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

工藤　聖 †　　トラン・コン・マン †　　中村　康弘 †

† 防衛大学校
239-8686 神奈川県横須賀市　走水 1-10-20
em54036@nda.ac.jp, manhtc@gmail.com, yas@nda.ac.jp

あらまし　不正リダイレクトによる Drive-By Download 等，Web アクセスを用いたサイバー攻撃が問題となっている．本稿では，個別クライアントが発する HTTP リクエストに含まれる特徴的な URL シーケンスを検出することにより，不正リダイレクトの候補を検出する方法を提案する．クライアントが発する全てのリクエストを対象にした場合，その組み合わせの数が膨大となり現実的ではない．そこで，特定クライアントが特定時間内に連続的に発する HTTP リクエストシーケンスの類似性を特徴量とする．組織の Proxy ログを用いた検証実験の結果を述べる．

Detection of suspicious redirection using HTTP request sequence

Sei Kudo†　　Tran Cong Manh†　　Yasuhiro Nakamura†

†National Defense Academy
1-10-20 Hashirimizu, Yokosuka-city, Kanagawa-Pref 239-8686, JAPAN
em54036@nda.ac.jp, manhtc@gmail.com, yas@nda.ac.jp

Abstract　A cyber attack using Web access such as Drive-By Download by the unjust redirection becomes the problem. In this report, technique to detect candidates of the suspicious redirection by detecting the characteristic URL sequence that an individual client gives is proposed. When all combinations of the request that a client gives become target of the analysis, the number of combination become enormous and it is not realistic. Therefore, characteristic quantity is assumed to the similarity of the HTTP request sequence that an authorized client gives continually before specific time. Result of inspection experiment using the Proxy log of the organization is described.

1　はじめに

近年，改竄された Web サイトから不正なリダイレクトを経てマルウェアをダウンロードさせられる，Drive-By Download（以下「DBD」という.）攻撃が問題になっている．マルウェアによる被害を局限するためには，DBD 攻撃が実行された場合に直ちにこれを検知し，アラートを発する，あるいは通信を遮断する等の迅速な処置が重要となる．

そこで，特定クライアントが DBD 攻撃を受けた場合，特定時間内に連続的に発する HTTP リクエストのシーケンスに特徴的なパターンが現れることを利用し，攻撃を検知する手法を提案する．

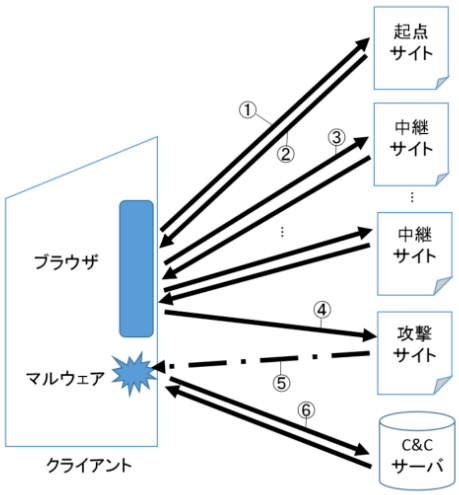

図 1: Drive-By Download の動作概要

2 Drive-by Download

DBD は，Web サイトにアクセスしたクライアントに強制的にマルウェアをインストールさせる攻撃である．概要を図 1 に示す．

DBD 攻撃は，ユーザが使用するコンピュータ等（以下「クライアント」という.）が，攻撃の起点となる Web サイト（以下「起点サイト」という.）にアクセスすることにより開始される．クライアントは通常，Web ブラウザ等のアプリケーション（以下「エージェント」という.）を用いて，Web サーバに HTTP リクエストを送信し，Web サイト閲覧等，所望のサービスを利用する．DBD 攻撃者は，外部から改竄した正規の Web サイトや，最初から悪意をもって作成した Web サイトを起点サイトとして用意し，クライアントからのアクセスを待ち受ける．クライアントが Web サーバにアクセスし，HTTP リクエストが送信される（図 1①）と，起点サイトは iframe タグや JavaScript コードを含む悪意ある HTTP レスポンスをクライアントに送信する（図 1②）．

