

IIJR

Internet
Infrastructure
Review

Aug.2015

Vol. 28

インフラストラクチャセキュリティ

機械学習とセキュリティ

ブロードバンドトラフィックレポート

ブロードバンドとモバイルの トラフィックを比較

技術トレンド

ソーシャル・ビッグデータ

IIJ

Internet Initiative Japan

Internet Infrastructure Review

August 2015 Vol.28

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	13
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	17
1.4 フォーカスリサーチ	18
1.4.1 機械学習とセキュリティ	18
1.4.2 猛威を振るうAngler Exploit Kit	21
1.4.3 ID管理技術～オンライン認証にパスワードを使わない方法へ～	24
1.5 おわりに	27
2. ブロードバンドトラフィックレポート	28
2.1 概要	28
2.2 データについて	28
2.3 利用者の1日の使用量	29
2.4 ポート別使用量	31
2.5 まとめ	33
3. 技術トレンド	34
3.1 ビッグデータの現状	34
3.2 ソーシャル・ビッグデータ	36
3.2.1 ソーシャル・ビッグデータの性質	36
3.3 ビッグデータとしてのWikipedia	37
3.3.1 ソーシャル・ビッグデータとしてのWikipedia	38
3.3.2 Tobias Preisの研究	38
3.4 まとめ	39

エグゼクティブサマリ

今回より、浅羽からエグゼクティブサマリを引き継ぐことになりましたサービスオペレーション本部長の山井美和と申します。今後ともどうぞよろしくお願いいたします。

日本年金機構の個人情報流出事件は記憶に新しい事件だと思います。この事件をきっかけに、世の中の情報セキュリティに対する認識が高まり、インターネットを使う上での安全対策が、インフラ側だけでなく個人や組織自身の防御にも必要であることがより一層強く認知されたのではないのでしょうか。インフラを作って運用する側としてできることは、サービスという形での提供を目指していますが、それを凌ぐスピードで脅威が迫ってきていることは明らかです。提供者と利用者が一体となったセキュリティオペレーションにより、インターネットを守る機運が高まることを期待する今日この頃です。

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章の「インフラストラクチャセキュリティ」では、インターネット内で起こっているHacktivismに基づく攻撃やソーシャルメディアに関連する不正アクセスの事例などを取り上げ解析しています。また、フォーカスリサーチでは、機械学習とセキュリティ、猛威を振るうAngler Exploit Kit、ID管理技術の実際の利用に焦点を当てて解説しています。

2章の「ブロードバンドトラフィックレポート」では、直近のインターネットを流れるトラフィックを分析していますが、ここ数年で伸びてきているモバイルトラフィックについて、IIJの個人及び法人向けモバイルサービスによるトラフィックを取り出し、ブロードバンドトラフィックとの比較を行っています。モバイルトラフィックの傾向を探る興味深い掘り下げになっています。

3章の「技術トレンド」では、ビッグデータを取り扱いました。IoTやCPSというコンセプトが脚光を浴びるようになってきましたが、ソーシャルネットワーキングの世界ではこれらを含む多様なデータが発信されています。一方的に発せられるツイートをはじめ、ソーシャルな集合知とも言えるWikipediaも有力なビッグデータのソースと言えるでしょう。ここでは、研究所で取り組んでいる研究内容をご紹介します。

インターネットインフラストラクチャは、単なる物理的なインフラだけでなくソーシャルな基盤という意味でのインフラへと変化してきていると言えます。本レポートでは今後もこうした広い側面から技術という切り口でレビューしていきたいと思います。

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)

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サービスの運用サポート部門を統括し現在に至る。

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。

機械学習とセキュリティ

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IJJ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、IJJ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2015年4月から6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、SNSアカウントの乗っ取りやWebサイト改ざんなどの攻撃も多発しています。標的型攻撃によるマルウェア感染も多く発生しており、日本年金機構の事件では最大で125万件の個人情報漏えいした可能性が指摘されています。不正アクセスによる情報漏えいも継続して発生しており、米国の人事管理局が不正アクセスを受け、約400万人分の職員の情報が漏えいする事件などが発生しています。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、2015年4月から6月までの期間にIJJ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。

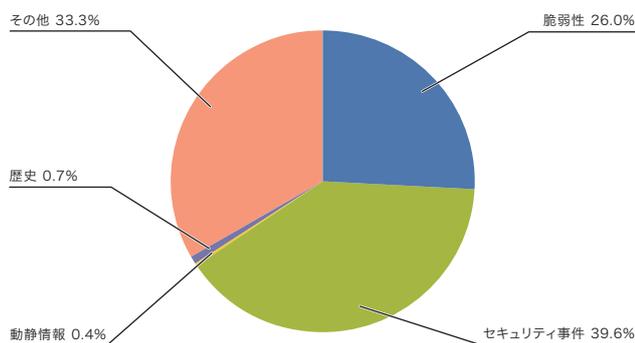


図-1 カテゴリ別比率(2015年4月~6月)

■ 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。

ISILもしくはその理念に共感していると考えられる個人や組織による、Webサイトの改ざんやSNSアカウントの乗っ取りなどが世界中で発生しています。4月にはフランスのTV局が大規模な攻撃を受け、番組放送が中断する事件が発生しました。この攻撃では不正アクセスによる局内のシステムへの攻撃*2と同時にFacebookやTwitterなどのSNSアカウントの乗っ取りと不正な投稿も行われました。対抗してAnonymousによるISILに関連していると考えられるWebサイトやSNSアカウントなどのリストを公開し、停止や削除を促すなどの活動も継続して行われています(OplSIS)。更に、パレスチナ系のAnonymousとして以前から活動していたグループがISILを支持していたとして他のAnonymousから攻撃対象とされるなど、混乱した状況が続いています。

サウジアラビア政府が行ったイエメンへの空爆への抗議などから、サウジアラビア系の新聞社に対するDDoS攻撃や、複数の政府機関に対する不正侵入とそれによる機密情報の漏えいが発生しています(OpSaudi)。更に、イスラエル軍によるガザ空爆に反対し、イスラエルの複数の企業で不正アクセスによる個人情報やカード情報の漏えいが発生するなどの攻撃も継続しています(OplIsrael)。このように中東各国では、紛争や外交などの情勢に応じて、インターネット上でも攻撃が続いています。

カナダでは、情報当局の権限を大幅に拡大・強化し、プライバシーの問題が指摘されている対テロ法案が6月に可決成立したことから、複数の政府機関に対するDDoS攻撃や不正侵入によるWebサイトの改ざんや内部情報の漏えいなどの攻撃が発生

*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:ワ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。

*2 この攻撃についてはRATが使われた可能性が指摘されている。トレンドマイクロ社、「仏テレビ局へのサイバー攻撃、『Njw0rm』の改訂版「Kjw0rm」が関与か」(<http://blog.trendmicro.co.jp/archives/11271>)。

