Vol.2011-CSEC-54 No.33
Vol.2011-SPT-1 No.33
2011/7/13

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

安藤槙悟 [1)]　寺田真敏 [1)2)]　菊池浩明 [3)]　趙晋輝 [1)]

誘導リンクを埋め込むホームページ改ざんとマルウェア配信サイトとの組合せにより，ユーザがホームページを閲覧しただけでマルウェアに感染してしまう Web 感染型マルウェアによる脅威が増加しており，対策が急務となっている．本稿では，ホームページのリンクの深さと広がりを指標とした判定方法を導入することにより，Gumblar ウイルスのような Web 感染型マルウェアの感染活動に関わる異常な通信を検知する手法を提案する．ここで，ホームページのリンクの深さとは，ブラウザによるファイルの自動読み込みやリダイレクションの呼び出しの段数である．広がりとは，異なるドメインへのリンク跨ぎの段数である．プロトタイプ実装では，自動読み込みやリダイレクションによるホームページのリンクの深さを通信階層の段数として HTTP ヘッダに記録することで，提案方式を悪性 Web サイト検知システムとして実現すると共に，その有効性を評価した．

Proposal of Malicious Websites Detection System by Transition of Communications based on Unauthorized Redirection

Shingo Ando [1)], Masato Terada [1)2)], Hiroaki Kikuchi [3)], Jinhui Chao [1)]

In order to protect the web malware activities such as called a 'Gumblar', which uses Website manipulation to automatically lead users to go to a malware download site, it is necessary to improve the countermeasure environment. To solve this problem, we propose a method to detect malicious websites by watching the depth and width of links in home page. This paper described the overview of malicious websites detection system by transition of communications based on unauthorized redirection. Also, we evaluated effectiveness of the proposed method by using the prototype system we have implemented.

1. はじめに

インターネットが普及し利便性が向上する一方で，悪意のあるソフトウェアであるマルウェアによる被害が深刻化している．これと共に，マルウェアの攻撃手法の多様化・巧妙化は進んできており，感染形態にも大きな変化がみられる．2001 年頃から流布し始めたワームや 2004 年頃から出現した遠隔操作可能なボットの感染形態は，感染対象のホストに対してマルウェア自身が攻撃コードを送信する能動型感染マルウェアが主流であった．2008 年頃からは，ブラウザが利用するプラグインやアプリケーションの脆弱性を利用し，気が付かないうちにマルウェアのダウンロードと実行をおこなう攻撃手法，ドライブバイダウンロード(drive-by-download)を用いた Web 感染型マルウェアが主流となっている．特に，2009 年 5 月に出現した，マルウェア配信サイトへの誘導リンクを埋め込むホームページ改ざんとドライブバイダウンロード攻撃とを組み合わせた Gumblar ウイルス(ウイルス感染被害を発生させる Web 感染型マルウェアの俗称)には，対策を困難にさせる 3 つの要因が存在する．1 つは新たなマルウェアの出現頻度が非常に早いため，ウイルス対策ソフトによるパターンファイルの更新が追いつかないこと．2 つ目は，攻撃に用いられる JavaScript コードの難読化によるシグネチャ型検知の回避である．3 つ目は，ホームページや HTTP プロトコルというインターネットに欠かせない仕組みと通信プロトコルを巧妙に利用して感染活動を行っている点である．

本研究の目的は，上述の対策を困難にさせる要因を踏まえた，Gumblar ウイルスに代表される Web 感染型マルウェアの対策を検討することにある．そこで，本稿では，Gumblar ウイルスの感染活動の中に，(1) 誘導リンクを埋め込まれた正規サイト，攻撃コード配布サイトやマルウェア配布サイトなどの悪性 Web サイトが関わっていること，(2) ブラウザによるファイルの自動読み込みやリダイレクションを利用しているという特徴に着目した悪性 Web サイト検知システムを提案する．提案方式は，これら特徴に，ホームページのリンクの深さと広がりを指標とした判定を導入することにより，Gumblar ウイルスのような Web 感染型マルウェアの感染活動に関わる異常な通信を検知する手法である．また，提案方式を実装した悪性 Web サイト検知システムのプロトタイプの評価を通して有効性を示す．

