

IIJR

Internet
Infrastructure
Review

Dec.2016

Vol. 33

インフラストラクチャセキュリティ

Mirai Botnetの検知と対策

コンテンツ配信

業界が挑む ストリーミング形式の統合

技術トレンド

言語リソースとしてのWikipedia

IIJ

Internet Initiative Japan

Internet Infrastructure Review

December 2016 Vol.33

エグゼクティブサマリ	3
1. インフラストラクチャセキュリティ	4
1.1 はじめに	4
1.2 インシデントサマリ	4
1.3 インシデントサーベイ	11
1.3.1 DDoS攻撃	11
1.3.2 マルウェアの活動	12
1.3.3 SQLインジェクション攻撃	16
1.3.4 Webサイト改ざん	16
1.4 フォーカスリサーチ	18
1.4.1 Mirai Botnetの検知と対策	18
1.4.2 SSL/TLSにまつわるエトセトラ	24
1.5 おわりに	29
2. コンテンツ配信	30
3. 技術トレンド	34
3.1 はじめに	34
3.2 「Unix考古学」の執筆	34
3.3 Drilldown Search	34
3.4 固有表現認識(NER)	35
3.5 Natural Language Toolkit(NLTK)で利用できるNER	36
3.6 統計的自然言語処理と"コーパス"	36
3.7 WikipediaからNEコーパスを生成する	38
3.8 おわりに	39

エグゼクティブサマリ

他国のこととは言え、米国大統領選挙の結果がトランプ氏に決まったことで世界中が大騒ぎになり、金融、経済、外交、と騒がしくなっているのは、その結果が良かったのか悪かったのか評価に困るところです。次は、安倍首相とロシアのプーチン大統領との会談が控えており、益々国際政治からも目が離せなくなりました。こんな最中であっても、インターネットのトラフィックはどんどん大きくなっています。米国大統領選挙についてもネットによる中継が地球規模で視聴され、日本の真夜中にトラフィックが増えるなど、世界の動きを支える重要な社会インフラとなっていることを改めて実感しています。

本レポートは、このような状況の中で、サービスプロバイダとしてのIIJが、インターネットやクラウドの基盤を支え、お客様に安心・安全に利用し続けていただくために継続的に取り組んでいる様々な調査・解析の結果や、技術開発の成果、ならびに、重要な技術情報を定期的にとりまとめ、ご提供するものです。

1章では、7月から9月までのインシデントを中心に解説しています。毎年、歴史的特異日がある8月から9月は警戒を強めていますが、ここ数年は平穏に過ぎるようになってきています。一部宗教関連での攻撃もありましたが、歴史的背景や国際情勢にも気を配りながら、この3カ月を振り返りました。また、マルウェアとしては、Mirai Botnetを取り上げました。いわゆるIoT機器は、単体では大きな処理能力は持っていませんが、セキュリティ管理が適切に行われないと感染しやすく、大量の機器が巨大なネットワークを形成して攻撃に加わることが可能であることから、今後のインターネットインフラには注意が必要であると考えて取り上げています。

2章では、インターネット配信業界が挑むストリーミング形式の統合について、現状方式での解決策を解説しています。ライブストリーミングは実時間より若干の遅延が発生します。ライブイベントをリアルタイムに近い形で再生させたい業界関係者は多く、その改善に向けた動きが加速されることを期待しています。また、携帯回線の通信容量制限に絡む問題から利用が進んでいないオフライン再生についても、Wi-Fi接続時に動画をモバイル機器にダウンロードさせ、オフラインでも動画を視聴できるようにする技術など、標準化が期待されています。今後益々、発展が期待されるストリーミングビジネスの細部を深く掘り下げました。

3章では、Wikipediaを“科学”してみました。分かりにくいかもしれませんが、Wikipediaの研究を手がけている研究者が、歴史的な事実のDrilldownの機械化を日々考えていて、自然言語研究の分野にまで踏み込んだ結果、Wikipedia分析のために作った基盤技術が応用できそうであるという結論に至った過程を紹介しています。ある意味でビッグデータの解析やDeep Learningにも似た観点で、Wikipediaに蓄積されている情報の中から歴史的な事実を発掘して行く、1つの手段になるかもしれません。

IIJでは、このような活動を通じて、インターネットの安定性を維持しながらも、日々改善し発展させていく努力を続けております。今後も、お客様の企業活動のインフラとして最大限に活用していただけるよう、様々なサービス及びソリューションを提供し続けて参ります。



山井 美和 (やまい よしかず)

IIJ 常務執行役員 サービス基盤本部長。

1999年6月IIJに入社と同時に株式会社クロスウェイコミュニケーションズへ出向し、WDM・SONET網構築、広域LANサービスの企画、データセンター建設に従事し、2004年6月に帰任。帰任後は、IIJのサービス運用部門を担当。2016年4月からはインフラ運用部門を加え、IIJの法人ITサービス全般の運用を統括。同時にIIJのデータセンター事業を統括し、国内初の外気冷却を用いたコンテナ型の「松江データセンターパーク」の立ち上げを主導している。

Mirai Botnetの検知と対策

1.1 はじめに

このレポートは、インターネットの安定運用のためにIJ自身が取得した一般情報、インシデントの観測情報、サービスに関連する情報、協力関係にある企業や団体から得た情報を元に、IJが対応したインシデントについてまとめたものです。今回のレポートで対象とする2016年7月から9月までの期間では、依然としてAnonymousなどのHacktivismによる攻撃が複数発生しており、DDoS攻撃や不正アクセスによる情報漏えい、Webサイト改ざんなどの攻撃が多発しています。またIoT機器に感染するマルウェアから構成されたボットネットによる過去最大規模のDDoS攻撃も発生しています。このように、インターネットでは依然として多くのインシデントが発生する状況が続いています。

1.2 インシデントサマリ

ここでは、2016年7月から9月までの期間にIJが取り扱ったインシデントと、その対応を示します。まず、この期間に取り扱ったインシデントの分布を図-1に示します*1。

■ Anonymousなどの活動

この期間においても、Anonymousに代表されるHacktivistによる攻撃活動は継続しています。様々な事件や主張に応じて、多数の国の企業や政府関連サイトに対するDDoS攻撃や情報漏えい事件が発生しました。

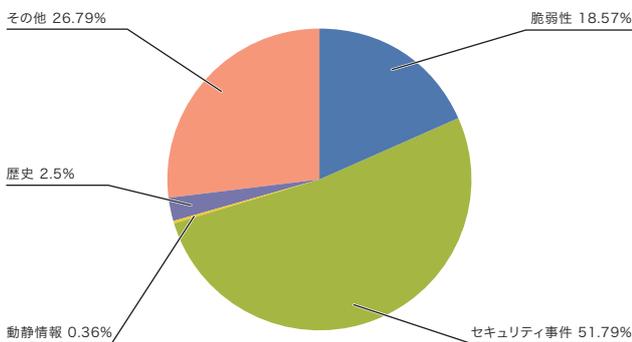


図-1 カテゴリ別比率(2016年7月~9月)

日本で行われているイルカや小型クジラの追い込み漁への抗議活動として、2013年からAnonymousによると考えられるDDoS攻撃が断続的に行われていますが、今年も9月1日の漁解禁日に合わせて攻撃キャンペーンの継続が宣言されました(OpKillingBay/OpWhales/OpSeaWorld)。攻撃対象は一部で異なる部分があるものの、概ね昨年までの対象をそのまま踏襲しています。9月に入りこれらのターゲットを狙ったDoS攻撃が活発に行われましたが、これまでと同様に攻撃対象のリストに記載がないWebサイトへの攻撃も数多く発生しました。この結果、9月の1ヵ月間に国内の様々なサイトに対して30件を越えるDoS攻撃が観測されました。10月以降も攻撃活動は衰えておらず、攻撃キャンペーンへの参加者も増加していると思われることから、引き続き警戒が必要な状況です。

8月下旬から9月上旬にかけて、国内の複数のサイトにおいて同時多発的なDoS攻撃が発生し、多くのサイトでサービスへの接続障害などの影響が出ました。これらのサイトが狙われた理由や攻撃者の目的については不明な点が多くはっきりしたことは分かりませんが、恒心教に関連する掲示板サイトなどが複数含まれており、攻撃者の目的と何らかの関連があるのではないかと考えられます。

9月18日は満州事変の発端となった柳条湖事件が起こった日であることから、この前後の期間において日中間でサイバー攻撃が発生しやすい「歴史的特異日」としてよく知られています。特に2010年には尖閣諸島付近で発生した中国漁船衝突事件を契機としたデモ活動、2012年には活動家による尖閣諸島上陸事件及び日本政府による尖閣諸島の国有化をきっかけとしたデモ活動がそれぞれ大規模に発生し、同時期に日本国内の多数のサイトに対しても様々なサイバー攻撃が行われました。2013年以降はこの時期の攻撃活動は年々沈静化する傾向にありましたが、今年も特に目立った攻撃活動は観測されませんでした。このように歴史的背景を伴った現実世界での出来事と連動してサイ

*1 このレポートでは取り扱ったインシデントを、脆弱性、動静情報、歴史、セキュリティ事件、その他の5種類に分類している。

脆弱性: インターネットや利用者の環境でよく利用されているネットワーク機器やサーバ機器、ソフトウェアなどの脆弱性への対応を示す。
動静情報: 要人による国際会議や、国際紛争に起因する攻撃など、国内外の情勢や国際的なイベントに関連するインシデントへの対応を示す。
歴史: 歴史上の記念日などで、過去に史実に関連して攻撃が発生した日における注意・警戒、インシデントの検知、対策などの作業を示す。
セキュリティ事件: フォームなどのマルウェアの活性化や、特定サイトへのDDoS攻撃など、突発的に発生したインシデントとその対応を示す。
その他: イベントによるトラフィック集中など、直接セキュリティに関わるものではないインシデントや、セキュリティ関係情報を示す。

バー攻撃が発生することは多いため、歴史的特異日や国際情勢などの動静情報にも注意を払う必要があります。

■ 脆弱性とその対応

この期間中では、Microsoft社のWindows^{*2*3*4*5*6*7}、Internet Explorer^{*8*9*10}、Edge^{*11*12*13}、Office^{*14*15*16}などで多数の修正が行われました。Adobe社のAdobe Flash Player、Adobe Acrobat及びReaderでも修正が行われています。Oracle社のJava SEでも四半期ごとに行われている更新が提供され、多くの脆弱性が修正されました。これらの脆弱性のいくつかは修正が行われる前に悪用が確認されています。

サーバアプリケーションでは、データベースサーバとして利用されているOracleを含むOracle社の複数の製品で四半期ごとに行われている更新が提供され、多くの脆弱性が修正されました。同じくOracle社のデータベースサーバであるMySQLでは、SQLインジェクション攻撃によって設定ファイルを書き換えることにより、リモートから権限昇格や任意のコード実行

が可能となる脆弱性が見つかり、Linuxの各ディストリビューションなどで修正の対応が行われました。DNSサーバのBIND9でも、lightweight resolverの問い合わせの処理の不具合や、DNS応答を作成する処理の不具合によって、外部からDoS攻撃が可能となる脆弱性などが見つかり、修正されています。SSL/TLSの実装においても、3DESなどの64ビットブロック暗号において同じ鍵で暗号化されたデータを大量に傍受することで暗号文の衝突を見つけ出し平文回復を行う攻撃手法(SWEET32)が公開されました。このためOpenSSLなどの実装において、デフォルト設定での3DESの使用を制限するなどの対応が行われました。

またLinuxカーネルのTCP実装における脆弱性によって、TCPの通信内容を盗聴できないOff-Pathの攻撃者がTCPのセッションをハイジャックできるサイドチャネル攻撃が可能であることを研究者らが示し、最新のLinuxカーネルにおいてこの脆弱性が修正されました^{*17}。

- *2 「マイクロソフト セキュリティ情報 MS16-086 - 緊急 JScript および VBScript 用の累積的なセキュリティ更新プログラム (3169996)」(<https://technet.microsoft.com/ja-jp/library/security/MS16-086>)。
- *3 「マイクロソフト セキュリティ情報 MS16-087 - 緊急 Windows 印刷スプーラー コンポーネント用のセキュリティ更新プログラム (3170005)」(<https://technet.microsoft.com/ja-jp/library/security/MS16-087>)。
- *4 「マイクロソフト セキュリティ情報 MS16-097 - 緊急 Microsoft Graphics コンポーネント用のセキュリティ更新プログラム (3177393)」(<https://technet.microsoft.com/library/security/MS16-097>)。
- *5 「マイクロソフト セキュリティ情報 MS16-102 - 緊急 Microsoft Windows PDF ライブラリ用のセキュリティ更新プログラム (3182248)」(<https://technet.microsoft.com/library/security/MS16-102>)。
- *6 「マイクロソフト セキュリティ情報 MS16-106 - 緊急 Microsoft Graphics コンポーネント用のセキュリティ更新プログラム (3185848)」(<https://technet.microsoft.com/library/security/MS16-106>)。
- *7 「マイクロソフト セキュリティ情報 MS16-116 - 緊急 VBScript スクリプト エンジン用の OLE オートメーションのセキュリティ更新プログラム (3188724)」(<https://technet.microsoft.com/library/security/MS16-116>)。
- *8 「マイクロソフト セキュリティ情報 MS16-084 - 緊急 Internet Explorer 用の累積的なセキュリティ更新プログラム (3169991)」(<https://technet.microsoft.com/ja-jp/library/security/MS16-084>)。
- *9 「マイクロソフト セキュリティ情報 MS16-095 - 緊急 Internet Explorer 用の累積的なセキュリティ更新プログラム (3177356)」(<https://technet.microsoft.com/ja-jp/library/security/ms16-095>)。
- *10 「マイクロソフト セキュリティ情報 MS16-104 - 緊急 Internet Explorer 用の累積的なセキュリティ更新プログラム (3183038)」(<https://technet.microsoft.com/library/security/MS16-104>)。
- *11 「マイクロソフト セキュリティ情報 MS16-085 - 緊急 Microsoft Edge 用の累積的なセキュリティ更新プログラム (3169999)」(<https://technet.microsoft.com/ja-jp/library/security/MS16-085>)。
- *12 「マイクロソフト セキュリティ情報 MS16-096 - 緊急 Microsoft Edge 用の累積的なセキュリティ更新プログラム (3177358)」(<https://technet.microsoft.com/library/security/MS16-096>)。
- *13 「マイクロソフト セキュリティ情報 MS16-105 - 緊急 Microsoft Edge 用の累積的なセキュリティ更新プログラム (3183043)」(<https://technet.microsoft.com/library/security/MS16-105>)。
- *14 「マイクロソフト セキュリティ情報 MS16-088 - 緊急 Microsoft Office 用のセキュリティ更新プログラム (3170008)」(<https://technet.microsoft.com/ja-jp/library/security/MS16-088>)。
- *15 「マイクロソフト セキュリティ情報 MS16-099 - 緊急 Microsoft Office 用のセキュリティ更新プログラム (3177451)」(<https://technet.microsoft.com/library/security/MS16-099>)。
- *16 「マイクロソフト セキュリティ情報 MS16-107 - 緊急 Microsoft Office 用のセキュリティ更新プログラム (3185852)」(<https://technet.microsoft.com/library/security/MS16-107>)。
- *17 25th USENIX Security Symposium, "Off-Path TCP Exploits: Global Rate Limit Considered Dangerous"(<https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/cao>)。

7月のインシデント

1	脆 7日:Adobe Acrobat及びReaderに不正終了や任意のコード実行の可能性がある複数の脆弱性が見つかり、修正された。 「Adobe Acrobat および Reader に関するセキュリティアップデート公開」(https://helpx.adobe.com/jp/security/products/acrobat/apsb16-26.html)。
2	
3	セ 7日:東京ズーネットのWebサイトが不正アクセスを受けて改ざんされ、メールマガジンに登録されていたメールアドレス 21,688件が流出した。 「都立動物園・水族園のホームページに不正アクセス発生 東京都」(http://www.metro.tokyo.jp/INET/OSHIRASE/2016/07/20q78400.htm)。「都立動物園・水族園ホームページの不正アクセス(続報) 東京都」(http://www.metro.tokyo.jp/INET/OSHIRASE/2016/07/20q78600.htm)。「都立動物園・水族園の公式ウェブサイト「東京ズーネット」復旧のお知らせとお詫び 東京ズーネット」(http://www.tokyo-zoo.net/topic/topics_detail?kind=&inst=&link_num=23843)。
4	
5	他 8日:観光庁が第1回目の「旅行業界情報流出事案検討会」を開催し、旅行業界で相次いで発生した情報流出事案に関する報告書の検証や論点整理などが行われた。 「第1回「旅行業界情報流出事案検討会」を開催します 2016年 報道発表 報道・会見 観光庁」(http://www.mlit.go.jp/kankocho/news06_000283.html)。
6	
7	他 10日:Bitcoinが42万ブロックに到達し、マイナーへの報酬が25BTCから12.5BTCに半減したが、価格に大きな変動はなかった。
8	
9	セ 11日:Twitterの共同創業者であるJack Dorsey氏やYahoo!のCEOであるMarissa Mayer氏のTwitterアカウントがOurMineチームによって乗っ取られた。
10	脆 12日:Adobe Flash Playerに、不正終了や任意のコード実行の可能性がある複数の脆弱性が見つかり、修正された。 「Adobe Flash Player に関するセキュリティアップデート公開」(https://helpx.adobe.com/jp/security/products/flash-player/apsb16-25.html)。
11	
12	脆 13日:Microsoft社は、2016年7月のセキュリティ情報を公開し、MS16-084など6件の緊急と5件の重要な更新を含む合計11件の修正をリリースした。 「2016年7月のマイクロソフト セキュリティ情報の概要」(https://technet.microsoft.com/ja-jp/library/security/ms16-jul.aspx)。
13	
14	他 14日:警察庁は、海外サーバに開設された偽サイトなどによる詐欺などの被害を抑制するため、Webブラウザ事業者などが加盟する国際的な団体であるAPWG(フィッシング対策ワーキンググループ)への情報提供を開始したことを発表した。 警察庁、「APWGに対する海外偽サイト等の情報提供の開始について」(http://www.npa.go.jp/cyber/pdf/APWG.pdf)。
15	
16	セ 16日:Ubuntu ForumsのWebサイトから約200万人分のユーザ情報(ユーザ名、メールアドレス、IPアドレス)が流出した。vBulletinのプラグインであるForumrunnerにSQLインジェクションの脆弱性があり、これを悪用された。 "Notice of Ubuntu Forums breach; user passwords not compromised Ubuntu Insights"(https://insights.ubuntu.com/2016/07/15/notice-of-security-breach-on-ubuntu-forums/)。
17	
18	脆 18日:Apple社はiOS 9.3.3とOS X El Capitan 10.11.6及びセキュリティアップデート2016-004をリリースし、リモートの攻撃者によって任意のコードを実行される可能性があるなどの複数の脆弱性を修正した。また、併せてtvOS 9.2.2とwatchOS 2.2.2もリリースされた。 「iOS 9.3.3 のセキュリティコンテンツについて - Apple サポート」(https://support.apple.com/ja-jp/HT206902)。「OS X El Capitan v10.11.6 およびセキュリティアップデート 2016-004 のセキュリティコンテンツについて - Apple サポート」(https://support.apple.com/ja-jp/HT206903)。「tvOS 9.2.2 のセキュリティコンテンツについて - Apple サポート」(https://support.apple.com/ja-jp/HT206905)。「watchOS 2.2.2 のセキュリティコンテンツについて - Apple サポート」(https://support.apple.com/ja-jp/HT206904)。
19	
20	
21	脆 19日:Oracle社は四半期ごとの定例アップデートを公開し、Java SEやOracle Database Serverなどを含む複数製品について、合計276件の脆弱性を修正した。 "Oracle Critical Patch Update Advisory - July 2016"(http://www.oracle.com/technetwork/security-advisory/cpujul2016-2881720.html)。
22	
23	脆 19日:CGI(Common Gateway Interface)を利用するWebサーバに広く影響するhttpoxy脆弱性が公表された。また警察庁では、定点観測システムにおいて当該脆弱性を標的としてと考えられるアクセスを観測したことから、注意喚起を行った。 httpoxy,"A CGI application vulnerability for PHP, Go, Python and others"(https://httpoxy.org/)。警察庁、「CGI等を利用するウェブサーバの脆弱性(httpoxy)を標的としたアクセスの観測について」(http://www.npa.go.jp/cyberpolice/detect/pdf/20160720.pdf)。
24	
25	他 20日:位置情報を利用したスマートフォン向けゲーム「ポケモンGO」が国内でもリリースされ、先行してリリースされていた米国などにおける状況を考慮して、内閣サイバーセキュリティセンターから注意喚起が出された。 「位置情報ゲーム「ポケモンGO」に関する注意喚起について」(http://www.nisc.go.jp/active/kihon/pdf/reminder_20160721.pdf)。
26	
27	他 21日:6月に仮想通貨の投資ファンドであるThe DAOから不正に引き出されたETHを取り戻すために、Ethereumがハードフォークを実施した。 "Hard Fork Completed - Ethereum Blog"(https://blog.ethereum.org/2016/07/20/hard-fork-completed/)。
28	
29	セ 23日:WikiLeaksが米国の民主党全国委員会(DNC)の内部メールおよそ2万通をWebサイトで公開した。これを受けてDNCの全国委員長など複数の幹部が辞任した。またDNCに侵入したとされるGUCCIFER 2.0は公開されたメール情報も自分が提供したものと主張した。 "WikiLeaks - Search the DNC email database"(https://wikileaks.org/dnc-emails/)。
30	
30	セ 25日:韓国のチケット予約サイトINTERPARKにおいて、外部からの不正アクセスによって約1,030万人分の個人情報が流出した。
31	セ 28日:米国の民主党議会選挙対策委員会(Democratic Congressional Campaign Committee: DCCC)が外部からの不正アクセスを受けた可能性があり、FBIが捜査しているとReutersが報道した。また後にDCCCもサイバー攻撃の標的となったことを認めた。

