3Z-8

HTTP 通信の時間軸解析を用いた Web 感染型マルウェア検知
Detecting Web-based Malware by Analysis of HTTP Traffic Timeline

永井 信弘† 千葉 大紀‡ 後藤 滋樹‡

† 早稲田大学 基幹理工学部 情報理工学科
‡ 早稲田大学 基幹理工学研究科 情報理工学専攻

概要

2009 年に Gumblar が出現してから，Web ページを閲覧しただけで感染する Web 感染型マルウェアの脅威が継続している．マルウェアに感染すると個人情報の漏洩や Web ページの改ざんなどの問題が引き起こされる．本研究は HTTP 通信を時間軸に沿って解析して，Web 感染型マルウェアの自動的なダウンロードと，ユーザによる正常な実行ファイルの手動によるダウンロードを識別する．この識別を用いて Web 感染型マルウェアの検知が可能となる．提案手法は Web ページの攻撃コードや実行ファイルの中身に依存しないため，既存の手法では検知が困難な，未知の Web 感染型マルウェアの検知に対して優位性がある．

1 提案手法

本手法では，実行ファイルのダウンロードに対応する 2 つの**差分時間**を算出することで Web 感染型マルウェアの検知を行う．具体的には，(1) Web ページにアクセスしてから，検査対象実行ファイルのダウンロードを開始するまでの差分時間，(2) 検査対象実行ファイルと直近の実行ファイルとのダウンロード開始時刻の差分時間を算出し，どちらかの差分時間が閾値よりも短い場合に Web 感染型マルウェアであると判定する．(1) の差分時間を t_{dw}，(2) の差分時間を t_{de} としたときの，t_{dw} および t_{de} の時間軸上での関係を図 1 に示す．

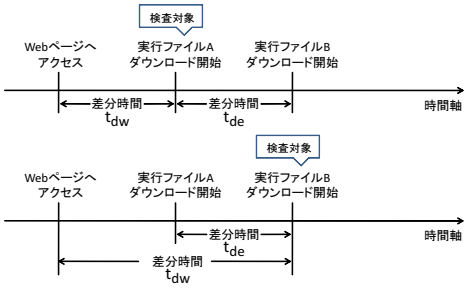

図 1 2 つの差分時間の関係.

以降では，本手法が Web 感染型マルウェア検知に用いた 2 つの差分時間と，差分時間の算出の際に適用した例外について順番に説明する．

1.1 Web ページアクセスからの差分時間

ユーザが Web ブラウザを用いて実行ファイルを取得する際は，以下に示す手順でダウンロードが開始される．

- Web ページへアクセス→実行ファイルへのリンクを探す→リンクをクリック→ダウンロード開始

一方，Web 感染型マルウェアがダウンロードされる場合，悪性 Web ページへアクセスした後には，ユーザは一切操作を行わない．そのため，Web 感染型マルウェアがダウンロードされる際には，ユーザによるリンクの探索とクリックという手順が省略される．

よって，Web ページへアクセスしてから実行ファイルのダウンロードを開始するまでの差分時間が著しく短かった場合には，対象の実行ファイルは Web 感染型マルウェアの可能性が高い．

なお，本研究では Web ページを構成するファイルのうち，表 1 に記載した種類のファイルのダウンロードが完了した時刻を，Web ページへアクセスした時刻と定義する．これらのファイルのダウンロードが完了すると，Web ブラウザ上に Web ページの外観が表示され，ユーザが実行ファイルへのリンクの探索を開始できるようになるためである．

表 1 Web ページの外観表示に必要なファイル.

・CSS
・HTML
・Java Archive
・Java Class File
・JavaScript
・PDF
・SWF
・XHTML
・XSL
・XSLT

1.2 直近の実行ファイルダウンロードとの差分時間

悪性 Web ページは，一度の攻撃で複数の Web 感染型マルウェアを配布する場合がある [1]．この場合，Web

Copyright ©2013 Information Processing Society of Japan.
All Rights Reserved.