しています(OpC51)。インドでもインターネットに対する規制を強化しようとする政府に対する抗議として、電気通信規制庁を含む複数の政府機関へのDDoS攻撃や不正アクセスによるアカウント情報の漏えいなどの被害が発生しています。日本では5月に、クジラやイルカなどの海洋動物の商業利用への抗議からAnonymousによると考えられる不正アクセスにより、日本動物園水族館協会(JAZA)から登録者の電子メールアドレスなどの登録情報の漏えいが発生していたことが報道されています(OpSeaWorld)。

またSyrian Electronic Armyを名乗る何者かによるアカウントの乗っ取りやWebサイトの改ざんも継続して発生しており、被害を受けた企業としては米国の新聞社であるWashington Post社や米国陸軍なども含まれていました。これ以外にも、イタリア、ブラジル、中国など世界各国の政府とその関連サイトに対して、AnonymousなどのHacktivist達による攻撃が継続して行われました。また、政府機関など著名なSNSアカウントの乗っ取り事件も継続して発生しています。

■ 脆弱性とその対応

この期間中では、Microsoft社のWindows^{*3*4*5*6*7}、Internet Explorer^{*8*9*10}、Office^{*11}などで修正が行われました。Adobe社のAdobe Flash Playerでも修正が行われました。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。なお、Java 7については、今回の修正を持ってサポート終了の予定となっていることから、Java 8への移行が推奨されています。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る更新が提供され、多くの脆弱性が修正されました。CMSとして利用されるWordPressについても、複数のバージョンで任意のコード実行が可能なXSSの脆弱性が見つかり、修正されています。また、WordPressについてはプラグインでもXSSの脆弱性が見つかり、修正されています^{*12}。同じくCMSとして利用されるMovable Typeでも入力値のチェックが不十分なことからリモートでコード実行が可能な脆弱性が見つかり、修正されています^{*13}。

5月には米国のセキュリティ企業であるCrowdStrike社から、QEMUの仮想フロッピードライブコントローラにバッファオーバーフローの脆弱性(CVE-2015-3456)が発表され、修正されています。この脆弱性は、発見者によりVIRTUALIZED ENVIRONMENT NEGLECTED OPERATIONS MANIPULATION (VENOM)と名づけられましたが、クラウドサービスなどで

-
- *3 「マイクロソフト セキュリティ情報 MS15-034 - 緊急 HTTP.sysの脆弱性により、リモートでコードが実行される(3042553)」(<https://technet.microsoft.com/ja-jp/library/security/ms15-034.aspx>)。
 - *4 「マイクロソフト セキュリティ情報 MS15-035 - 緊急 Microsoft Graphicsコンポーネントの脆弱性により、リモートでコードが実行される(3046306)」(<https://technet.microsoft.com/ja-jp/library/security/ms15-035.aspx>)。
 - *5 「マイクロソフト セキュリティ情報 MS15-044 - 緊急 Microsoft フォントドライバーの脆弱性により、リモートでコードが実行される(3057110)」(<https://technet.microsoft.com/ja-jp/library/security/ms15-044.aspx>)。
 - *6 「マイクロソフト セキュリティ情報 MS15-045 - 緊急 Windows Journalの脆弱性により、リモートでコードが実行される(3046002)」(<https://technet.microsoft.com/ja-jp/library/security/ms15-045.aspx>)。
 - *7 「マイクロソフト セキュリティ情報 MS15-057 - 緊急 Windows Media Playerの脆弱性により、リモートでコードが実行される(3033890)」(<https://technet.microsoft.com/ja-jp/library/security/ms15-057.aspx>)。
 - *8 「マイクロソフト セキュリティ情報 MS15-032 - 緊急 Internet Explorer用の累積的なセキュリティ更新プログラム(3038314)」(<https://technet.microsoft.com/ja-jp/library/security/ms15-032.aspx>)。
 - *9 「マイクロソフト セキュリティ情報 MS15-043 - 緊急 Internet Explorer用の累積的なセキュリティ更新プログラム(3049563)」(<https://technet.microsoft.com/ja-jp/library/security/ms15-043.aspx>)。
 - *10 「マイクロソフト セキュリティ情報 MS15-056 - 緊急 Internet Explorer用の累積的なセキュリティ更新プログラム(3058515)」(<https://technet.microsoft.com/ja-jp/library/security/ms15-056.aspx>)。
 - *11 「マイクロソフト セキュリティ情報 MS15-033 - 緊急 Microsoft Officeの脆弱性により、リモートでコードが実行される(3048019)」(<https://technet.microsoft.com/ja-jp/library/security/ms15-033.aspx>)。
 - *12 詳細については、次の米国のセキュリティ企業であるSucuri社のBlogに詳しい。"Security Advisory:Persistent XSS in WP-Super-Cache"(<https://blog.sucuri.net/2015/04/security-advisory-persistent-xss-in-wp-super-cache.html>)。
 - *13 シックス・アパート株式会社、「[重要] 6.0.8、5.2.13 セキュリティアップデートの提供を開始」(<http://www.sixapart.jp/movabletype/news/2015/04/15-1045.html>)。