[1)] 中央大学大学院 理工学研究科
Graduate School of Science Engineering, Chuo University
[2)] (株)日立製作所
. Hitachi Ltd.
[3)] 東海大学
Tokai University

Vol.2011-CSEC-54 No.33
Vol.2011-SPT-1 No.33
2011/7/13

2. 関連研究

本章では，Web 感染型マルウェアの検知と難読化 JavaScript コードの解析それぞれの関連研究について述べる．

2.1 Web 感染型マルウェアの検知

(1) 誘導リンクの埋め込まれた正規サイトの検知

文献[1]では誘導リンクとして埋め込まれたコードの記述パターンを元に，文献[2]ではコードの特徴をパラメータ化し，重み付けによる総合判定でホームページ改ざんを検知する方法を提案している．また，文献[3]では，WAF(Web Application Firewall)を利用し，Web サーバからの応答データに不正なコードが含まれている場合には，検知と共に，不正なコードそのものを機能しないようにする方法を提案している．

(2) 悪性 Web サイトの検知

悪性 Web サイトの URL やドメインの公開ブラックリスト化[4]，高対話型の Web クライアントハニーポットを用いたマルウェア検体の収集[5][6]などを通して，Web サイトの評判をデータベース化し，アクセスする際に，サイトの危険度判定情報として利用する Web レピュテーションの利用が広がりつつある．

(3) 不正なリダイレクトの検知

文献[7]では機械学習を用いて悪性 Web サイトへのリダイレクトを検出する方法を，文献[8]では不正なリダイレクトの検知の応用として，複数の Web サイトからのリダイレクトの集中度合いと個々の Web サイトのリダイレクト変更頻度に注目し，これらを相補的に用いてホームページ改ざん検知方式を提案している．

2.2 難読化 JavaScript コードの解析

Gumblar ウイルスでは，JavaScript コードに難読化を施し，他の URL に誘導する手法を使用している．難読化 JavaScript コードの解析についてはオンライン解析サイトとしてのサービス提供[9]や，文献[10]のように，悪性 Web サイトに誘導するスクリプトファイルだけではなく，PDF ファイル内の JavaScript の解析に着目した難読化 JavaScript コードの動的解析システムが検討されている．

3. 悪性 Web サイト検知システムの提案

本章では，ホームページのリンクの深さと広がりを指標とした判定方法により，Gumblar ウイルスのような Web 感染型マルウェアの感染活動に関わる異常な通信を検知する手法について述べる．

3.1 既存方式の課題と解決のアプローチ

前述の関連研究で示している通り，Web 感染型マルウェアの対策は，埋め込まれたコードに基づき正規サイトでのホームページ改ざんを検知したり，ブラックリストに基づき悪性 Web サイトであるか否かを判定したりする方式が主流である．これらの方式では，改ざんの検知対象になっていない正規サイトにアクセスした場合や，ブラックリストに登録されていない悪性 Web サイトにアクセスした場合には，被害に遭う可能性がある．

そこで，本提案方式では，Gumblar ウイルスに代表される Web 感染型マルウェアの感染活動の特徴に着目すること，ホームページのリンクの特徴に基づき調査することで，Web 感染型マルウェアの感染活動に関わる異常な通信の検知をおこなう．これにより，ホームページ改ざん検知やブラックリストのような登録手順を必要とすることなく，Web 感染型マルウェアの感染活動に関わる異常な通信の検知を可能とする．

3.2 Gumblar ウイルスの感染活動の特徴

本節では，Web 感染型マルウェアのひとつである，Gumblar ウイルスの感染活動の特徴について示す．

Gumblar ウイルスの感染活動には，正規サイトに埋め込まれた誘導リンク，悪性 Web サイトが提供する攻撃コードとマルウェアが関与している．はじめに，正規サイトに存在する脆弱性を悪用するなどして，ホームページに誘導リンクが埋め込まれる(図 1 の①②)．埋め込まれる誘導リンクは，HTML の meta タグ，iframe タグ，JavaScript などで転送(リダイレクト)を目的としたコードである．さらに，JavaScript だと，escape 文，replace 文などを用いて難読化される場合が多い．誘導リンクの埋め込まれた正規サイトにアクセスしたユーザは，攻撃コード配布サイトに転送(リダイレクト)された後，プラグインやアプリケーションの脆弱性が攻撃される(図 1 の③～⑥)．攻撃コードには難読化された JavaScript だけではなく，Flash や PDF ファイルなどが用いられる．脆弱性を悪用して実行された攻撃コードは，マルウェア配布サイトから悪意のあるプログラムをダウンロードした後，実行する(図 1 の⑦)．これにより，ユーザがホームページを閲覧しただけでマルウェアに感染してしまうことになる．

