Exploit Kit で作成された悪性コンテンツの類似性調査

今野 由也† 角田 裕†

†東北工業大学
982-8577 宮城県仙台市太白区香澄町 35-1
m142803@st.tohtech.ac.jp , tsuno@m.ieice.org

あらまし 近年，Webブラウザの脆弱性を悪用し，Webサイトにアクセスしたユーザを自動的にマルウェアに感染させるDrive-by-Download攻撃が脅威となっている．この攻撃で利用されるWebサイト群の構築にはExploit Kitと呼ばれるツールが利用される．このツールはWebコンテンツや攻撃コードをテンプレートの形で提供することで専門的知識なしにWebサイト群の構築を可能にする．従って，Exploit Kitがテンプレートによって作成したコンテンツは互いに類似性が存在すると考えられる．そこで，本研究ではD3Mを対象にDrive-by-Download攻撃で利用される悪性コンテンツの類似性を分析した．分析の結果，同一のExploit Kitによって作成されたとみられる類似したコンテンツが確認されたため報告する．

Analysis of Similarities among Malicious Contents Generated by Exploit Kit

Yuya Konno† Hiroshi Tsunoda†

†Tohoku Institute of Technology
35-1, Yagiyama Kasumi-cho, Taihaku-ku, Sendai, Miyagi, 982-8577, JAPAN
m142803@st.tohtech.ac.jp, tsuno@m.ieice.org

Abstract In recent years, much attention has been focused on the threat of drive-by-download attacks. In this attack, attackers produce malicious web sites and a victim who accesses the web sites are forced to download malware. To reproduce the malicious web sites easily, attackers use a malicious software called exploit kits. Since exploit kits provide templates of web contents and exploit codes, we presume that the similarity of contents is the important characteristics in order to find exploit kits. In order to validate this hypothesis, we analyze the similarities of malicious contents in D3M dataset. As a result we found some group of contents that have great similarity. By investigating those contents in detail, we show that those contents are maliciousness and generated by exploit kits.

1 はじめに

近年 Drive-by-Download (DbD)攻撃が猛威を振るっている[1]. この攻撃は Web ブラウザやそのプラグインの脆弱性を利用し，ユーザの意図に関わらずマルウェアをダウンロードさせる．脆弱性を利用されるソフトウェアは Java Runtime Environment , Adobe Flash

Player，Microsoft Internet Explorer など様々存在している．特に Flash Player には 2014 年から 2015 年 8 月 24 日までに 242 件もの脆弱性が発見されており[2]，攻撃者に狙われやすいソフトウェアのひとつとなっている．

DbD 攻撃を行う悪性サイト群を容易に構築するために Exploit Kit と呼ばれるツールが利用される．近年では，Angler Exploit Kit や Blackhole Exploit Kit など数多くの Exploit Kit の存在[3]が確認されており，これらは新しい脆弱性を利用したコードを次々に追加するなど日々更新されている[3]．例えば，2015 年 7 月に明らかになった Flash Player の脆弱性 (CVE-2015-5119) [4]を利用した攻撃コードは，脆弱性の発表当日に Angler Exploit Kit と Nuclear Exploit Kit に追加されていたことが知られている[5]．このように DbD 攻撃の対策においては，脆弱性対策のみならず Exploit Kit への対策も重要となっている．

Exploit Kit には Web コンテンツや攻撃コードなどのテンプレートが用意されており，それらを活用することで攻撃者は容易に悪性サイトを構築できる．この特徴から，同種の Exploit Kit から生成された悪性サイト群の URL や JavaScript の関数名などには共通した特徴的なパターンが存在すると指摘されている[6]．

本研究では，悪性サイトのコンテンツがテンプレートによって作られることに注目し，コンテンツ間の類似性に着目した．悪性コンテンツ間に類似性がみられれば，それは Exploit Kit によって生成された可能性が高く，Exploit Kit の検知の手掛かりになるはずである．そこで，DbD 攻撃の通信データである D3M データセット[7]に含まれるコンテンツの類似性を分析し，類似したコンテンツの内容を詳細に調査した．調査の結果，類似したコンテンツからは Exploit Kit によって作成されたとみられる特徴が確認された．