※ 日付は日本標準時

【凡例】

脆 脆弱性	セ セキュリティ事件	動 動静情報	歴 歴史	他 その他
--------------	-------------------	---------------	-------------	--------------

■ 過去最大規模のDDoS攻撃

この期間では、過去最大規模のDDoS攻撃が相次いで観測されました。セキュリティ業界では著名な専門家であるBrian Krebs氏が運営するブログ"Krebs on Security"は9月8日^{*18}と10日^{*19}に相次いでイスラエルのvDOSサービスに関する記事を掲載し、その後に140Gbps程度のDoS攻撃を受けました。vDOSはbooter/stresserなどとも呼ばれるDDoS-for-hireサービス(DDoS攻撃代行サービス)の1つで、Krebs氏がその内情を調査し実態を暴く記事を掲載したこと、またイスラエルでvDOSサービスを運営していた18才の2人の男性が逮捕されたことなどから、その報復としてDoS攻撃を受けたと考えられています。攻撃はその後も激しさを増し、9月20日に別のbooter関連の記事^{*20}が掲載された後に約620Gbpsの大規模な攻撃を受けたとされています^{*21}。Krebs on Securityに対してはProlexic(Akamai)が過去4年間にわたってDDoS攻撃対策サービスを無償提供していましたが、攻撃規模があまりにも大きく他の有償顧客への影響も避けられないことから、サービス提供が中止されることになり、Krebs on Securityは一時的にアクセスできなくなりました。その後Googleが提供するProject Shield^{*22}による保護を受けられることとなり、9月25日に復旧しています。Krebs氏は9月20日に公開した記事の中で、ルータ、IPカメラ、デジタルビデオレコーダーなどのいわゆるIoT機器に感染するマルウェアによって構成されたボットネットが攻撃元と考えられることを明らかにしました。IoT機器によるボットネットからのDDoS攻撃はこれまでもセキュリティベンダーなどから多数報告されています^{*23}。Arborの観測によると、リオ五輪期間中に発生した約540GbpsのDDoS攻撃もIoT機器のボットネットによるものであり、五輪開催の数カ月前からtelnet(23/tcp)への通信が急増していて、IoT機器へのマルウェア感染活動が活発化していたと報告しています^{*24}。

またKrebs on Securityへの攻撃とほぼ同時期に、フランスのホスティングサービスであるOVHも同様のDDoS攻撃を受け

ており、ピーク時には1Tbpsを越える攻撃トラフィックが観測されています^{*25}。なお今回の攻撃元と推測されるマルウェアの1つMiraiのソースコードが10月に入ってインターネット上の掲示板サイトで公開されています。Miraiの詳細については「1.4.1 Mirai Botnetの検知と対策」も併せてご参照ください。

IoT機器については、telnetなどの管理用ポートが開いているだけでなく、初期パスワードが変更されずにそのまま使われていたり、管理用に変更不可能なパスワードが設定されているなど、多くの問題点が指摘されています。そのためこうした脆弱なIoT機器をターゲットとしたマルウェアの感染活動が近年拡大しており、攻撃インフラとして悪用されているのが現状です。

■ 大量のパスワード情報漏えい

この期間においても引き続きインターネット上の様々なサービスから過去に大量のパスワード情報が漏えいしていたことが相次いで発覚しました。Mail.ruから約2,500万件、Dropboxから約6,800万件、Last.fmから約4,300万件、Rambler.ruから約9,800万件、Qip.ruから約3,300万件、Yahoo!から約5億件など、いずれも漏えい件数が非常に大規模でした。特にYahoo!については単一のサービス事業者からの漏えい件数としては過去最大規模です。これらの中には平文のパスワード情報が含まれていたものや、ソルトなしのMD5ハッシュによるパスワード情報が含まれていたものもあり、パスワード解析が比較的容易であるため、すぐに悪用される危険性が高いものでした。これらは一部を除いて、既に漏えいが確認されているMySpaceやLinkedInなどのサービスと同様に、ロシア人ハッカーグループが2011年から2014年頃に侵入して取得したデータが元になっていると考えられ、今年になってからロシアの掲示板サイトやDark Web上の販売サイトで一般に売りはじめられたことによって情報流出が分かったものです。

*18 "Israeli Online Attack Service 'vDOS' Earned \$600,000 in Two Years — Krebs on Security"(<http://krebsonsecurity.com/2016/09/israeli-online-attack-service-vdos-earned-600000-in-two-years/>)。

*19 "Alleged vDOS Proprietors Arrested in Israel — Krebs on Security"(<http://krebsonsecurity.com/2016/09/alleged-vdos-proprietors-arrested-in-israel/>)。

*20 "DDoS Mitigation Firm Has History of Hijacks — Krebs on Security"(<http://krebsonsecurity.com/2016/09/ddos-mitigation-firm-has-history-of-hijacks/>)。

*21 "KrebsOnSecurity Hit With Record DDoS — Krebs on Security"(<https://krebsonsecurity.com/2016/09/krebsonsecurity-hit-with-record-ddos/>)。

*22 "Google | Project Shield | Free DDoS protection"(<https://projectshield.withgoogle.com/public/>)。

*23 例えば、FlashpointとLevel 3はBASHLITEボットネットについて8月末にブログで報告している。"Attack of Things! - Beyond Bandwidth" (<http://blog.level3.com/security/attack-of-things/>)。

*24 "Rio Olympics Take the Gold for 540gb/sec Sustained DDoS Attacks!"(<https://www.arbornetworks.com/blog/asert/rio-olympics-take-gold-540gbsec-sustained-ddos-attacks/>)。

*25 "OVH News - The DDoS that didn't break the camel's VAC"(<https://www.ovh.com/us/news/articles/a2367.the-ddos-that-didnt-break-the-camels-vac>)。

8月のインシデント

1	他	2日:内閣サイバーセキュリティセンター(NISC)において、セキュリティマインドを持った企業経営ワーキンググループ第3回会合が開催され、同ワーキンググループにて策定された「企業経営のためのサイバーセキュリティの考え方」が公開された。
2		「セキュリティマインドを持った企業経営ワーキンググループ」(http://www.nisc.go.jp/conference/cs/jinzai/wg/index.html)。
3	セ	3日:香港のBitcoin取引所であるBitfinexがBitcoinの盗難被害に遭い取引を一時停止した。被害額は約12万BTC(およそ60億円相当)で、この影響によりBitcoinの価格が1割程度下落した。
4		"Security Breach - Bitfinex blog"(http://blog.bitfinex.com/announcements/security-breach/)。
5	セ	5日:リオデジャネイロで開催されたオリンピックに合わせてAnonymousによる抗議活動が行われた(OpOlympics/OpOlympicHacking/OpR10)。開催期間中に地元のリオデジャネイロ州のWebサイトへのDoS攻撃などが発生した。
6		
7	脆	8日:Check Point Software Technologiesは、数多くのAndroid端末に搭載されているQualcomm社製のチップセットに「QuadRooter」と名付けられた4件の脆弱性があることを公表した。これらの脆弱性は他の脆弱性と併せ、9月6日に公開された"Android Security Bulletin—September 2016"(https://source.android.com/security/bulletin/2016-09-01.html)で修正されている。
8		"QuadRooter:New Android Vulnerabilities in Over 900 Million Devices Check Point Blog"(http://blog.checkpoint.com/2016/08/07/quadrooter/)。
9	脆	10日:Microsoft社は、2016年8月のセキュリティ情報を公開し、MS16-095など5件の緊急と4件の重要な更新を含む合計9件の修正をリリースした。
10		「2016年8月のマイクロソフト セキュリティ情報の概要」(https://technet.microsoft.com/ja-jp/library/security/ms16-aug.aspx)。
11	セ	13日:The Shadow Brokersと名乗る何者かが、Equation Groupのものだとするファイル群の一部を公開、残りをオークションにかけると発表した。
12		
13	セ	14日:世界アンチ・ドーピング機関(World Anti-Doping Agency:WADA)が管理するAnti-Doping Administration and Management System(ADAMS)において、ロシアのYuliya Stepanova選手のアカウントへの不正ログインが確認された。またこの事件の発生前に複数のユーザがフィッシングメールを受信していたことが分かった。
14		"WADA confirms illegal activity on Yuliya Stepanova's ADAMS account World Anti-Doping Agency"(https://www.wada-ama.org/en/media/news/2016-08/wada-confirms-illegal-activity-on-yuliya-stepanovas-adams-account)。
15	セ	19日:Anonymousによる攻撃キャンペーン#OpKillingBay 2016のターゲットリストが公開された。
16		
17	他	19日:Twitterがこの半年間において、テロ活動に関連したとして235,000アカウントを停止したことを発表した。既に2月に125,000アカウントの停止を発表しており、合わせて360,000アカウントが停止されたことになる。
18		"An update on our efforts to combat violent extremism Twitter Blogs"(https://blog.twitter.com/2016/an-update-on-our-efforts-to-combat-violent-extremism)。
19	脆	25日:INRIAグループから、3DESなどの64ビットブロック暗号において同じ鍵で暗号化されたデータを大量に傍受することで暗号文の衝突を見つけ出し平文回復を行う攻撃手法(SWEET32)が公開された。OpenSSLなどの実装では、デフォルト設定での3DESの使用を制限するなどの対応が行われた。
20		"Sweet32:Birthday attacks on 64-bit block ciphers in TLS and OpenVPN"(https://sweet32.info/)。
21	セ	25日:Mail.ruの掲示板サイトから約2,500万人分のユーザ情報が流出したことが分かった。vBulletinの既知のSQLインジェクション脆弱性が悪用された。
22		"LeakedSource Analysis of *.mail.ru Hack"(https://www.leakedsource.com/blog/mailru/)。
23	セ	28日:さくらインターネットのDNSサーバが外部からのDoS攻撃を受けて障害が発生した。またその後も数日間にわたって断続的にレンタルサーバやDNSサーバに対するDoS攻撃が発生した。
24		「外部からのDoSトラフィックによるネームサーバ障害」(http://support.sakura.ad.jp/mainte/mainteentry.php?id=20072)。
25	セ	28日:技術評論社のWebサイトに対して外部からのDoS攻撃が発生し、サービスが利用できなくなった。またその後も数日間にわたって断続的にDoS攻撃が発生した。
26		「DoS攻撃による断続的な接続障害についてのお詫び」(https://gihyo.jp/news/info/2016/09/0901?ard=1472782622)。
27	セ	30日:米国のイリノイ州及びアリゾナ州において、選挙システムに対する外部からの不正アクセスがあったことを受けて、FBIから注意喚起情報が出されていたことが分かった。
28		
29	セ	31日:Dropboxから2012年時点のユーザのメールアドレスとパスワード情報およそ6,800万人分が流出していたことが分かり、該当ユーザのパスワードがリセットされた。
30		"Resetting passwords to keep your files safe Dropbox Blog"(https://blogs.dropbox.com/dropbox/2016/08/resetting-passwords-to-keep-your-files-safe/)。
31	セ	31日:中国の認証局WoSignの証明書発行システムに問題があり、不正な証明書が発行可能な状態であることが分かった。
		"The story of how WoSign gave me an SSL certificate for GitHub.com Schrauger.com"(https://www.schrauger.com/the-story-of-how-wosign-gave-me-an-ssl-certificate-for-github-com)。"WoSign Incidents Report(September 4th 2016)"(https://www.wosign.com/report/wosign_incidents_report_09042016.pdf)。

※ 日付は日本標準時

【凡例】

脆 脆弱性 セ セキュリティ事件 動 動静情報 歴 歴史 他 その他

こうして流出したパスワード情報は、パスワードを複数のサービスで使い回しているユーザを狙って、他サイトへのなりすましによる不正ログイン攻撃や、著名人など特定のSNSアカウントの乗っ取りに悪用される危険性が高いことが過去の事例から分かっています。そのため、ユーザは自分のパスワード情報が漏えいする可能性をあらかじめ考慮して、複数のサービスで同じパスワードを使い回さないようにするなどの自衛策が求められます。

■ 政府機関の取り組み

内閣官房内閣サイバーセキュリティセンター(NISC)において、サイバーセキュリティ戦略本部第9回会合が8月31日に開催され、前回会合で出された案にパブリックコメントの結果を踏まえた見直しを行い、「サイバーセキュリティ2016」が決定されました^{*26}。これは、サイバーセキュリティ政策の基本的な方向性を示す新たな国家戦略として2015年9月に閣議決定された「サイバーセキュリティ戦略」に基づく2期目の年次計画であり、政府が2016年度に実施する具体的な取り組みを戦略の体系に沿って示したものです。わが国の経済の持続的な発展や安全で安心な国民生活の実現、国際社会における平和の維持などを目的とした様々な施策が盛り込まれています。政府機関だけでなく、重要インフラ事業者や企業、個人などとも連携してこれらの取り組みを推進していくことが求められます。

観光庁では、今年6月に旅行業界において情報流出事案が相次いで発生したことを受け、問題点の検証及び再発防止策を取りまとめるため「旅行業界情報流出事案検討会」を設置しましたが、その第1回目の会合が7月8日に開催されました。その後の7月22日の第2回会合において中間取りまとめを行い、9月16日の第3回会合で進捗状況などを確認しています。また旅行業界内への情報共有や、情報管理の徹底及び情報流出の再発を防止するための情報共有会議も観光庁と旅行業界と共同で6月から不定期に開催しています^{*27}。中間取りまとめでは旅行業者や観光庁が再発防止に向けて今後取るべき対策についての提言がなされ、情報共有会議を通じて一般にもその内容が公開されています。

■ その他

8月13日にThe Shadow Brokersと名乗る何者かが、Equation Groupのものとして主張するファイル群の一部を公開し、残りをオークションにかけると発表しました。Equation GroupというのはKaspersky Labs社が2015年に報告した攻撃グループの名称であり^{*28}、その攻撃の特徴と過去の別の攻撃との類似性や、Edward Snowden氏によって持ち出されたNSAの機密文書に記載されている内容との一致などから、米国家安全保障局(NSA)において主に攻撃を担当する部局Tailored Access Operations(TAO)だと考えられてきました。

これらの公開されたファイルの中にはCiscoやJuniperなどのファイアウォール製品を対象としたエクスプロイトコードやマルウェアなどが含まれており、セキュリティ研究者やベンダーによる検証の結果、実際に動作するものだと分かりました。これらのコードの中には、それまで未知のゼロデイの脆弱性を悪用するものも含まれていました。対象機器のベンダーはそれぞれ脆弱性を修正する対応に追われましたが、中でもBENIGNCERTAINというコードネームのエクスプロイトコードはVPNの暗号化に利用されるRSA秘密鍵をリモートから取得可能なもので、当初対象と考えられていた古いPIX Firewallだけでなく、Cisco IOSソフトウェアを搭載する多数のネットワーク機器に影響が及ぶことがCisco社の検証により明らかとなりました^{*29}。

なお、過去に公開済みのSnowden氏によってリークされた文書の中に記載されている複数のコードネームが今回のファイルに含まれていることや、The Interceptから新たに公開されたNSAの機密文書の中に識別子として含まれているMSGIDが、The Shadow Brokersから公開されたバイナリにも使われていることなどから^{*30}、The Shadow Brokersによって公開されたファイル群はEquation Groupのものというだけでなく、NSAのTAOによって作成されたものである可能性が非常に高くなっています。The Shadow Brokersの正体やファイルを公開した目的などについても様々な憶測が飛び交いましたが、詳細は不明ではっきりしたことは何も分かっていません。

*26 NISC、「サイバーセキュリティ戦略本部第9回会合」(<http://www.nisc.go.jp/conference/cs/index.html#cs09>)。

*27 「第2回 観光庁・旅行業界情報共有会議」の概要について | 2016年 | トピックス | 報道・会見 | 観光庁」(http://www.mlit.go.jp/kankochu/topics06_000080.html)。

*28 "Equation: The Death Star of Malware Galaxy - Securelist"(<https://securelist.com/blog/research/68750/equation-the-death-star-of-malware-galaxy/>)。

*29 "IKEv1 Information Disclosure Vulnerability in Multiple Cisco Products"(<https://tools.cisco.com/security/center/content/CiscoSecurityAdvisory/cisco-sa-20160916-ikev1>)。

*30 "The NSA Leak Is Real, Snowden Documents Confirm"(<https://theintercept.com/2016/08/19/the-nsa-was-hacked-snowden-documents-confirm/>)。