このような Gumblar ウイルスの感染活動には，誘導リンクを埋め込まれた正規サイト，攻撃コード配布サイトやマルウェア配布サイトなどの悪性 Web サイトが関わっており，これらのサイトをつなげるために，ブラウザによるファイルの自動読み込みやリダイレクションが利用されているという構成上の特徴がある．

Vol.2011-CSEC-54 No.33
Vol.2011-SPT-1 No.33
2011/7/13

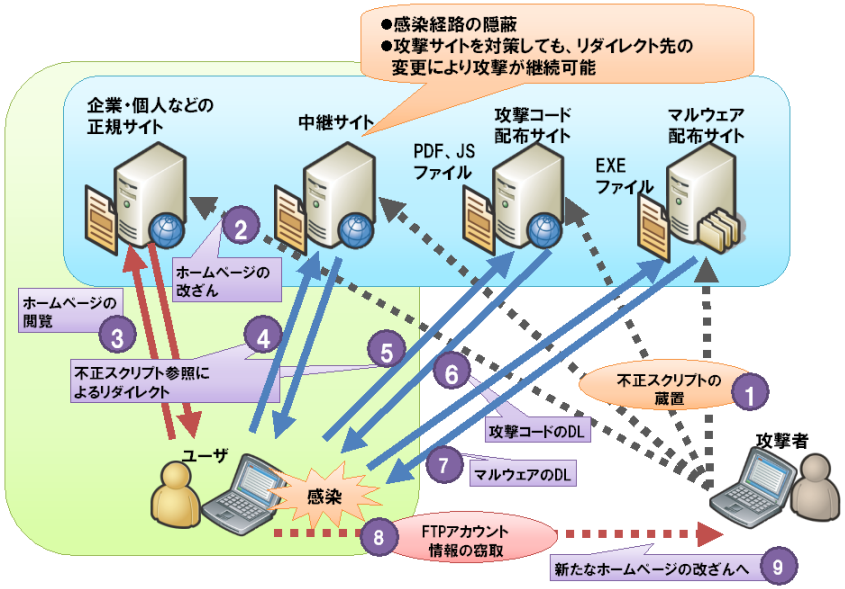

図 1：Gumblar ウイルスによる感染活動の流れ

3.3 ホームページのリンクの深さと広がり

ブラウザ上で 1 ページとして表示されるホームページは，リンクによって結びつけられた複数のファイルによって構成されている．本提案方式では，この点に着目し，「ホームページのリンクの深さ」「ホームページのリンクの広がり」というホームページのリンクに基づき特徴を調査するための新たな指標を導入する．

(1) リンクの深さ

ブラウザによるファイルの自動読み込みやリダイレクションに対する指標であり，自動読み込みやリダイレクションの呼び出しの段数と定義する．例えば，図 2 において，指定した URL からダウンロードした HTML ファイルを第 0 層とし，その HTML ファイルから呼び出される JavaScript ファイル，スタイルシートを第 1 層とする．さらに，第 1 層のファイルから呼び出される画像ファイルを第 2 層とする．第 1 層以降は，ブラウザによる自動読み込みとなるため，段数が増すほど，第三者が介在できる余地が広がるため，ファイルに対する信頼度は低くなる．

(2) リンクの広がり

誘導リンクを埋め込まれた正規サイト，攻撃コード配布サイトやマルウェア配布サイトなどの悪性 Web サイトが関わりに対する指標であり，異なるドメインへのリンク跨ぎの段数と定義する．リンクの広がりは，リンクの深さに依存する部分もあるが，異なるドメインへのリンク跨ぎの段数が増すほど，リンクの深さ同様，ファイルに対する信頼度は低くなる．

