロードがされる．ウイルスのダウンロードは，秘密裏で行われるため，ユーザが気付くことは難しい．このようなドライブバイダウンロード攻撃には，スクリプトコードが利用される．ドライブバイダウンロード攻撃に利用されるスクリプトコードは，攻撃者によって難読化されていて，攻撃解析者が簡単に解析できないように細工してある．このようにドライブバイダウンロード攻撃は近年巧妙化する傾向にある．

表 1　2013 年検知会社数年間ランキング

| 順位 | ウイルス | 件数 |
|---|---|---|
| 1 | JS/Redirector.ar | 3328 |
| 2 | Generic!atr | 2571 |
| 3 | Generic Downloader.z | 2152 |
| 4 | JS/Exploit!JNLP.c | 2092 |
| 5 | JS/Exploit-Blacole.ht | 1693 |
| 6 | JS/Blacole-Redirect.ae | 1644 |
| 7 | W32/Conficker.worm!inf | 1533 |
| 8 | JS/Iframe.gen.k | 1518 |
| 9 | JS/Exploit!JNLP | 1455 |
| 10 | JS/Exploit-Blacole.le | 1178 |

ドライブバイダウンロード攻撃の対策として，難読化スクリプトコードの解析，悪性 URL の収集，HTTP 通信の情報解析など，様々な対策手法が提案されている．にもかかわらず，依然ドライブバイダウンロード攻撃による，マルウェアの感染が報告されている．ドライブバイダウンロード攻撃は Web サイトの改ざんによって引き起こされる．Web サイトの改ざんを長期間検出できない場合，ドライブ

†1 公立はこだて未来大学大学院システム情報科学研究科
Graduate School of Systems Information Science, Future University Hakodate
†2 公立はこだて未来大学システム情報科学部
School of Systems Information Science, Future University Hakodate

バイダウンロード攻撃の発見が遅れる場合が考えられる．そこで本研究では，クライアントであるユーザが Web サイトにアクセスしたタイミングでドライブバイダウンロード攻撃の疑いのある Web ページを検出し，ドライブバイダウンロード攻撃の防止を既存手法より高精度に行う手法の提案を行う．

## 2. 攻撃手法と技術

### 2.1 ドライブバイダウンロード攻撃と攻撃フロー

ドライブバイダウンロード攻撃は，マルウェア感染攻撃の一種である．ユーザが Web サイトをアクセスした際，ユーザの意図に関わらず，ユーザに悪意あるソフトウェアをダウンロードさせる[2].