9月のインシデント

1	セ	1日: Anonymousによる#OpKillingBayの攻撃キャンペーンにおいて、北欧などの多数のWebサイトへの攻撃に関わったとして、20才のデンマーク人がデンマーク、ノルウェー、フィンランド、FBIの合同捜査によって逮捕された。TwitterではRektFaggotの名前で活動しており、2015年には日本国内へも活発に攻撃活動を実施したことを示唆していた。
2	セ	2日: ヨドバシカメラのWebサイトに対して外部からのDoS攻撃が発生し、サービスが利用できなくなった。またその後も数日間にわたって断続的にDoS攻撃が発生した。
3	セ	5日: 株式会社ティノス・セシールの通販サイト「セシールオンラインショップ」において、第三者のなりすましによる不正ログインが発生した。「弊社「セシールオンラインショップ」への不正アクセスについて」(http://www.cecile.co.jp/fst/information/20160905.pdf)。
4	セ	9日: 米政府は初代の連邦最高情報セキュリティ責任者(CISO)を任命したと発表した。CISOは米政府機関におけるサイバーセキュリティのポリシー、計画及び実装を推進する役割を担うこととなった。 "Announcing the First Federal Chief Information Security Officer whitehouse.gov"(https://www.whitehouse.gov/blog/2016/09/08/announcing-first-federal-chief-information-security-officer)。
5	他	9日: 米政府は初代の連邦最高情報セキュリティ責任者(CISO)を任命したと発表した。CISOは米政府機関におけるサイバーセキュリティのポリシー、計画及び実装を推進する役割を担うこととなった。 "Announcing the First Federal Chief Information Security Officer whitehouse.gov"(https://www.whitehouse.gov/blog/2016/09/08/announcing-first-federal-chief-information-security-officer)。
6	セ	10日: DDos-for-hireサービスの1つであるvDOSの2人の創業者がイスラエルで逮捕された。 "Alleged vDOS Proprietors Arrested in Israel — Krebs on Security"(http://krebsonsecurity.com/2016/09/alleged-vdos-proprietors-arrested-in-israel/)。
7	セ	10日: DDos-for-hireサービスの1つであるvDOSの2人の創業者がイスラエルで逮捕された。 "Alleged vDOS Proprietors Arrested in Israel — Krebs on Security"(http://krebsonsecurity.com/2016/09/alleged-vdos-proprietors-arrested-in-israel/)。
8	脆	12日: MySQLにおいて、リモートからコード実行・権限昇格が可能な脆弱性が見つかり、研究者によって公開された。 "MySQL <= 5.7.14 Remote Root Code Execution / Privilege Escalation (Oday) (CVE-2016-6662)"(http://legalhackers.com/advisories/MySQL-Exploit-Remote-Root-Code-Execution-Privesc-CVE-2016-6662.html)。
9	脆	12日: MySQLにおいて、リモートからコード実行・権限昇格が可能な脆弱性が見つかり、研究者によって公開された。 "MySQL <= 5.7.14 Remote Root Code Execution / Privilege Escalation (Oday) (CVE-2016-6662)"(http://legalhackers.com/advisories/MySQL-Exploit-Remote-Root-Code-Execution-Privesc-CVE-2016-6662.html)。
10	脆	13日: Apple社はiOS 10及び10.0.1をリリースし、リモートの攻撃者によって任意のコードを実行される可能性があるなどの複数の脆弱性を修正した。また、併せてtvOS 10とwatchOS 3もリリースされた。 「iOS 10 のセキュリティコンテンツについて - Apple サポート」(https://support.apple.com/ja-jp/HT207143)。「iOS 10.0.1 のセキュリティコンテンツについて - Apple サポート」(https://support.apple.com/ja-jp/HT207145)。「tvOS 10 のセキュリティコンテンツについて - Apple サポート」(https://support.apple.com/ja-jp/HT207142)。「watchOS 3 のセキュリティコンテンツについて - Apple サポート」(https://support.apple.com/ja-jp/HT207141)。
11	脆	13日: Apple社はiOS 10及び10.0.1をリリースし、リモートの攻撃者によって任意のコードを実行される可能性があるなどの複数の脆弱性を修正した。また、併せてtvOS 10とwatchOS 3もリリースされた。 「iOS 10 のセキュリティコンテンツについて - Apple サポート」(https://support.apple.com/ja-jp/HT207143)。「iOS 10.0.1 のセキュリティコンテンツについて - Apple サポート」(https://support.apple.com/ja-jp/HT207145)。「tvOS 10 のセキュリティコンテンツについて - Apple サポート」(https://support.apple.com/ja-jp/HT207142)。「watchOS 3 のセキュリティコンテンツについて - Apple サポート」(https://support.apple.com/ja-jp/HT207141)。
12	脆	13日: Adobe Flash Playerに、不正終了や任意のコード実行の可能性がある複数の脆弱性が見つかり、修正された。 「Adobe Flash Player に関するセキュリティアップデート公開」(https://helpx.adobe.com/jp/security/products/flash-player/apsb16-29.html)。
13	脆	13日: Adobe Flash Playerに、不正終了や任意のコード実行の可能性がある複数の脆弱性が見つかり、修正された。 「Adobe Flash Player に関するセキュリティアップデート公開」(https://helpx.adobe.com/jp/security/products/flash-player/apsb16-29.html)。
14	セ	13日: Fancy Bears' Hack Teamを名乗る何者かが、世界アンチ・ドーピング機関(World Anti-Doping Agency: WADA)が管理するAnti-Doping Administration and Management System(ADAMS)のデータを不正に取得して公開した。またWADAもこれを確認した。 "American Athletes Caught Doping 2016-09-13"(http://fancybear.net/pages/1.html)。"WADA Confirms Attack by Russian Cyber Espionage Group World Anti-Doping Agency"(https://www.wada-ama.org/en/media/news/2016-09/wada-confirms-attack-by-russian-cyber-espionage-group)。
15	脆	13日: Fancy Bears' Hack Teamを名乗る何者かが、世界アンチ・ドーピング機関(World Anti-Doping Agency: WADA)が管理するAnti-Doping Administration and Management System(ADAMS)のデータを不正に取得して公開した。またWADAもこれを確認した。 "American Athletes Caught Doping 2016-09-13"(http://fancybear.net/pages/1.html)。"WADA Confirms Attack by Russian Cyber Espionage Group World Anti-Doping Agency"(https://www.wada-ama.org/en/media/news/2016-09/wada-confirms-attack-by-russian-cyber-espionage-group)。
16	セ	13日: Fancy Bears' Hack Teamを名乗る何者かが、世界アンチ・ドーピング機関(World Anti-Doping Agency: WADA)が管理するAnti-Doping Administration and Management System(ADAMS)のデータを不正に取得して公開した。またWADAもこれを確認した。 "American Athletes Caught Doping 2016-09-13"(http://fancybear.net/pages/1.html)。"WADA Confirms Attack by Russian Cyber Espionage Group World Anti-Doping Agency"(https://www.wada-ama.org/en/media/news/2016-09/wada-confirms-attack-by-russian-cyber-espionage-group)。
17	脆	14日: Microsoft社は、2016年9月のセキュリティ情報を公開し、MS16-104など7件の緊急と7件の重要な更新を含む合計14件の修正をリリースした。 「2016年9月のマイクロソフト セキュリティ情報の概要」(https://technet.microsoft.com/ja-jp/library/security/ms16-sep.aspx)。
18	脆	14日: Microsoft社は、2016年9月のセキュリティ情報を公開し、MS16-104など7件の緊急と7件の重要な更新を含む合計14件の修正をリリースした。 「2016年9月のマイクロソフト セキュリティ情報の概要」(https://technet.microsoft.com/ja-jp/library/security/ms16-sep.aspx)。
19	セ	14日: DC LeaksがColin Powell氏のGmailのデータを不正に取得し、2014年から2016年の約2年間分のメールを公開した。 "DC Leaks Colin Luther Powell"(http://dcleaks.com/index.php/portfolio_page/colin-luther-powell/)。
20	セ	14日: DC LeaksがColin Powell氏のGmailのデータを不正に取得し、2014年から2016年の約2年間分のメールを公開した。 "DC Leaks Colin Luther Powell"(http://dcleaks.com/index.php/portfolio_page/colin-luther-powell/)。
21	セ	14日: DC LeaksがColin Powell氏のGmailのデータを不正に取得し、2014年から2016年の約2年間分のメールを公開した。 "DC Leaks Colin Luther Powell"(http://dcleaks.com/index.php/portfolio_page/colin-luther-powell/)。
22	歴	18日: 毎年、歴史的な要因により、この日の前後に発生していた攻撃について、今年は組織的な攻撃は特に見られなかった。
23	脆	20日: Apple社はmacOS Sierra 10.12をリリースし、リモートの攻撃者によって任意のコードを実行される可能性があるなどの複数の脆弱性を修正した。 「macOS Sierra 10.12 のセキュリティコンテンツについて - Apple サポート」(https://support.apple.com/ja-jp/HT207170)。
24	脆	20日: Apple社はmacOS Sierra 10.12をリリースし、リモートの攻撃者によって任意のコードを実行される可能性があるなどの複数の脆弱性を修正した。 「macOS Sierra 10.12 のセキュリティコンテンツについて - Apple サポート」(https://support.apple.com/ja-jp/HT207170)。
25	セ	21日: Krebs on Securityが約620GbpsのDDoS攻撃を受け、一時アクセスできなくなった。 "KrebsOnSecurity Hit With Record DDoS — Krebs on Security"(https://krebsonsecurity.com/2016/09/krebsonsecurity-hit-with-record-ddos/)。
26	セ	21日: Krebs on Securityが約620GbpsのDDoS攻撃を受け、一時アクセスできなくなった。 "KrebsOnSecurity Hit With Record DDoS — Krebs on Security"(https://krebsonsecurity.com/2016/09/krebsonsecurity-hit-with-record-ddos/)。
27	セ	23日: 米Yahoo!から2014年後半時点のユーザのメールアドレスやパスワード情報などが少なくとも約5億人分流出していたことが分かった。 "An Important Message About Yahoo User Security Yahoo"(https://yahoo.tumblr.com/post/150781911849/an-important-message-about-yahoo-user-security)。
28	脆	28日: BIND9に、DNS応答を作成する処理に不具合があるため、外部からDoS攻撃可能な脆弱性が見つかり、修正された。 "CVE-2016-2776: Assertion Failure in buffer.c While Building Responses to a Specifically Constructed Request Internet Systems Consortium Knowledge Base"(https://kb.isc.org/article/AA-01419)。
29	脆	28日: BIND9に、DNS応答を作成する処理に不具合があるため、外部からDoS攻撃可能な脆弱性が見つかり、修正された。 "CVE-2016-2776: Assertion Failure in buffer.c While Building Responses to a Specifically Constructed Request Internet Systems Consortium Knowledge Base"(https://kb.isc.org/article/AA-01419)。
30	他	30日: 内閣サイバーセキュリティセンター(NISC)は、政府のサイバーセキュリティに関する予算額を発表した。平成29年度予算概算要求額は601.4億円で、平成28年度当初予算額498.3億円から約103億円の増加となった。また平成28年度第2次補正予算として72.2億円を計上した。 「政府のサイバーセキュリティに関する予算(平成29年度予算概算要求及び平成28年度予算第2次補正)」(http://www.nisc.go.jp/active/kihon/pdf/yosan2017.pdf)。

※ 日付は日本標準時

【凡例】

脆	脆弱性	セ	セキュリティ事件	動	動静情報	歴	歴史	他	その他
---	-----	---	----------	---	------	---	----	---	-----

1.3 インシデントサーベイ

1.3.1 DDoS攻撃

現在、一般の企業のサーバに対するDDoS攻撃が、日常的に発生するようになっており、その内容は、多岐にわたります。しかし、攻撃の多くは、脆弱性などの高度な知識を利用したのではなく、多量の通信を発生させて通信回線を埋めたり、サーバの処理を過負荷にしたりすることでサービスの妨害を狙ったものになっています。

■ 直接観測による状況

図-2に、2016年7月から9月の期間にIJ DDoSプロテクションサービスで取り扱ったDDoS攻撃の状況を示します。

ここでは、IJ DDoSプロテクションサービスの基準で攻撃と判定した通信異常の件数を示しています。IJでは、ここに示す以外のDDoS攻撃にも対処していますが、攻撃の実態を正確に把握することが困難なため、この集計からは除外しています。

DDoS攻撃には多くの攻撃手法が存在し、攻撃対象となった環境の規模(回線容量やサーバの性能)によって、その影響度合いが異なります。図-2では、DDoS攻撃全体を、回線容量に対する攻撃^{*31}、サーバに対する攻撃^{*32}、複合攻撃(1つの攻撃対象に対し、同時に数種類の攻撃を行うもの)の3種類に分類しています。

この3か月間でIJは、392件のDDoS攻撃に対処しました。1日あたりの対処件数は4.26件で、平均発生件数は前回のレポート期間と比べて増加しています。DDoS攻撃全体に占める割合は、サーバに対する攻撃が62.76%、複合攻撃が31.89%、回線容量に対する攻撃が5.36%でした。

今回の対象期間で観測された中で最も大規模な攻撃は、複合攻撃に分類したもので、最大89万ppsのパケットによって2.97Gbpsの通信量を発生させる攻撃でした。攻撃の継続時間は、全体の79.34%が攻撃開始から30分未満で終了し、19.13%が30分以上24時間未満の範囲に分布しており、24時間以上継続した攻撃は1.53%でした。なお、今回最も長く継続した攻撃は、複合攻撃に分類されるもので5日と3時間16分(123時間16分)にわたりました。

攻撃元の分布としては、多くの場合、国内、国外を問わず非常に多くのIPアドレスが観測されました。これは、IPスプーフィング^{*33}の利用や、DDoS攻撃を行うための手法としてのボットネット^{*34}の利用によるものと考えられます。

■ backscatterによる観測

次に、IJでのマルウェア活動観測プロジェクトMITFのハニーポット^{*35}によるDDoS攻撃のbackscatter観測結果を示します^{*36}。backscatterを観測することで、外部のネットワーク

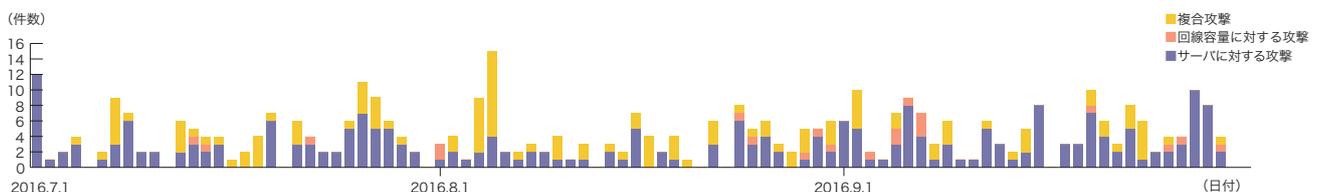


図-2 DDoS攻撃の発生件数

*31 攻撃対象に対し、本来不必要な大きなサイズのIPパケットやその断片を大量に送りつけることで、攻撃対象の接続回線の容量を圧迫する攻撃。UDPパケットを利用した場合にはUDP floodと呼ばれ、ICMPパケットを利用した場合にはICMP floodと呼ばれる。

*32 TCP SYN floodやTCP connection flood、HTTP GET flood攻撃など。TCP SYN flood攻撃は、TCP接続の開始の呼を示すSYNパケットを大量に送付することで、攻撃対象に大量の接続の準備をさせ、対象の処理能力やメモリなどを無駄に消費させる。TCP Connection flood攻撃は、実際に大量のTCP接続を確立させる。HTTP GET flood攻撃は、Webサーバに対しTCP接続を確立した後、HTTPのプロトコルコマンドGETを大量に送付することで、同様に攻撃対象の処理能力やメモリを無駄に消費させる。

*33 発信元IPアドレスの詐称。他人からの攻撃に見せかけたり、多人数からの攻撃に見せかけたりするために、攻撃パケットの送出時に、攻撃者が実際に利用しているIPアドレス以外のアドレスを付与した攻撃パケットを作成、送出すること。

*34 ボットとは、感染後に外部のC&Cサーバからの命令を受けて攻撃を実行するマルウェアの一種。ボットが多数集まって構成されたネットワークをボットネットと呼ぶ。

*35 IJのマルウェア活動観測プロジェクトMITFが設置しているハニーポット。「1.3.2 マルウェアの活動」も参照。

*36 この観測手法については、本レポートのVol.8 (http://www.ij.ad.jp/development/ijr/pdf/ijr_vol08.pdf)の「1.4.2 DDoS攻撃によるbackscatterの観測」で仕組みとその限界、IJによる観測結果の一部について紹介している。

で発生したDDoS攻撃の一部を、それに介在することなく第三者として検知できます。

2016年7月から9月の間に観測したbackscatterについて、発信元IPアドレスの国別分類を図-3に、ポート別のパケット数推移を図-4にそれぞれ示します。

観測されたDDoS攻撃の対象ポートのうち最も多かったものはWebサービスで利用される80/TCPで、全パケット数の48.7%を占めています。また、DNSで利用される53/UDP、HTTPSで利用される443/TCP、SSHで利用される22/TCPへの攻撃、ゲームの通信で利用されることがある27015/UDPへの攻撃、通常は利用されない19108/TCP、8370/TCP、3306/UDPなどへの攻撃が観測されています。

2014年2月から多く観測され前期間の5月25日に収束した53/UDPは、9月20日頃から再び観測されるようになり、1日5,000パケット程度の水準に戻っています。

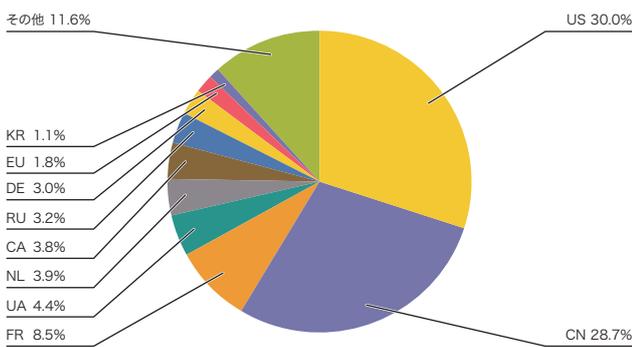


図-3 DDoS攻撃のbackscatter観測による攻撃先の国別分類

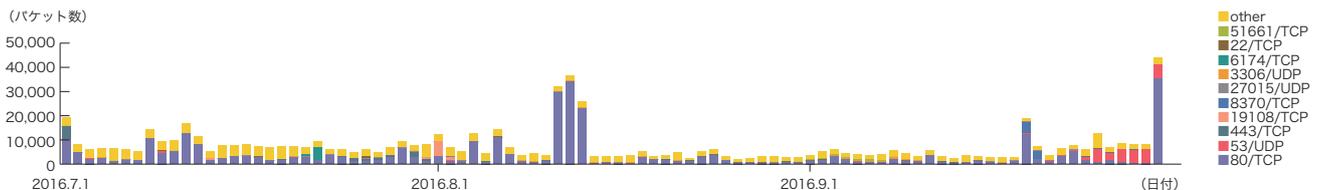


図-4 DDoS攻撃によるbackscatter観測 (観測パケット数、ポート別推移)

図-3で、DDoS攻撃の対象となったIPアドレスと考えられるbackscatterの発信元の国別分類を見ると、米国の30.0%が最も大きな割合を占めています。その後中国の28.7%、フランスの8.5%といった国が続いています。

特に多くのbackscatterを観測した場合について、攻撃先のポート別にみると、Webサーバ(80/TCP及び443/TCP)への攻撃としては、7月25日から28日にかけてブルガリアの調査報道サイトへの攻撃、7月28日から30日にかけて中国のオンラインゲーム関連サイトへの攻撃、8月11日から13日にかけて米国のCDN事業者がもつ多数のサーバへの攻撃、9月19日と30日には中国にある電気街の公式サイトへの攻撃を観測しています。他のポートへの攻撃としては、7月21日から22日にかけて中国の特定のIPアドレスに対する6174/TCPへの攻撃、7月31日から8月2日にかけて中国の特定のIPアドレスに対する19108/TCPへの攻撃、9月19日から20日にかけて中国の特定のIPアドレスに対する8370/TCPへの攻撃を観測しています。

また、今回の対象期間中に話題となったDDoS攻撃のうち、IJのbackscatter観測で検知した攻撃としては、7月6日にOurMine Teamを名乗るグループによるWikiLeaksへの攻撃、7月7日にAnonymousによるジンバブエ与党サイトへの攻撃、9月3日に国内家電量販店サイトへの攻撃をそれぞれ検知しています。

1.3.2 マルウェアの活動

ここでは、IJが実施しているマルウェアの活動観測プロジェクトMITF^{*37}による観測結果を示します。MITFでは、一般利用者と同様にインターネットに接続したハニーポット^{*38}を利用して、インターネットから到着する通信を観測しています。その

*37 Malware Investigation Task Forceの略。MITFは2007年5月から開始した活動で、ハニーポットを用いてネットワーク上でマルウェアの活動の観測を行い、マルウェアの流行状況を把握し、対策のための技術情報を集め、対策につなげる試み。

*38 脆弱性のエミュレーションなどの手法で、攻撃を受けつけて被害に遭ったふりをし、攻撃者の行為やマルウェアの活動目的を記録する装置。

また9月に入ってから、2323/TCPのアクセスが増加しています。図-7に2323/TCPの国別のアクセス数を示します。7月、8月にはほとんど通信が発生していないのに対し、9月6日に増加し始め、9月14日以降に急増しているのがわかります。MiraiBotは10回に1回の割合で、2323/TCPに通信する特徴があることが知られており、増加し始めた時期と合わせると、これらのほとんどはMiraiBotによるものと考えられます。国別に見ると、ベトナム、中国、ブラジル、コロンビア、韓国など、幅広い国に割り当てられたIPアドレスから受信していました。

7月初旬にSSDPプロトコルである1900/UDPが増加しています。主に米国、中国、フランス、韓国、ドイツなどに割り当てられたIPアドレスからSSDPの探査要求を受けています。これらは、SSDPリフレクターを使ったDDoS攻撃に利用可能な機器を探査する通信であると考えられます。

1433/TCPについても引き続き通信が増加しています。調査したところ、中国に割り当てられたIPアドレスを中心とした多数のIPアドレスからの通信でした。

本レポート期間中も引き続き53413/UDPが増加しています。調査したところ、Netis、Netcore製のルータの脆弱性を狙った攻撃の通信でした。この脆弱性は、2014年8月にトレンドマイクロによって報告されており^{*47}、JPCERT/CCが2015年4月から6月にかけて攻撃が増加したことを報告しています^{*48}。

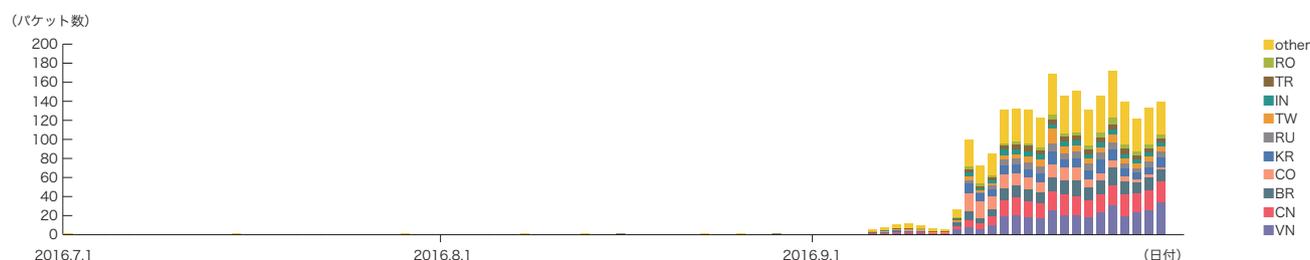


図-7 ハニーポットに到着した通信の推移(日別・2323/TCP・1台あたり)