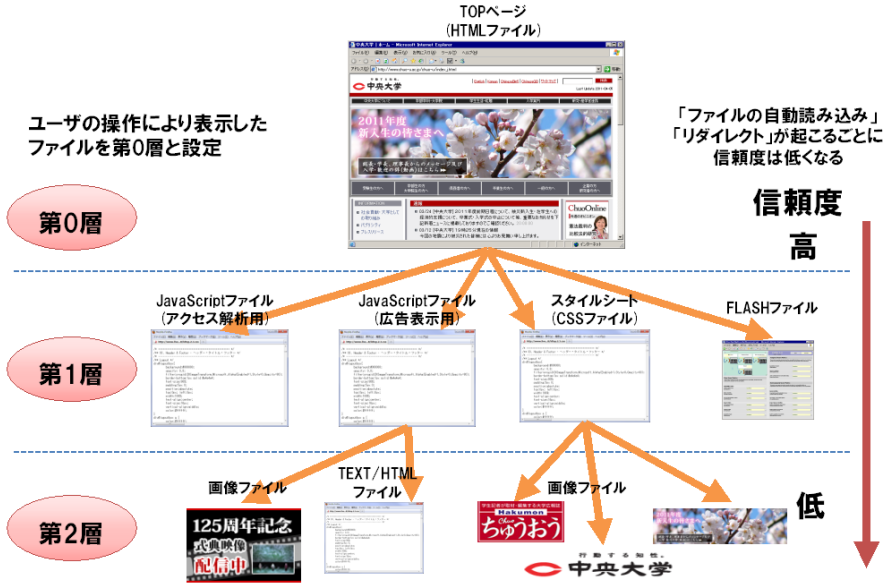

図 2：ホームページのリンクの深さ

3.4 提案方式

ホームページのリンクの深さと広がりという新たな指標を用いて感染活動に関わる異常な通信を検知する方法の基本的な考え方は，次の通りである．

- ユーザ操作による HTTP 通信(以降，ユーザ操作による通信)とブラウザの自動読み込みやリダイレクションによる HTTP 通信(以降，ブラウザによる自動通信)とを区別する．
- ブラウザによる自動通信に対して，ホームページのリンクの深さと広がりの点から制限を設定することで，感染活動に関わる異常な通信を検知する．

Vol.2011-CSEC-54 No.33
Vol.2011-SPT-1 No.33
2011/7/13

- ブラウザによる自動通信に対して，ダウンロードするファイル形式毎に，ホームページのリンクの広がりから制限を設定することで，感染活動に関わる異常な通信を検知する．

例えば，図 3 の場合には，感染活動に関わる異常な通信は，次の通りとなる．

- ホームページのリンクの深さが第 2 層以降で，かつリンクの広がりが 1 段目以降の PDF，EXE ファイルへのアクセス
- ホームページのリンクの深さが第 3 層以降で，かつリンクの広がりが 1 段目以降の HTML，JavaScript ファイルへのアクセス
- ホームページのリンクの深さが第 4 層以降のファイルへのアクセス