以下，第 2 章では関連研究について述べ，第 3 章ではコンテンツの類似性を求める手法ついて述べる．第 4 章ではコンテンツの類似性を調査した事例を述べ，第 5 章では調査結果の事例よりコンテンツの類似性を利用した手法について考察する．第 6 章では本研究で確認した事例についてまとめ，今後の課題について述べる．

2 関連研究

本章では DbD 攻撃の原理を説明し，Exploit Kit を利用するメリットについて触れる．さらに Exploit Kit によって構築された悪性サイトの検知に関する既存手法について述べる．

2.1 Drive-by-Download 攻撃と Exploit Kit の役割

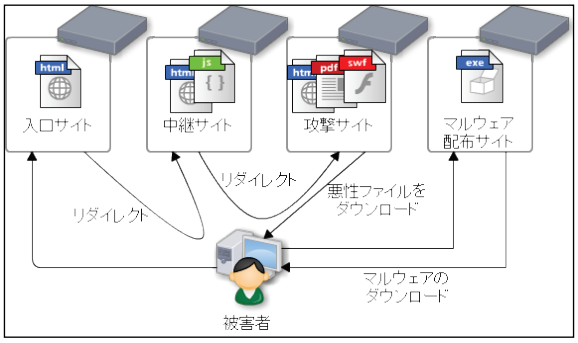

図 1 Malware Distribution Network の概要

DbD 攻撃は図 1 の MDN (Malware Distribution Network) [8]を構築し攻撃を行うことが実態調査[9]によって明らかになった．MDN とはそれぞれ役割の異なる複数のサイトが連携し攻撃を行うネットワークであり，通常は入口サイト，中継サイト，攻撃サイト，マルウェア配布サイトで構成されている．入口サイトはユーザのアクセスを中継サイトに誘導(リダイレクト)する役割を担う．アクセス数の多い Web サイトが改ざんされて入口サイトとなる場合がある．中継サイトはユーザが利用しているブラウザやそのプラグインの種別・バージョンなどから，利用可能な脆弱性を判断し，それに応じた攻撃サイトへアクセスをリダイレクトする．攻撃サイトは脆弱性を突く攻撃コードを含んだコンテンツをダウンロードさせる．そのコードがユーザをマルウェア配布サイトにアクセスさせマルウェアに感染させる．

MDN の構築には専門的な知識が必要であり，構築後の維持にかかるコストが高い．そこで，容易に MDN を構築するために Exploit Kit が利用される．Exploit Kit では，攻撃コードなどがテンプレートとして準備されており，攻撃者は専門的知識がなくとも MDN の構築が可能になる．また，構築した悪性サイトの URL がブラックリストに掲載された場合，攻撃者は攻撃の解析妨害を意図して対象のサイトを削除するため，悪性サイトは短命の傾向があると言われている[10]．この背景には新たな MDN を素早く構築できる Exploit Kit の存在があると考えられる．

以上のことから，DbD 攻撃の対策には，Exploit Kit やそれによって生成される悪性サイトの特徴を捉え検知につなげることが重要である．

2.2 Exploit Kit により作成された悪性サイトの検知手法

悪性サイトによる被害を防ぐため，危険性が高いサイトの URL などを収集しブラックリストとして共有する取り組みが行われている[11][12]．孫ら[13]は，Exploit Kit で作成された悪性コンテンツの URL が類似していることに着目し，ブラックリスト中の URL と類似した特徴を有する URL を，未知の悪性 URL として発見する手法を提案している．

また，笠間ら[6]は，使用される URL やそのクエリ引数，使用される JavaScript の関数名，難読化手法，リダイレクトの特徴に着目することを提案している．そして，同種の Exploit Kit が生成する悪性サイトの URL には共通した特徴的なパターンが存在することを示している．北野ら[14]は DbD 攻撃のメカニズムに依存した定性的な特徴（HTTP の status コード，User-Agent，MIME-Type など）を利用し悪性サイトを検知している．柴原ら[15]は Exploit Kit を利用し作成した MDN に対してブラウザやプラグイン，IP を変え繰り返しアクセスすることで得られた通信遷移の特徴をシグネチャとして自動的に作成する手法を提案している．