■ ネットワーク上のマルウェアの活動

同じ期間中でのマルウェアの検体取得元の分布を図-8に、マルウェアの総取得検体数の推移を図-9に、そのうちのユニーク検体数の推移を図-10にそれぞれ示します。このうち図-9と図-10では、1日あたりに取得した検体^{*49}の総数を総取得検体数、検体の種類をハッシュ値^{*50}で分類したものをユニーク検体数としています。また、検体をウイルス対策ソフトで判別し、上位10種類の内訳をマルウェア名称別に色分けして示しています。なお、図-9と図-10は前回同様に複数のウイルス対策ソフトウェアの検出名によりConficker判定を行いConfickerと認められたデータを除いて集計しています。

期間中の1日あたりの平均値は、総取得検体数が124、ユニーク検体数が20でした。未検出の検体をより詳しく調査した結果、台湾、インド、ベトナムなどに割り当てられたIPアドレスで複数のSDBOTファミリー(IRCボットの一種)やビットコインマイニングツールのダウンロードなどが観測されています。

未検出の検体の約61%がテキスト形式でした。これらテキスト形式の多くは、HTMLであり、Webサーバからの404や403によるエラー応答であるため、古いワームなどのマルウェアが感染活動を続けているものの、新たに感染させたPCが、マルウェアをダウンロードしに行くダウンロード先のサイトが既に閉鎖させられていると考えられます。また、本レポート期間にハニーポット環境をメンテナンスし、HTTP経由やFTP経

*47 「UDPポートを開放した状態にするNetis製ルータに存在する不具合を確認」(<http://blog.trendmicro.co.jp/archives/9725>)。

*48 「インターネット定点観測レポート(2015年4~6月)」(<https://www.jpccert.or.jp/tsubame/report/report201504-06.html>)。

*49 ここでは、ハニーポットなどで取得したマルウェアを指す。

*50 様々な入力に対して一定長の出力をす一方方向関数(ハッシュ関数)を用いて得られた値。ハッシュ関数は異なる入力に対しては可能な限り異なる出力を得られるよう設計されている。難読化やパディングなどにより、同じマルウェアでも異なるハッシュ値を持つ検体を簡単に作成できてしまうため、ハッシュ値で検体の一意性を保証することはできないが、MITFではこの事実を考慮した上で指標として採用している。

由での攻撃を検知できるようにした結果、PHP形式のボットや.htaccessによるリダイレクタを取得できるようになったことも影響しています。MITF独自の解析では、今回の調査期間中に取得した検体は、ワーム型28.9%、ボット型61.3%、ダウンロード型9.8%でした。ボットの割合が大幅に増加していますが、これは前述のとおり、ハニーポット環境のメンテナンスにより、HTTPやFTP経由での攻撃を検知できるようになり、PHP形式のボットが数多く取得されたことによります。また解析により、54個のボットネットC&Cサーバ^{*51}と110個のマルウェア配布サイトの存在を確認しました。前回のレポート期間

中から大幅に増加していますが、これはハニーポットのメンテナンスにより、取得するマルウェアが増えた点や、解析環境が更新された影響と、DGA(ドメイン生成アルゴリズム)を持つマルウェアを検出したためです。

■ Confickerの活動

本レポート期間中、Confickerを含む1日あたりの平均値は、総取得検体数が4,185、ユニーク検体数は374でした。総取得検体数で97.0%、ユニーク検体数で94.8%を占めています。今回の対象期間でも支配的な状況が変わらないことから、Confickerを含む図は省略しています。本レポート期間中の総取得検体数は前回の対象期間と比較すると、約51%減少しています。これは前回のレポート期間の後半になるにしたがって減少傾向にあったのと、本レポート期間に切り替わる際にハニーポットのメンテナンスを行い、センサーのIPアドレスが変更になったためだと考えられます。Conficker Working Groupの観測記録^{*52}によると、2016年10月現在で、ユニークIPアドレスの総数は50万台とされています。2011年11月の約320万台と比較すると、約16%に減少したことになりますが、依然として大規模に感染し続けていることが分かります。

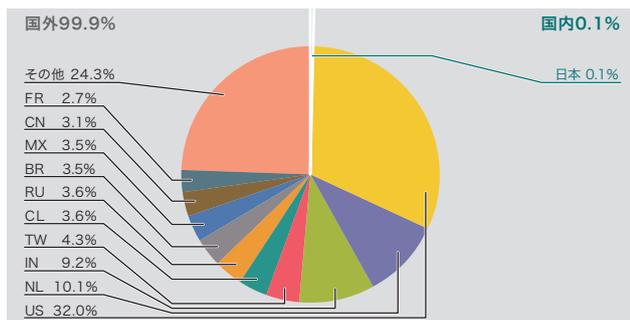


図-8 検体取得元の分布(国別分類、全期間、Confickerを除く)

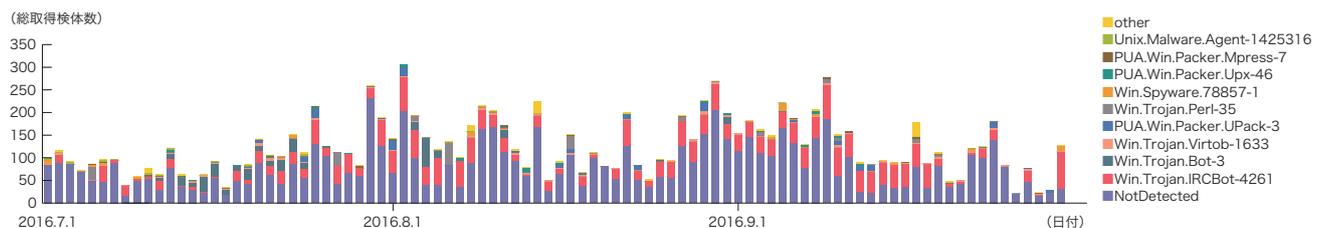


図-9 総取得検体数の推移(Confickerを除く)

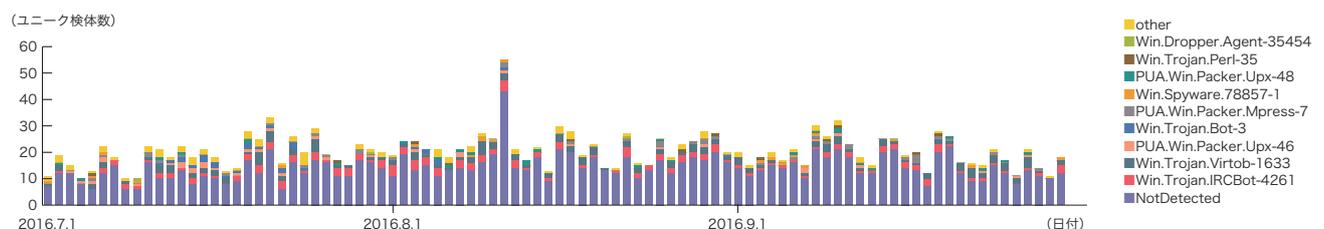


図-10 ユニーク検体数の推移(Confickerを除く)

*51 Command & Controlサーバの略。多数のボットで構成されたボットネットに指令を与えるサーバ。

*52 Conficker Working Groupの観測記録(<http://www.confickerworkinggroup.org/wiki/pmwiki.php/ANY/InfectionTracking>)。本レポート期間中、数値のデータが1月7日以降表示されていないため、10月前半の最高値をグラフから目視で確認して採用している。

1.3.3 SQLインジェクション攻撃

IJでは、Webサーバに対する攻撃のうち、SQLインジェクション攻撃^{*53}について継続して調査を行っています。SQLインジェクション攻撃は、過去にもたびたび流行し話題となった攻撃です。SQLインジェクション攻撃には、データを盗むための試み、データベースサーバに過負荷を起こすための試み、コンテンツ書き換えの試みの3つがあることが分かっています。

2016年7月から9月までに検知した、Webサーバに対するSQLインジェクション攻撃の発信元の分布を図-11に、攻撃の推移を図-12にそれぞれ示します。これらは、IJマネージドIPSサービスのシグネチャによる攻撃の検出結果をまとめたものです。発信元の分布では、米国35.7%、中国23.7%、日本9.8%となり、以下その他の国々が続いています。Webサーバに対するSQLインジェクション攻撃の合計値は前回と比べてほぼ横ばいの傾向にありますが、中国は増加傾向、また、日本は減少傾向にあります。

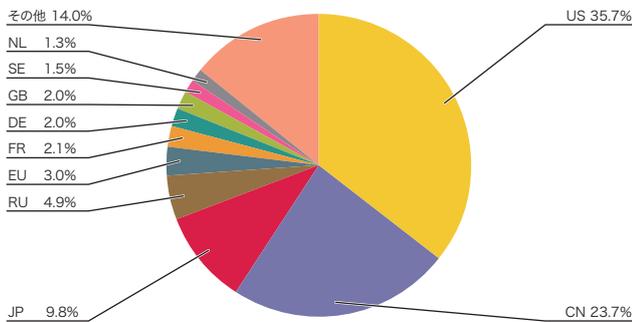


図-11 SQLインジェクション攻撃の発信元の分布

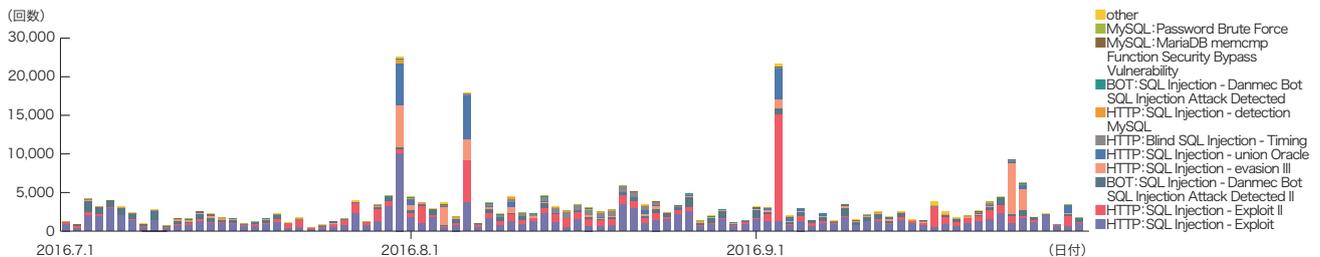


図-12 SQLインジェクション攻撃の推移(日別、攻撃種類別)

この期間中、7月31日には中国の複数の攻撃元から複数の攻撃先に対する攻撃が発生しています。8月6日には米国の特定の攻撃元から特定の攻撃先に攻撃が発生しています。9月3日には中国の特定の攻撃元から特定の攻撃先に攻撃が発生しています。9月24日には様々な地域の攻撃元から特定の攻撃先に攻撃が発生しています。また、9月25日にはインドネシアから特定の攻撃先に攻撃が発生しています。これらの攻撃はWebサーバの脆弱性を探る試みであったと考えられます。

ここまで示したとおり、各種の攻撃はそれぞれ適切に検出され、サービス上の対応が行われています。しかし、攻撃の試みは継続しているため、引き続き注意が必要な状況です。

1.3.4 Webサイト改ざん

MITFのWebクローラ(クライアントハニーポット)によって調査したWebサイト改ざん状況を示します^{*54}。

このWebクローラは国内の著名サイトや人気サイトなどを中心とした数十万のWebサイトを日次で巡回しており、更に巡回対象を順次追加しています。また、一時的にアクセス数が増加したWebサイトなどを対象に、一時的な観測も行っています。一般的な国内ユーザによる閲覧頻度が高いと考えられるWebサイトを巡回調査することで、改ざんサイトの増減や悪用される脆弱性、配布されるマルウェアなどの傾向が推測しやすくなります。

2016年7月から9月までの期間は、検知した受動的攻撃の大部分をNeutrino ExploitKitによるドライブバイダウンロー

*53 Webサーバに対するアクセスを通じて、SQLコマンドを発行し、その背後にいるデータベースを操作する攻撃。データベースの内容を権限なく閲覧、改ざんすることにより、機密情報の入手やWebコンテンツの書き換えを行う。

*54 Webクローラによる観測手法については本レポートのVol.22 (http://www.ij.ad.jp/company/development/report/iir/pdf/iir_vol22.pdf)の「1.4.3 WebクローラによるWebサイト改ざん調査」で仕組みを紹介している。

ド攻撃が占めました(図-13)。2016年6月にAngler ExploitKitが消滅して以来継続していた傾向です。しかし、9月下旬頃にはNeutrinoは一切観測されなくなり、代わりにRig ExploitKitが大規模に観測されるようになりました^{*55*56}。これらのペイロードとしては、Locky、Cerber、Ursnifなどを確認しています。

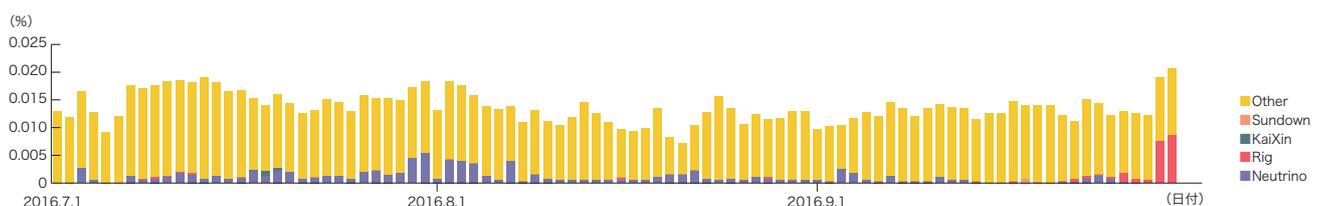
Exploit Kitへの誘導元として悪用されているWebサイトについて、規模やコンテンツなどに共通点は確認できませんが、WordPressで運用されているWebサイトが複数確認されています。また、AnglerやNeutrinoの誘導元として悪用されていたWebサイト、他の詐欺行為などに悪用されていたWebサイトがRigへの誘導元として機能するようになったケースを多数観測しています。なお、これらの誘導元となっているWebサイトにMac OS Xクライアントでアクセスした場合は、Landing pageへ誘導されない、あるいはLanding pageが応答を返さないことを確認しています。本期間中、Mac OS Xを対象としたドライブバイダウンロード攻撃は観測していません^{*57}。

また、ブラウザ画面にマルウェア感染などを仄めかす偽のダイアログなどを表示して、PUA^{*58}のインストールや偽のサポートセンターへの電話を促す詐欺サイトへの誘導の観測数が高

い値で継続しています。これらの詐欺サイトはMac OS X環境でも機能します。

誘導元WebサイトからLocationヘッダなどを利用してランサムウェアやPUAなどの実行ファイルを単にダウンロードさせようとするケースも複数見受けられました。このような場合、ファイルが自動的に実行されることはありませんが、例えばブラウザがInternet Explorerであれば実行の是非を確認するダイアログが表示されるため、誤って実行してしまう可能性が考えられます。このように実行ファイルを直接ダウンロードさせようとするケースでは、実行ファイルの配置場所として、一部のクラウドストレージサービスを利用しているものや、誘導元となったWebサイトと同じサーバを利用しているものなどがありました。

9月下旬からドライブバイダウンロード攻撃が急増しているため、ブラウザ利用環境ではOS、アプリケーションやプラグインのバージョン管理やEMET導入などの脆弱性対策の徹底しておくことを推奨します^{*59}。Webサイト運営者は利用しているWebアプリケーションやフレームワーク、プラグインの脆弱性管理による脆弱性対策、及び、広告や集計サービスなど外部から提供されるマッシュアップコンテンツの管理が必須です。



※調査対象は日本国内の数十万サイト。近年のExploitKitによるドライブバイダウンロードは、クライアントのシステム環境やセッション情報、送信元アドレスの属性、攻撃回数などのノルマ達成状況などによって攻撃内容や攻撃の有無が変わるよう設定されているため、試行環境や状況によって大きく異なる結果が得られる場合がある。
※詐欺サイトや実行ファイルへの直リンクなど、ExploitKit以外の受動的攻撃による脅威はOtherに分類している。

図-13 Webサイト閲覧時の受動的攻撃発生率(%)(Exploit Kit別)

*55 9月29日にRig EKの検知精度向上を目的としたWebクロウラの機能強化を実施した。9月29日以降のRig EK観測数急増の直接の原因はこの機能強化によるものと考えられる。ただしRig EKの観測数はこの機能強化の少し前から増加傾向にあった。また、例えばMalwarebytes Labs「RIG exploit kit takes on large malvertising campaign」(<https://blog.malwarebytes.com/cybercrime/exploits/2016/09/rig-exploit-kit-takes-on-large-malvertising-campaign/>)などでもRigの大規模な攻撃キャンペーンが紹介されている。このため、少なくとも8月中旬から9月中旬頃を起点としてRigによる攻撃が増加していたものと推測している。

*56 2016年9月末から10月中旬までのRig EKの観測状況については、IJ-SECT「Rig Exploit Kit観測数の拡大に関する注意喚起」(<https://sect.ij.ad.jp/d/2016/10/178746.html>)で速報している。

*57 MITF Webクロウラシステムでは、Windows環境のクライアントハニーポットで巡回した際に受動的攻撃の可能性を示す挙動が観測されたWebサイトに対して、Mac OS X環境のクライアントハニーポットによる追加調査を行っている。

*58 Potentially Unwanted Applicationの略。一般的な業務に不要と思われたり、用途によってはPCユーザやシステム管理者にとって不適切な結果を招く可能性があると考えられりするアプリケーションの総称。

*59 例えば管理者権限の分離やアプリケーションホワイトリストの適用などが考えられる。詳細は本レポートのVol.31 (<http://www.ij.ad.jp/company/development/report/iir/031.html>)の「1.4.3 マルウェアに感染しないためのWindowsクライアント要塞化」参照。

1.4 フォーカスリサーチ

インターネット上で発生するインシデントは、その種類や規模が時々刻々と変化しています。このため、IJでは、流行したインシデントについて独自の調査や解析を続けることで対策につなげています。ここでは、これまでに実施した調査のうち、Mirai Botnetの検知と対策、SSL/TLSにまつわるエトセトラの2つのテーマについて紹介します。

1.4.1 Mirai Botnetの検知と対策

■ Mirai Botnetとは

Mirai Botnet(以下、Mirai)は、ネットワーク接続可能なカメラやデジタルビデオレコーダーなどのIoT機器に感染し、ボットネットを構築するマルウェアです。IoT機器は単体では大きな処理能力は有していませんが、大量の機器が攻撃に利用された場合、非常に強力な攻撃を起こすことができます。多くのIoT機器は適切にセキュリティ管理がされておらず巨大なボットネットが形成されやすいと言えます。

2016年9月下旬に、アメリカのセキュリティ情報サイト「Krebs on Security」やフランスのインターネットサービスプロバ

イダであるOVHに対して大規模なDDoS攻撃が発生しました*60*61が、この攻撃にMiraiが使用されたと言われていきます。Krebs on Securityでは最大665Gbps、OVHでは最大1Tbpsという、かつて発生したことがない規模のDDoS攻撃を受けました。その後、10月上旬にAnna-senpaiと名乗るMiraiの作者が突如ソースコードを公開しました。現在、Anna-senpaiが公開したソースコードにはアクセスすることはできませんが、ソースコードは各所にミラーリングされ、誰でもアクセスすることができる状態にあります。

本稿では、公開されたソースコードを基にMiraiの動作を解説した上で、Miraiの検知や感染有無の判断方法と対策を解説します*62。

■ Miraiのシステム構成と動作

Miraiの最小のシステム構成は図-14のようになっています。IoT機器はx86以外のCPUを使用していることも多いため、MiraiではARMやMIPSのバイナリをクロスコンパイルするためのシェルスクリプトも付属しています。このシェルスクリプトにはいくつかのビルドオプションを指定することができ

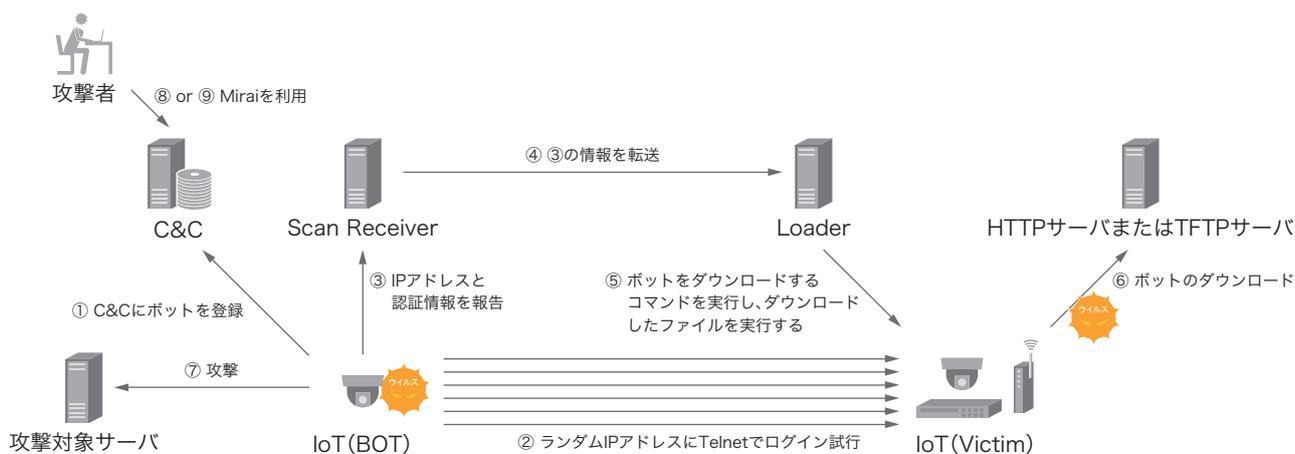


図-14 Mirai Botnetシステム構成

*60 Krebs on Security, "KrebsOnSecurity Hit With Record DDoS" (<https://krebsonsecurity.com/2016/09/krebsonsecurity-hit-with-record-ddos/>).