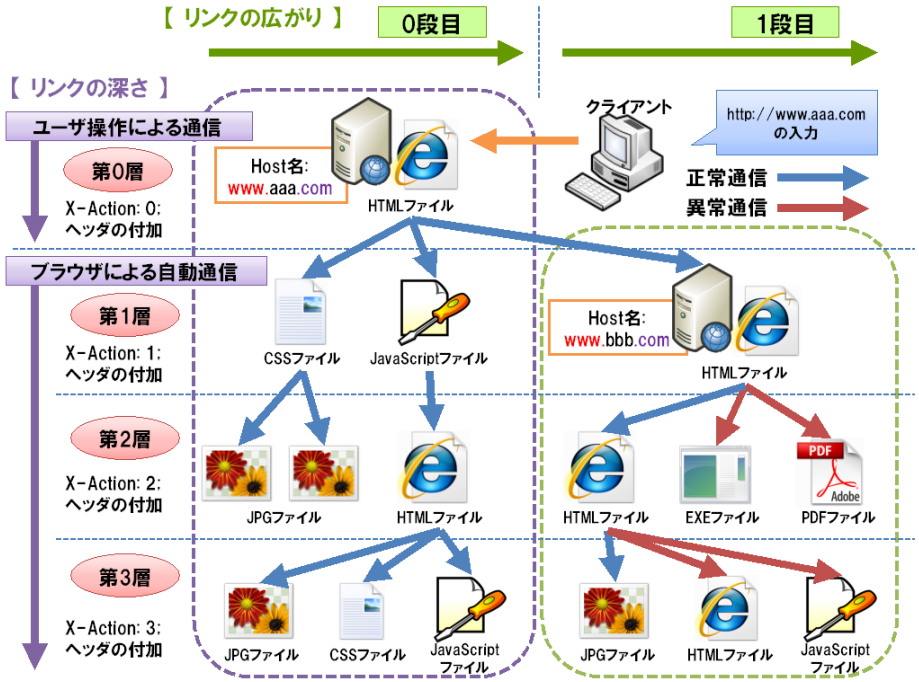

図 3：感染活動に関わる異常な通信の検知の例

4. プロトタイプ実装

本章では，ブラウザである Firefox の拡張機能(Firefox アドオン)を用いた悪性 Web サイト検知システムの実装について述べる．なお，ブラウザ上に悪性 Web サイト検知システムのプロトタイプを実装した理由は，ユーザ操作による通信とブラウザによる自動通信とを区別する必要があったことに加え，HTTPS 通信への対応，検知後の対策につなげることができるからである．

図 4 は，実装したプロトタイプのブロック図で，3 つの機能から構成される．プロトタイプの特徴は次の通りである．

- ユーザ操作による通信とブラウザによる自動通信とを区別し，ブラウザによる自動通信を，ユーザ操作による通信を基点として段数をカウントし，HTTP ヘッダ(X-Action ヘッダ)に記録する．
- ブラウザによる自動通信をホームページのリンクの深さと広がりの指標から判定する検知ルールを用いて通信遮断する．なお，リンクの深さ(X-Action ヘッダ)と広がり(Host ヘッダ)に関する情報は，HTTP ヘッダ上に記録しているため，通信遮断はブラウザだけではなく，IDS(侵入検知システム)，プロキシサーバでの実装が可能となる．

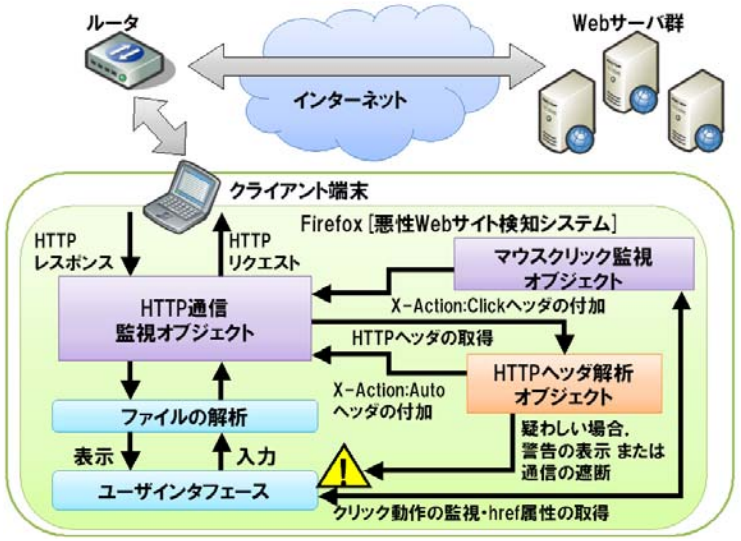

図 4：悪性 Web サイト検知システムのプロトタイプの概要

Vol.2011-CSEC-54 No.33
Vol.2011-SPT-1 No.33
2011/7/13

(1) HTTP **通信監視オブジェクト**

ホームページのリンクの深さを記録するため，nsIHttpChannel インタフェースを用いて，HTTP リクエスト／レスポンスの監視・取得と HTTP ヘッダ(X-Action ヘッダ)の付加をおこなう．取得した HTTP リクエスト／レスポンスは HTTP ヘッダ解析オブジェクトへ渡し，内容の解析をおこなう．また，マウスクリック監視オブジェクトから HTTP ヘッダの付加を指示された場合，指定された URL の HTTP リクエストの HTTP ヘッダに"X-Action: 0; Click"ヘッダを追加する．HTTP ヘッダ解析オブジェクトから HTTP ヘッダの付加を指示された場合には，指定された URL の HTTP リクエストの HTTP ヘッダに"X-Action: [1,2,3,...]; Auto"ヘッダを追加する(第 1 層の場合は"X-Action: 1; Auto").