Exploit Kit による悪性サイトの既存検知手法では，URL や HTTP ヘッダなど MDN にアクセスする際に現れる特徴に着目しており，実際の攻撃コードの情報に関係なく検出する手法がほとんどである．

一方，本研究ではコンテンツ自体の類似性に着目し調査する．Exploit Kit は Web コンテンツや攻撃コードがテンプレートとして準備されており，同一の Exploit Kit で作られたコンテンツであればテンプレートによる影響を受け，コンテンツに共通の特徴が得られると考えられる．

次節ではコンテンツに含まれる共通の特徴を捉えるため文字の出現頻度に着目した関連研究を述べる．

2.3 文字の出現頻度に着目した検知手法

コンテンツ中の文字や文字列の出現頻度は，マルウェアの検知によく利用される特徴である[16][17]．

DbD 攻撃に関するデータに対して出現頻度に着目した手法として次の研究がある．西田ら[18]は，難読化 JavaScript の文字出現頻度が一般の JavaScript とは異なることに着目し，悪質な難読化 JavaScript の検知手法を提案している．Otsuki ら[19]はマルウェアの種類ごとの特徴量を評価し，特定の ASCII 文字コードの出現頻度に着目することでマルウェアを検知できる可能性を明らかにしている．prophiler[20]では悪性サイト検知の特徴量の 1 つとしてソフトウェアのセキュリティホールを突く shellcode の検知が行われている．不可視文字や 16 進数のみで構成される文字列，繰り返し出現する文字により悪性サイトの検知を実現している．

また，他の攻撃に関する研究として，メールを媒体とした標的型攻撃の検知に関する分野ではメールアドレスや件名，言語など様々な文字列の類似性に着目し標的型攻撃を自動分類する手法が提案されている[21]．また，辻ら[17]は拡散型ワームのフローに含まれる文字列が類似していることに着目し，フローに含まれる文字の出現頻度を算出しフロー間の類似性によ

って拡散型ワームを検出している.

本研究では, フロー間の類似性を分析した辻らの手法を参考にし, DbD 攻撃のコンテンツの類似性を分析する. 次章では類似性の調査手法について詳述する.

3 コンテンツの類似性調査

本章では本調査で利用した DbD 攻撃に関するデータについて述べ, 本研究で利用した類似性の評価手法を説明する.

3.1 調査対象データ

本調査で利用したコンテンツはマルウェア対策研究人材育成ワークショップ[7]提供の D3M を利用した. D3M は高対話型の Web クライアント型ハニーポットによって悪性 URL を巡回して採取した DbD 攻撃に関する通信データであり, 次のデータを含んでいる.

- 攻撃通信データ
- マルウェア
- マルウェア通信データ

本研究では D3M (2010～2015) の攻撃通信データを利用し, 攻撃通信データに含まれる HTML ファイルや JavaScript ファイルをコンテンツとして抽出した. これらを対象とした理由は, Exploit Kit によって必ず作成されるコンテンツであり, 難読化処理など Exploit Kit の特徴が出やすいと考えたためである. なお, jQuery ファイルなど URL から正常コンテンツと判断できるものについては調査対象外とした.

3.2 類似性の評価手法

図 2 にコンテンツ間の類似性を評価する手法の概要を示す. まず, 各コンテンツを 8bit ごとのコードに分割し, 各コードの出現頻度を表す 256 クラスのヒストグラムに変換する. そして, 任意の 2 つのコンテンツに対応する$\boldsymbol{h_i}$, $\boldsymbol{h_j}$間のユークリッド距離$D(\boldsymbol{h_i}, \boldsymbol{h_j})$を(1)式より求める.

$$D(\boldsymbol{h_i}, \boldsymbol{h_j}) = \sum_{k=0}^{255} \sqrt{\left(h_{i,k} - h_{j,k}\right)^2} \quad \cdots \quad (1)$$

ここで, $h_{i,k}$はヒストグラムh_iの第kクラスを指している. コンテンツ間の類似性が高ければそのヒストグラム間の距離は短くなるため, 一定以下の距離を持つコンテンツを類似していると判断する.