*61 OVHに対するDDoSの様子を伝えるツイート (<https://twitter.com/olesovhcom/status/779297257199964160>).

*62 本稿では次のリポジトリの10月末時点のソースコードを参照している。GitHub - jgamblin/Mirai-Source-Code: Leaked Mirai Source Code for Research/loC Development Purposes (<https://github.com/jgamblin/Mirai-Source-Code>).

ますが、「ssh」オプションは実装されていないため「telnet」オプションのみが意味を持ちます。また、Miraiに関連する通信はすべて平文で行われています。

■ IoT(Bot)

Miraiに感染し、ボットとなったIoT機器です。MiraiがIoT機器に感染すると、watchdog timerを無効にし、自動的にリブートしないようにします。また、ローカルホストの48101/tcpへの接続可否で他のMiraiボットのインスタンスが動作しているか判断します。動作している場合、該当するプロセスを終了させます。次に自身のプロセス名をランダムな文字列に変更します。ここまでの処理が成功すると、起動成功の目印として標準出力に「listening tun0」を出力します。後述する通信⑤の処理でも、この文字列をチェックすることでダウンロードしたボットの起動が成功したことを確認します。その後、22/tcp、23/tcp、80/tcpをバインドしているプロセスを終了させ^{*63}、Miraiがこれらのポートをバインドすることで管理インタフェースにアクセスできないようにします。他にも、実行パスに「.anime」を含むプロセス、実行ファイルが削除されているプロセス、Qbot/Zollard/Remaitenのプロセ

スなどを終了させます。このように管理系のプロセスやマルウェアなどを終了させることで、Miraiに感染したIoT機器が復旧させられたり、他のマルウェアに乗っ取られることを防ぎます。また、C&CサーバやScan Receiverのドメイン名やポート番号、他のIoT機器へのログイン試行で使用する認証情報などは、難読化のため、キーである0xdeadbeefを1バイトごとに分割した値を用いて、1バイトずつXORされています。

通信①: ボットにハードコーディングされたC&Cサーバのドメイン名とポート番号に接続します(デフォルト: cnc.changeme.com, 23/tcp)。TCPセッションが確立した後に、表-1のフォーマットのデータをC&Cサーバに送信すると、ボットがC&Cサーバに登録されます。また、60秒ごとに表-2のデータフォーマットでハートビートがボットから送信されます。一方、C&Cサーバからボットに対しては、必要に応じて表-3のデータフォーマットで攻撃命令が送信されます。このTCPセッションは永続的で切断されても自動的に再接続されます。

表-1 C&C接続時のデータフォーマット

パケット	データ長(バイト)	意味
1	4	0x00 0x00 0x00 0x01(固定バイト列)
2	1	パケット3で送るデータ長(*1)
3	(*1)で指定した長さ	ボット起動時の第1引数

表-2 C&Cへのハートビートデータフォーマット

データ長(バイト)	意味
2	0x00 0x00(固定バイト列) C&Cがハートビートを受信すると、C&Cからもハートビートが送信される。

表-3 攻撃命令データフォーマット

データ長(バイト)	意味
2	データ長(最大4096)
4	攻撃実行時間
1	攻撃ID
1	攻撃先の数(この後ろに指定した数だけ攻撃先IPアドレスとネットマスクの組み合わせが続く)
4	攻撃先IPアドレス
1	ネットマスク
1	フラグの数(この後ろに指定した数だけフラグの組み合わせが続く)
1	フラグID
1	フラグに指定するデータの長さ(*2)
(*2)で指定した長さ	フラグに指定するデータ(文字列)

*63 80/tcpは付属しているシェルスクリプトのコンパイルオプションでは終了しない。

通信②: ランダムなIPv4アドレスの23/tcpに対してSYNスキャンを行います。ただし、10回に1回の割合で2323/tcpに対するSYNスキャンを行います。これは感染対象機器の一部がTelnetを23/tcpではなく、2323/tcpで提供しているためであると考えられます。なお、IPアドレスが表-4のものと合致した場合、再度ランダムなIPアドレスを選択します。SYNスキャンに反応があったIPアドレスに接続後、ボットにハードコーディングされた認証情報を使って、ログインできるか確認します。認証情報は表-5の61種類ありますが、それぞれに重みづけがされているため、すべての認証情報が均等に使われるわけではありません。この通信は、Miraiに感染している間、発生し続けます。

通信③: ②でログイン試行に成功すると、ボットにハードコーディングされたScan Receiverに接続します

(デフォルト:report.changeme.com, 48101/tcp)。TCPセッション確立後、表-6のデータフォーマットで認証情報を送信します。

通信⑦: C&Cサーバから攻撃命令を受け取ると、指定された攻撃対象に攻撃を行います。攻撃方法の詳細については、表-7を参照してください。なお、HTTP floodに関しては、DOSarrestとCloudFlareを認識するコードが実装されています。DOSarrestについてはDDoS対策サービスに対抗するためと思われるコードが実装されていますが、CloudFlareについては認識をするのみで特に対抗するためのコードは実装されていません。更にソースコードには、「Proxy knockback connection」という実装予定と思われる攻撃の名前や、実装はされているものの呼び出しが行われない攻撃も存在します。

表-4 攻撃対象外IPアドレス

IPアドレス	割り当て先
127.0.0.0/8	Loopback
0.0.0.0/8	Invalid address space
3.0.0.0/8	General Electric Company
15.0.0.0/7	Hewlett-Packard Company
56.0.0.0/8	US Postal Service
10.0.0.0/8	Internal network
192.168.0.0/16	Internal network
172.16.0.0/14	Internal network
100.64.0.0/10	IANA NAT reserved
169.254.0.0/16	IANA NAT reserved
198.18.0.0/15	IANA Special use
224.*.*+	Multicast
第1オクテットが以下のIPアドレス: 6, 7, 11, 21, 22, 26, 28, 29, 30, 33, 55, 214, 215	Department of Defense

表-6 Scan Receiverに送信する認証情報のデータフォーマット

データ長(バイト)	意味
1	0x00(固定)
4	IPアドレス
2	ポート番号
1	ユーザ名の長さ(*3)
(*3)で指定した長さ	ユーザ名(文字列)
1	パスワードの長さ(*4)
(*4)で指定した長さ	パスワード(文字列)

表-5 ボットにハードコーディングされた認証情報

ユーザ名	パスワード	ユーザ名	パスワード
root	xc3511	admin1	password
root	vizxv	administrator	1234
root	admin	666666	666666
admin	admin	888888	888888
root	888888	ubnt	ubnt
root	xmhdipc	root	klv1234
root	default	root	Zte521
root	juantech	root	hi3518
root	123456	root	jvbsd
root	54321	root	anko
support	support	root	zlxx.
root	(none)	root	7ujMko0vizxv
admin	password	root	7ujMko0admin
root	root	root	system
root	12345	root	ikwb
user	user	root	dreambox
admin	(none)	root	user
root	pass	root	realtek
admin	admin1234	root	00000000
root	1111	admin	1111111
admin	smcadmin	admin	1234
admin	1111	admin	12345
root	666666	admin	54321
root	password	admin	123456
root	1234	admin	7ujMko0admin
root	klv123	admin	1234
Administrator	admin	admin	pass
service	service	admin	meinsm
supervisor	supervisor	tech	tech
guest	guest	mother	fXXXXr(一部伏字)
guest	12345		

■ Scan Receiver

48101/tcpでボットのスキャン結果を受け取るサーバです。受け取った情報をパースして標準出力に出力するだけの簡単な機能を持っています。この出力結果を後述するLoaderに読み込ませます。

処理④: ボットから受け取ったログイン可能なIoT機器の情報をファイルコピーするなどして、Loaderの標準入力に入力します。公開されたソースコードでは情報の転送を自動的に行う仕組みは用意されていません。

■ Loader

入力されたログイン可能なIoT機器の情報を基に、実際の感染活動を行うサーバです。起動後、大量のスレッドを作成し、複数のIoT機器を同時に感染させることができます。

通信⑤: Scan Receiverから受け取った情報を基にIoT機器にログインします。ログイン後、busyboxのwgetコマンドまたはtftpコマンドを使用して、ボットのバイナリをダウンロードして実行します。この時、IoT機器のCPUアーキテクチャに対応したボットのバイナリをダウンロードする必要がありますが、LoaderはIoT機器内の「/bin/echo」バイナリのELFヘッダを解析することでCPUアーキテクチャを判別します。wgetやtftpが使えない場合、IoT機器内の/bin/echoコマンドを使用して、ボットダウンロードを送りこんで実行します。感染に成功すると、「IoT(Bot)」と同様の動作を開始します。

通信⑥: Loaderにハードコーディングされた情報にしたがって、busyboxのwgetコマンドやtftpコマンドでボットのバイナリをダウンロードします(wgetデフォルト:100.200.100.100, 80/tcp, ファ

表-7 DDoS攻撃一覧

攻撃ID	コマンド	攻撃内容	攻撃詳細
0	udp	UDP flood	UDPパケットを大量に送り付ける。
1	vse	Valve source engine specific flood	Source Engine用のUDP Floodを行う。
2	dns	DNS resolver flood using the targets domain, input IP is ignored	指定したドメイン名に対してDNS水責め攻撃を行う。IoT機器にキャッシュDNS設定がされていない場合、以下のキャッシュDNSサーバを使う。8.8.8.8, 74.82.42.42, 64.6.64.6, 4.2.2.2
3	syn	SYN flood	SYNパケットを大量に送り付ける。
4	ack	ACK flood	ACKパケットを大量に送り付ける。
5	stomp	TCP stomp flood	DDoS対策機器をバイパスすることを意図した攻撃。TCPセッション確立後に、大量のACKパケットを送り付ける。
6	greip	GRE IP flood	GREでカプセル化したIP-UDPパケットを大量に送り付ける。
7	greeth	GRE Ethernet flood	GREでカプセル化したETH-IP-UDPパケットを大量に送り付ける。
8	なし	Proxy knockback connection	未実装のため、詳細不明。
9	udpplain	UDP flood with less options. optimized for higher PPS	設定項目を少なくし、高速化を図ったUDP Flood。
10	http	HTTP flood	HTTP GETなどのリクエストを大量に送り付ける。User-Agentは表-8からランダムに選ばれる。

表-8 HTTP floodに使用されるUser-Agent

User-Agent
Mozilla/5.0 (Windows NT 10.0; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/51.0.2704.103 Safari/537.36
Mozilla/5.0 (Windows NT 10.0; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/52.0.2743.116 Safari/537.36
Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/51.0.2704.103 Safari/537.36
Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/52.0.2743.116 Safari/537.36
Mozilla/5.0 (Macintosh; Intel Mac OS X 10_11_6) AppleWebKit/601.7.7 (KHTML, like Gecko) Version/9.1.2 Safari/601.7.7

イルパス:/bins/mirai.[arch]^{*64}) (tftpデフォルト: 100.200.100.100, ファイル名:mirai.[arch])。⑤で送り込まれたボットダウンロードも同様に、ハードコーディングされた接続先のHTTPサーバからボットをダウンロードします(デフォルト:127.0.0.1, 80/tcp, ファイルパス:/bins/mirai.[arch])。

■ C&C

Miraiボットネットを利用するために、管理者やユーザがログインするサーバです。ユーザのアカウントや攻撃実行履歴、攻撃先のホワイトリストがデータベースを使って管理されています。ユーザごとに利用可能なボットの数、最大攻撃継続時間、クールダウンタイムが設定されており、攻撃が終了してもクールダウンタイムで設定した時間が経過しないと次の攻撃を行うことができません。なお、攻撃の停止やボットの活動停止などのコマンドは用意されていません。

通信⑧: 管理者またはユーザはC&CサーバにTelnetで接続してボットネットを利用します。認証はユーザ名

とパスワードで行われます。ボットもC&Cサーバに23/tcpで接続しますが、23/tcpに接続直後のTelnet IACによって識別を行っています。使用できるコマンドは表-9を参照してください。攻撃コマンド例は図-15を参照してください。また、C&Cへのログイン時の一部のメッセージはロシア語になっており、作者またその関係者がロシア語圏の人物であることを窺わせませす(図-17)。

通信⑨: 管理者またはユーザはTelnetの代わりに、101/tcpに接続することでAPI経由でボットネットを使用することもできます。攻撃コマンド例は図-16を参照してください。

■ 感染試行と感染有無の判断

■ ファイアウォールログ

ファイアウォール外部から内部に対して、23/tcpと2323/tcpへのアクセス数がおおよそ9対1になっているようであれば、Miraiの感染活動に晒されていると考えられます。送信先IPアドレスはランダムに選択されるため、一部を除きほとんどの組織

表-9 C&Cサーバで使用できるコマンド

コマンド	詳細
adduser	ユーザを追加する。管理者のみ実行可能。
botcount	C&Cサーバに接続しているボットの数を表示する。管理者のみ実行可能。
-<BotCount>	攻撃に使用するボットの数指定する。
@<string>	攻撃に使用するボットのカテゴリ指定する。
?	ヘルプを表示する。
source	攻撃のフラグとしてのみ指定可能。送信元IPアドレスを指定する。管理者のみ実行可能。

```
[-<BotCount>] <atk cmd> <ip_addr>[/<mask>].<ip_addr>[/<mask>]]...<br><duration> [<flags>]
```

図-15 攻撃コマンド例(CLI)

```
<API Key>[[-<BotCount>] <atk cmd> <ip_addr>[/<mask>].<ip_addr>[/<mask>]]...<br><duration> [<flags>]
```

図-16 攻撃コマンド例(API)

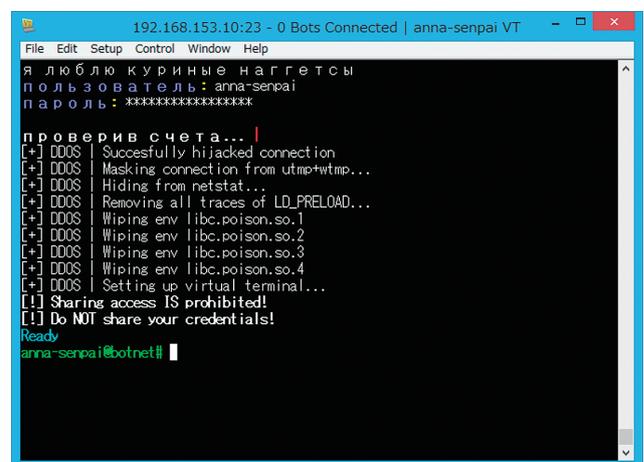


図-17 C&Cログイン画面

*64 [arch]はCPUアーキテクチャを表す。

が攻撃対象となります。逆にファイアウォール内部から外部に対して、23/tcpと2323/tcpへのアクセス数がおおよそ9対1になっているのであれば、内部の機器がMiraiに感染しているのはほぼ間違いないと考えられます。スキャン時の送信元IPアドレスは詐称されないため、感染している機器は簡単に判別することが可能です。

■ IDS/IPS

Miraiのシステム間の通信はすべて平文で行われているため、IDS/IPSで監視することができます。図-18に、Miraiの通信を検知するSnortシグネチャの一例を示します。公開されているソースコードを基にしているため、新バージョンのMiraiや亜種の通信は検知できない可能性があります。

■ 対策

■ IoT機器メーカー

IoT機器のセキュリティを考える上で、メーカーの役割は非常に重要です。Miraiに関する一連の事件を通して、多くのニュースやレポートでは対策として、IoT機器のパスワードを変更する

ように伝えていますが、確かに、非常に大量のIoT機器が感染していることの直接的な原因の多くはユーザがパスワードをデフォルトから変更していないことですが、一部の機種においてはパスワードがハードコーディングされており、変更はもちろん、アカウントの無効化などもできない状態であったことが伝えられています。このような状況では、ユーザ側で対応を行うことができないため、メーカー側に責任があると考えべきではないでしょうか。メーカーは以下のような観点でセキュリティ設計の見直しを行うべきでしょう。

- ・ バックドアアカウントを作成しない。
- ・ パスワードをハードコーディングしない。
- ・ パスワードに使用できる文字種を極端に制限しない。
- ・ すべての管理インタフェースをマニュアルに明記する（バックドアを実装しない）。
- ・ TelnetやHTTPなど平文での通信は避ける。
- ・ ユーザが最初にログインしたときにパスワードを変更させ、デフォルトパスワードは使えないようにする。

・ボット登録とハートビート

```
alert tcp any any -> any 23 (msg:"Mirai Botnet: Register Bot with C&C"; flow:to_server,established; content:"|00 00 00 01|"; depth:4; sid:1000000; rev:1)
alert tcp any any -> any 23 (msg:"Mirai Botnet: Send Heartbeat from Bot to C&C"; flow:to_server,established; content:"|00 00|"; depth:2; pcre:"/\x00\x00$/m"; sid:1000001; rev:1)
alert tcp any 23 -> any any (msg:"Mirai Botnet: Reply Heartbeat from C&C to Bot"; flow:from_server,established; content:"|00 00|"; depth:2; pcre:"/\x00\x00$/m"; sid:1000002; rev:1)
```

・ボットダウンローダのダウンロード

```
alert tcp any any -> any [23,2323] (msg:"Mirai Botnet: Download Bot Downloader via Telnet (echo)"; flow:to_server,established; content:"echo -ne "; content:"> upnp|3b| /bin/busybox ECCHI"; sid:1000060; rev:1)
```

・ボットバイナリのダウンロードコマンド実行

```
alert tcp any any -> any [23,2323] (msg:"Mirai Botnet: Download Bot binary via Telnet (wget)"; flow:to_server,established; content:"/bin/busybox wget http://"; content:"/bins/mirai."; content:"-O -> dvrHelper|3b| /bin/busybox chmod 777 dvrHelper|3b| /bin/busybox ECCHI"; sid:1000070; rev:1)
alert tcp any any -> any [23,2323] (msg:"Mirai Botnet: Download Bot binary via Telnet (tftp)"; flow:to_server,established; content:"/bin/busybox tftp "; content:"-g -l dvrHelper -r mirai."; content:"/bin/busybox chmod 777 dvrHelper|3b| /bin/busybox ECCHI"; sid:1000071; rev:1)
```

・ボットバイナリのダウンロード通信

```
alert tcp any any -> any 80 (msg:"Mirai Botnet: Download Bot binary via HTTP"; flow:to_server,established; content:"GET /bins/mirai."; pcre:"/^GET /bins/mirai\.(\arm|arm7|m68k|mips|mips|ppc|sh4|spc|x86) HTTP/1\.[01]|0d 0a|$/mi"; sid:1000080; rev:1)
alert udp any any -> any 69 (msg:"Mirai Botnet: Download Bot binary via TFTP"; flow:to_server; content:"|00 01|mirai."; pcre:"/\x00\x01 mirai\.(\arm|arm7|m68k|mips|mips|ppc|sh4|spc|x86)\x00.+$/mi"; sid:1000081; rev:1)
```

・ボット実行

```
alert tcp any any -> any [23,2323] (msg:"Mirai Botnet: Run Bot binary (upnp & dvrHelper)"; flow:to_server,established; content:"./upnp|3b| ./dvrHelper telnet."; content:"/bin/busybox IHCCE"; pcre:"/\^.\upnp\.; \.\dvrHelper telnet\.(\arm|arm7|m68k|mips|mips|ppc|sh4|spc|x86)\.; \bin\busybox IHCCE/m"; sid:1000090; rev:1)
alert tcp any any -> any [23,2323] (msg:"Mirai Botnet: Run Bot binary (dvrHelper)"; flow:to_server,established; content:"./dvrHelper telnet."; content:"/bin/busybox IHCCE"; pcre:"/\^.\dvrHelper telnet\.(\arm|arm7|m68k|mips|mips|ppc|sh4|spc|x86)\.; \bin\busybox IHCCE/m"; sid:1000091; rev:1)
```