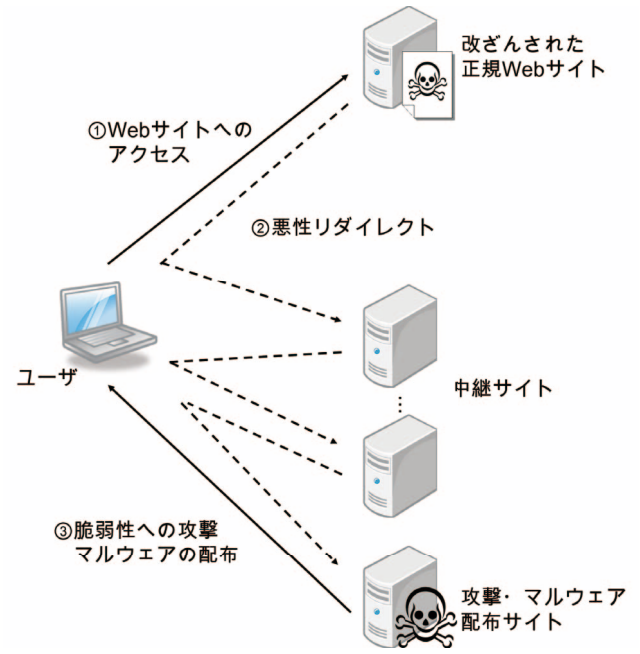


図 1　ドライブバイダウンロード攻撃のフロー

ドライブバイダウンロード攻撃の典型的なフローは図 1 のようになっている．攻撃者は，正規 Web サイトの Web ページから攻撃者が用意する攻撃 Web サイトへのリダイレクトを目的として，正規 Web サイトのページを改ざんする．改ざんされた Web ページにアクセスしたユーザは，攻撃者が改ざんによって仕掛けたリダイレクトにより，攻撃者が用意した攻撃サイトへ誘導させられる．一般にドライブバイダウンロード攻撃におけるこのリダイレクトは複数存在することが多い．リダイレクトが複数ある理由は，攻撃者が攻撃検出を回避あるいは困難にするためと考えらえる．攻撃サイトでは，ユーザの使用する OS，ブラウザ，ブラウザのアドオンの脆弱性を突く攻撃が行われ，ユーザの制御が攻撃者に奪われる．その後，ユーザはマルウェア配布サイトへ誘導され，悪意あるソフトウェアをダウンロードさせられる．

### 2.2 HTTP ヘッダ

ドライブバイダウンロード攻撃は Web を介してなされる．Web 上の HTTP 通信には HTTP メッセージが利用されており，クライアントからサーバへの要求である HTTP リクエストとサーバからクライアントへの応答である HTTP レスポンスの 2 つから成り立つ[3]. HTTP リクエストには，要求する Web ページ URI 情報も含まれる. HTTP レスポンスには，HTTP 通信の状態を示すステータスコードが含まれる．

HTTP リクエストのヘッダには，クライアントに関する情報が含まれる．HTTP レスポンスのヘッダには，サーバに関する情報，コンテンツの情報が含まれる．要求ヘッダのパラメータとして，要求先のホストを示す Host，HTTP リクエストの発生元を参照する Referer などが含まれる．応答ヘッダのパラメータとして，クライアントが取得するコンテンツの型を示す Content-Type，要求する Web ページの URL 以外の Web ページを提供するために利用される Location などが含まれる．Referer は HTTP リクエストの発生元の情報を含む．それにより Referer の情報から，クライアントの Web ページの遷移がわかる場合がある．ユーザの個人情報が Referer に含まれる場合などのユーザのプライバシーが守られない場合，RFC2068 では，Referer を HTTP ヘッダに付加させないことが推奨されている．

### 2.3 攻撃で利用される技術

#### 2.3.1 フィンガープリンティング

近年のドライブバイダウンロード攻撃では，フィンガープリンティングが利用されている[4]．フィンガープリンティングとは，サーバである Web サイトが Web サイトにアクセスしたクライアント環境を識別する手法である．一般的に，ユーザの使用する環境に合わせたコンテンツを提供するために使用されるものであるが，ドライブバイダウンロード攻撃者は JavaScript を利用して，ユーザの使用するブラウザやプラグインの環境情報を取得し，その環境情報をもとにリダイレクト先 URL を変更する攻撃を行う．攻撃者は，このフィンガープリンティングによって，攻撃の成功率を向上させている．

#### 2.3.2 難読化スクリプトコード

ドライブバイダウンロード攻撃では，図 2 のような難読化されたスクリプトコードを利用するものがある．攻撃者は，スクリプトコード内の文字列の置き換えなど難読化を施し，攻撃解析をする第 3 者にスクリプトコードの挙動を簡単に解析できないようにしている．

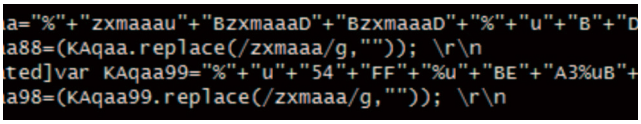


図 2　難読化スクリプトコードの例

## 3. 関連研究

### 3.1 難読化スクリプトコードの解析

難読化スクリプトコードの解析手法は，静的にコードを解析する方法とコードを実行させることによって解析する方法の 2 つある．

神薗らは JavaScript のコードを擬似的な環境で実行することにより難読化スクリプトコードを解析する手法を提案している[5]．難読化スクリプトコードの実行環境を監視しながら，JavaScript の関数 eval を使用することで難読化を解除したスクリプトを抽出する方法である．

また，神薗らは動的解析だけでは，JavaScript の挙動をすべて解析することはできないとして，抽象構文解析木を用いて，スクリプトコード自体ではなく，スクリプトコードの挙動を解析する方法を提案している[6]．この方法は，JavaScript コードに含まれる関数名や引数の値が異なっていても，JavaScript 構造を把握することができる．同じ攻撃パターンの JavaScript の構文解析木は全て同じ構造ではないが，似た構造をしているので，それを利用し JavaScript コードを悪性とみなす．つまり，抽象構文解析木をシグネチャとして悪性の難読化された JavaScript コードを見つける手法である．

