MWS Cup 2013

事前課題１　「Drive-by Download 攻撃解析」　解答例

１．出題の意図

MWS[1]で研究用データセット[2]として提供している D3M (Drive-by Download Data by Marionette)[3]には、ドライブバイダウンロード攻撃を行う悪性通信や、その際に感染するマルウェアおよびマルウェアが行う通信に関するデータセットが含まれている。本事前課題では、実際に Web 空間から上記マルウェア感染に関連する悪性な Web サイトを発見させるものであり。よって、単に攻撃検知精度だけでなく、どのように Web 空間から効率的に発見するかを問うものであり、総合的な解決策が求められる。データセットの分析や本事前課題を通じて、攻撃手法の正確な理解と、効果的かつ実用的な対策の研究が促進されることを期待する。

２．出題内容

Drive-by download 攻撃を行う悪性な Web サイトについて、設問に答えよ。

[設問 1]　近年、正規サイトが改ざんされることにより drive-by download 攻撃を行う悪性サイトの踏み台となる事例が継続的に発生している。このような改ざんされたサイトを Web 空間から効率的に発見するためにはどのようなクローリング方法、観測方法を行うのが良いか 500 字以内で述べよ。

[設問 2]　Web 空間の中から改ざんされて drive-by download 攻撃の踏み台にされた Web サイトを発見し、検知の根拠および攻撃に関連する URL 情報を答えよ。可能であれば設問 1 で述べた方法を用いること。なお、発見できなかった場合であっても、具体的な手法や試行錯誤の過程を説明することで加点対象となる。

３．解答例

MWS Cup2013 参加チームから提出された事前課題レポートより、解答内容を抜粋・要約する。【】内は参加チームを名を表す。

[設問 1]

【Alkaneters】

サイトに挿入される攻撃コードに含まれる特徴的な文字列を検索エンジンで検索することで、Drive-by Download 攻撃の入口サイトを発見する。発見した入口サイトを仮想マシン上の Windows で閲覧することで動的解析を行う。

【人海戦術チーム】

クローリング手法は、(ア)被害を受けやすいサイト（著名サイト、特定の CMS）を収集、(イ)改ざんされたサイトの情報（コメントや関数名など特徴的な文字列）を検索エンジン等でリスト化して収集するという方法が挙げられる。踏み台判定手法は、(I)クローリングで蓄積したコンテンツを解析する手法と(II)通信内容を解析する手法が挙げられる。コンテンツ解析は、HTML を各パーツに分解して検索可能なデータベース化を行い、ブラックリスト等と照合する。通信解析は、シグネチャを定義しパターンマッチングする。

【SecCap の愉快な仲間たち】

クローリング方法として、サーバの脆弱性を考慮し、取得した情報の中に存在する URL を辿っていくことで悪性サイトの発見を行う。サーバの脆弱性を知ることで改ざんの可能性を数値化でき、数値が高いほど改ざんされている可能性が高いとする。数値化には OS や Web サーバのバージョン、脆弱性診断ツールなどを用いた結果を利用する。

【WeightAnkle teamby Res9】

①攻撃者がよく利用するリダイレクト手法である HTML 内の<script>または<iframe> タグの埋め込みを検索エンジンにより検出し、改ざんサイトの候補とする。②ISP におけるバックボーンでのセキュリティ対策でマスユーザからアクセスされる悪性 URL への HTTP Referer ヘッダの URL を改ざんサイトの候補とする。上記①②の対象 URL へクローリングし、リダイレクトの観測とリダイレクト先 URL の URL ブラックリストとの突き合わせを持って改ざんサイトと判定する。

【Team Enu】

(1) ブラックリストの活用：ブラックリストに掲載されている既知の悪性サイト URL や URL に含まれるドメインやパス名、ファイル名を検索することで、悪性サイトにリンクしている怪しいサイトを発見することができる。(2) 脆弱性情報の活用：Apache、CMS 等といったサービスの脆弱性を含んだ古いバージョンを用いているサイトは改ざんされている可能性が高いと考えた。(3) 攻撃コード情報の活用：Exploit kit や PoC といった具体的な攻撃コード内の特徴的な文字列を検索する方法である。

【セキュリティ讃歌】
検索エンジンを使用したクローリングを行う．検索エンジンに指定する検索ワードは，改ざんされたサイトに挿入される可能性の高い不正なスクリプトの特徴を検索ワードとする．不正スクリプトの特徴は，D3M 2013 や Exploit Kit を使用した際に付与される文字列から取得する．