悪意ある HTTP レスポンスを受信したクライアントは，記述されたタグやコードに従って別の Web サイト（以下「中継サイト」という.）にリダイレクトする（図 1③）．中継サイトは同様の方法によって他の中継サイトにリダイレクトさせ，最終的にマルウェアの配布を目的としたサイト（以下「攻撃サイト」という.）にリダイレクトさせる（図 1④）．

攻撃サイトは，クライアントの脆弱性を利用してマルウェアをダウンロードさせる（図 1⑤）．ダウンロードされたマルウェアは，指示等を行うために攻撃者が設置したサーバ（以下「C&C サーバ」という.）との交信や，取得した情報の送信を目的として，外部との通信を行う（図 1⑥）．

3 先行研究

DBD 攻撃対策は，近年活発に研究が行われている．

攻撃サイトに着目した研究としては，悪性サイトを観測してドメインをブラックリスト化する方法 [1]，既知の悪性 URL と類似した URL を検索する方法 [2] がある．

DBD によって発生する通信を検出する研究としては，ペイロードの内容を解析するものと，ヘッダ情報のみに基づいて判定するものがある．ペイロードを解析する研究については，ページ間の遷移及び参照関係に基づく判定基準を設定した松中らの手法 [3] がある．ヘッダ情報を利用するものは，パラメータを特徴量に変換し機械学習により判定する手法が提案されている．松本ら [4] は，ファイルの種類，サイズ，時刻等を特徴量として決定木学習（Weka）による識別を行うと共に，GeoLocation に基づく可視化と併用する手法を提案している．また，進藤ら [5] は，HTTP リクエストで取得するファイルタイプ遷移を，Content-Type 情報を特徴量として SVM により判定している．

機械学習を用いない方法としては，通信の定性的な特徴（メソッド，サイズ，ステータスコード，ユーザーエージェント等）に基づいて検知ルールを設定する大谷らの研究 [6]，レスポンスヘッダに PHP のバージョン情報が含まれているかを基準とする酒井らの研究 [7] がある．

これらの手法のうち，攻撃サイトに着目した手法については，把握されてない新たな攻撃サイトには完全に対応できない．通信を検知する手法のうち，ペイロードを解析の対象とする手

法は，解析対象のデータが膨大になる．酒井らの手法は単純な規則を用いているが，False Positive（以下「FP」という.），False Negative（以下「FN」という.）の確率が比較的高く，検知は確実ではない．大谷らの手法は，Exploit Kitの特性を分析して軽量かつ確実な検知ルールを提案しているが，未知のパターンには対応できない．

また，アクセス先の遷移等，アクセス間の相互関係を考慮する場合，情報を保持するメモリの容量が膨大になる可能性があり，その点を考慮した手法が求められる．そこで，

- 未知のExploit Kit等によるDBD攻撃でも検知できる
- FNの確率が低い
- 少ないオーバーヘッド及びメモリ消費で実装できる

といった条件を満たす検知手法として，次に述べる方法を提案する．

4　提案手法

単一のクライアントに対し，以下のパターンに該当したリクエスト群（以下「シーケンス」という.）をDBD攻撃候補と判定する．

条件1 HTTPレスポンスのステータスコードが300番台

条件2 条件1の生起後1秒以内に特定のファイルタイプをダウンロード

条件3 条件2の生起後120秒以内にユーザーエージェントが変化

提案手法の各条件の意義について説明する．

4.1　条件1

DBD攻撃の際は起点サイトから中継サイト，中継サイトから攻撃サイトへのリダイレクトが発生する．その際，HTTPレスポンスのステータスコードは300番台となる．よって，これをDBD攻撃を検知する条件とする．