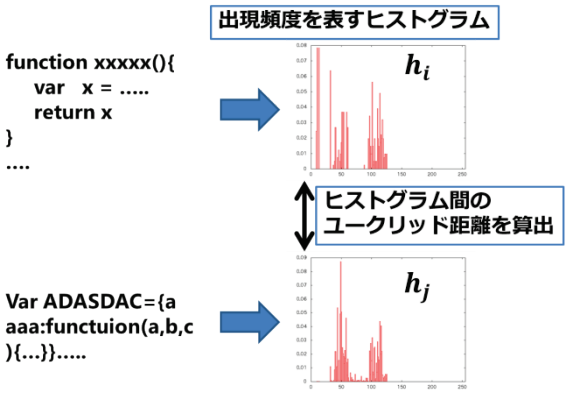

図 2 コンテンツ間の類似性を調査する手法

4 調査結果

本章では前章で述べた手法によりコンテンツの類似性を調査し, Exploit Kit によって作成されたとみられるコンテンツが類似していたか確認する.

事前調査として 1,024Byte のテキストファイルに含まれる A という 1 文字を a に置き換え, 置き換え前後のコンテンツから作成したヒストグラムの距離を求めた. このときの距離は 0.0312 であったため, それより短い 0.02 を類似性の判断する閾値として採用した.

D3M に含まれるコンテンツは全部で 1,087 個あり, 互いの距離が 0.02 以下のコンテンツのグループは 18 (124 個)存在した. この 18 グループからは Apache HTTP サーバのエラーページなど, 攻撃と関連性の低い一般的なコンテンツは除いてある. さらに DbD 攻撃との関連性が確認されたグループは 6 グループ存在していた. DbD 攻撃への関連性が確認された事例の一部を次節で説明する.

4.1 MDAC の脆弱性を利用した類似したコンテンツ

図 3 は MDAC (Microsoft Data Access Components) に関する脆弱性 CVE-2006-0003[22]を利用する攻撃を含んだコンテンツであり，これと類似した 7 つのコンテンツを発見した．表 1 にそのコンテンツの URL リストを示す．

```
<script>
var Vg='a06d04937ccdc754e9ebc1c93e37da1309ac.....
var HJN = '';
var q = Vg.slice ( 38, 14236 );
for ( K = 38 ; K < 14236 ; K += 2 )
{
        HJN += '%' + Vg.slice ( K, K + 2 );
}
document.write(unescape(HJN));
</script>
```

図 3 CVE-2006-0003 を含んだコンテンツ

表 1 MDAC の脆弱性を利用したコンテンツの URL

1	http://AAAAA .br/vxap.htm
2	http://BBBBB.com/sewfs.htm
3	http://CCCCC.de/page/
4	http://CCCCC.de/page/idie.htm
5	http://DDDDD.es/estimulates/qlufuh.htm
6	http://DDDDD.es/lubricantes/ynfjh.htm
7	http://EEEEE.pl/eqyb.htm

表 1 にあるすべてのコンテンツにおいて図 3 の CVE-2006-0003 の脆弱性を突く攻撃コードが含まれており，これらは同種の Exploit Kit によって作成された可能性が高い．この脆弱性を突く攻撃コードを提供している Exploit Kit の一部[23]を以下に示す．

- Blackhole Exploit Kit 1.2.5, 2.0
- Elenore 1.8.91
- Sakura1.1
- Crime Pack
- Mpack

[6][13][24]において Exploit Kit によって作成されたコンテンツの URL には共通したパターンが存在すると述べられている．しかし表 1 の URL を見るとホスト名はすべて異なり，特にファイル名はランダムな文字列で共通したパターンは見られない．今回の事例の場合共通したパターンが存在しないため既存の検知手法で捕捉できない可能性も考えられる．