(2) **マウスクリック監視オブジェクト**

ユーザ操作による通信とブラウザによる自動通信とを区別するため，addEventListener オブジェクトを用いて，ブラウザ内のマウスクリックの監視をおこなう．具体的には，イベントターゲット(ブラウザ)にイベントリスナ(左クリック)を登録することによって監視をおこない，クリックされた箇所のプロパティに href(Hypertext Reference)属性が存在した場合，HTTP 通信監視オブジェクトに URL を渡す.

(3) HTTP **ヘッダ解析オブジェクト**

HTTP ヘッダに記録されているリンクの深さ(X-Action ヘッダ)と広がり(Host ヘッダ)，コンテンツ(Content-Type ヘッダ)，リダイレクション(Location ヘッダ)，呼び出し元(Referrer ヘッダ)などの情報の解析による異常通信の検知や通信遮断をおこなう.

- HTTP リクエスト処理
  - Referrer ヘッダが存在した場合には，X-Action ヘッダの値+1 を指示する．
  - リンクの深さ(X-Action ヘッダ)が設定値以上の場合，以降のブラウザによる自動通信を遮断する．
- HTTP レスポンス処理
  - Location ヘッダが存在した場合には，X-Action ヘッダの値+1 を指示する．
  - Content-Type ヘッダに格納されているファイル形式が，EXE，PDF ファイルであった場合，リンクの深さ(X-Action ヘッダ)と広がり(Host ヘッダ)が設定値以上の場合，以降のブラウザによる自動通信を遮断する．

5. 評価

本章では，提案方式の有効性を確認するために，悪性 Web サイト検知システムのプロトタイプを用いて実施した評価について述べる.

5.1 評価項目と条件

(1) **評価項目**

- 正規サイトへアクセスした際の誤検知率
- 改ざんされた正規サイトへアクセスした際の検知率

(2) **評価にあたっての条件：ブラウザによる自動通信を遮断する判定ルール**

- リンクの深さ(X-Action ヘッダ)が 3 以上であった場合
- Content-Type ヘッダが application/pdf(PDF ファイル)で，かつリンクの深さ(X-Action ヘッダ)が第 0 層ではない場合
- Content-Type ヘッダが application/octet-stream，application/x-download，application/x-msdownload，application/x-msdos-program のいずれかで，HTTP リクエストのリンクの広がり(Host ヘッダ)が 1 段目以降の場合(第 0 層の HTTP リクエストの Host ヘッダに格納されているドメイン名が異なる場合)

5.2 評価結果

(1) **正規サイトへアクセスした際の誤検知率**

評価をおこなう正規サイトには，Google ディレクトリに登録されている Web サイトを使用した．評価には 12 のカテゴリ(アート，ニュース，オンラインショップ，ビジネス，家庭，ゲーム，社会，コンピュータ，健康，科学，スポーツ，各種資料)を用い，各カテゴリの上位に表示されていた Web サイトの中からランダムに 50 サイトを選び，合計 600 サイトを正規サイトとして評価をおこなった．プロトタイプでの誤検知率を表 1 に示す.