【GOTO Love and 初代森研】
検索エンジンを利用して収集した脆弱性のありそうなサイトへのクローリングが効果的である。Google Dorks のような脆弱性に関連するクエリを用いて検索する。収集した脆弱性のありそうなサイト、すなわち攻撃を受ける可能性が高いサイトの URL に対してクローリングを行う。

【KIT-CUT】
カラーコード攻撃などに見られる特定の文字列を一定間隔で Google 等の検索サイトで検索を行い、結果を収集する。

【TDUISL in 親方】
リダイレクト先の URL をブラックリストと照合し、一致したらリダイレクト元が改ざんされたサイトとする。

【頑張ります】
クローリングを行う際の起点となる URL は、リンク先に別ドメインの URL を含んでおり、かつ、それが特定のトップレベルドメイン（com, info, vn, cm 等）を含んでいる場合、優先的にアクセスを行うこととする。

【NU14】
(1) CMS (Wordpress, MobileType, Apache Struts 等) サイト：手軽さから様々な場面で利用される一方で管理者のセキュリティ知識が伴わない場面が多く、脆弱性の放置によって攻撃者に改竄されうる。(2)レンタル VPS サーバや個人向け ISP の IP アドレス帯：安価にそして手軽に設置できることから、セキュリティが脆弱な Web サーバが運用されているのではないかと考えた。(3) SNS サイトの URL：SNS では本人認証の曖昧のままユーザ同士が結びつく傾向があり、悪意のあるユーザも入り込みやすい。これらの URL を順に収集し観測する。

【Olab 2013】
クローリング等において、対象のページに JavaScript があれば SpiderMonkey 等の JavaScript エンジンで実行させる。動的解析ツールがヒープ汚染を検知したら、実行中の JavaScript を含んでいるページは Drive-by Download を行うページである可能性が高いものとする。

[設問 2]
発見できたチームの回答のみ掲載する。なお、解答で記述されているドメイン名の一部は伏字にする。

【人海戦術チーム】
(1) 踏み台として改ざんされたサイトの特徴的な文字列として、HTML コメントで「81a338」という文字列が埋め込まれるという情報を入手。
(2) 検索エンジンにて(1)の文字列を含むサイトをリストアップ。
(3) 「http://www.AAAAAA.com/peye042.html 」というサイトを発見。
(4) JavaScript を抽出して、解析したところ、iframe タグで「http://BBBBBB.com/exit.php」にアクセスすることが判明した。
(5) また、(3)の URL をブラックリスト提供サイトで照合したところ、Sucuri.net で既知の JavaScript マルウェアであると判定される。
(6) 動的解析環境で、(3)の URL にアクセスしたところ、最終的に広告サイトに遷移した。これは、解析環境が攻撃条件に該当せず、攻撃行為は行われず、結果的にマルウェアのダウンロードが行われてなかったと判断した。

URL（アクセス順）	URL 種別
http://www.AAAAAA.com/peye042.html	踏み台の URL
http://BBBBBB.com/exit.php	攻撃を行う URL
http://CCCCCC.com/plastic.html	リダイレクト
http://DDDDDD.ru/	広告サイト（攻撃条件に該当せず、マルウェアのダウンロードは行われてない）

【WeightAnkle teamby Res9】
以下で解答した「攻撃を行う URL」は、URL ブラックリストと突合させた結果、検知された既知の攻撃を行う URL です。URL パターンが BlackHole v2.0 exploit kit の URL パターンと類似しており、ダイナミック DNS サービスにより、IP アドレスが日々変化しているサーバです。「こちらの攻撃を行い URL」へのリダイレクトが複数の Web サイトより同日一斉に検知されておりまして、以下の踏み台 URL はその一部です。
おそらく攻撃ツールを用いた Web 改ざんにより攻撃サイトへリダイレクトされたと推測で

き、第三者機関の URL レピュテーションを考慮しても、攻撃サイトの検知結果は正しいと考えられ、その関連性が見られない攻撃サイトへリダイレクトを行っている以下の踏み台 URL も正しいと考えます。

URL（アクセス順）	URL 種別
hxxp://www.EEEEEE.com/	踏み台の URL
hxxp://delivery.FFFFFF.com/ 7f01baa99716452bda5bba0572c58be9/afr-zone.php	攻撃を行う URL

【Team enu】

まず、入口の URL にアクセスすると、攻撃を行う URL へ転送される。攻撃を行う URL にアクセスすると、難読化された JavaScript がダウンロードされ、JavaScript を分析すると .jar ファイルをダウンロードさせようとする記述が見つかった。

一方で、wireshark のログには、IIIIII.ru というドメインに対する名前解決を行おうとする挙動が見られたが、レコードが削除されたためかドメインの名前解決ができなかったので、アクセスすることができずマルウェアをダウンロードするには至らなかった。