4.2 ソフトウェアのバージョン情報を取得しリダイレクト先を判断するコンテンツ

表 2 に示した 2 つの URL はいずれもソフトウェアやそのバージョン情報によってリダイレクト先を選択している類似したコンテンツである．

表 2 ソフトウェアのバージョン情報を判別するコンテンツの URL

8	http://FFFFF.co.cc/bl
9	http://X.X.X.X/index.php

これらのコンテンツでは同じホスト上にある PluginDetect.js を読み込み，以下に列挙した脆弱性を突く攻撃が可能か判断している．

8, 9 の両方で使われている脆弱性

- CVE-2010-0188(Adobe Acrobat)
- CVE-2010-1297(Adobe Acrobat)
- CVE-2010-2884(Adobe Acrobat)
- CVE-2008-2992(Adobe Acrobat)
- CVE-2010-0842(Java)
- CVE-2010-3552(Java)
- JavaSignedApplet

8 のみで使われている脆弱性

- CVE-2006-0003(MDAC)
- CVE-2010-3183(Firefox)

図 4 は表 2 のコンテンツから起きたリダイレクトの流れを表した図である．今回実際にアクセスが行われた脆弱性を含むコンテンツは 4 つ存在し，8 の URL のみ MDAC の脆弱性を悪用した攻撃コードを含むコンテンツがダウンロードされている．

8 と 9 の URL を MALWARE DOMAIN LIST[11]上で検索したところ，8 の種別は不明だが何らかの Exploit Kit で作成されたコンテンツとされており，9 は BLEEDING LIFE

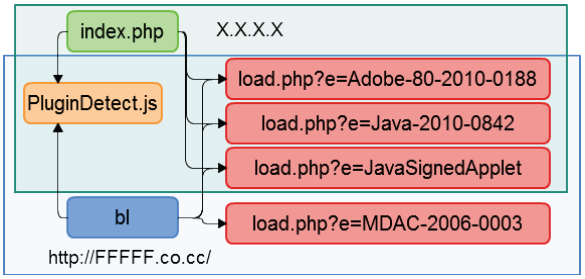

図 4 複数のソフトウェアのバージョン情報を取得しリダイレクトを判断するコンテンツ

Exploit Kit から作成されたコンテンツの可能性が指摘されている. 8 と 9 の URL のコンテンツは類似していることから, 8 の URL のコンテンツも BLEEDING LIFE Exploit Kit により作成された可能性が極めて高い.

よって, 今回のようにどの Exploit Kit から作成されたコンテンツか判断できない場合においてもコンテンツの類似性に着目することで Exploit Kit 特定の手掛かりを得られると考えられる.

5 考察

4.1 節においては Exploit Kit によって作成されたとみられる類似したコンテンツが確認されたが, URL の共通のパターンは確認されず URL の共通パターンに着目した検知手法では捕捉のできない可能性がある事例が明らかとなった. URL の共通パターンに着目した検知手法において URL の共通パターンが変化しにくい要因を以下のようにまとめている.

- Exploit Kit の作成者がプログラム保護を目的に暗号化を施す
- リダイレクト先のファイル名を変更する場合は難読化処理を施されたファイルを可読化する必要がある

攻撃者が暗号を解読した場合や可読化によりリダイレクト先のファイル名を変更した場合は検知が困難になることが予想される. 本調査で利用したコンテンツの類似性に着目することでそのような場合おいても対応が可能になると考えられる.

ただし, 今回利用した類似性の調査手法は 8bit コードの出現する順番を考慮していないため誤検知を引き起こすことも考えられる. また, 通常のネットワークを流れるコンテンツと D3M のコンテンツ間の類似度は未評価であるため今後検討を行う必要がある.

4.2 節で確認された類似したコンテンツについては攻撃対象の脆弱性を突く攻撃へのリダイレクト処理の数が異なっていた. よって, Exploit Kit に脆弱性を突く攻撃コードが追加された程度のアップデートであればコンテンツの類似性により Exploit Kit 特有の特徴を捉え, Exploit Kit から作成されたコンテンツを検出できる可能性がある.

6 まとめ

本稿では DbD 攻撃の通信データに含まれるコンテンツの類似性を調査し, Exploit Kit により作成されたと思われるコンテンツが確認された. また, 今回確認された類似したコンテンツの中には同じ Exploit Kit から作成されたとみられるコンテンツ同士に URL の共通したパターンを持たない事例も発見した. 既存の手法では捕捉のできない事例の可能性があり, 今後はどれほどのコンテンツでそのような事例が確認されるか追加調査を行う予定である.