4月のインシデント

1	セ 1日:3月に発生したGitHubに対する大規模なDDoS攻撃について、中国への通信が関連していることを示すレポートが複数発表された。この攻撃の詳細については次のNETRESECの発表などに詳しい。"China's Man-on-the-Side Attack on GitHub"(http://www.netresec.com/?page=Blog&month=2015-03&post=China%27s-Man-on-the-Side-Attack-on-GitHub)。また、セキュリティ研究者のRobert Graham氏は次のレポートで中間者攻撃がどこで行われているかの考察を行っている。"Pin-pointing China's attack against GitHub"(http://blog.erratasec.com/2015/04/pin-pointing-chinas-attack-against.html#VahhVfntnK4)。
2	
3	
4	セ 8日:複数の上場企業が利用していた株主向けサービスサイトに登録している会員情報が漏えいした可能性があることが公表された。この事件では漏えいした情報を元にメールや電話による投資勧誘などが行われたことから発覚したが、その後の調査で不正アクセスなどではなく内部犯行であったことが公表されている。
5	他 8日:FBIより、ISILを名乗る何者かによるWordPressプラグインの脆弱性を悪用したWeb改ざんが相次いでいるとして注意喚起が行われた。The Internet Crime Complaint Center(IC3)、「ISIL DEFACEMENTS EXPLOITING WORDPRESS VULNERABILITIES」(http://www.ic3.gov/media/2015/150407-1.aspx)。
6	
7	
8	セ 9日:フランスのTV局の1つであるTV5MONDEグループがISILを名乗る何者かによる不正アクセスを受け、WebサイトやSNSサイトが乗っ取られるなどの被害が発生した。また、この攻撃ではTV放送自体が7時間中断するなどの影響を受けた。「TV5MONDEサイバー攻撃について」(https://japon.tv5monde.com/Resources/Articles/Events/2015/04/TV5MONDE%E3%82%B5%E3%82%A4%E3%83%8F%E3%83%BC%E6%94%BB%E6%92%83%E3%81%AB%E3%81%A4%E3%81%84%E3%81%A6?lang=ja-JP)。
9	
10	セ 10日:欧州刑事警察機構(Europol)は米連邦捜査局(FBI)など各国の捜査機関やセキュリティ企業と連携して、Beebone botnet(別名AAEHなど)を摘発したことを発表した。「International Police Operation Targets Polymorphic BEEBONE Botnet」(https://www.europol.europa.eu/content/international-police-operation-targets-polymorphic-beebone-botnet)。
11	
12	セ 10日:警視庁は、総務省やTelecom-ISAC Japan、民間業者と連携し、国内初となるインターネットバンキングに係るマルウェア(VAWTRAK)のテイクダウンを実施した。「ネットバンキングウイルス無力化作戦の実施について」(http://www.keishicho.metro.tokyo.jp/haiteku/haiteku/haiteku504.htm)。
13	
14	セ 14日:INTERPOLは、各国の捜査機関や民間企業と連携してSIMDAボットネットのテイクダウンを実施したことを発表した。この作戦は4月9日に、サイバー犯罪対策に関する研究開発や捜査支援を目的としてインターポールがシンガポールに設立したThe INTERPOL Global Complex for Innovation(IGCI)が統括して実施された。「INTERPOL coordinates global operation to take down Simda botnet」(http://www.interpol.int/News-and-media/News/2015/N2015-038)。
15	セ 14日:マレーシアのドメインである.myで、何者かによる不正アクセスにより、GoogleやYahooなど著名な企業のドメインがDNSハイジャックされる事件が発生した。MYNIC、「GOOGLE & YAHOO MALAYSIA DOMAIN BACK TO NORMAL WITHIN 24 HOURS」(https://www.mynic.my/upload_media/press_release.pdf)。
16	
17	
18	
19	脆 15日:Microsoft社は、2015年4月のセキュリティ情報を公開し、MS15-034など4件の緊急と7件の重要な更新を含む合計11件の修正をリリースした。「2015年4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/ms15-apr)。
20	
21	脆 15日:Adobe Flash Playerに、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。「APSB15-06: Adobe Flash Player用のセキュリティアップデート公開」(https://helpx.adobe.com/jp/security/products/flash-player/apsb15-06.html)。
22	脆 15日:Oracle社はOracleを含む複数製品について、四半期ごとの定例アップデートを公開し、Java SEの14件の脆弱性を含む合計98件の脆弱性を修正した。なお、Java7は今回の更新でサポートが終了となった。「Oracle Critical Patch Update Advisory - April 2015」(http://www.oracle.com/technetwork/topics/security/cpuapr2015-2365600.html)。
23	
24	他 15日:欧州委員会(EC)は、Google社が検索サービスでの支配的な地位を濫用して、自社のショッピングサービスを一般の検索結果より有利に扱っていたとして、EUの反トラスト法の規則に違反している疑いから異議告知書(Statement of Objection)を送付したことを発表した。また、併せて、Android mobile operating systemについても個別に調査することを発表している。詳細については次の"Antitrust:Commission sends Statement of Objections to Google on comparison shopping service"(http://europa.eu/rapid/press-release_MEMO-15-4781_en.htm)を参照のこと。Android mobile operating systemについては、「Antitrust:Commission opens formal investigation against Google in relation to Android mobile operating system」(http://europa.eu/rapid/press-release_MEMO-15-4782_en.htm)を参照のこと。
25	
26	
27	
28	脆 22日:WordPressに、権限のないユーザによる任意のコード実行が可能なXSS脆弱性を含む複数の脆弱性が見つかり、更新された。「WordPress 4.1.2 セキュリティリリース」(https://ja.wordpress.org/2015/04/22/wordpress-4-1-2/)。
29	
30	脆 28日:WordPressに、コメントの入力からサーバ上で任意のコードを実行することが可能なXSS脆弱性が見つかり、更新された。「WordPress 4.2.1 セキュリティリリース」(https://ja.wordpress.org/2015/04/28/wordpress-4-2-1/)。

※ 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
--------------	-------------------	---------------	-------------	--------------

利用されているXenやKVMなどの仮想化ソフトウェアにおいて、ゲスト側からホストへのDoS攻撃や任意のコード実行が可能だったことから影響範囲が広く話題となりました。

同じく5月には、TLSプロトコルのDH鍵交換に、中間者攻撃により暗号強度の低い暗号に格下げされることで、通信内容の盗聴や改ざんが可能となる脆弱性(CVE-2015-4000)が見つかり、修正されました。この脆弱性はLogjamと呼ばれ、多くのサーバやブラウザが影響を受ける可能性があることから、3月に公表されたSSL/TLSの実装における脆弱性であるFREAKと同様に話題となりました。

■ 標的型攻撃によるマルウェア感染と情報漏えい

この期間では、組織内部の端末へのマルウェア感染とそれによる情報漏えいなどの事件が相次ぎました。6月には、日本年金機構でメールに添付されていたマルウェアに感染したことで、101万人分の個人情報が入りこみました^{*14}。この事件がきっかけとなって、その後、複数の企業や団体、病院や大学などで昨年9月より継続して同様の事件が起きていたことが分かり、公表されています。これらの事件では共通して、フリーメールなどを利用して発信元を詐称し、マルウェアが添付されたいわゆる標的型メールによって感染したとされており、感染した後に端末自身のファイルや共有サーバなどのファイルを外部に送信していたことなどが分かっています。

同じようにメールによるマルウェア感染としては複合機からのメールを装った事例も報告されています^{*15}。また、これ以外にも複数の地方公共団体で、組織内部の端末へのマルウェア感染とそれによる情報漏えいが発生しています。これらの事件では、改ざんされたWebサイトへのアクセスによるいわゆる水飲み場型攻撃により感染した可能性が指摘されています。

IPAから、5月にサイバー情報共有イニシアティブ(J-CSIP)^{*16}の活動をまとめた、「サイバー情報共有イニシアティブ(J-CSIP)

2014年度 活動レポート」が公表されましたが、この中では、31ヵ月にもわたり、同一の攻撃者によると考えられる攻撃が継続して観測されていることが報告されています。このように、新たな手口・手法の出現や、使われるマルウェアの高機能化などにより攻撃は多様化しており、アンチウイルスソフトの導入などといった単体の対策だけでは対応が困難な状況となっています。このような既存のセキュリティ対策で対処しきれない攻撃への対策については、例えば、IPAから2014年9月に公開された、「『高度標的型攻撃』対策に向けたシステム設計ガイド」^{*17}などをにおいて検討されています。