### 3.2 悪性 URL の収集

Web クローラによる Web 巡回を行うことにより，悪性 URL を収集しブラックリストで悪性 Web ページの通信を遮断する手法がある．

ブラックリストの作成には，基本的に 3 つの処理が行われる．まず悪性と考えられる URL を収集する．収集には，Web クローラを用いて，Web ページに含まれるリンクを辿り，URL を収集する．次に，その URL に対してフィルタを利用し，解析者が実際に Web ページにアクセスし HTML，JavaScript のコードを評価し，点数をつける．この点数を用いて，Web ページが悪性であるかどうか判断し，ブラックリストを作成する．

### 3.3 HTTP ヘッダ解析

HTTP ヘッダ，IP アドレス，ドメイン情報を利用して，悪性と考えらえる HTTP 通信を検知し，攻撃を防ぐ手法がある．

安藤らは Web ページに含まれるリダイレクトの深さ，広がりを指標とし，ドライブバイダウンロード攻撃の悪性 Web ページとの通信を遮断し，攻撃を防ぐ手法を提案している[7]．リダイレクトの深さは，ブラウザによるファイルの自動読み込みやリダイレクトに対する指標であり，リダイレクトの段数と定義している．任意の Web ページの段数を 1 とする．その Web ページが読み込む Web ページの段数を 2 とする．Web ページの段数が 2 の読み込む Web ページの段数を 3 とする．3 段以上深い段数も同様にカウントする．リダイレクトによる通信は，段数が増すほど第 3 者のコンテンツが含まれ，Web ページに対する信頼度が低くなると仮定している．また，リダイレクトの深さは，異なるドメイン名へのリンク跨ぎの段数と定義している．ユーザがアクセスした Web サイトのドメインと異なる場合や広さの段数が増すほど，リダイレクトの深さ同様に Web ページに対する信頼度が低くなると仮定している．安藤らは， HTTP リクエスト，HTTP レスポンスの監視，取得，ヘッダの付加を行い，ユーザの操作による通信とブラウザによる自動通信を区別している．ユーザがクリックした URL の HTTP リクエストの HTTP ヘッダに”X-Action: 1; Click”ヘッダを付加する．ブラウザが自動で発生する HTTP リクエストには X-Action: [2,3,4,…]; Auto ヘッダを付加する．第 2 層の場合は X-Action:1; Auto となる．リダイレクトの深さ，段数をカウントする方法は，Referer 情報と Location 情報を利用する．HTTP リクエストについて，HTTP ヘッダに Referer が存在した場合，X-Action ヘッダの値に 1 を足す．また，HTTP レスポンスについて，Location 情報が存在した場合には，X-Action ヘッダの値に 1 を足す．これを HTTP 通信の行われる際に，処理を行っていく．X-Action ヘッダの値が 4 以上であった場合，通信を悪性とみなし，通信遮断する．また，Content-Type が application/pdf で，X-Action ヘッダの値が 0 でない場合，Content-Type が application/octet-stream ， application/x-download ， application/x-msdownload，application/x-msdos-program のいずれかでユーザが手動でアクセスした Web ページのドメイン名とリダイレクト先のドメイン名が異なる場合，発生する通信を悪性とみなし，通信遮断することでドライブバイダウンロード攻撃の攻撃検知する手法がある．

### 3.4 関連研究の問題点

難読化スクリプトコードの解析は，攻撃とみなされるパターンが存在すれば，攻撃を検知できる．しかし，未知の難読化スクリプトコードパターンに対応できない場合に攻撃を正確に検知できないという問題がある．

ブラックリストは，悪性 URL を収集し，配布することで，悪性 Web ページであるかどうか判断する．しかし，ドライブバイダウンロード攻撃では，フィンガープリンティングが利用される．フィンガープリンティングによって，収集される悪性 URL は Web クローラのブラウザ，プラグインに有効な攻撃を含む悪性 URL しか収集することができない．悪性 URL の収集だけでは，ドライブバイダウンロード