参考文献

[1] IIJ , Internet Infrastructure Review(IIR) Vo.25, http://www.iij.ad.jp/company/development/report/iir/025/01_03.html

[2] CVE Detailes , http://www.cvedetails.com/product/6761/Adobe-Flash-Player.html?vendor_id=53

[3] contagio: An Overview of Exploit Packs (Update 25) May 2015 , http://contagiodump.blogspot.jp/2010/06/overview-of-exploit-packs-update.html

[4] CVE , CVE-2015-5119 , https://cve.mitre.org/cgi-bin/cvename.cgi?name=C

VE-2015-5119
[5] TREND MICRO, "Hacking Team Flash Zero-Day Integrated Into Exploit Kits", http://blog.trendmicro.com/trendlabs-security-intelligence/hacking-team-flash-zero-day-integrated-into-exploit-kits/
[6] 笠間貴弘，神薗雅紀，井上大介，”Exploit Kit の特徴を用いた悪性サイト検知手法の提案”，CSS2013
[7] 神薗雅紀，秋山満昭，笠間貴弘，村上純一，畑田充弘，寺田真敏，”マルウェア対策のための研究用データセット~MWS Datasets 2015~”，情報処理学会研究報告，Vol.2015-CSEC-70，No.6，pp.1-8，2015
[8] 八木毅，秋山満昭，村山純一，”コンピュータネットワークセキュリティ”，コロナ社
[9] N. Provos, P. Mavrommatis, M . A. Rajab, and F. Monrose, “All Your iFrames Points to Us,” Proc. USENIX Conference on Security Symposium, Feb. 2008
[10] L . Bilege , E . Kirda , C . Kruegel and M . Balduzzi , EXPOSE: Finding Malicious Domains Using Passive DNS Analysis , Journal ACM Transaction on Information System Security , Vol.16 , Issue 4 , No.14 、2014
[11] MALWARE DOMAIN LIST , http://www.malwaredomainlist.com/
[12] hpHosts，http://www.hosts-file.net/
[13] 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”，CSS2014
[14] 北野美紗，大谷尚通，宮本久仁夫，“Drive-by-Download 攻撃における通信の定性的特徴とその遷移を捉えた検知方式”，CSS2013
[15] 柴原健一，笠間貴弘，神薗雅紀，吉岡克成，松本勉，“Exploit Kit 検知用シグネチャの動的解析に基づく自動作成”，情報処理学会研究報告，Vol.2015-CSEC-64，No.35，pp.1-7，2015
[16] 岩本舞，小島俊輔，中島卓雄，“マハラノビス距離を用いた静的解析によるマルウェアの検出”，情報処理学会研究報告，Vol.2015-CSEC-68，No.49 ，pp.1-7，2015
[17] 辻雅史，和泉勇治，角田裕，根元義章，フローペイロードの類似性に基づく拡散型ワーム検出に関する一検討，信学技報，NS，Vol.105，No.405，pp.9-12，Nov，2005
[18] 西田雅太，星澤裕二，笠間貴弘，後藤将史，井上大介，中尾康二，“文字列出現頻度をパラメータとした機械学習による悪質な難読化 JavaScript の検出”，情報処理学会，Vol.2014-CSEC-64，No.21，pp.1-7
[19] Y. Otsuki, M. Ichino, S.Kimura, M. Hatada, H . Yosiura, “Evaluating payload features for malware infection detection,” Journal of information processing, Vol.55, No.2, 2014
[20] D. Canali, M. Cova, G. Vigna, and C. Kruegel, “Prophiler: A fast filter for the large-scale detection of malicious web pages,” in Proc. WWW, pp. 197-206, ACM, 2011
[21] 北條考佳，松浦幹太，“文字列類似性を考慮した標的型攻撃のグループ化手法”，CSS2014
[22] CVE-2006-0003，http://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2006-0003
[23] SERT Report Exploit Kits – v1.0，https://www.solutionary.com/_assets/pdf/sert-exploit-kit-overview-1174sr.pdf
[24] F. Howard , ”Exploring the Blackhole exploit kit,” Sophos Technical Paper、March，2012