2Y-7

Web 感染型マルウェアのリダイレクト解析
Redirect analysis of web-based malware

髙田 雄太† 森 達哉†‡ 後藤 滋樹†

† 早稲田大学 基幹理工学部 情報理工学科
‡NTT サービスインテグレーション研究所

概要

Web 感染型マルウェアの攻撃手段の一種である Drive-by-Download (DBD) 攻撃は，複数のサイトへの自動リダイレクトを用いることによって検出を回避する仕組みを備えている．また個々のリダイレクトには難読化が施されており，リアルタイムでの検出は困難である．本研究は実際の DBD 攻撃通信データを分析し，悪性リダイレクトの検出に有効な特徴を明らかにする事を狙いとする．

1 はじめに

Web 感染型マルウェアによる被害が拡大している．Web 感染型マルウェアの大きな特徴の一つは，Web ブラウザやプラグインの脆弱性を突いて，利用者が気づかぬうちにマルウェアをダウンロード，インストールさせる DBD 攻撃 [1] を仕掛ける点にある．DBD 攻撃は，複数のサイトへのリダイレクトを実施することで攻撃サイトの検出を困難にしている (図 1)．さらに各々のリダイレクトは難読化が施されたスクリプトによって実施されるため，DBD 攻撃に関わる踏台サイト，攻撃サイト，配布サイトに対応する悪性 URL の抽出は困難な課題である．本研究では，実際の DBD 攻撃を，制御された環境で実行して解析することにより，リダイレクトの性質を明らかにする事を目標とする．このようにして明らかにした性質を活用すると，未知の難読化が施されたリダイレクトの検出が可能となる．

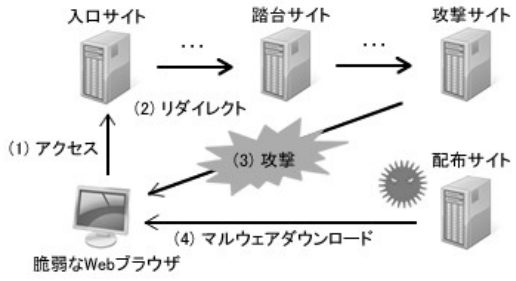

図 1 Drive-by-Download 攻撃の概要.

2 リダイレクト解析

通常の HTTP 通信におけるリダイレクトは，GET 要求に含まれる Referer フィールドや，サーバーが返答するステータスコードを参照する事で追跡できる．すなわち，GET 要求の Referer フィールドに過去に参照した URL が存在した場合は，その URL をリダイレクト元と判定出来る．また，HTTP ステータスコードがリダイレクトに対応する 300 番台である場合には，HTTP 応答ヘッダーの Location フィールドにリダイレクト先 URL が入る (後述の URL 推定)．

しかしながら DBD が用いるリダイレクトは，難読化した JavaScript 等によって Referer を隠蔽する手口が多数を占めている．実際，後に示すように本研究で検出した悪性リダイレクトの内で，Referer フィールドが存在したものはわずかに約 10% であった．本研究では DBD 攻撃通信のみが存在する，制御された環境で計測した通信を詳細に分析するアプローチを採用する．すなわち DBD 攻撃の入口 URL をブラウザに投入し，各種の踏台サイト，攻撃サイトを経由し，マルウェア配布サイトからマルウェアのダウンロード至るまでの通信データを分析する．

2.1 リダイレクト検出方法

リダイレクトとその他の通信を区別するために，一連の攻撃通信が有効である時間の範囲を経験的に決定した．すなわち，解析の始点となるある HTML オブジェクトの GET 要求から，現在解析している HTML オブジェクトの HTTP 応答 + 約 120 秒以内に発生した GET 要求および HTTP 応答に対して，以下で述べる方法を適用することにより，悪性リダイレクトを推定する．ここでは，リダイレクトを特定することで時間範囲は広くなる．

■URL 推定 URL 推定では HTTP オブジェクトおよび Location フィールドに含まれる URL を用いる事によって，リダイレクトを検出する．第一の方法では，はじめに HTTP で送信された HTML オブジェクトの文字列解析を行い，URL を抽出する．ここで分析の対象

となる HTML オブジェクトは，いずれも DBD 攻撃の入口サイトから発生したものであり，後続するすべての HTTP 通信は攻撃によって自動的に引き起こされたものであると仮定する．次に上記で構築した URL リストに対する GET 要求を検索し，存在した場合にリダイレクトによって引き起こされたものと判定する．第二の方法では 300 番台のステータスコードを含む HTTP 応答ヘッダーにおいて Location フィールドに記録された URL をリダイレクト先 URL として抽出する．

■Host **推定**　Host 推定では，はじめに通信データに出現したドメイン名 D と IP アドレス A のペア (D,A) をすべて保持しておく．次に再び最初から通信データを読みこみ，入口サイト毎に Referer や URL 推定で特定できたペアを保持しておく．最後にそのペアに対する新規の HTTP 通信が発生していた場合にはその HTTP 通信をリダイレクトとみなす．今回の実験では最初にブラウザに URL を与える後は自動的に通信の遷移が行われるため，このような仮定ができる．