■ 不正アクセスなどによる情報漏えい

不正アクセスによる情報漏えいも引き続き発生しています。5月には米国の美容関連企業のWebサイトが不正アクセスを受け、クレジットカード情報などが漏えいする事件が発生しています。また、同じく米国の保険会社では、内部のデータベースへの不正アクセスを受け、110万人分の個人情報が入りこみました。6月には、米国の内国歳入庁(IRS)から約10万人分の納税者に関する情報が漏えいする事件が発生しています。更に米国人事管理局が不正アクセスを受け、約400万人分の連邦政府職員の情報が漏えいする事件が発生しています。また、同じく6月には米国のパスワード管理サービスが不正アクセスを受け、認証情報の一部が漏えいした可能性があるとしてマスターパスワードの更新を呼びかける事件も発生しました。

日本でも、5月に、大手ISPが不正アクセスを受け、FTPアカウントなどが漏えいする事件が発生しています。4月には大学から内部資料がインターネット上に公開される事件も発生していますが、これはサーバ設定の不備により、公開されてしまったことが原因でした。6月には国立情報学研究所で利用していた動作検証用サーバへの不正アクセスが発生し、DDoS攻撃の踏み台として悪用される事件が発生していますが、こちらはパスワードを安易なものに設定したことによるものでした。株主向けサービスサイトから登録情報が漏えいした事件では、4月の

*14 日本年金機構、「日本年金機構の個人情報が流出したお客様へのお詫びについて」(<http://www.nenkin.go.jp/n/data/service/0000028648uArRENS1eQ.pdf>)。

*15 複合機からのメールを装ったメールについては昔から確認されているが、トレンドマイクロ社のセキュリティブログなどで6月に大幅に増加していることが報告されている。「複合機の通知を偽装したメールがマクロ型不正プログラムを頒布、日本でも被害」(<http://blog.trendmicro.co.jp/archives/11776>)。

*16 2011年に発足した重要インフラ企業などを中心とした参加組織間での情報共有と早期対応を行うための取り組み。「サイバー情報共有イニシアティブ(J-CSIP(ジェイシップ))」(<https://www.ipa.go.jp/security/J-CSIP/>)。

*17 IPA、「『高度標的型攻撃』対策に向けたシステム設計ガイド」の公開」(<https://www.ipa.go.jp/security/vuln/newattack.html>)。

5月のインシデント

1	脆 7日:WordPressに、権限のないユーザによる任意のコード実行が可能なXSS脆弱性を含む複数の脆弱性が見つかり、更新された。 「WordPress 4.2.2 セキュリティとメンテナンスのリリース」(https://ja.wordpress.org/2015/05/07/wordpress-4-2-2/)。
2	
3	他 12日:IPAより、SSL/TLSサーバの構築者や運営者が適切なセキュリティを考慮した暗号設定ができるようにするためのガイドラインとして、SSL/TLS暗号設定ガイドラインが公表された。 「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)～」(http://www.ipa.go.jp/security/vuln/ssl_crypt_config.html)。
4	
5	脆 13日:Microsoft社は、2015年5月のセキュリティ情報を公開し、MS15-043やMS15-044、MS15-045など3件の緊急と10件の重要な更新を含む合計13件の修正をリリースした。 「2015年5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/ms15-may)。
6	
7	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。 「APSB15-09: Adobe Flash Player用のセキュリティアップデート公開」(https://helpx.adobe.com/jp/security/products/flash-player/apsb15-09.html)。
8	
9	脆 14日:XenやKVMなどで使用されているQEMUのフロッピーディスクコントローラ(FDC)に、エラーによる異常終了や任意のコードを実行される可能性のある脆弱性(CVE-2015-3456)が見つかり、修正された。この脆弱性は発見者によりVENOMとの名称が付けられている。 詳細については次の発見者であるCrowdStrike社のWebサイトを参照のこと。「VENOM VIRTUALIZED ENVIRONMENT NEGLECTED OPERATIONS MANIPULATION」(http://venom.crowdstrike.com/)。
10	
11	
12	他 15日:IETFで2月に承認されたHTTP/2について、RFCとして公開された。 「Hypertext Transfer Protocol Version 2(HTTP/2)」(http://www.rfc-editor.org/rfc/rfc7540.txt)。
13	
14	脆 21日:TLSプロトコルのDH鍵交換に中間者攻撃により暗号強度の低い暗号を選択させることで通信内容の盗聴や改ざんが可能な脆弱性が見つかり、修正された。 詳細については次の発見者による解説を参照のこと。「The Logjam Attack」(https://weakdh.org/)。
15	セ 22日:複数の金融機関やオンラインショップなどに対し、脅迫を伴うDDoS攻撃が発生し、システムにアクセスしにくくなるなどの影響が出た。
16	他 22日:総務省は、情報セキュリティアドバイザーボードにて取りまとめられたサイバーセキュリティ政策推進に関する提言を公表した。 「『サイバーセキュリティ政策推進に関する提言』の公表」(http://www.soumu.go.jp/menu_news/s-news/01ryutsu03_02000093.html)。
17	
18	他 22日:IPAとJPCERT/CCは、脆弱性関連情報の円滑な流通及び対策の普及を図るための取り組みである、情報セキュリティ早期警戒パートナーシップについて、脆弱性関連情報を届出や通知を受けた際に参考となる、「情報セキュリティ早期警戒パートナーシップガイドライン」の2015年版を公開した。 IPA、「情報セキュリティ早期警戒パートナーシップガイドライン」(http://www.ipa.go.jp/security/ciadr/partnership_guide.html)。
19	
20	セ 25日:警察庁は、オンラインゲームのユーティリティツールを対象とした探査活動の増加について注意喚起を行った。 このツールにはプロキシ機能が実装されており、第三者が外部から利用できるいわゆるオープンプロキシの探査活動と考えられる。詳細については「特定のポートを対象としたプロキシ探査の増加について」(http://www.npa.go.jp/cyberpolice/detect/pdf/20150525.pdf)を参照のこと。
21	
22	他 25日:サイバーセキュリティ戦略本部の第2回会合が行われ、サイバーセキュリティ施策の基本的な方針について定める、新たなサイバーセキュリティ戦略(案)について決定した。 NISC、「第2回会合(平成27年5月25日)」(http://www.nisc.go.jp/conference/cs/index.html#cs02)。
23	
24	セ 26日:警察庁は、産業制御システムで使用される特定のPLCソフトウェアにおいて、脆弱性を狙った探査活動が増加しているとして注意喚起を行った。 「産業制御システムで使用されるPLCの脆弱性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/20150526.pdf)。
25	
26	セ 26日:JPCERTコーディネーションセンターより、Webサイトのコンテンツ改ざんと、それによる攻撃サイトへの誘導によるランサムウェアに感染させる被害が発生しているとして注意喚起を行った。 「ランサムウェア感染に関する注意喚起」(http://www.jpcert.or.jp/at/2015/at150015.html)。
27	
28	セ 27日:米国内国歳入庁(IRS)が不正アクセスを受け、社会保障番号など約10万人分の納税者に関する情報が漏えいする事件が発生した。 「IRS Statement on the "Get Transcript" Application」(http://www.irs.gov/uac/Newsroom/IRS-Statement-on-the-Get-Transcript-Application)。
29	
30	
31	