攻撃を防ぐことができるとは言えない．

また，現在の HTTP ヘッダ解析手法は，リダイレクトによって取得されるファイルが良性であっても，誤検知によってウイルスのダウンロードが引き起こる場合があるという問題が存在する．また，段数が低い悪性の通信を遮断できない場合が発生する場合があり，正確に攻撃検出することができていないという問題がある．

## 4. 提案手法

### 4.1 Web 階層

Web 階層とは，任意の Web ページがリダイレクトによって読み込む Web ページの構造と定義する．

任意の Web ページの階層を 1 とする．その Web ページが読み込む Web ページの階層を 2 とする．Web ページの階層が 2 の読み込む Web ページの階層を 3 とする．3 層以上深い階層も同様に階層をカウントする．

Web 階層を木構造で表すと図 3 のようになる．階層 1 の Web ページを木構造の根とする．階層 1 の Web ページが読み込む階層 2 の Web ページは，根の子となる．

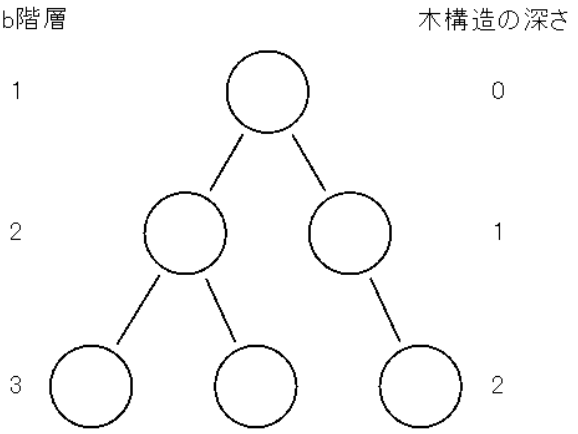


図 3　Web 階層

Web 階層は，HTTP ヘッダの Referer，Location を利用して，カウントする．

Referer は，任意の Web ページの遷移元を保持する情報である．任意の Web ページ A が存在し，A は Web ページ B を読み込むと仮定する．ユーザのブラウザが A を読み込んだ際，B を取得するため，ブラウザは，要求である HTTP リクエストを B に行う．その際，B の読み込みに細工がなければ，B の HTTP リクエストのヘッダに B は A から発生した通信であるという情報が付加される．この付加された情報が Referer である．

Location は，任意の Web ページを読み込んだ際，別の URL の読み込みを命令する情報である．任意の Web ページ C は Web ページ D へのリダイレクトが発生すると仮定する．ユーザのブラウザが C を読み込んだ際，C の HTTP レスポンスには別の Web ページ D の URL が記載されており，ブラウザは D を自動で読み込む．Location は，要求の応答である HTTP レスポンスに付加される情報である．

### 4.2 PageRank

PageRank は，ある論文の重要性は他の論文からの引用数によって評価されるという学術論文の考えを Web に適用したものである[8]．論文の重要度は，被引用数で決まり，PageRank は被リンク数が影響する．PageRank のアルゴリズムは，「任意の Web ページ A の PageRank は Web ページ A にリンクしている各ページの PageRank を，そのページからの外向きのリンク数で割った値の総和」として定義される． PageRank は Google から見た Web ページの重要度であり，0 から 10 の 11 段階でランク付けされる．PageRank の値が高ければ高いほど，Google から見た Web ページの重要度は高いとされる．Google は，自動化されたプログラムなどによって，リンクを作成するなど，不自然なリンクについてペナルティを与える場合もある．Google は，PageRank を上げるには，インターネットコミュニティで自然に人気を獲得すること，関連性の高い独自のコンテンツを作成することと述べている．本研究では，この PageRank を利用し，ユーザがアクセスする Web ページの信頼度のひとつの指標とする．

### 4.3 提案アルゴリズム

HTTP 通信の Web 階層と PageRank を用いてドライブバイダウンロード攻撃による悪性リダイレクトを防止する．HTTP 通信の通信遷移を利用し，Web ページの Web 階層をカウントする．条件を満たす HTTP 通信の情報と，PageRank を利用し，通信が悪性リダイレクトかどうかを判別する．