図-18 Snort用シグネチャ

■ ユーザ

Miraiはファイルを残さず、メモリ上のみで動作するため、IoT機器を再起動すれば感染から復旧させることができます。しかし、デフォルトパスワードのままではすぐに再感染してしまうため、パスワードを変更する必要があります。

IoT機器に限りませんが、パスワードはデフォルトから必ず変更しなければなりません。デフォルトの認証情報はマニュアルなどに記載されているため、攻撃者には既知の情報になっていると考えるべきです。そのため、デフォルトパスワードのままでは認証がない状態と同等であると言っても過言ではありません。パスワードを変更する場合、よく言われるように、可能な限り複雑で長いパスワードを設定することが望ましいでしょう。また、必要がないのであればIoT機器を直接インターネットに接続してはいけません。インターネットからアクセスする必要がある場合、IoT機器自身やファイアウォールなどで適切にアクセス制限を行ってください。

1.4.2 SSL/TLSにまつわるエトセトラ

ここ数年、SSL/TLSプロトコルやその実装に対する新たな攻撃が多く公開されただけでなく、暗号アルゴリズムの危殆化による移行要請、更に新バージョンであるTLS1.3がほぼ完結に向かっているなど大きな動向変化が見られます。そこで本節ではSSL/TLSに関する動向変化を報告すると共に、IoT時代に向けた課題について取り上げます。

■ SSL/TLS バージョン推移の経緯

Netscape Communicationsから1995年に公開されたSSL2.0はいくつかの拡張と問題点が修正され、SSL3.0^{*65}として翌年公開されています。SSL2.0はHandshakeメッセージ部分に改ざん防止機能を有しない(データ完全性を保証していない)ため中間者攻撃が可能であり、プロトコルそのものが脆弱であると認識されています^{*66}。またSSL3.0は2014年10月に発覚したPOODLE攻撃^{*67}によりメッセージの暗号化にCBC暗号モードを利用した場合にPadding Oracle攻撃が可能になることが分かり、現在は利用しないことが推奨されています^{*68}。

SSLの後継であるTLSは現在3つのバージョン:TLS1.0(1999年策定)、TLS1.1(2006年策定)、TLS1.2(2008年策定)があり、いずれも未だに広く利用されているプロトコルです。TLS1.0がSSL3.0をベースにIETFで策定された後、TLS1.1にてCBC暗号モード利用時に露呈するBEAST攻撃やその亜種への対策などを予め仕様に組み入れるなどの安全性強化を図っています。更にTLS1.2は認証暗号(AEAD: Authenticated Encryption with Associated Data)^{*69}の利用が可能となりました。しかしこれらのプロトコルに対しここ数年で多くの攻撃がなされてきました。2015年2月に発行されたRFC7457^{*70}は2014年頃までに公知となったTLSに対する攻撃の歴史がまとめられています。次に紹介するRC4ストリーム暗号に関する脆弱性が取り上げられているほか、利用者の想定よりも低

*65 1996年にSSL3.0が公開された時点ではIETF主導で策定されている仕様ではなかったが、RFC6101として史料という位置づけでRFC化されている(<https://datatracker.ietf.org/doc/rfc6101/>)。

*66 RFC6176: Prohibiting Secure Sockets Layer (SSL) Version 2.0(<https://datatracker.ietf.org/doc/rfc6176/>)。

*67 POODLE攻撃に関する解説は2014年11月発行の本レポートVol.25(<http://www.ij.ad.jp/company/development/report/iir/025.html>)の「1.4.2 POODLE attack」にて紹介している。

*68 RFC7568: Deprecating Secure Sockets Layer Version 3.0(<https://datatracker.ietf.org/doc/rfc7568/>)。

*69 RFC5116: An Interface and Algorithms for Authenticated Encryption(<https://datatracker.ietf.org/doc/rfc5116/>)。

*70 RFC7457: Summarizing Known Attacks on Transport Layer Security (TLS) and Datagram TLS (DTLS)(<https://datatracker.ietf.org/doc/rfc7457/>)。前年の2014年3月にはCRYPTRECからCRYPTREC暗号技術ガイドライン(SSL/TLSにおける近年の攻撃への対応)(http://www.cryptrec.go.jp/report/c13_kentou_giji02_r2.pdf)が発行されており、RFC7457に列挙されている攻撃の一部について参照できる。

いTLSバージョンを使用させるダウングレード攻撃や、圧縮機能を有効にしている場合に起こるタイミング攻撃など多岐にわたる既知の攻撃が取り上げられています。また、それ以降についてのSSL/TLSの主な脆弱性のポータルサイトとしてはCELLOS^{*71}などで参照できますが、それぞれの攻撃に対して知識を積み上げ、その都度対処していくことは大変難しい状況になりました。

一例として暗号アルゴリズムのRC4とTripleDESに起こった事例を見ていくことにしましょう。RC4はSSL/TLSにおいてCipherSuitesが定義され、これまで多く利用されてきたストリーム暗号の代表格です。暗号アルゴリズムへの攻撃としては様々な攻撃モデルが考えられますが、SSL/TLSのような暗号プロトコルにおいて現実的な状況としてBroadcast settingと呼ばれる条件があります。これは同じ平文(暗号化される前のデータ)に対して複数の異なる鍵で暗号化された大量の暗号文が入手できる、という仮定でありSSL/TLSの利用形態を考えると十分に想定可能な条件です。この攻撃モデルのもと2001年から多くの暗号解読研究がなされており^{*72}、RC4が生成するストリーム鍵の僅かな不均衡(バイアス)から平文を復元する攻撃が公開されています。具体的にはブラウザ上で不正なJavaScriptを動かせることによって大量の暗号文を生成するという手法が述べられており、USENIX Security 2015で発表された論文によると 9×2^{27} の暗号文を入手することで

94%の確率でcookieを搾取可能と報告されています^{*73}。IETFとしてもアカデミアによる複数の研究結果を鑑み、現実的な脅威として捉え2015年2月にRC4を排除する旨のRFCが発行されています^{*74}。

一方でTripleDESが脆弱であると再認識されたSWEET32攻撃^{*75}は、暗号アルゴリズムそのものに対する攻撃手法ではなく、SSL/TLSにてCBC暗号モードを利用する際に起こり得る潜在的な攻撃であり、完全に防ぐことはできないものとなっており、各ベンダーの対策もTripleDESの利用を制限もしくは格下げするというものでした。その一方で注意したいのはその攻撃が成功に至るまでに必要なリソースです。ACM CCS'16で発表された論文によると、2ブロック分のCookieを復元するためには785ギガバイトの暗号文をおよそ38時間もの間キャプチャし続ける必要があると報告されています^{*76}。RC4バイアス攻撃はストリーム暗号RC4そのものに対する攻撃ですが、SWEET32攻撃は64ビットブロック暗号への攻撃ではあるもののCBC暗号モードの利用が必須であり、暗号アルゴリズムそのものの攻撃とは言い難い一面もあります。SWEET32攻撃が登場した際に、RC4を排除するRFC7465と同様にTripleDESの排除に関するインターネットドラフト^{*77}が掘り起こされ、これを策定すべきかどうかについてCFRG(Crypto Forum Research Group)にて議論がありましたが、RFC化するには至りませんでした。

*71 CELLOS consortium, Publication (<https://www.cellos-consortium.org/index.php?Publication>)。)

*72 2001年から研究されているRC4バイアス攻撃については以下にまとめられている。Kenneth G. Paterson, "Big Bias Hunting in Amazonia: Large-scale Computation and Exploitation of RC4 Biases", ASIACRYPT2014 Invited Talk (http://des.cse.nsysu.edu.tw/asiacrypt2014/doc/8_1_Big%20Bias%20Hunting%20in%20Amazonia%20Large-scale%20Computation%20and%20Exploitation%20of%20RC4%20Biases.pdf)。)

*73 Mathy Vanhoef, Frank Piessens, "All Your Biases Belong To Us: Breaking RC4 in WPA-TKIP and TLS" (<https://www.rc4nomore.com/vanhoef-usenix2015.pdf>)。RC4 NOMORE (<https://www.rc4nomore.com/>)のサイトでもサマリーを参照可能である。

*74 RFC7465: Prohibiting RC4 Cipher Suites (<https://datatracker.ietf.org/doc/rfc7465/>)。)

*75 Sweet32: Birthday attacks on 64-bit block ciphers in TLS and OpenVPN (<https://sweet32.info/>)。)

*76 Karthikeyan Bhargavan and Gaëtan Leurent, "On the Practical (In-)Security of 64-bit Block Ciphers: Collision Attacks on HTTP over TLS and OpenVPN", ACM CCS'16 (<http://dl.acm.org/citation.cfm?id=2978423&CFID=697886415&CFTOKEN=82935453>)。)

*77 B. Kaduk et al., Deprecate 3DES and RC4 in Kerberos (<https://www.ietf.org/archive/id/draft-kaduk-kitten-des-des-des-die-die-die-00.txt>)。)

更に興味深いことをSWEET32攻撃は示唆しています。TripleDESもCBC暗号モードも未だにCRYPTREC暗号リストのうち電子政府推奨暗号リスト^{*78}に掲載されながらも、この「安全な」2つのプリミティブを同時に利用したために脆弱になるという状況を引き起こしてしまいました。この事例は暗号アルゴリズムのミスユースとまでは言い切れませんが、このようなミスユースを想定した暗号アルゴリズムも想定されておりAEADのコンペティションであるCAESAR^{*79}では、毎回異なるノンスを利用することが想定されているケースにおいて、誤って同じノンスを利用したとしても安全であることをメリットとするNonce Misuse-Resistantと呼ばれる性質を持つアルゴリズムも提案されています。そのほかのミスユース事例としてはDSA署名において同じパラメータを使って署名を行った場合に秘密鍵が漏えいすることが知られています⁸⁰。

以上のような暗号アルゴリズムのレイヤでの直接的な復元攻撃とは異なり、タイミング攻撃・サイドチャネル攻撃などにカテゴライズされる手法も進展があり、結果としてTLS1.0及びTLS1.1では使用できないAEADに対応するためTLS1.2への移行が望まれています。これはCBC暗号モードを利用していたとしてもTLS1.1以上のバージョンを利用することでBEAST攻撃やその亜種への対策が可能と考えられてきた一方で、TLS1.2に対してもタイミング攻撃が可能なLucky13^{*81}攻撃が登場したこともAEAD利用に拍車をかけることとなりました。更に、前述したようにRC4も使えないことからAEADへの移行が必要と認識されるようになりました。しかし各バージョンには実装必須のCipherSuitesがあることも考慮する必要があります。TLS1.0においては、TLS_DHE_

DSS_WITH_3DES_EDE_CBC_SHA、TLS1.1ではTLS_RSA_WITH_3DES_EDE_CBC_SHAが実装必須(Mandatory)のCipherSuitesとして挙げられています。共に共通鍵暗号としてTripleDESをCBC暗号モードで利用するアルゴリズムが選択されています。またTLS1.2ではTLS_RSA_WITH_AES_128_CBC_SHAが実装必須であり、AES利用となり、より安全なアルゴリズムにシフトしているものの、すべてのTLSバージョンにおいてCBC暗号モードの利用が必須となっています。これはRFC策定当時、CBC暗号モード利用時のパディングオラクル攻撃を想定していなかったと考えることができます。そのため、クライアント・サーバ共にCipherSuitesとしてCBC暗号モード利用のCipherSuitesを選択しない状況に置くことが必要です。特にサーバ側においては、選択されるべきCipherSuitesの優先度を正しく設定することが必要となります。また、CipherSuitesの選択としては公開鍵暗号アルゴリズムとしてForward Secrecy^{*82}対応のアルゴリズムを選ぶことが望まれている点も考慮すべきです。更にExport-grade暗号の設定についても注意が必要です。サーバ側のCipherSuites選択時により強い暗号アルゴリズムが選ばれることが通常の動作であるため、かつて使われていた弱いCipherSuitesを設定上残しておいたとしてもそれらは使われることはないだろうと考えられてきました。しかし2015年1月のFREAK攻撃や同年5月のLogjam攻撃の発表^{*83}によりExport-gradeな暗号アルゴリズムを設定しているがために起こってしまう攻撃が発覚しました。更に2016年3月にはDROWN攻撃^{*84}が公開され、Export-gradeな暗号アルゴリズムを利用しない環境においてもSSL2.0を有効にしてしまっている状況下で暗号文の復元攻撃が可能であることが分かりました。

*78 CRYPTREC、電子政府における調達のために参照すべき暗号のリスト(CRYPTREC暗号リスト)(<https://www.cryptrec.go.jp/list.html>)。ただしTripleDESについては「当面の利用を認める」という位置づけである。これを安全であると判断するかどうかは読み手に依る。

*79 CAESAR: Competition for Authenticated Encryption: Security, Applicability, and Robustness(<https://competitions.cr.yp.to/caesar.html>)。

*80 RFC6979: Deterministic Usage of the Digital Signature Algorithm (DSA) and Elliptic Curve Digital Signature Algorithm (ECDSA) (<https://datatracker.ietf.org/doc/rfc6979/>)。

*81 Nadhem AlFardan, Kenny Paterson, "Lucky Thirteen: Breaking the TLS and DTLS Record Protocols"(<http://www.isg.rhul.ac.uk/tls/Lucky13.html>)。

*82 Forward Secrecyに関する解説は2014年2月発行の本レポートVol.22(<http://www.iij.ad.jp/company/development/report/iir/022.html>)の「1.4.2 Forward Secrecy」にて紹介している。

*83 須賀、暗号と社会の素敵な出会い: 2. SSL/TLSと暗号プロトコルの安全性 -恒久的に噴出する脆弱性ととの戦い-、会誌「情報処理」Vol.56 No.11(<http://id.nii.ac.jp/1001/00145437/>)。

*84 The DROWN Attack(<https://drownattack.com/>)。

■ サーバが満たすべき設定基準の例

前パートにてすべてのSSL/TLSへの攻撃を加味してその都度対処していくことは大変難しい状況であることを述べました。本パートではWebサーバ管理者がどのような判断基準を持ってWeb設定すべきか記載されたいくつかのドキュメントを紹介したいと思います。その1つとして、2015年5月にCRYPTRECから発行されたSSL/TLSサイトの暗号技術に関わる設定ガイドライン^{*85}があります。プロトコルバージョン・サーバ証明書・CipherSuitesについて具体的にSSL/TLSサーバに要求される設定基準が示されています。このガイドラインは政府システムだけに特化して記載されているものではないため、民間で利用する一般的なWebサーバにおける設定改善のための参考書としても参照することができます。しかし公開後2年弱たったこともあり、現状とはそぐわない箇所も散見されるため、改訂や補遺が望まれます。更に各種アプライアンス製品の初期設定において、または手動で設定した場合にこのガイドラインの要件を満たすことができるかなどの詳細なレポートも発行され^{*86}、より現実的な対策方法を入手できるようになってきました。そのほか日本語でも参照できる暗号プロトコルの安全性評価に関しては、例えば非営利団体であるCELLOS(暗号プロトコル評価技術コンソーシアム)からも公開されてきましたが、CRYPTRECの重点課題検討タスクフォースで2015年度検討されました。その結果、2016年度から暗号プロトコル課題検討WG^{*87}が設置され、議論が行われています。

SSL/TLSブラウザベンダーからもWebサーバの設定に関する情報が集約されています^{*88}^{*89}。各論についてはここでは触れませんが、CipherSuitesの対応だけでなく、常時SSL/TLS化の

要となるHSTS(HTTP Strict Transport Security)などについても記載されています。また、SSL/TLSサーバのテストサイトを提供しているQualys SSL Labでも設定のベストプラクティスに関する文書^{*90}が提供されており、詳しい評価基準^{*91}についても記載されているため、レイティングの理由だけでなくサーバ設定の改善点を知ることができます。一般ユーザでも当該FQDNを入力するだけで評価結果を参照することができますため、SSL/TLSサーバの設定不備が公知となってしまいます。そのため、レイティング結果が芳しくない場合には、ただ単に設定ミスなのか、一部のリスクを許容して設定しているのか区別がつかないという課題もあります。弱い暗号アルゴリズムの移行を促進している点を考えると大変よい活動であると考えられています。

■ ブラウザベンダーの対応状況の変化

Webサイトのレイティングという観点では、ブラウザからも容易にサーバの状況を知ることができるようになりました。その1つがURLを記載するエリアにあるsecurity indicatorであり、例えばEV SSL証明書利用サーバにアクセスすると緑のバーで表記されるエリアになります。Chromeではこのsecurity indicatorに関してバージョン52(Macintoshデスクトップのみ。それ以外はバージョン53)から表記方法に変化がありました。これは2016年6月に開催されたユーザビリティセキュリティを扱う国際会議で発表が行われ^{*92}、この結果に基づいて改善されたインタフェースが導入されています。この論文ではSSL/TLS接続状況をsecurity indicatorを通じて表示するアイコンについていくつかのバリエーションを被験者に提示した上で、よりよいものを選択するというアプ

*85 CRYPTREC, SSL/TLS暗号設定ガイドライン～安全なウェブサイトのために(暗号設定対策編)～(https://www.ipa.go.jp/security/vuln/ssl_crypto_config.html)。

*86 情報処理推進機構、「SSL/TLSアプライアンス製品の暗号設定方法等の調査報告書」の公開(http://www.ipa.go.jp/security/fy28/reports/crypto_survey/)。

*87 CRYPTREC, CRYPTREC Report 2015(https://www.cryptrec.go.jp/report/c15_prom_web.pdf)。

*88 Google Developers - Web Fundamentals, "Enabling HTTPS on Your Servers"(<https://developers.google.com/web/fundamentals/security/encrypt-in-transit/enable-https/>)。

*89 Mozilla, "Security/Server Side TLS"(https://wiki.mozilla.org/Security/Server_Side_TLS)。

*90 Qualys SSL Lab, "SSL and TLS Deployment Best Practices Version 1.5 (8 June 2016)"(<https://github.com/ssllabs/research/wiki/SSL-and-TLS-Deployment-Best-Practices>)。

*91 Qualys SSL Lab, "SSL Server Rating Guide"(<https://www.ssllabs.com/projects/rating-guide/>)。執筆時の最新版はversion 2009k(14 October 2015)(https://www.ssllabs.com/downloads/SSL_Server_Rating_Guide.pdf)であるが2017年以降に更新されることが2016年11月にアナウンスされている: Announcing SSL Labs Grading Changes for 2017(<https://blog.qualys.com/ssllabs/2016/11/16/announcing-ssl-labs-grading-changes-for-2017>)。

*92 Adrienne Porter Felt et al., "Rethinking Connection Security Indicators", SOUPS2016(<http://research.google.com/pubs/pub45366.html>)。論文はUSENIXのサイトで参照可能(<https://www.usenix.org/system/files/conference/soups2016/soups2016-paper-porter-felt.pdf>)。

ローチを取っており、実際にChromeに実装されています。アイコン表示にはコネクションやサーバの信頼性に関する意味を持たせており、EV SSL証明書を持つ正しいHTTPS通信だけでなく、軽微なエラーを含むHTTPS通信、重大なエラーを含むHTTPS通信、などに応じて異なるアイコンが利用されています。このうち「軽微なエラーを含むHTTPS通信」については、HTMLコンテンツ内にHTTPで指し示されたコンテンツを有するMixed contents (HTTPコンテンツとHTTPSコンテンツの混在)という状況で表示されています。ブラウザベンダーはMixed contentsを起こしているサイトに改善を求めており^{*93}^{*94}、アップデート前のアイコンはニュートラルな印象を与えるアイコンでしたが、先の論文においてとても重大ではないがネガティブな印象を与えるアイコンを選択してユーザーに注意を意識させています。このブラウザベンダーの変更にWeb管理者が対応できていないために、Mixed contentsエラーの出るHTTPSコンテンツを意図せず発信してしまっているSSL/TLSサーバが多く散見されており、今一度確認されることをお勧めします。

OSやアプリケーションが保有・管理し、信頼すべきルート証明書や中間CA証明書のリストを証明書ストア(Certificate store)と表現します。ブラウザなどのSSL/TLSクライアントはこの証明書ストアを参照し、当該Webサーバが信頼できるかどうかを判断してアプリケーション側でその状態を表示することが通例です。証明書ストアに格納されている証明書の削除・追加は可能なことが多いですが、通常利用時にはユーザーは証明書を意識することはありません。そのため、SSL/TLSサーバはベンダーの証明書ストアの状況を把握して証明書を入手する必要があります。これは各OSやアプリケーションで証明書ストアに含まれている証明書群が微妙に異なっており、同じ証明書でもある証明書ストアにおいては安全と判断される

一方、他の証明書ストアではルートが辿れず証明書検証に失敗する事例が存在します。証明書を発行する認証機関においては、ベンダーの基準を満たすことができなかったために廃業に追い込まれた業者も存在しています。このようにトラストアンカーを創り上げる組織体に強く依存したモデルになっており、特にブラウザベンダーの動向が注目される事態を招いています。

証明書検証という観点ではもう1つ大きな問題があります。Androidにおいては証明書ストアにまつわる処理を正しく行わず、証明書検証時に検証処理を端折るアプリケーションの脆弱性がここ数年、大変多く報告されています。この例のように、利用ユーザーはPKIを意識することなくアプリケーションのユーザインタフェースのみを閲覧するだけで当該通信の状況を把握することになっています。実際、先に見たように主要ブラウザベンダーの努力により利用ユーザーはsecurity indicatorを通して通信状況を把握することができるようになっており、その他のチャネルで調査を行うことなく、ブラウザの表記を完全に信頼することが多くなっていると考えられます。SSL/TLSクライアントの1つであるWebブラウザは、通常のPCで閲覧するだけでなくタブレットやスマートフォンなど、より小さい表示領域しか持たないデバイスにおいて危険な通信であるかどうかをすぐに判断できるインタフェースを持つ必要があります。特に今後、表示能力が非力なIoT製品におけるsecurity indicatorについても考慮が必要です。

■ TLS1.3

TLS1.2の次期バージョンTLS1.3^{*95}については2016年11月のIETF meeting 97^{*96}にてTLS1.3ドラフトのTLS WGラストコールがアナウンスされ2017年2月のRFC発行に向けて佳境を迎えています。TLS1.3では特に実装必須のCipherSuitesは指定されてはいませんが、従来利用されていたCBCモード

*93 Google Developers - Web Fundamentals. Preventing Mixed Content (<https://developers.google.com/web/fundamentals/security/prevent-mixed-content/fixing-mixed-content>).