また、入口の URL は virustotal や Symantec Endpoint Protection のシグネチャにも検知されたため、悪性であると判断できる。

URL（アクセス順）	URL 種別
http://old.GGGGGG.bg/wp-enter.php	入口（踏み台）の URL
http://HHHHHH.com/closest/i9jfuhioejskveohnuojfir.php	攻撃を行う URL
http://IIIIII.ru/	マルウェアをダウンロードする URL？

【GOTO Love and 初代森研】

難読化された javascript が body タグの閉じる直前に挿入されており、この javascript により、別の javascript が読み込まれるようになっている。おそらく、読み込まれるはずだった javascript が、実際に攻撃を行う javascript を含んだ URL であるか、もしくは、次の攻撃の踏み台にされるはずの URL であったと考えられる。

URL（アクセス順）	URL 種別
http://JJJJJJ.net/archive/index.php/t-3759.html	踏み台の URL
http://KKKKKK.ru:8080/google.com/technorati.com/iciba.com.php	攻撃を行う URL

【KIT-CUT】

昨今急増しているカラーコード攻撃に用いられる値を利用する。今回は#d68107#を検索した。ページ下部に挿入されている、不自然な数値列の列挙がある、スクリプト内に fromCharCode が存在し、eval で実行されていることから、ドライブ・バイ・ダウンロー

ド攻撃の特徴と一致する。

URL（アクセス順）	URL 種別
http://LLLLLL.info/	踏み台の URL
http://MMMMMM.de/typo3/YN7MxKpZ.php	中継を行う URL(ただし 404)
http://www.NNNNNN.net/	踏み台の URL
http://www.OOOOOO.com/M2BbxNLq.php	中継を行う URL(ただし 302)

４． D3M (Drive-by Download Data by Marionette)

D3M はドライブバイダウンロード攻撃に関するデータセットであり、NTT セキュアプラットフォーム研究所が開発した Web クライアントハニーポット（Marionette）[4][5]によって収集されている。D3M には攻撃通信データ、マルウェア検体、およびマルウェア検体を動的解析した際の通信データが含まれている。なお、D3M は 2010 年から 2013 年の 4 年間にわたって提供されてきたため、ドライブバイダウンロード攻撃の時系列的な変化を分析する事も可能である。

攻撃通信データには各種悪性サイトに対する通信が含まれており、マルウェア検体は web 経由における特徴的な検体が含まれている。また、マルウェア検体は取得してから 24 時間以内にマルウェア動的解析器である BotnetWatcher[6]により解析されているため、実際に C&C サーバとの通信や二次検体のダウンロードなどの通信が含まれる。

５． 総評

本事前課題は、攻撃検知だけでなく巡回対象の選定と Web 巡回など、複数の技術的課題を総合的に解決するための思考力と実際のシステム構築が求められる課題であった。半数以上のチームが発見までいたらなかったが、しかしながら独自の Web クローラシステムを作成したチームや検知の効率化などで創意工夫がみられたチームがあった。脆弱性を内包している事が多い Web サイトを中心とした探索や、攻撃に利用される特徴的な文字列が検索エンジンで検索可能であることを利用する探索により、効率的に踏み台となる改ざんサイトを発見するアイデアが多くみられた。実際に Web 空間に存在する改ざんサイトを発見し、提案手法の実現可能性と有効性を示していたチームがあったことから、具体的な対策に資する技術である可能性がある。この事前課題を通じた取り組みが研究として深化し、MWS 等での発表へ、延いてはマルウェア対策技術の進展に繋がる事を期待する。

参考文献

[1] マルウェア対策研究人材育成ワークショップ(MWS: anti-Malware engineering WorkShop), http://www.iwsec.org/mws/2013/

[2] 神薗，他，マルウェア対策のための研究用データセット ～ MWS Datasets 2013 ～, http://www.iwsec.org/mws/2013/manuscript/1A1-1.pdf, MWS2013

[3] 秋山, MWS2013 意見交換会 D3M (Drive-by Download Data by Marionette) 2013 http://www.iwsec.org/mws/2013/files/D3M_Dataset_2013.pdf

[4] M. Akiyama, et al., Design and Implementation of High Interaction Client Honeypot for Drive-by-download Attacks, IEICE Transactions on Communication, Vol.E93-B,No.05,pp.1131-1139,May. 2010.

[5] M. Akiyama, et al., Scalable and Performance-Efficient Client Honeypot on High Interaction System, IEEE/IPSJ SAINT2012

[6] K. Aoki, et al., Controlling Malware HTTP Communications in Dynamic Analysis System using Search Engine, IEEE CSS2011