ユーザのクリックまたは URL バーに入力した Web ページの階層を 1 とし，その Web ページが読み込む Web ページの階層を Referer，Location を利用して，Web 階層をカウントする．階層が 4 以上の通信に対して，リクエスト URL の完全修飾ドメイン名 FQDN の PageRank を取得する．PageRank が 0 または，存在しない場合は，悪性リダイレクトとみなし，通信を遮断する．また，HTTP ヘッダは書き換えが可能なので，Referer が存在しない場合もある．このときは，リクエスト URL に含まれる FQDN の PageRank を取得し，PageRank が 0 または存在しない場合は同様に悪性リダイレクトとみなす．さらに階層が 2 以上で Content-Type が application/pdf の Web ページ，Web ブラウザが自動で読み込む Web ページの Content-Type が application/octet-stream, application/x-download , application/x-msdownload , application/x-msdos-program のいずれかで，階層 1 の FQDN とリダイレクト先の FQDN が異なる場合リダイレクト先の

FQDN の PageRank を取得し，悪性リダイレクトの判別を行う．文献[9]より悪性 Web サイトの生存期間は 1 日のものが多いとわかっている．生存期間が短い Web サイトに PageRank は存在しないため, Google に評価されてないを悪性 Web サイトと見なし，これらの Web サイトとの通信を遮断することで悪性リダイレクトを防ぐ．

## 5. 実験データと実験方法

### 5.1 実験概要

実験は，実際に行われた HTTP 通信を再現し，攻撃検出の精度を測る．提案手法，その他検知項目を変えた手法を適用し，実験を行う．

### 5.2 実験データ

#### 5.2.1 良性通信データ

Alexa Internet, Inc は，Web サイトのアクセス数の調査や統計をとっている．Alexa では，世界，国別のカテゴリでそれぞれアクセス数が高い Web サイト上位 500 件のランキングを公表している[10]．また，Web コンテンツのカテゴリ別で, それぞれのアクセス数が高い Web サイト最大上位 500 件を公表している．

本研究では，Alexa が公表する世界のアクセスランキングトップの Web サイトを実験対象とする．世界のアクセスランキングは Alexa から 2015 年 4 月 16 日に取得し, HTTP 通信を行う Web サイト 100 件を実験対象とする．

本実験では, HTTP 通信を行う Web サイト 100 件の URL を巡回して得られた通信データを良性通信データとし，本実験で使用する標本とする．実験対象の HTTP 通信を行う Web サイト 100 件は, 世界のアクセスランキングトップ 500 位以内にランクインしているが，必ずしも良性通信を行うとは限らない．しかし，本実験では，良性通信データの標本として定義する．また HTTP 通信を行う Web サイト 100 件の URL を実験で使用する．

クローラで Web サイトにアクセスし, tcpdump で良性通信データを収集した．

#### 5.2.2 悪性通信データ

NTT セキュアプラットフォーム研究所は，2010 年から Web クライアント型ハニーポットを使用し，ドライブバイダウンロード攻撃に関連するデータを収集している[11]．感染手法の検知，解析技術の研究のためにドライブバイダウンロード攻撃に関するデータ D3M(Drive-by Download Data by Marionette)2014 データセットを研究機関に提供している． D3M2014 データセットには過去に収集した通信データセット D3M2010，D3M2011，D3M2012，D3M2013 のデータセットが含まれる．

D3M データセットには，悪性 URL を巡回して得られたドライブバイダウンロード攻撃の攻撃通信データと巡回 URL のリスト，ドライブバイダウンロード攻撃によってクライアントにダウンロードされた実行形式ファイルのマルウェアのハッシュ値，マルウェアサンドボックス上で実行した際のマルウェア通信データが含まれる．本実験では，悪性 URL を巡回して得られたドライブバイダウンロード攻撃の攻撃通信データと巡回 URL のリストを実験対象とする．

攻撃通信データは，ハニーポットの通信を tcpdump でパケットキャプチャした PCAP 形式のファイルである．このファイルを悪性通信データとし, D3M2010 を除いたデータを本実験で使用する標本とする．また， 巡回で攻撃を検出した URL を実験で使用する．