表 1：プロトタイプの誤検知率(正規サイトへアクセス)

項目	スロット数	割合(%)	HTTP セッション数	割合(%)
有効 URL 数	600		23,147	
3 階層以上の通信検知	19	3.17	148	0.64
PDF ファイルの検知	0	0	0	0
EXE ファイルの検知	1	0.17	1	0.004
誤検知	20	3.33	149	0.64

(2) **改ざんされた正規サイトへアクセスした際の検知率**

プロトタイプが実際の Gumblar ウイルスを検知できるのか調査するため，ホームページ改ざんに関する情報提供サイト[11][12]の URL を元に，インターネット上の改ざんされた正規サイトへアクセスを試み，マルウェアの侵入を未然に防ぐことができるかどうか検証した．ただし，改ざんされた正規サイトから辿ることで，攻撃コードやマルウェア配布サイトまでの通信遷移を収集したが，攻撃コード配布サイトにアクセ

Vol.2011-CSEC-54 No.33
Vol.2011-SPT-1 No.33
2011/7/13

スしても応答が 404 Not Found であったり，Web 感染型マルウェアに起因した改ざんでなかったりするものが非常に多く，該当する改ざんされた正規サイトは 4 件であった．この 4 件を対象に実施したプロトタイプでの検知率を表 2 に示す．

表 2：プロトタイプの検知率(改ざんされた正規サイトへアクセス)

項目	スロット数	割合(%)
3 階層以上の通信検知	3	0
PDF ファイルの検知	3	0
EXE ファイルの検知	2	0
計	8	0

5.3 考察

正規サイトへアクセスした際の誤検知率は 600 サイト中 20 サイトという結果であった．誤検知が発生した Web サイトを確認してみたところ，一部の広告画像が表示されなくなるのみであった．このことから，通常利用においては，一部の広告が表示されなくなるものの，主要コンテンツの閲覧には問題ないことがわかった．また，EXE ファイルでの誤検知が 1 件発生していたが，これは JPG ファイルであるにも関わらず，Content-Type が octet-stream と観測されたものであった．

検知率においては，Web 感染型マルウェアに起因した改ざん件数が少なかったため評価が不十分であるが，いずれの攻撃においてもマルウェア感染を阻止することができた．ただし，悪性 Web サイトに直接アクセスしてしまった場合には，プロトタイプでは感染活動を阻止することはできないため，閲覧する Web サイト自体の信頼度を評価する仕組みとの併用などが必要となる．

6. おわりに

本稿では，ブラウザによるファイルの自動読み込みやリダイレクションに着目した悪性 Web サイト検知システムを提案し，実装した．また，評価を通して，提案方式のプロトタイプが Web 感染型マルウェアを検知し，マルウェア感染を未然に防ぐことを示した．今後の課題は，マウスクリック以外のユーザ操作の反映，攻撃活動と検出した URL に対する処理の改善，Firefox 以外のブラウザへのシステム実装，XSS(クロスサイトスクリプティング)などの他種の攻撃への対応などがあげられる．

参考文献

[1] 田中達哉 他：改ざんサイト自動検知システム DICE の開発と評価，情報処理学会 コンピュータセキュリティ シンポジウム 2010 (2010 年 10 月)
[2] KDDI：「ウェブ改ざん検知システム」の「Gumblar (ガンブラー)」対応について http://www.kddi.com/corporate/news_release/2010/0115a/index.html
[3] Scutum：http://www.securesky-tech.com/service/waf.html
[4] Malware Domain List：http://www.malwaredomainlist.com/
[5] Mitsuaki Akiyama, et.al., Design and Implementation of High Interaction Client Honeypot for Drive-by-download Attacks, IEICE Transactions on Communication, Vol.E93-B No.5 pp.1131-1139 (2010).
[6] Takahiro Matsuki, et.al., Proposal of Service Cooperative Model Client Honeypot to Analyze Web-based Malware, 4th IFIP WG 11.11 International Conference of Trust Management (IFIPTM 2010), (2010).
[7] 寺田剛陽 他：検知を目指した不正リダイレクトの分析，情報処理学会 コンピュータセキュリティ シンポジウム 2010 (2010 年 10 月)
[8] 上松晴信 他：相補的な Web 感染型マルウェア検知方式の提案，情報処理学会 CSEC 研究報告 Vol.2011-CSEC-52 No.53 (2011 年 3 月)
[9] jsunpack-n：http://jsunpack.jeek.org/
[10] 神薗雅紀 他：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，サイバークリーンセンター・情報処理学会，マルウェア対策研究人材育成ワークショップ 2010 (2010 年 10 月)
[11] Zone-H.org - Unrestricted information：http://zone-h.org/
[12] JP ドメイン Web 改竄速報：http://izumino.jp/Security/def jp.html