※ 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

公表時には、外部からの不正アクセスの可能性が疑われていましたが、5月に発表された最終報告では内部犯行だったことが公表されています。

■ 政府機関の取り組み

政府機関のセキュリティ対策の動きとしては、5月にはサイバーセキュリティ戦略本部の第2回会合が行われ、サイバーセキュリティ施策の基本的な方針について定める、新たなサイバーセキュリティ戦略(案)について決定しています。この中では、2020年のオリンピック・パラリンピック東京大会の開催も踏まえ、今後3カ年程度のサイバーセキュリティ政策について、目的達成に向けた研究開発の推進や人材育成も含めた各施策の方向性が示されています。今後、パブリックコメントによる修正などを経て、閣議決定が行われる予定です。

4月には総務省のICTサービス安心・安全研究会の下で開催されている「個人情報・利用者情報等の取扱いに関するWG」で検討が進められていた、電気通信事業における個人情報保護に関するガイドライン及び解説の改正について、改正案がまとめられました。この改正案はパブリックコメントによる修正を経て6月に改正が行われました。主な改正内容としては、個人情報の適正な取得における具体例が追記されたり、通信履歴について、接続認証ログの保存期間などの目安が示されたり、犯罪捜査時の位置情報の取得について、要件の削除が行われるなどしています。

■ その他

4月には、国際刑事警察機構(ICPO)がシンガポールに設置したサイバー犯罪対策専門組織IGCI(INTERPOL Global Complex for Innovation)が本格稼働しています^{*18}。これに先立ち、IGCIが統括し、各国の法執行機関や民間企業などが連携し、Simdaボットネットのテイクダウン作戦を実施しています。日

本でも警視庁や総務省、Telecom-ISAC Japan、民間企業などが連携してVAWTRAKのテイクダウン作戦が実施されました。

5月には、オープンソースソフトウェアのライブラリサイトであるSourceForgeで、公開されていたソフトウェアインストーラに許可なくサードパーティー製のソフトをバンドルして配布したことが分かり、開発プロジェクトから抗議される^{*19}と共に他の開発者やコミュニティから大規模な反発を受ける事態となりました。SourceForgeでは否定的な反応が多かったとしてこの方針を撤回しています^{*20}。

また、5月から6月にかけて複数の金融機関やオンラインショップなどに対し、脅迫を伴ったDDoS攻撃が発生しています。これらの攻撃では、脅迫の際にビットコインによる支払いを要求しており、拒否した場合には数百GbpsのDDoS攻撃を行う用意があることをほのめかしていたとされています。DDoS攻撃とビットコインによる支払いという点で、昨年と同様の攻撃を行っているDD4BCを名乗るグループが関与していると考えられ^{*21}、ヨーロッパや香港などの金融機関への攻撃事例もあることから今後も注意が必要です。

6月には、2014年12月に発生した出版社のWebサイトが不正アクセスを受け、別のWebサイトに誘導される事件に関連しているとして、未成年の少年が逮捕されています。この少年については7月に他人のクレジットカードを不正利用したとして再逮捕されています。

IP電話などの電話サービスが第三者に不正に利用され、高額な国際電話料金を請求される事例が多くなっているとして、総務省から注意喚起が行われました。その後、セキュリティ企業から特定のベンダー製品で、パスワードが初期設定のまま運用さ

*18 "International gathering marks inauguration of INTERPOL Global Complex for Innovation" (<http://www.interpol.int/News-and-media/News/2015/N2015-039>)。

*19 GIMP PROJECT, "GIMP PROJECT'S OFFICIAL STATEMENT ON SOURCEFORGE'S ACTIONS" (<http://www.gimp.org/>)。

*20 SourceForge Community Blog, "Third party offers will be presented with Opt-In projects only" (<http://sourceforge.net/blog/third-party-offers-will-be-presented-with-opt-in-projects-only/>)。

*21 DD4BCについては、例えば次のArbor Networks社のレポートを参照のこと。"DD4BC DDoS Extortion Threat Activity" (<https://asert.arbornetworks.com/dd4bc-ddos-extortion-threat-activity/>)。

6月のインシデント

1	セ 1日: 日本年金機構はマルウェア感染による不正アクセスにより最大で125万件の個人情報が流出した可能性があることを発表した。 「日本年金機構の個人情報流出について」(http://www.nenkin.go.jp/n/data/service/0000150601ndjlleouli.pdf)。
2	
3	他 3日: 米国家安全保障局(NSA)による監視活動を改革する内容が含まれている米国自由法案(USA Freedom Act)が米国上院で可決し、大統領が署名して成立した。この法律では米国民を含む通話記録の大規模な収集活動を世界中で行っていたとして問題となっていたNSAに対し、情報収集活動を制限する内容となっている。 Congress.gov, "H.R.2048 - USA FREEDOM Act of 2015"(https://www.congress.gov/bill/114th-congress/house-bill/2048/text)。
4	
5	セ 5日: 米人事管理局(OPM)が不正アクセスを受け、約400万人分の連邦政府職員の情報に漏えいする事件が発生した。 "OPM to Notify Employees of Cybersecurity Incident"(https://www.opm.gov/news/releases/2015/06/opm-to-notify-employees-of-cybersecurity-incident/)。
6	
7	セ 5日: 国立情報学研究所は、動作検証用サーバが不正アクセスを受け、DoS攻撃の踏み台になっていたことを発表した。 「動作検証用サーバへの不正アクセスについて」(http://www.nii.ac.jp/news/2015/0605)。
8	他 5日: 企業が過去に利用していたドメインが第三者に取得され、詐欺サイトなど別のサイトへ誘導される事例が確認されたことから、合併した新会社より注意喚起が行われた。
9	
10	脆 10日: Microsoft社は、2015年6月のセキュリティ情報を公開し、MS15-056とMS15-057の2件の緊急と6件の重要な更新を含む合計8件の修正をリリースした。 「2015年6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/ms15-Jun)。
11	
12	脆 10日: Adobe Reader及びAcrobat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。 「APSB15-10: Adobe ReaderおよびAcrobat用セキュリティアップデート公開」(https://helpx.adobe.com/jp/security/products/reader/apsb15-10.html)。
13	
14	他 12日: 総務省は、IP電話をはじめとする電話サービスが第三者に不正利用され、利用者に高額な国際電話料金の請求される事件が発生しているとして注意喚起を行った。 PBXやIP電話を利用するために接続するルータの脆弱性を悪用したり、安易なパスワードを使用していたことからIDとパスワードを盗まれて国際通話に悪用されている事例があるとしている。「第三者によるIP電話等の不正利用に関する注意喚起」(http://www.soumu.go.jp/menu_kyotsuu/important/kinkyu02_000191.html)。
15	
16	セ 16日: パスワードの一元管理ツールを提供している米国のLastPass社は、不正アクセスを受け、認証ハッシュやユーザアカウント情報の一部などが流出したことを発表した。 "LastPass Security Notice"(https://blog.lastpass.com/ja/2015/06/lastpass-security-notice.html/)。
17	
18	他 16日: フィッシング対策協議会は、SMS(ショートメッセージサービス)を使った、銀行のフィッシングサイトへ誘導する手口が5月下旬より複数の銀行で確認されているとして注意喚起を行った。 「【注意喚起】SMS(ショートメッセージサービス)で誘導される銀行のフィッシングサイトにご注意ください(2015/06/16)」(https://www.antiphishing.jp/news/alert/_sms_20150616.html)。
19	
20	
21	脆 17日: Samsung社製の携帯端末のSwiftKey SDKを用いたキーボード機能に、任意のコードを実行可能な複数の脆弱性が見つかり、修正された。 JVN、「JVNVU#94598171 Samsung Galaxy SにプリインストールされたSwiftKeyが言語バックのアップデートを正しく検証しない脆弱性」(https://jvn.jp/vu/JVNVU94598171/index.html)。
22	
23	セ 22日: LOTポーランド航空で、フライトプランコンピュータへ攻撃を受け、航空機が欠航や遅延となる事件が発生した。 LOT Polish Airlines, "TODAY AFTERNOON LOT ENCOUNTERED IT ATTACK, THAT AFFECTED OUR GROUND OPERATION SYSTEMS."(http://corporate.lot.com/pl/en/press-news?article=772922)。
24	
25	脆 24日: Adobe Flash Playerに、任意のコード実行の可能性がある脆弱性が見つかり、修正された。この脆弱性を悪用した限定的な標的型攻撃が既に発生していることが確認されている。 「APSB15-14: Adobe Flash Player用のセキュリティアップデート公開」(https://helpx.adobe.com/jp/security/products/flash-player/apsb15-14.html)。
26	
27	他 24日: 総務省は、電気通信事業における個人情報保護に関するガイドライン及び解説の改正を行った。 「電気通信事業における個人情報保護に関するガイドライン」(http://www.soumu.go.jp/main_sosiki/joho_tsusin/d_syohi/telecom_perinfo_guideline_intro.html)。
28	
29	セ 25日: ある銀行のオンラインシステムに脅迫を伴うDDoS攻撃が発生し、システムにアクセスしにくくなるなどの影響が出た。
30	