### 5.3 標本のデータ抽出

標本となる実験データには HTTP ヘッダ以外の雑音となる通信が含まれる．そこで標本の実験データから HTTP リクエストと HTTP レスポンスのパケットを抽出し，抽出した各パケットからフレーム番号, HTTP リクエストフラグ, HTTP レスポンスフラグ，送信元ポート番号，送信先ポート番号，HTTP ヘッダの HOST 情報，Referer，Location，Content-Type，リクエスト URI を抽出した．

### 5.4 実験方法

実験は標本である実験データからパケット情報，HTTP ヘッダ情報を抽出したテキストファイルを使用する．一般的にユーザは Web ブラウザでリンクをクリック，URL バーに URL を打ち，Web ページへアクセスする．この動作を再現するために，本実験では，実験データの取得の際，巡回した URL のテキストファイルを使用する．2 種類のテキストファイルを読み込み，提案手法などの評価手法を適用する．

### 5.5 実験項目

実験の検知項目は以下の通りである．

I. 階層 3，4 または，5 以上の Web ページ
II. 階層 2 以上で Content-Type が application/pdf の Web ページ
III. Web ブラウザが自動で読み込む Web ページの Content-Type が application/octet-stream，application/x-download，application/x-msdownload，application/x-msdos-program のいずれかで，階層 1 の FQDN とリダイレクト先の FQDN が異なる場合
IV. Web ページの発生元が不明な場合

いずれかの条件に当てはまる場合，PageRank を取得し，攻撃かどうかを判断する. 本論文では, 検知項目を変更し，9 つの実験を行なった．

I の検知項目について，検知する階層の数を変化させる

によって攻撃検出率に差が見られると考えられる.そこで,階層を 3，4，5 にした場合，どのような結果になるために実験を行なった．また，既存手法では，用いられない IV の適用，PageRank を適用することで，どのように検知結果が変わるのかを実験を行なった.

#### 5.5.1 階層別の実験

提案手法 I の階層数 3, 4 または 5 にして実験を行なった.さらにそれぞれの階層数で，IV を適用しない場合，PageRank を適用しない場合の実験を行なった.

### 5.6 評価方法

真陽性率，偽陰性率，真陰性率，偽陽性率，全体の攻撃検出率の 5 つの評価項目によって各手法を評価する.

悪性通信データはセッション単位で評価を行う．ここでのセッションとは,ユーザが任意の 1 件の URL にアクセスした際に行われる通信を指す．本研究では，アクセス時間を考慮せず, ある Web サイトへの 1 回のアクセスを 1 セッションとする．真陽性率は，攻撃が発生したセッションにリダイレクト先の攻撃を未然に防いだ場合の割合である.(1)の計算式で真陽性率を評価する.

**真陽性率**

$$= \frac{\textit{マルウェアのダウンロードを未然に防いだ件数}}{\textit{悪性通信データに含まれるマルウェアの数}} \quad (1)$$

偽陰性率は，攻撃が発生したセッションに対して，リダイレクト先の攻撃を防げなかった場合の割合である.(2)の計算式で偽陰性率を評価する.

**偽陰性率**

$$= \frac{\textit{マルウェアをダウンロードした件数}}{\textit{悪性通信データに含まれるマルウェアの数}} \quad (2)$$

良性通信データは Web ページごとに評価を行う.真陰性率は，良性 Web ページを良性 Web ページとみなした場合である．(3)の計算式で真陰性率を評価する.

**真陰性率**

$$= \frac{\textit{良性Webページを良性とみなした件数}}{\textit{良性通信HTTPリクエスト・レスポンス組全数}} \quad (3)$$

偽陽性率は，良性 Web ページを悪性リダイレクトとみなした場合である．(4)の計算式で偽陽性率を評価する

**偽陽性率**

$$= \frac{\textit{良性Webページを悪性とみなした件数}}{\textit{良性通信HTTPリクエスト・レスポンス組全数}} \quad (4)$$

また全体の攻撃検出率として，(5)の計算式を用いる.