4.2　条件2

リダイレクトを繰り返した結果，脆弱性を利用してマルウェアがダウンロードされる．利用される脆弱性はJava，Adobe Flash Player，Adobe Reader等が報告されている．ここでは，exe，swf，pdf形式のファイルを検出の対象とした．iframeタグやJavaScriptによるリダイレクトは，極めて短時間に実行されると考えられる．アクセスログのタイムスタンプは1秒単位が多いため，リダイレクトからダウンロードが1秒以内に発生していれば，DBDによるマルウェアの可能性が高いと判断する．

4.3　条件3

ダウンロードされたマルウェアは，クライアントのコンピュータ内部で，外部との通信を行うと考えられる．この際，記録されるHTTPリクエストのユーザーエージェントは，ユーザーが普段使用しているブラウザとは異なるものになる．よって，HTTPリクエストを時系列で見た時，ダウンロードから一定時間内に異なるユーザーエージェントからのリクエストが発生していれば，マルウェアの可能性が高いと判断する．なお，マルウェアのダウンロードから通信開始までの時間は，経験的に120秒以内と想定した．

5　検証実験

5.1　実装

提案手法をPerlスクリプトで実装し，Linuxサーバ上で実行した．実装の際は，オーバーヘッドの低減のため以下の処置を行った．

- htmlファイルと同時にダウンロードされる画像ファイルに対するリクエストを解析対象から除外
- アプリケーションの定期的な更新のためのアクセスのうち，WindowsUpdate，ウィルス対策ソフト等，既知のドメインに対するリクエストをホワイトリストにより除外

表 1: Proxy ログ記録内容

項目	例
タイムスタンプ	03/Jul/2015:14:12:34 +0900
クライアント IP アドレス	10.19.0.xx
レスポンスステータスコード	403
メソッド	GET
URI	http://xxx
受信サイズ	4054
送信サイズ	428
ユーザーエージェント	Mozilla/5.0(Windows NT 6.1; Trident/7.0; rv:11.0) like Gecko