※ 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

れていたり、インターネットからのアクセスを許可していたことにより被害が多く発生しているとの指摘が行われましたが、ベンダーは否定しています。この問題については7月になって、総務省より、電気通信事業におけるサイバー攻撃への適正な対処のあり方に関する研究会での議論を踏まえ、不正利用による被害を未然に防止し、被害の拡大を防ぐために、電気通信事業者関係団体に対し、利用者との間で国際電話サービスを提供する契約を締結している電気通信事業者等が適切な対応を講じるよう周知することへの協力要請が行われています*22。

SSL/TLSサーバでは、相次いで複数の脆弱性が明らかとなるなどしています。一方で最新の対策を行うことで古い機器からの接続ができなくなるなどの問題もあり、安全性と必要となる相互接続性とのトレードオフを考慮した設計や運用が求められています。これに答える形で、5月にIPAより、SSL/TLSサーバの構築者や運営者が適切なセキュリティを考慮した暗号設定ができるようにするためのガイドラインである、「SSL/TLS暗号設定ガイドライン」が公表されています。

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、2015年4月から6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*23、サーバに対する攻撃*24、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。

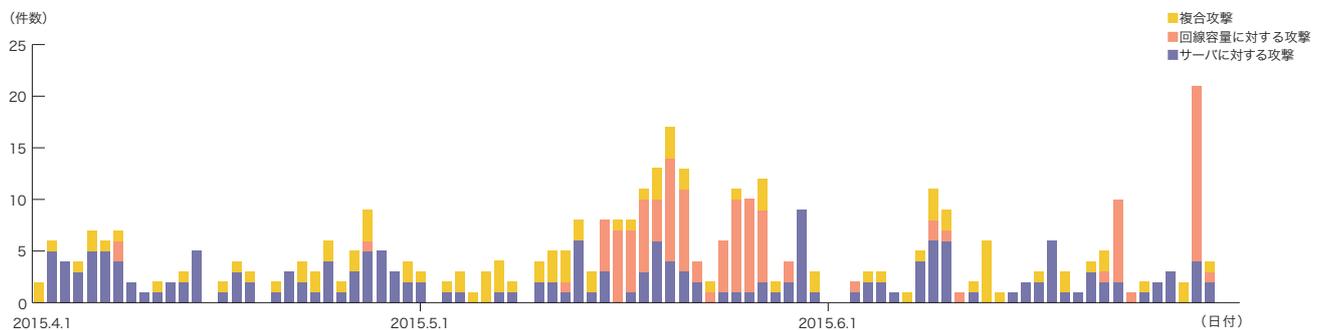


図-2 DDoS攻撃の発生件数

*22 総務省、「第三者によるIP電話等の不正利用への対策について(要請)」(http://www.soumu.go.jp/menu_news/s-news/01ryutsu03_02000096.html)。
 *23 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と呼ばれる、ICMPパケットを利用した場合にはICMP floodと呼ばれる。
 *24 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利用させる。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。

この3ヵ月間でIJJは、361件のDDoS攻撃に対処しました。1日あたりの対処件数は3.97件で、平均発生件数は前回のレポート期間と比べて減少しました。DDoS攻撃全体に占める割合は、サーバに対する攻撃が52.1%、複合攻撃が14.9%、回線容量に対する攻撃が33%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271万3千ppsの packets によって8.25Gbps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77.3%が攻撃開始から30分未満で終了し、22.7%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ありませんでした。なお、今回もっとも長く継続した攻撃は、複合攻撃に分類されるもので12時間39分にわたりました。