*94 Mozilla support. Mixed content blocking in Firefox (<https://support.mozilla.org/en-US/kb/mixed-content-blocking-firefox>).

*95 The Transport Layer Security (TLS) Protocol Version 1.3 (<https://tswg.github.io/tls13-spec/>). IETFで管理されているInternet-Draftは執筆時点でVersion 18 (<https://datatracker.ietf.org/doc/draft-ietf-tls-tls13/18/>)である。

*96 Eric Rescorla. TLS 1.3 (draft-ietf-tls-tls13-18) (<https://www.ietf.org/proceedings/97/slides/slides-97-tls-tls-13-00.pdf>).

の利用は想定されておらず暗号化とMAC(データ改ざんがなされていないことを示すデータ)を同時に付与するAEADのみがCipherSuitesとして列挙されています。AEADとしてはAEAD_AES_128_GCM、AEAD_AES_256_GCM^{*97}、AEAD_AES_128_CCM^{*98}、AEAD_CHACHA20_POLY1305^{*99}、AEAD_AES_128_CCM_8の5つがそれに該当し、CBC暗号モード利用のブロック暗号やストリーム暗号の代替アルゴリズムとして利用されます。TLS1.3は、それ以前のバージョンで生成されていたMAC(メッセージ認証用データ)用の鍵データを導出するインターフェースがないため、実質的にはAEADの利用だけに制限されることとなります。

IETF meeting 97にて相互接続性結果が報告されたようにTLS1.3実装がブラウザ・サーバ共に進められており、一部の製品においてはユーザが手にとって確認することができるようになりました^{*100*101}。TLS1.2とは互換性がないことからバージョンのメジャーアップデート(TLS2.0、TLS2、TLS4)も

提案されていましたが、会合ではTLS1.3支持派が多かったため、このままTLS1.3として普及されるでしょう。更に来年RFC化されることで、多くの実装がTLS1.3に対応すると考えられます。一方で可搬性のあるデータ形式でブラウザ間を持ち歩くような利用状況も想定されていることから、この機能を突いた攻撃が発表されるかもしれません。そのほか、TLSへの耐量子暗号^{*102}の適用も話題となりました^{*103}。今後もレガシーデバイスにおけるバージョン移行の問題と新機能追加という正反対の動きで混沌とした状況が続いていくものと考えられます。

1.5 おわりに

このレポートは、IJJが対応を行ったインシデントについてまとめたものです。今回は、Mirai Botnetの検知と対策、SSL/TLSにまつわるエトセトラについて紹介しました。IJJでは、このレポートのようにインシデントとその対応について明らかにして公開していくことで、インターネット利用の危険な側面を伝えるように努力しています。



執筆者:
齋藤 衛 (さいとう まもる)

IJJ セキュリティ本部 本部長、セキュリティ情報統括室 室長兼務。法人向けセキュリティサービス開発などに従事後、2001年よりIJJグループの緊急対応チームIJJ-SECTの代表として活動し、CSIRTの国際団体であるFIRSTに加盟。ICT-ISAC Japan、日本セキュリティオペレーション事業者協議会など、複数の団体の運営委員を務める。

根岸 征史 (1.2 インシデントサマリ)

小林 直、永尾 慎啓、鈴木 博志、小林 稔、梨和 久雄 (1.3 インシデントサーベイ)

小林 稔 (1.4.1 Mirai Botnetの検知と対策)

須賀 祐治 (1.4.2 SSL/TLSにまつわるエトセトラ)

IJJ セキュリティ本部 セキュリティ情報統括室

協力:

桃井 康成、平松 弘行 IJJ セキュリティ本部 セキュリティ情報統括室

*97 RFC5288: AES Galois Counter Mode (GCM) Cipher Suites for TLS (<https://datatracker.ietf.org/doc/rfc5288>).

*98 RFC6655: AES-CCM Cipher Suites for Transport Layer Security (TLS) (<https://datatracker.ietf.org/doc/rfc6655>).

*99 RFC7539: ChaCha20 and Poly1305 for IETF Protocols (<https://datatracker.ietf.org/doc/rfc7539/>).

*100 Chrome Platform Status, TLS 1.3 (<https://www.chromestatus.com/feature/5712755738804224>).

*101 CloudFlare, Introducing TLS 1.3 (<https://blog.cloudflare.com/introducing-tls-1-3/>).

*102 耐量子暗号に関する解説は2016年6月発行の本レポートVol.31 (<http://www.ijj.ad.jp/company/development/report/iir/031.html>)の「1.4.3 耐量子暗号の動向」にて紹介している。

*103 Chrome Platform Status, "CECPQ1 in TLS" (<https://www.chromestatus.com/feature/5749214348836864>).

業界が挑むストリーミング形式の統合

ストリーミング配信を構成する技術は、いくつかの階層に分類することができます(表-1)。これらの技術は様々な会社から提案され、その多くは国際標準規格となっています。各々の技術を組み合わせて最適なストリーミング配信を達成するには、規格のすり合わせが必要です。現在、この中の「ストリーミング形式」に大きな動きが出てきています。

Appleが提唱しているHTTP Live Streaming(HLS)は2009年に発表されました。以降現在に至るまで、ストリーミング形式の一翼を担う重要なフォーマットです。コンテナにセグメント化されたMPEG2-TSを採用し、映像・音声データを収納した細かいファイルを連続的に配信することが特色です。また、配送プロトコルにHTTP/1.1を採用したのも大きなインパクトがありました。HLSは当初iOSを対象としたフォーマットでしたが、その後macOSやtvOS、Androidでも広く使われるようになってきました。

AppleはHLSの規格案をインターネットの標準化を推進するInternet Engineering Task Force(IETF)へ対し、Internet-Drafts(I-D)として提出しています。文書名は"draft-pantos-http-live-streaming"です。I-Dは作業中の文章という位置付けですが、RFCとして完成するものもあれば、提案だけで終わるものもあります。Appleは2009年5月1日に発行された初版以来、Roger Pantos氏の名前でこのI-Dを提出し続けており、2016年9月現在では20版を数えます。

HLSの構造は非常にシンプルです。以下に再生のための指示内容を記したファイル(マニフェストファイル)の例を示します(図-1)。

階層	技術
プレゼンテーション	HTML5, Flashなど
コーデック、メタデータ	H.264, H.265(HEVC)、AAC、WebVTT
コンテナ	MPEG2-TS, MP4
ストリーミング形式	HLS, MPEG-DASH, CMAF
配送プロトコル	HTTP/1.1
トランスポートプロトコル ネットワークプロトコル	TCP/IP

表-1 ストリーミング配信の構成

本質的なのは、URIで示されるリソースです。この例では、サーバmedia.example.com上には「first.ts」「second.ts」「third.ts」という3つのメディアファイルが設置される必要があります。このファイルは静的に配置されても、あるいは動的に配置されても構いません。そしてスキーム名を見て分かるように、このメディアファイルはhttpにより配信されることが示されています。

このマニフェストファイルもやはりWebサーバ上に設置され、再生プレイヤーに渡されます。再生プレイヤーは参考情報を読み込みつつ、URIで示されたメディアファイルに「上から順番に」アクセスしていきます。コメント行には参考情報が記されており、図-1の例だとEXT-X-TARGETDURATIONはメディアファイルの最長の長さを、EXTINFは次に続くメディアファイルの実際の長さを秒で表しています。つまり、この再生指示ファイルによって指定されているメディアは、合計で21.021秒の再生時間であることが分かります。

このI-Dの標準化活動は実施されていません。RFCとして発行するためにはIETFのワーキンググループでの議論が必要になります。しかしこのI-DはどのWGでも継続的な課題として取り上げられておらず、また議論もされていません。いわば「私教版」の扱いです。

これに対抗する意味合いもあり、MPEG-DASH(Dynamic Streaming over HTTP)がInternational Organization for Standardization(ISO)のMPEG(Moving Picture Experts Group)で規格化され、2012年にISO/IEC 23009-1として出版されています。MPEGは1988年より活動している専門家集団であり、標準規格化作業のためのワーキンググループです。

```
#EXTM3U
#EXT-X-TARGETDURATION:10
#EXTINF:9.009,
http://media.example.com/first.ts
#EXTINF:9.009,
http://media.example.com/second.ts
#EXTINF:3.003,
http://media.example.com/third.ts
#EXT-X-ENDLIST
```

図-1 HLS用マニフェストファイルの例

このグループはこれまで映像・音声圧縮フォーマットの規格化作業を主な領域としていましたが、インターネットを用いた動

画配信が広がりを見せたため、ストリーミング形式の策定にも乗り出してきました。

```
<MPD xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns="urn:mpeg:dash:schema:mpd:2011"
xsi:schemaLocation="urn:mpeg:dash:schema:mpd:2011
DASH-MPD.xsd" mediaPresentationDuration="PT0H1M6.1S"
minBufferTime="PT1.5S"
profiles="urn:mpeg:dash:profile:isoff-on-demand:2011"
type="static">
  <Period duration="PT0H1M6.1S" start="PT0S">
    <AdaptationSet>
      <Representation bandwidth="4000000"
codecs="avc1.4d401e" height="1080" id="1"
mimeType="video/mp4" width="1920">
        <BaseURL>video_4000.mp4</BaseURL>
        <SegmentBase indexRange="860-1023">
          <Initialization range="0-859" />
        </SegmentBase>
      </Representation>
      <Representation bandwidth="2400000"
codecs="avc1.4d401e" height="720" id="2"
mimeType="video/mp4" width="1280">
        <BaseURL>video_2400.mp4</BaseURL>
        <SegmentBase indexRange="859-1022">
          <Initialization range="0-858" />
        </SegmentBase>
      </Representation>
    </AdaptationSet>
    <AdaptationSet>
      <Representation bandwidth="128000" codecs="mp4a.40.2"
id="5" mimeType="audio/mp4">
        <BaseURL>audio_128.mp4</BaseURL>
        <SegmentBase indexRange="783-946">
          <Initialization range="0-782" />
        </SegmentBase>
      </Representation>
    </AdaptationSet>
  </Period>
</MPD>
```

図-2 MPDの例

MPEG-DASHの規定ではマニフェストファイルのことをMPD (Media Presentation Description) と呼びます。次にMPDの例を示します(図-2)。

XMLで構造化されているのが分かります。AdaptationSetで定義される複数のRepresentationは、クライアントが動的に切り替えて再生することを想定しています。この例ではサーバ側で4Mbpsと2.4Mbpsのビデオストリームが用意されていることが示されており、クライアントは自らの環境(回線状況やCPU消費など)に応じて最適なストリームを選択することができますようになっています。

現在広く使われているHLSやMPEG-DASH, Smooth StreamingやHTTP Dynamic Streamingは、配信プロトコルとしてHTTP/1.1を採用していたところに大きなメリットがあります。既に広く普及していたHTTP/1.1は、ストリーミング専用のプロトコルを用いるよりもスケラビリティの面で優れていました。HTTPに載せて運ばれるストリーミング形式はデータをセグメント化して、細切れにしてサーバからクライアントに配送するというアイデアは共通していました。

しかしその設計に際してHLS, MPEG-DASH, Smooth Streaming, HDSは各々マニフェストファイルとコンテナ形式を採用し、その結果すべての組み合わせが異なることになりました(表-2)。このような事態は、コンテンツ制作やCDN事業者などの現場を混乱させました。幅広くクライアントをサポートしようとする

名称	規格化	マニフェストファイル	コンテナ形式	標準化の有無
HTTP Live Streaming	Apple	m3u8 m3uの規格を独自拡張	セグメント化されたMPEG2-TS 最新版でMP4をサポート	標準化されず ただしApple以外にも広汎に利用されている
MPEG-DASH	ISO/IEC	MPD XMLで記述	MP4, MPEG2-TS (MP4が多く使われている)	国際標準 詳細定義は各所で行われている (MPEG Industry Forumなど)
Smooth Streaming	Microsoft	isml	MP4	なし
HTTP Dynamic Streaming	Adobe	f4m XMLで記述	f4f	なし

表-2 代表的なストリーミング形式

と、すべてのマニフェストファイルとコンテナ形式をあらかじめ作っておかなければなりません。仮にHLS、MPEG-DASH、Smooth Streaming、HDSを利用すると、4種類のファイルを作成し、管理しなければなりません。制作現場は工数が増大しますし、4倍のストレージを用意しなければなりません。またCDN事業者の配信システムにおいてキャッシュの効率低下を招くこととなります。再生デバイス側でも、どの方式を実装すれば良いのかは難しい問題です。実際にはSmooth StreamingとHDSが採用されることは減少しており、HLSとMPEG-DASHの2種類が選ばれることが多いと思いますが、いずれにせよこれは非効率です。

HLSは最新の20版で、大きな構造の変化がありました。MP4のサポートです。これまで指定されていたフラグメントMPEG2-TSに加え、フラグメントMP4のサポートが追加されました。またPacket AudioとWebVTTも新たなマルチメディア形式として追加されています。フラグメントMP4(fMP4)はやはりISO/IECにより標準化されている形式で、一連のデータが順序付けられた複数のファイルのことで。

このサポートにより、HLSとMPEG-DASHのメディアライブラリを共通化できる可能性が生まれました。フラグメントMP4で用意した一種類のメディアライブラリがあれば、複数のシステムに配信できるようになります。また配信事業者側の観点から言えば、キャッシュ運用の効率化が図られることとなります。HLSであってもMPEG-DASHであっても、データ容

量の多い部分は動画データを格納しているフラグメントMP4ファイル群になるからです。

Appleのこのような動きは、ストリーミング形式の統一を目指す流れに合致しています。そこで提案されているのがCommon Media Application Format(CMAF)です。原案はAppleとMSにより提案され、MPEG(The Moving Picture Experts Group)で議論されています。このようなストリーミング形式の標準化作業は、IETFではなくMPEGで討議する方がより現実的です。コンテナ形式やコーデック分野のエキスパートが集まっているコミュニティがMPEGだからです。

CMAFの規格案には、サブタイトルとして次のような文言があります。

Media Application Format optimized for large scale delivery of a single encrypted, adaptable multimedia presentation to a wide range of devices; compatible with a variety of adaptive streaming, broadcast, download, and storage delivery methods

網羅的に技術をカバーしているように思えますが、CMAFはこれまでのHLSやMPEG-DASHでの成果を受け継ぐ形になっています。逆に言えば、これまでストリーミング形式で課題になってきた事項はすべてカバーされています。

名称	マニフェストファイル	コンテナ
HTTP Live Streaming	m3u8	CMAF <CMAFの内部構造> + CMAF Presentation + CMAF Selection Set (同一コンテンツの異なる要素を複数収容できる;カメラ映像、コーデック、多言語など) + CMAF Switching Set (同一コンテンツを異なるエンコード形式で複数収容できる) + CMAF Track + CMAF Fragments + CMAF Header
MPEG-DASH	MPD	

表-3 HLS、MPEG-DASHとCMAFの関係

CMAFはマニフェストファイルやプレイヤー、配送プロトコルを定義しません。HLSやMPEG-DASHがそれぞれ持つマニフェストファイルから呼び出せる形になっています。

CMAFの構造は階層化されており、複数の言語や異なるビットレートなどに対応したものとなっています(表-3)。

CMAFでメディアライブラリの作成が一本化されれば、コンテンツ作成の手間が大幅に減ります。ユーザに直接的なメリットをもたらすものではありませんが、ストリーミングメディアのさらなる普及のためには必要不可欠な作業だと受け止められているようです。

CMAFの意義は、AppleやMicrosoftが独自路線によらず、業界として大きな流れを生み出すことができるかどうかにかかっているでしょう。このようなフレームワーク、そして議論の際に培われた人的交流があれば、次の課題に対してもよりよい動きができるようになるでしょう。統一されたプロトコルやフォーマットありきではなく、いま動いているものをベースにしつつ更なる改善を目指す形は、イノベーションを起こしやすいとも言えます。

最後に、今後の課題を2つ述べます。

まず、動画視聴開始までのタイミングを短くすることです。現在のライブストリーミングは実時間より概ね30秒かそれ以上遅延が発生しています。これは複合的な要因があり、エンコードにかかる時間やネットワークへのアップロード、そしてプレ

イヤーが再生初期にバッファするセグメントの数などが影響しています。エンコーダや配信サーバ設定で簡単に変えられるものではないため、ストリームを生成する側では制御しにくいのです。しかし、やはりライブイベントはリアルタイムに近い形で再生させたいというニーズがあります。この課題は業界にも広く認識されており、将来的に改善のための提案がなされるのではないのでしょうか。

もう1つがオフライン再生です。特にモバイルでの動画視聴を考えたときに、キャリアによるダウンロード容量制限は大きな壁となっています。モバイルでは動画配信を控える方も多いでしょう。そこで、Wi-Fi接続時に一旦動画をモバイル機器にダウンロードさせてしまい、オフラインでも動画を視聴できるようにするための技術です。

GoogleはYouTubeアプリでオフライン再生するための仕組みを加えました。特定の動画を保存して、48時間までオフライン再生することができるようになっています。この機能は全ユーザには解放されておらず、主に通信インフラが発展途上の国でのみ有効とされています。またHLSもオフライン再生のための仕組みを整えました。こちらはiOSアプリで対応コードを作り込む必要があります。

いずれも一部のアプリや仕組みで対応が始まっていますが、本格的に普及が始まった段階で標準化される必要があるものです。こうした動きをフォローアップできるかどうか、業界に問われているとも言えるでしょう。



執筆者:

山本 文治 (やまもと ぶんじ)

IJ 経営企画本部 配信事業推進部 シニアエンジニア。

1995年にIJメディアコミュニケーションズに入社。

2005年よりIJに勤務。主にストリーミング技術開発に従事。同技術を議論するStreams-JP Mailing Listを主催するなど、市場の発展に貢献。

言語リソースとしてのWikipedia

3.1 はじめに

世界最大のオンライン百科事典であるWikipediaを活用しておられる方々が多いのではないかと思います。信頼性の問題が指摘されてはいるものの質・量共に史上最大の百科事典から得られる情報量の多さ故に、筆者も調べ物の手段として日常的に活用しています。また、筆者の場合はソーシャルデータとしてのWikipediaの研究を手がけている関係でWikipediaのPageview Countに大いに関心を持っているところです。

このようにWikipediaを研究対象として捉えている研究者はたくさんいらっしゃるようです。例えばDBPedia^{*1}やその日本語版^{*2}はWikipediaの情報を抽出し、LOD(Linked Open Data)に変換して公開する試みでして、セマンティック・ウェブといった研究などに活用されているそうです。あるいは自然言語処理の研究分野でもWikipediaは膨大な文例を収蔵する言語リソースとして認識されており、様々な活用方法が提案されています。そこで本稿では、筆者の研究からは離れて、言語リソースとしてのWikipediaをテーマに紹介します。