**全体の攻撃検出率**

$$= \frac{\textit{真陽性の件数} + \textit{真陰性の件数}}{\textit{真陽性} + \textit{偽陰性} + \textit{真陰性} + \textit{偽陽性の件数}} \quad (5)$$

## 6. 実験結果と考察

### 6.1 実験結果

実験結果を図 4~6 に示す．提案手法である実験 4 の全体の攻撃検出率が 96.28%と最も高い結果となった. 関連手法である HTTP ヘッダ解析の既存手法は, 真陽性率は, 99.21%, 真陰性率は, 77.03%, 全体の攻撃検出率は 96.23%であった.

### 6.2 考察

#### 6.2.1 提案手法の検出精度

提案手法は，関連手法である HTTP ヘッダ解析の既存手法よりも，真陽性率は，0.05 ポイント，真陰性率は，18.60 ポイント，全体の攻撃検出率は 0.05 ポイント上回った．わずかではあるが, PageRank を適用した場合の方が既存手法よりも，正確に検出できている．このことから良性データに関しては, PageRank は重要な攻撃検知項目であると考えられる.

図 4～6 から真陽性率について，実験 1，3，4，6，7，9 が高いことがわかる．これらの実験では，リダイレクトで発生する発生元が不明な Web ページを悪性とみなす手法，リダイレクトで発生する発生元が不明な Web ページの PageRank を取得して攻撃かどうか判別する手法を適用している. 発生元が不明な Web ページを悪性とみなすことで, ドライブバイダウンロード攻撃に対して高い検出率を達成できている．実験 1，3，4，6，7，9 はすべて IV を適用している．IV を適用していない実験 2，5，8 と比べ，高い真陽性率を達成できている．よって Referer が存在しない発生元が不明な Web ページを悪性かどうか判別することで高い真陽性が達成できると考えらえる.

実験 1，3 では，悪性リダイレクトとみなすことで，多くの悪性リダイレクトを悪性とみなすことができた．また, HTTP 通信において，Referer，Location で Web 階層の発生元が不明な Web ページのリクエスト URL の FQDN の PageRank を取得し, 悪性リダイレクトを検知することで悪性リダイレクトの検知率が高い結果になったと考えられる. 良性データに関して, 真陰性率が実験 4～9 に比べ低くなっている．多くの良性 Web ページを悪性と見なしている. PageRank を検知項目とした場合は, 良性データに含まれるトップサイトでも，PageRank が 0 または，存在しない Web サイトから Web ページを読み込む場合があるからであると考えられる．この現象は，階層別でなくても PageRank を検知項目とするとこの現象が見られる.

実験 4，6 では，多くの悪性リダイレクトを悪性とみなすことができた．

実験 7，9 に関し，同様，多くの悪性リダイレクトを悪性とみなすことができた．

IV を適用することで，高い真陽性率が達成できることがわかった．また，PageRank を適用することで，全体の攻撃検出率が 96.28%と最も高い検出率を達成している．よって，提案手法は有効であると考えらえる．