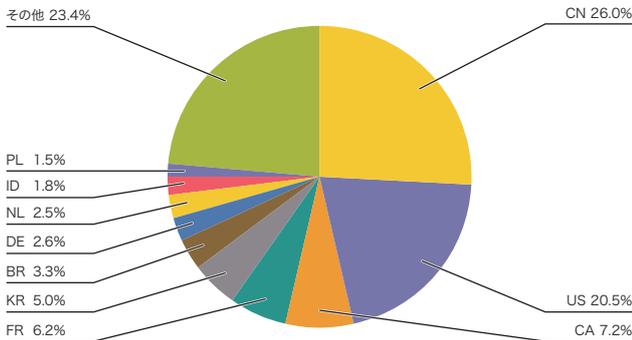


図-3 DDoS攻撃のbackscatter観測による攻撃先の国別分類

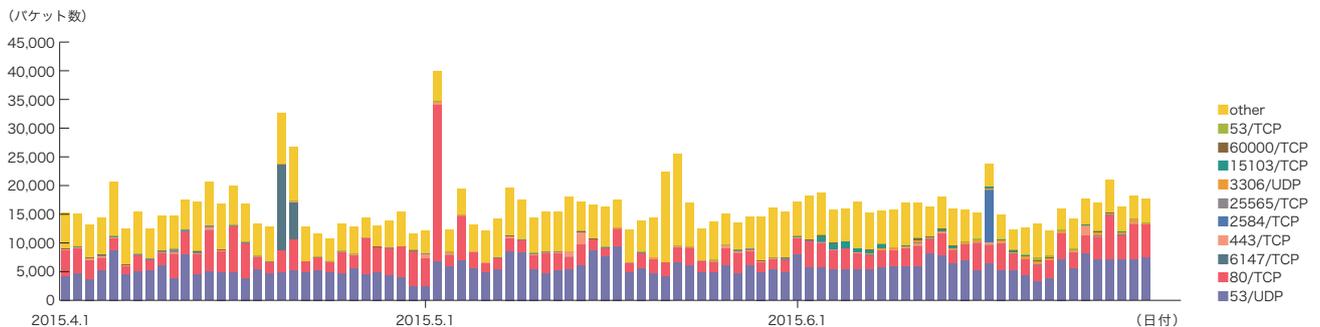


図-4 DDoS攻撃によるbackscatter観測(観測パケット数、ポート別推移)

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^{*25}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^{*26}の利用によるものと考えられます。

■ backscatterによる観測

次に、IJJ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^{*27}によるDDoS攻撃のbackscatter観測結果を示します^{*28}。backscatterを観測することで、外部のネットワークで発生した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。

2015年4月から6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のはDNSで利用される53/UDPで、全パケット数の34.5%を占めています。次いでWebサービスで利用される80/TCPが22.3%を占めており、上位2つで全体の56.8%に達しています。また、HTTPSで利用される443/TCP、DNSで利用される53/TCPへの攻撃、ゲームサーバで利用されることがある25565/TCP、通常は利用されない6147/TCPや2584/TCPなどへの攻撃が観測されています。

*25 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。

*26 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。

*27 IJJ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。

*28 この観測手法については、本レポートのVol.8 (http://www.ijj.ad.jp/development/iir/pdf/iir_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、IJJによる観測結果の一部について紹介している。

2014年2月から多く観測されている53/UDPは、1日平均のパケット数を見ると、前回の約6,200から減少して約5,600になりましたが、高止まりの状態にあります。

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、中国の26.0%が最も大きな割合を占めています。その後に米国の20.5%、カナダの7.2%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)への攻撃としては、4月26日から5月1日にかけて米国ホスティング事業者への攻撃、5月2日にはカナダのホスティング事業者のサーバ群に対する攻撃を観測しています。他のポートへの攻撃としては、4月19日から20日にかけて6147/TCPへの攻撃が観測されましたが、backscatterの発信元IPアドレスがプライベートアド

レスであるため、攻撃対象は不明です。6月17日には米国ISPのサーバに対する2584/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、IJJのbackscatter観測で検知した攻撃としては、4月13日から15日にかけてイギリスのオンラインカジノへの攻撃、5月5日にフランスの原子力関連企業への攻撃、6月18日と28日に複数のカナダ政府機関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、IJJが実施しているマルウェアの活動観測プロジェクトMITF^{*29}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^{*30}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。そのほとんどがマルウェアによる無作為に宛先を選んだ通信か、攻撃先を見つけるための探索の試みであると考えられます。

■ 無作為通信の状況

2015年4月から6月の期間中に、ハニーポットに到着した通信の発信元IPアドレスの国別分類を図-5に、その総量(到着パケット数)の推移を図-6に、それぞれ示します。MITFでは、数多くのハニーポットを用いて観測を行っていますが、ここでは1台あたりの平均を取り、到着したパケットの種類(上位10種類)ごとに推移を示しています。また、この観測では、MSRPCへの攻撃のような特定のポートに複数回の接続を伴う攻撃は、複数のTCP接続を1回の攻撃と数えるように補正しています。

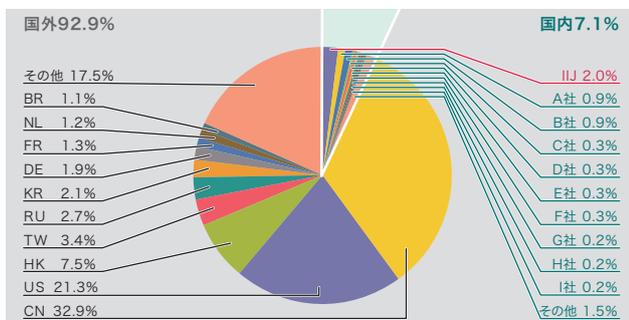


図-5 発信元の分布(国別分類、全期間)

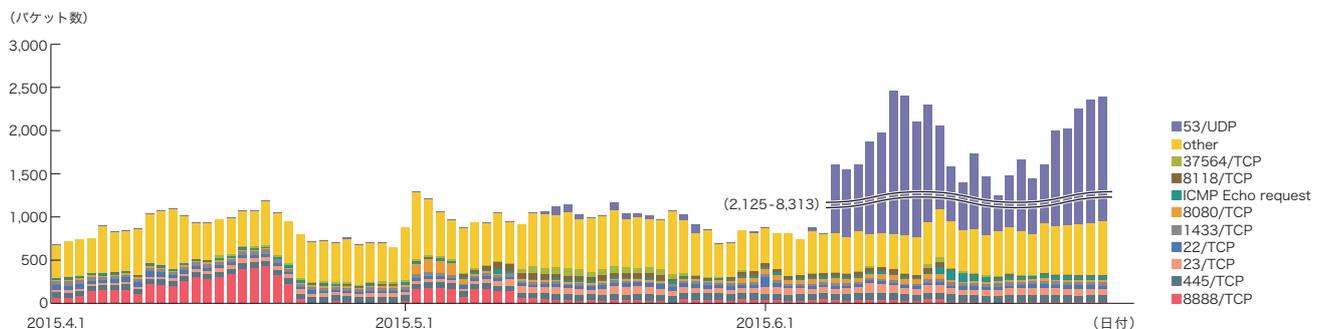


図-6 ハニーポットに到着した通信の推移(日別・宛先ポート別・1台あたり)

*29 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。

*30 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。

本レポートの期間中にハニーポットに到着した通信の多くは、多くがDNSの水責め攻撃を試みるものや、パケットの応答を増幅してDoS攻撃を行う機器を探査する通信、Proxyサーバの探査を行う通信、Microsoft社のOSで利用されているTCPポートに対する探索行為でした。6月7日から6月30日にかけて、ハニーポットの1つに対してDNSサーバで使用される53/UDPへの大量の通信が発生しています。これらの通信の発信元のIPアドレス分布は米国、中国が中心であるものの、ほとんどが重複しておらず広範囲だったことから、botが使われたか発信元が詐称された可能性があります。この通信内容を調査したところ、ランダム文字列、www.example.comのような存在しないAレコードの問い合わせが繰り返し行われていました。このことから、DNS水責め攻撃(DNS Water Torture)^{*31}であると考えられます。攻撃対象となったドメインのいくつかは中国のサーバ(.comドメイン)でした。