2.2 リダイレクトの悪性判定

DBD 攻撃の目的はマルウェアをクライアントにダウンロードさせることにある．従って，リダイレクトによってマルウェア (実行ファイル) がダウンロードされれば，そのリダイレクトを悪性と判定出来る．本研究で分析したデータは通常の通信が含まれないように制御しているので，前述の推定法で検出したリダイレクトによって途中あるいは最終的に実行ファイルをダウンロードした場合はそれらのリダイレクトを悪性と判定する．

3 分析結果

本研究では MWS2010 データセットに含まれる Web 感染型マルウェアの攻撃通信データである D3M2010 [2] を利用し，前節で提案したリダイレクト解析を適用する．D3M2010 は公開されている URL ブラックリスト [3] に登録されている URL の中から，別途 DBD 攻撃を検知した URL を抽出し，入口サイトとしてブラウザ上で実行している．データは 2010 年 3 月 8 日，9 日，11 日に収集した 3 日分である．D3M2010 の全データセットにおいて GET 要求に対する応答が返ってきた入口サイトの数は 523 であった．その通信の中から 264 の悪性リダイレクトが検出された．残り約半数の悪性と判断されなかったリダイレクトは，サーバーが解析対策を施したことによって実行ファイルが削除されていたり，URL やドメインが無効になっていたり，GET 要求を拒否したりするものであった．

本手法で抽出した悪性リダイレクトの特徴を分析した結果，表 1 を得た．個々の詳細は紙面の都合上割愛するが，これらの特徴は未知の悪性リダイレクトを検知する上で有効な指標になると考えられる．また，実行ファイルをダウンロードした URL をマルウェア配布 URL とし，その URL を抽出した推定方法の割合を表 2 に示す．

表 1　D3M2010 における各特徴を利用したリダイレクト数

	3 月 8 日	3 月 9 日	3 月 11 日
悪性リダイレクト総数	98	89	77
難読化 JS	98 (100%)	89 (100%)	77 (100%)
Referer 無し	89 (90.8%)	79 (88.8%)	75 (97.4%)
短時間のダウンロード	67 (68.4%)	69 (77.5%)	52 (67.5%)
PDF+JS	35 (35.7%)	33 (37.1%)	25 (32.5%)
特定の GET 変数	38 (38.8%)	30 (33.7%)	24 (31.2%)
Vary: User-Agent	29 (29.6%)	27 (30.3%)	25 (32.5%)
Server: nginx	29 (29.6%)	24 (27.0%)	21 (27.3%)
エフェメラルポート	8 (8.2%)	5 (5.6%)	10 (13.0%)
偽装ファイル	7 (7.1%)	6 (6.7%)	4 (5.2%)

表 2　マルウェア配布 URL を抽出した推定方法の割合

	3 月 8 日	3 月 9 日	3 月 11 日
URL 総数	202	205	158
Referer	13 (6.4%)	13 (6.3%)	6 (3.8%)
URL 推定	12 (5.9%)	10 (4.9%)	10 (6.3%)
Host 推定	177 (87.6%)	182 (88.8%)	142 (89.9%)

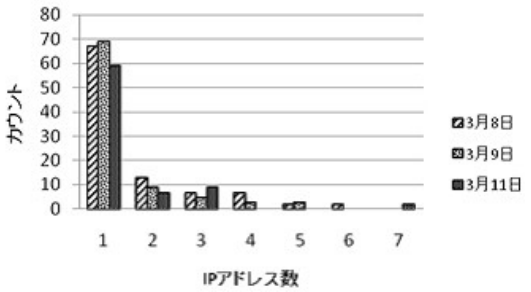

図 2　リダイレクト中の出現ユニーク IP アドレス数

90% 弱が Host 推定による抽出であることから，DBD 攻撃では難読化によって悪性 URL を隠し，Referer が残らないような遷移をしていることが分かる．最後にリダイレクト中の出現ユニーク IP アドレス数を集計した結果，図 2 を得た．リダイレクトは同一の IP アドレスに集中していることがわかる．悪性サイトはバーチャルドメインによって多数のドメインを少数の IP アドレスで運用するケースが多い．このような性質もリダイレクトを検出するために有効な尺度である．

4 まとめ

DBD 攻撃通信のみが存在する環境下で収集した通信データを分析し，リダイレクトを解析した．この結果，悪性リダイレクトに特徴的な性質を発見した．これらの特徴を利用することによって解析が困難なリダイレクトを発見することが期待出来る．得られた特徴によるリダイレクト検出の精度評価，実ネットワークで収集した攻撃通信データへの適用が今後の課題である．

参考文献

[1] D. Harley, P. M. Bureau, “Drive-by Downloads from the Trenches,” Proc. 3rd International Conference on MALWARE, October 2008.

[2] 畑田, 中津, 秋山, 三輪, “マルウェア対策のための研究用データセット～MWS 2010 Datasets～”, コンピュータセキュリティシンポジウム 2010, 2010 年 10 月.

[3] MDL, http://malwaredomainlist.com