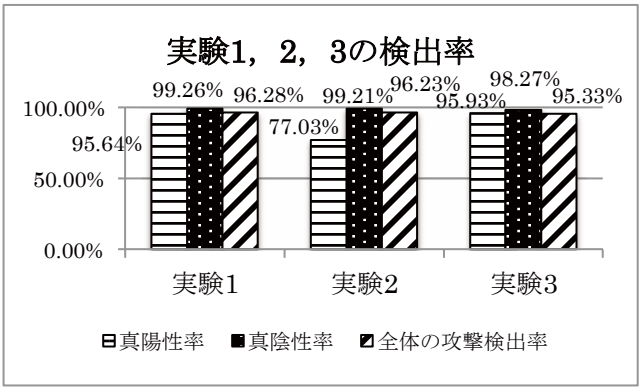


図 4　実験 1，2，3 の検出率

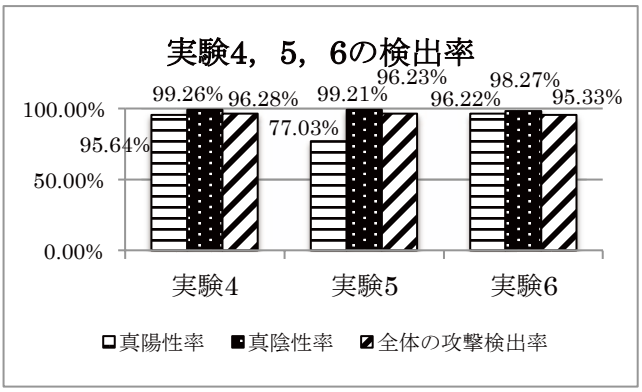


図 5　実験 4，5，6 の検出率

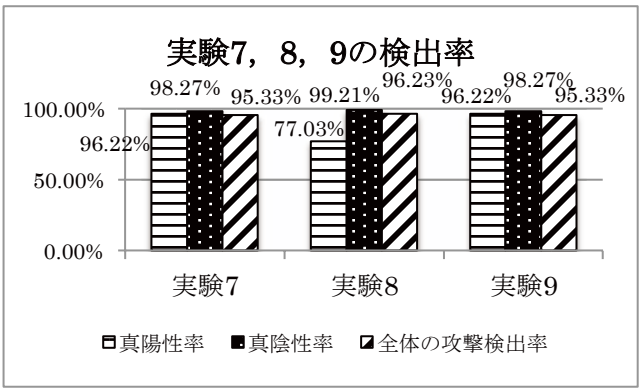


図 6　実験 7，8，9 の検出率

## 7. おわりに

### 7.1 まとめ

本研究では，Web を通じたマルウェア感染攻撃であるドライブバイダウンロード攻撃対策の既存手法よりも正確に攻撃を検知することを目的とした．この目的を達成するために HTTP ヘッダから Web ページの Web 階層，Web ページの通信遷移を明らかにして，PageRank を用いることで悪性リダイレクトの防止を行う手法を提案した．また，この手法の評価を行った．この評価として，Referer が存在しない発生元が不明な Web ページを悪性かどうか判別することで，高い攻撃検知率が得られた．この結果から提案手法は，ドライブバイダウンロード攻撃対策として既存手法よりも有効な対策手法であることが確認できた．また，提案手法の検知項目を変化させることによって，攻撃検出に有効な項目かどうかを調査することができた．

### 7.2 今後の展望

本研究により，Web ページの通信遷移を明らかにすることで，ドライブバイダウンロード攻撃の悪性リダイレクトを防ぎ，既存手法よりも正確に攻撃を検知することができたが，PageRank が悪性データに対して大きく有効的であるとは言えない結果となった．今後は PageRank ではない検知項目を利用し，マルウェアのダウンロードを防ぐ手法を提案したいと考えている．

## 参考文献


[1] マカフィー株式会社, "マンスリー ウイルスレポート | セキュリティ情報", http://www.mcafee.com/jp/threat-center/monthly/index.aspx

[2] IPA 独立行政法人 情報処理推進機構, "コンピュータウイルス・不正アクセスの届出状況[2010 年 11 月分]について", https://www.ipa.go.jp/security/txt/2010/12outline.html

[3] "RFC INDEX", http://www.rfc-editor.org/rfc-index.html

[4] C.Kolbitsch, B.Livshits, B.Zorn, C.Seifert, "Rozzle: De-cloaking Internet Malware", Proceedings of the 2012 IEEE Symposium on Security and Privacy (SP'12), pp. 443-457, 2012.

[5] 神薗雅紀, 西田雅太, 星澤裕二, "動的解析を利用した難読化 JavaScript コード解析システムの実装と評価", http://www.iwsec.org/mws/2010/presentation/2A3-1.pdf

[6] 神薗雅紀, 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 の特徴点抽出手法の提案", CSS2011

[7] 安藤慎悟, 寺田真敏, 菊池浩明, 趙晋輝, "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", 情報処理学会研究報告, Vol.2011-CSEC-54, No. 32, pp. 1-6, 2011.

[8] L.Page, S.Brin, R.Motwani, T. Winograd, "The pagerank citation ranking: Bringing order to the web", 1998

[9] 秋山満昭, 八木 毅, 針生剛男, "改ざん Web サイトのリダイレクトに基づく悪性 Web サイトの生存期間測定", 情報処理学会研究報告, Vol. 2014-SPT-8, No.32, pp. 1-6, 2014.

[10] Alexa, "Actionable Analytics for the Web", http://www.alexa.com/

[11] 秋山満昭, 神薗雅紀, 松木隆宏, 畑田光弘, "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, Vol. 2014-CSEC-66, No. 19, pp. 1-7, 2014.