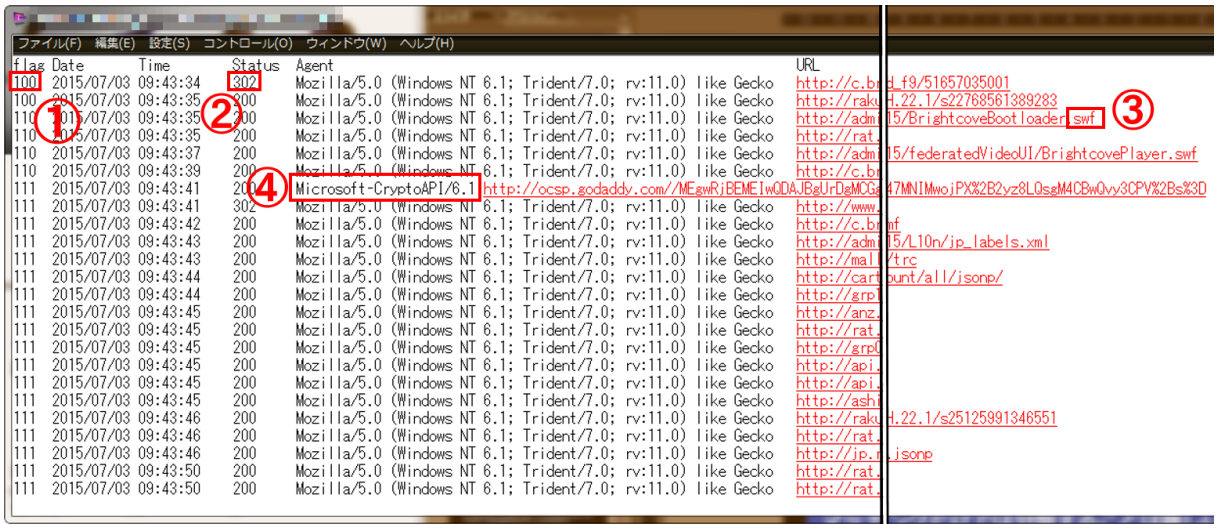

図 2: 抽出シーケンス例

5.2 検証結果

5.2.1 検証 1

提案手法を実際のアクセスに適用した場合の，一致件数と具体例を検証するため，組織の Proxy サーバのログを用いて検証を行った．期間は 2015 年 7 月 1 日〜31 日であり，行数は 96902 行である．ログの記録項目を表 1 に示す．検証の結果，16 件のリクエストシーケンスが DBD 攻撃候補として抽出された．なお，今回検出されたシーケンスの中に実際の DBD 攻撃は含まれていなかった．検出されたシーケンスの一例を図 2 に示す．flag の列は，上位の桁から順に条件 1，条件 2，条件 3 に該当したことを示している（図 2①）．1 行目でステータスコードが 302 であったため条件 1 に該当（図 2②），3 行目で 1 秒後に swf ファイルへのリクエストが行われたため条件 2 に該当（図 2③），7 行目で異なるエージェント（Microsoft-CryptoAPI/6.1）からリクエストが発生したことにより条件 3 に該当（図 2④）し，一連のシーケンスが抽出されている．

5.2.2 検証 2

組織においては，実際の DBD 攻撃時のログを保有していないため，実験が行えない．しかし，メールで送付された URL を直接クリックすることによりマルウェアをダウンロードさせられた事例があったため，当該時刻のログに対

して提案手法を適用した．リダイレクトが行われていないため，条件 1〜3 を適用した場合は検出されなかったが，条件 2 及び 3 のみを適用した場合は検出された．

5.2.3 考察

検証実験の結果，FP が現実的な範囲に収まることが確認できた．また，少なくとも DBD 攻撃によってマルウェアがダウンロードされてから外部との通信を開始するプロセスを検知できることが確認できた．

6 まとめ

本稿では，DBD 攻撃の特性に基づき，HTTP リクエストから攻撃を検知する手法を提案した．組織のログを使って検証した結果，DBD 攻撃の可能性があるリクエストシーケンスを抽出できることを確認した．今後は，アルゴリズムを改良し，FP の確率を低減させる方法を検討する．また，実際の DBD 攻撃時のログの入手または擬似的な DBD 攻撃の再現等により，検知性能の確認を行っていく．また，運用中の Proxy サーバに実装し，リアルタイムでの DBD 攻撃の検知を目指す．

参考文献

[1] 須藤 年章: 悪性サイトドメインの長期観測結果に基づくブラックリスト利用の有効性に関する一考察, コンピュータセキュリティシンポジウム 2013 論文集 2013(4), 376-381, 2013-10-14

[2] 孫 博, 秋山 満昭, 八木 毅, 森 達哉: 既知の悪性 URL 群と類似した特徴を持つ URL の検索, コンピュータセキュリティシンポジウム 2014 論文集 2014(2), 1-8, 2014-10-15

[3] 松中 隆志, 山田 明, 窪田 歩: 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, 研究報告コンピュータセキュリティ（CSEC）2015-CSEC-68(48), 1-8, 2015-02-26

[4] 松本 浩明, 石井 啓之, 薄羽 大樹, 菊池 浩明: Drive-by-Download 攻撃通信の可視化システム, コンピュータセキュリティシンポジウム 2014 論文集 2014(2), 9-16, 2014-10-15

[5] 進藤 康孝, 佐藤 彰洋, 中村 豊, 飯田 勝吉: マルウェア感染ステップのファイルタイプ遷移に基づいた 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 2014 論文集 2014(2), 575-582, 2014-10-15

[6] 大谷 尚通, 益子 博貴, 重田 真義: 実環境におけるサイバー攻撃検知システムの有効性評価および検知範囲の拡大に向けた検討, コンピュータセキュリティシンポジウム 2014 論文集 2014(2), 583-589, 2014-10-15

[7] 酒井 裕亮, 佐々木 良一: 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） 2013-CSEC-60(29), 1-6, 2013-03-07