3.2 「Unix考古学」の執筆

筆者が言語リソースとしてのWikipediaに関心を持つには少々訳がありまして、実は今年の4月に「Unix考古学^{*3}」という本を出版しました。UNIXに関わる歴史的な事実を文献を辿りながら紹介していくこの本は、強いて言えば歴史書にカテゴリズされると思います。

同書はもともと、月刊誌「UNIX USER」に2003年から2005年にわたって掲載された26編の記事を整理・再編集して単行本にまとめたものなのですが、取材や図書館通いをすることなく、Googleの検索エンジンで集めた文献だけで執筆された書籍であるところにも特徴があります。今日ならいざ知らず、1998年にGoogleの検索エンジンが公開されてから数年しか経っていない当時としては、これはかなり無謀な執筆スタイルでした。

もっとも、当時の筆者には(今考えれば根拠が極めて薄弱な)確信がありました。それは1980年代に現れた新しい職業の中に「データベース検索技術者」、通称「サーチャー」という職業がありました。それを紹介するテレビ番組のなかで、小説家とサーチャーがタッグを組んで新作を執筆するという事例が語られたのですが、そこではサーチャーが小説家に対し「あなたは執筆するための資料集めを一切する必要はありません。あなたが必要だと思う情報を教えてくれれば、私がすべてデータベースの中から引き出してみせます」と豪語したエピソードが紹介されていました。まだ就職したての新人だった、当時の私は「そんなことができるのか?」と非常に懐疑的な印象を持ったのでした。それから十数年経過して、連載の執筆依頼をもらった際、このエピソードを思い出した筆者は「Googleの検索エンジンが使える今なら、私にもこのアプローチが可能かもしれない」と考えたのでありました。

実は、執筆依頼を受諾してから最初の記事が発表されるまで、半年程度の猶予がありました。編集部としてはこの期間を使って数回分の記事を書きためることを期待していたのですが、筆者はほとんどの時間をGoogleを使って欲しい文献を探し当てるキーワードの選び方と並べ方の順序、つまり検索パターンを発見するために費やしました。結果、半年後に完成していたのは第1回と第2回のみで、後々、厳しいツケを払い続けることになりました。このようにして検索エンジンを片手にトピックを「どこまでも掘り下げて」文章を書き加えていく、ちょっと風変わりな執筆スタイルに定着しました。

3.3 Drilldown Search

今では「どこまでも掘り下げて」を表現するのにピッタリなDrilldownという用語があります。Wikipediaの"Drill down"^{*4}のページによれば「何かに注目しながら詳細なデータを得るために、ある場所から別の場所へと移動することを意味する」と定義されています。筆者のケースはこのページで例示されているオン

*1 DBPedia(<http://wiki.dbpedia.org/>)。

*2 DBPedia日本語版(<http://ja.dbpedia.org/>)。

*3 Unix考古学(<http://asciidwango.jp/post/142281038535/unix%E8%80%83%E5%8F%A4%E5%AD%A6-truth-of-the-legend>)。

*4 Drill down(https://en.wikipedia.org/wiki/Drill_down)。

ラインユーザやウェブサーファの事例が最も近いのだろうと想像していますが、注目するポイントが「計算機の歴史に関わる、あまり認知されていない文書」を発掘することにあるので、作業の中身は非常に特殊なケースなのではないかと考えています。

例えば・・・

7th Edition Unixではカーネルも全面的にC言語で書き換えられた。これは一般にも割と有名な話。この作業にあたったのはKen ThompsonとDennis RitchieとSteve Johnsonの3人であること、またSteve JohnsonはPortable C Compiler (PCC)の開発者であったことからこの作業に加わったことはDennis Ritchieが残した複数の文献(論文、講演資料、インタビュー)から確認することはできる。では、Steve Johnson自身が一連の開発作業について言及した文献はないか？

あるいは・・・

BSD UNIXの開発において、4BSDをリリースした際、主に3BSDとの比較でパフォーマンスが低下したことを指摘する批判が出て、その対応のためにUCBのCSRGはパフォーマンスチューニングを施した4.1BSDをリリースした。ここまではUNIXの歴史を語る多くの文献で言及されている。しかし、その実態はDavid Kashtanという人物の非常に攻撃的な批判記事にBill Joyが激怒し反撃を行った事実があることがKirk McKusickの文書で明らかになっている。それではKashtanの批判記事、及びそれに対するBill Joyの反論の内容が分かる文献はないか？

・・・といった歴史的な事実に関するDrilldownを試みるとなると、検索は非常に厄介な作業になってきます。探し当てた文献の関連箇所を読み込み、その中から新たな文献を探索するキーワードを拾い出す一連の検索サイクルを繰り返して、毎月6~10ページの記事を書くことは非常にエネルギーと集中力を要する作業で、小説家の執筆というよりはジャーナリストの記事作成に近い仕事なのではないかと想像しています。

事実、連載終了後から12年が経過して、単行本としての出版が決まった際、相当の分量の書き直しや新たな書き下ろしが必要

になったので、12年前を思い出して同様の作業を再度試みたのですが・・・歴史的な事実のDrilldownは、年老いた今の筆者にはとても耐えられない程のハードな作業に変化していました。「この作業の一部だけでもコンピュータにやらせられないだろうか？」ふと、そんなことを考えるようになっていきました。

3.4 固有表現認識 (NER)

以来、筆者は「歴史的な事実のDrilldownの機械化」を考え続けています。これは当初、研究というよりは執筆活動のための道具立てとしての側面が強かったのです。もちろん「これさえあれば執筆依頼を受けるハードルが下げられる」との少々ヨコシマな考えもありました(笑)。

機械化、つまり筆者の文献探索作業を模倣するソフトウェアを作るためには、まず「どのようにして文献を探していたのか？」を明らかにする必要があります。前述の記事執筆の経験から体得した「検索エンジンによる文献探索と事実情報の抽出」で最も重要だったノウハウは、固有名詞に着目することでした。在り来たりに見えるでしょうが人名、組織名、それから計算機の歴史的事実の場合は過去のコンピュータの型番や愛称などが該当します。これらの固有名詞をできるだけ拾い集めその組み合わせを変えて検索キーワードを指定すると、検索結果を望む方向に大きく絞れることが多かったのです。そこで英語テキストから固有名詞を抽出する技術を探すところから手を付けることにしました。

当然のことながら自然言語処理の研究領域の話になるのですが、調べ始めてすぐに気づいたことが2つありました。

- (1) 現在は統計的自然言語処理が主流
- (2) 日本の自然言語処理研究と英語圏の自然言語処理研究では研究の様相がかなり違う

まず(1)に関しては、現在の筆者とは研究の目的もゴールもまったく違うのですが、いずれも統計学に基づく技術をベースに研究が進められていることです。特に、筆者の場合はデータ分析の基盤技術も研究の対象となっているので、基盤技術だけを見れば多くの共通点が見つけられるように感じました。

次に(2)については、これまで自然言語処理と言うと暗黙のうちに「日本語の」と仮定していたところがあり、形態素解析や機

械翻訳といった日本語に関わり深い研究が主要なテーマだと
筆者は理解していたのですが、英語の自然言語処理研究では
(当然のことながら)取り組まれているテーマがだいぶ違う
との印象を受けました。

筆者が欲していた「英語テキストから固有名詞を抽出する技
術」も固有表現認識(NER:Named Entity Recognition)と呼
ばれる英語での自然言語処理研究では主要な研究テーマの1つ
になっているそうです。

3.5 Natural Language Toolkit(NLTK) で利用できるNER

今日ではNatural Language Toolkit(NLTK)にNERが搭載さ
れているので、図-1のようなPythonスクリプトを使って比較
的お手軽に固有表現の抽出を試すことができます。

このスクリプトは英文のテキストファイル名を渡すとNERを
実行して図-2のような出力をします。

出力は各行単語、品詞、NEタグが表示されています。NEタグの
先頭がB-の場合は固有表現の先頭、I-の場合は継続語を意味し
ます。PERSONは人名、ORGANIZASIONは組織名、GSPは"Geo-

Socio-Political group"の略です。"Hillary Clinton"や"Donald
Trump"が人名として認識されていることが分かります。

次に、もう少し大きなテキストでの事例として「Unix考古学」の
執筆でもお世話になったMarshall Kirk McKusickの"Twenty
Years of Berkeley Unix"のテキストから人名を抽出してみまし
た(図-3)。

誤認識もありますが、Ken ThompsonやDennis Ritchie、Bill
Joyと言った有名人だけでなく、Kirk Mckusick以外のBSD
UNIXの開発メンバーであったOzalp Babaoglu、Sam Leffler、
Mike Karels、Keith Bosticの名前も正しく認識されています。
このように、特定の目的(例えばUNIXの歴史に関するトピック)
のために特別なチューニングをしなくても、それなりの精度で
固有表現を抽出できるのが、最近の統計的自然言語処理の研究
成果なのでしょう。

3.6 統計的自然言語処理と"コーパス"

さて「特別なチューニングなし」に「それなりの精度で抽出でき
る」不思議さを知るには統計的自然言語処理の研究を勉強する
しかないのですが、解説書を読むと頻りに登場するのが
"コーパス"という用語です。

```
#!/usr/bin/env python
# -*- coding: utf-8 -*-

import nltk
import sys

if __name__ == "__main__":
    param = sys.argv
    args = param[1]

    with open(args, 'r') as f:
        sample = f.read()

    sentences = nltk.sent_tokenize(sample)
    tokenized_sentences = [nltk.word_tokenize(sentence) for sentence in sentences]
    tagged_sentences = [nltk.pos_tag(sentence) for sentence in tokenized_sentences]

    # for NLTK 2.7
    # chunked_sentences = nltk.batch_ne_chunk(tagged_sentences)

    # for NLTK 3.0
    chunked_sentences = nltk.ne_chunk_sents(tagged_sentences)

    entity_names = []
    for tree in chunked_sentences:
        print nltk.chunk.tree2conllstr(tree)
```

図-1 named_entity_recognition.py

Wikipedia英語版によると、「コーパス」とは言語学において使用される「巨大で構造化されたテキストのセット」で「統計的な分析や仮説検定、頻度のチェック、特定の言語のルールに基づく正当性の確認に使われる」そうです。筆者のような門外漢にはこの説明だけではピンと来ないのですが、一般に広く知られているジップの法則 (Zipf's law) 「出現頻度がk番目の単語が全体に占める割合は1/kに比例する」という経験則を思い出し、「人手が書いた文章と統計学の不思議な関係」を少しイメージしやすくなるのではないかと思います。つまり「人間は作文において無意識のうちに統計学的に偏りのある振る舞いをしている」。それが統計的自然言語処理がうまく機能する理由なんだろうと想像しています。

実は、この統計学的に偏りのある振る舞いを、筆者は「Unix考古学」の執筆経験を通して直感的に体験していました。例えば、Dennis Ritchieはホームページ^{*5}でUNIXの開発経緯に関して述べた多数の文書を公開しています。執筆の際には大変お世話になったのですが、複数の文章をなんども読み直すうちに「あれ、この言い回しは他のどこかでも見たことがある・・・」と気

```
$ cat NHK-short.txt
A US poll shows that Democratic presidential nominee Hillary Clinton's lead
over her Republican rival, Donald Trump, has shrunk to 3 percentage
points.
$ ./named_entity_recognition.py NHK-short.txt
A DT O
US NNP B-GSP
poll NN O
shows VBZ O
that IN O
Democratic JJ B-ORGANIZATION
presidential JJ O
nominee NN O
Hillary NNP B-PERSON
Clinton NNP I-PERSON
's POS O
lead NN O
over IN O
her PRP$ O
Republican JJ B-ORGANIZATION
rival NN O
.. O
Donald NNP B-PERSON
Trump NNP I-PERSON
.. O
has VBZ O
shrunk NN O
to TO O
3 CD O
percentage NN O
points NNS O
.. O
$
```

図-2 サンプルの英小文とNLTKの固有表現認識の実行結果

```
Alan Nemeth
Babaoglu
Beranek
Berkeley
Berkeley Software Design
Berkeley Software Distribution
Berkeley Unix
Bert Halstead
Bill Jolitz
Bill Joy
Bob Baker
Bob Fabry
Bob Guffy
Bob Kridle
Bostic
Casey Leedom
Chuck Haley
DARPA
Dan Lynch
David
Dennis Ritchie
Dickinson R. Debevoise
District Judge
Domenico Ferrari
Duane Adams
Eugene Wong
Fabry
Fateman
Ferrari
Freely Redistributable Marshall Kirk McKusick Early History Ken Thompson
Haley
Hibler
Jeff Schriebman
Jerry Popek
Jim Kulp
John Reiser
Jolitz
Joy
Karels
Keith
Keith Bostic
Keith Lantz
Keith Standiford
Ken Thompson
Laura Tong
Leffler
Linux
Lucasfilm
Math
Michael Stonebraker
Mike
Mike Karels
Mike Muuse
Networking Release
Ozalp Babaoglu
Pascal
Pauline
Peter Kessler
Professor Domenico Ferrari
Professor Richard Fateman
Professors Michael Stonebraker
Ray Noorda
Rick Macklem
Rick Rashid
Rob Gurwitz
Robert Elz
Sam Leffler
Schriebman
Schwartz
Statistics
Support Meanwhile
Susan Graham
System
System III
System Manual
System V
Tahoe
Thompson
Tom London
Unix
Unix Early
Utah
```

図-3 PERSON of "Twenty Years of Berkeley Unix"

*5 Dennis Ritchie (<https://www.bell-labs.com/usr/dmr/www/>)。

づくことが度々ありました。つまり、読者には「執筆者の癖」と理解できる記述が、統計的なテキスト分析では「統計学的な偏り」として把握できるようです。

ちなみにDennis Ritchieの最も分かりやすい癖は、PCCの開発者をSteve Johnsonと記していたところです。Ritchieの記述にはすべてこの名前が使われていたのですが、実は彼の本名はStephen Johnsonで、彼自身が執筆した論文やWikipediaページなどではStephenと記述されています。Ritchieの誤解だったのか、それとも彼はベル研ではSteveという愛称で呼ばれていたのか定かではありませんが、いずれにせよ、事実の後追いをしている筆者には「どんな条件で検索しても文献がまったく引かからない」非常に困惑させられる経験をするようになりました。

どうやら統計的自然言語処理とは人間が書いたテキストを対象に統計学的な偏りから様々な新たな知見を得る研究のようです。

3.7 WikipediaからNEコーパスを生成する

NLTKの話に戻ります。NLTKのNERの概要は"入門 自然言語処理"(原書:"Natural Language Processing with Python")の第7章「テキストからの情報抽出」で解説されていますが、その実装はNLTK Corpora^{*6}の1番目にリストされている"ACE Named Entity Chunker (Maximum entropy)"です。

Linked article texts:



Article classifications:

misc. person person

図-4 Deriving training sentences from Wikipedia text

この「最大エントロピー」と呼ばれる機械学習モデル(同書の「6.6 最大エントロピー分類器」に解説があります)は、固有表現タグを注釈したNEコーパスを使用します。単語に「人名」や「組織名」といった固有表現タグを付与するNEコーパスの作成は人間が介在せざるを得ない非常に手間のかかる作業で、故に無料配布されているNEコーパスはほとんどないそうです。実際、NLTKの場合もAutomatic Content Extraction(ACE)が作成したコーパスを使って学習したチャンカーのみをpickleファイルとして配布しています。コーパスそのものは含まれていません。

このNEコーパスをWikipediaのページデータを使って機械的に生成しようという試みがあります。論文"Transforming Wikipedia into Named Entity Training Data^{*7}"では、固有表現の注釈付きのコーパスを作成するために、Wikipediaの利用を提案しています。

Wikipediaの記事の中で記載された用語や名前は、多くの場合、適切な他の記事にリンクされています。そのような記事間のリンクを固有表現の注釈に変換することがこの論文の基本的なアイデアです。例えば「イアン・フレミング」の小説「サンダーボール」において「ジェームズ・ボンド」というキャラクターについて紹介する文章は、各々の固有表現について相互にリンクが張られていることが期待できます。「イアン・フレミング」の記事は、その表現が人物であることを示してくれるでしょうし、「サンダーボール作戦」の記事から、それが小説であることが分かります。このようにリンクを辿ると元の文章には自動的に注釈を付けることができるわけです(図-4)。

論文ではWikipediaからNERをトレーニングするための何百万もの文章を抽出して、巨大なコーパスを形成することができますと述べています。そのプロセスとして次の4つのステップを掲げています。

1. すべての記事をエンティティクラスに分類する
2. Wikipediaの記事から文章を取り出す
3. リンクターゲットに応じて固有表現をラベル付ける
4. コーパスに収録すべき文章を選択する

^{*6} NLTK Corpora(http://www.nltk.org/nltk_data/)。

^{*7} Transforming Wikipedia into Named Entity Training Data(<https://www.aclweb.org/anthology/U/U08/U08-1016.pdf>)。

この手順を使って、論文では標準的なCONLLカテゴリのエンティティクラス(LOC、ORG、PER、MISC)に基づくNEコーパスの生成を試みています。

現在では英語版は500万以上、日本語版でも100万を超える記事を収蔵するWikipediaを対象にこの手順を実行するのは明らかにビッグデータ処理になります。特に論文が記事の分類のために示すブートストラッピング、すなわち一部は人手でクラスを判別する方法はコーパスを完全に機械的に生成する上での困難な課題でしょう。

3.8 おわりに

図らずも、筆者が考える「歴史的な事実のDrilldown文献探索」の機械化は、最新の自然言語処理研究の成果を使えばかなり前進するのではないかと、筆者が研究として取り組んでいるWikipedia分析のために開発した基盤技術がこの取り組みにも活用できる可能性があることに少し驚きを感じているところです。

文献から人名やシステム名を抽出する仕事はNERを使えば実現可能ですが、その認識精度を上げるためにはNEコーパスの入手とそれを使った学習が欠かせません。独自のNEコーパスを生成するためのリソースとして、日頃から慣れ親しんでいるWikipediaが活用できるというのは朗報でした。残る課題は

500万ページ以上ある英語版Wikipediaから人物のページを拾い上げること・・・。

この問題は、筆者が手がける歴史書の執筆などではあまり問題にはならないように思います。歴史書の執筆とは筆者が定めたテーマ、すなわち時代や人物、事件について深く掘り下げていく行為にほかならないからです。英語版Wikipediaから執筆テーマに関連する人物のページを収集する作業などは日常茶飯事です(笑)。ノンフィクションや報道記事を手がける記者や社会科学の研究者は、人物に限らず固有表現やその分類に精通しています。彼らの知識を集合知として集め共有する方法を考えるのが、問題解決の早道だと思います。

以上、筆者は思いもよらぬことから今日の自然言語処理研究を学べる機会を得ましたが、この研究成果は、現在研究で手がけているソーシャルデータの分析にも活用できるのではないかと考えている次第です。例えば「要因分析」。ソーシャルデータの時系列分析を行うと急激な変動、いわゆるバーストを検知することができますが、このバーストを引き起こした原因がなんであったかを調べるためにはバースト発生時の社会的動向を追跡する必要があります。筆者は検索エンジンなどを使って報道記事を集めることを考えていたのですが、そのための検索条件を決める手段として本稿で紹介した方法が使えるのではないかと考えてます。



執筆者：
藤田 昭人 (ふじた あきと)

株式会社IIJイノベーションインスティテュート(IIJ-II)企画開発センター チーフアーキテクト。2008年IIJ入社。
構造化オーバーレイ研究の知見を活用したクラウドコンピューティング技術の研究開発に従事している。



Internet Initiative Japan

株式会社インターネットイニシアティブ(IIJ)について

IIJは、1992年、インターネットの研究開発活動に関わっていた技術者が中心となり、日本でインターネットを本格的に普及させようという構想を持って設立されました。

現在は、国内最大級のインターネットバックボーンを運用し、インターネットの基盤を担うと共に、官公庁や金融機関をはじめとしたハイエンドのビジネスユーザに、インターネット接続やシステムインテグレーション、アウトソーシングサービスなど、高品質なシステム環境をトータルに提供しています。

また、サービス開発やインターネットバックボーンの運用を通して蓄積した知見を積極的に発信し、社会基盤としてのインターネットの発展に尽力しています。

本書の著作権は、当社に帰属し、日本の著作権法及び国際条約により保護されています。本書の一部あるいは全部について、著作権者からの許諾を得ずに、いかなる方法においても無断で複製、翻案、公衆送信等することは禁じられています。当社は、本書の内容につき細心の注意を払っていますが、本書に記載されている情報の正確性、有用性につき保証するものではありません。

©2008-2016 Internet Initiative Japan Inc. All rights reserved.
IIJ-MKTG019-0033

株式会社インターネットイニシアティブ

〒102-0071 東京都千代田区富士見2-10-2 飯田橋グラン・ブルーム
E-mail: info@ij.ad.jp URL: <http://www.ij.ad.jp/>