感染型マルウェアのダウンロードを開始してから，次の Web 感染型マルウェアのダウンロードを開始するまでにはユーザの操作は一切必要としない．一方，ユーザが複数個の実行ファイルを手動でダウンロードする場合，クリック等の操作を行うことで各実行ファイルへのリンクにアクセスする必要がある．すなわち，複数の実行ファイルのダウンロード開始時刻の間隔が著しく短い場合には，Web 感染型マルウェアである可能性が高い．よって検査対象実行ファイルのダウンロード開始時刻と，直近にダウンロードされた実行ファイルのダウンロード開始時刻との差分時間が著しく短い場合には Web 感染型マルウェアであると判定できる．

1.3 良性リダイレクトによる例外

ユーザが手動で実行ファイルのダウンロードを行う際に，Drive-by-Download 攻撃と同じように複数の Web ページへの遷移が行われる場合がある．具体的には，Web ページ内に埋め込まれた実行ファイルへのリンクが直リンクではなく，リダイレクトを利用して複数の Web ページを経由する場合である．この場合，以下のような遷移で実行ファイルのダウンロードが開始される．

- Web ページ A →クリック→ Web ページ B
  →リダイレクト→実行ファイルのダウンロード開始

上記のような遷移でダウンロードが行われる場合，t_{dw} を用いた検知方式では，正常な実行ファイルを Web 感染型マルウェアであると誤検知する可能性がある．なぜなら，Web ページ B から実行ファイルのダウンロード開始までの遷移はリダイレクトにより自動的に行われるためである．本研究では，このような誤検知を防ぐために，Web ページを構成する HTML ファイルもしくは Javascript ファイルに表 2 のリダイレクト用コードが含まれていた場合には，例外としてその Web ページを t_{dw} を算出する対象から除外することとした．これらのコードは Web 感染型マルウェアの強制ダウンロードに必要ではなく，悪性 Web ページに含まれる可能性は低い．

表 2　良性リダイレクト用コード.

・http-equiv=”refresh”
・location.assign
・location.href
・location.replace
・window.location
・window.open

2 性能評価

2.1 データ

悪性通信のデータとして，Drive-by-Download 攻撃の通信データを集めた D3M データセット (D3M2010, D3M2011, D3M2012) [2] を用いる．同様に良性通信のデータとして，ある実ネットワークで観測された HTTP 通信のキャプチャデータを用いる．各データセット内でダウンロードされた実行ファイルの個数を表 3 に示す．なお良性通信データは，HTTP 応答内の Accept と User-Agent フィールドを参照することで，あらかじめ Web ブラウザ以外による HTTP 通信を除外している．

表 3　データセットの検体数.

データセット	検体数
悪性	709
良性	164

2.2 評価結果

本手法を 2.1 節のデータセットに適用した結果を表 4 に示す．ただし，表 4 には精度の良い閾値 1 秒付近の結果を抜粋した．なお，表中の TPR (True Positive Rate) は Web 感染型マルウェアを正しく Web 感染型マルウェアであると判定した割合，FPR (False Positive Rate) は正常な実行ファイルを Web 感染型マルウェアであると誤検知した割合である．

実験結果から，閾値を 1 秒付近に設定することで，本手法により低 FPR で 45% 程度の TPR を実現できた．また，TPR と FPR はトレードオフの関係になっている．

表 4　提案手法の性能.

閾値 [s]	TPR	FPR
0.7	43.4%	1.2%
0.8	45.4%	1.8%
0.9	46.8%	1.8%
1.0	47.3%	3.7%
1.1	47.8%	4.3%
1.2	47.8%	4.3%
1.3	48.2%	4.9%

3 まとめ

本研究では，HTTP 通信を時間軸に沿って解析することで Web 感染型マルウェアを検知する手法を提案した．性能評価の結果，低 FPR で Web 感染型マルウェアの検知が可能であることを示した．本手法は，Web ページ内の悪性コードや実行ファイルの中身に依存しない手法であるため，未知の Web 感染型マルウェアの検知において有効である．今後の課題は，実ネットワーク上での本手法による Web 感染型マルウェア検知の性能評価である．

参考文献

[1] 寺田 剛陽, 古川 忠延, 東角 芳樹, 鳥居 悟, 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 3F1, October 2010.

[2] MWS2012 実行委員会, 研究用データセット MWS2012 Datasets について, http://www.iwsec.org/mws/2012/about.html

Copyright ©2013 Information Processing Society of Japan.
All Rights Reserved.