また、いくつかのIPアドレスからはDNSのTXTレコードの問い合わせと共に、multicast DNSで使用される5353/UDPや、NTPで使用される123/UDP、NAT-PMPで使われる5351/UDPなどを探査する行為も行われていました。これらは攻撃したいサイトを送信元に詐称してパケットを送信し、応答パケットを増幅するタイプのリフレクション攻撃を行うことが可能な機器を探査していたと考えられます。また、5351/UDPに関しては、一部の機器が外部からNAT-PMPを操作可能であったとする報告^{*32}がなされており、脆弱な機器を探査する行為であった可能性もあります。

4月中、及び5月初旬にかけて、8888/TCPの通信が増加しています。この通信は中国に割り当てられた2つのIPアドレスから広範囲のIPアドレスに対して繰り返し行われており、8888/TCP以外にもオンラインゲーム用ユーティリティツール KanColleViewer で使われる37564/TCP^{*33}、Privoxyと呼ばれるProxyサーバで使用される8118/TCP^{*34}、Squidなど、いくつかのProxyサーバの初期設定である3128/TCP、Polipoと呼ばれるProxyサーバで使用される8123/TCP、8080/TCPの代替として使われる8090/TCPなどのポートに対する探査行為も同時に行われていたため、Open Proxyサーバを探査する行為であったと考えています。

■ 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-7に、マルウェアの総取得検体数の推移を図-8に、そのうちのユニーク検体数の推移を図-9にそれぞれ示します。このうち図-8と図-9では、1日あたりに取得した検体^{*35}の総数を総取得検体数、検体の種類をハッシュ値^{*36}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-8と図-9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91、ユニーク検体数が17でした。未検出の検体をより詳しく調査した結

*31 Secure64 Software Corporation, "Water Torture: A Slow Drip DNS DDoS Attack" (<https://blog.secure64.com/?p=377>)。日本語での解説としては、株式会社日本レジストリサービス 森下氏による次の資料が詳しい。「DNS水責め(Water Torture)攻撃について」(http://2014.secon.jp/dns/dns_water_torture.pdf)。MITFハニーポットはDNSの問い合わせパケットを受信しても、権威サーバやキャッシュサーバに問い合わせに行かないため、攻撃には加担していない。

*32 「JVNVU#99291862複数のNAT-PMPデバイスがWAN側から操作可能な問題」(<https://jvn.jp/vu/JVNVU99291862/>)。また、過去にはこのポートに対する探査行為についても報告されている。「外部からNAT-PMPの操作が可能である機器の探索行為について」(<http://www.npa.go.jp/cyberpolice/detect/pdf/20141030.pdf>)。

*33 KanColleViewerに関しては、初期設定がOpen Proxyになっていたとして、JVNなどが注意喚起をしている。「JVNVU#98282440「提督業も忙しい!」(KanColleViewer)がオープンプロキシとして動作する問題」(<https://jvn.jp/vu/JVNVU98282440/>)。また、警察庁がこのポートに対する探査の増加を報告している。「特定のポートを対象としたプロキシ探索の増加について」(<http://www.npa.go.jp/cyberpolice/detect/pdf/20150525.pdf>)。

*34 過去にはこのポートのアクセス増加が報告されている。「インターネット観測結果等(平成27年1月期)」(<https://www.npa.go.jp/cyberpolice/detect/pdf/20150304.pdf>)。

*35 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。

*36 様々な入力に対して一定長の出力をする一方向性関数(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に設計されている。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。

果、中国、米国、インド、台湾、オーストリアなどに割り当てられたIPアドレスでWormなどが観測されました。また、未検出の検体の約55%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。

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90.0%、ポット型1.8%、ダウンロード型8.2%でした。また解析により、106個のポットネットC&Cサーバ^{*37}と9個のマルウェア配布サイトの存在を確認しました。ポットネットのC&Cサーバの数が以前よりも高くなってはいますが、これはDGA(ドメイン生成アルゴリズム)を持つ検体が期間中に出現したためです。

■ 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27,999、ユニーク検体数は549でした。短期間での増減を繰り返しながらも、総取得検体数で99.7%、ユニーク検体数で97.0%を占めています。このように、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し、約44%増加し、ユニーク検体数は前号から約10%減少しました。総取得検体数の増加は、本レポートの期

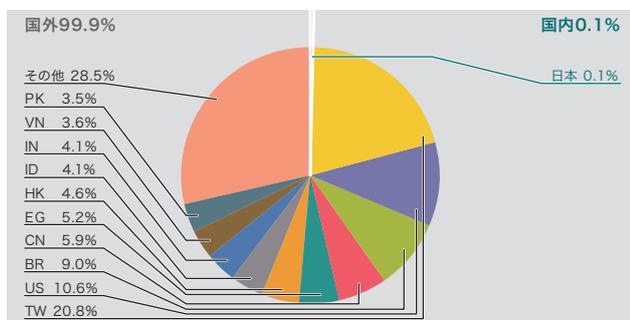


図-7 検体取得元の分布 (国別分類、全期間、Confickerを除く)

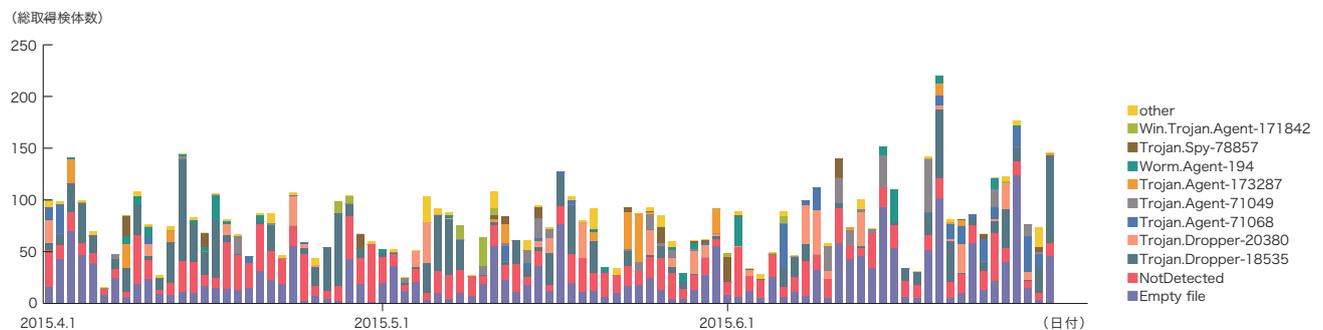


図-8 総取得検体数の推移 (Confickerを除く)

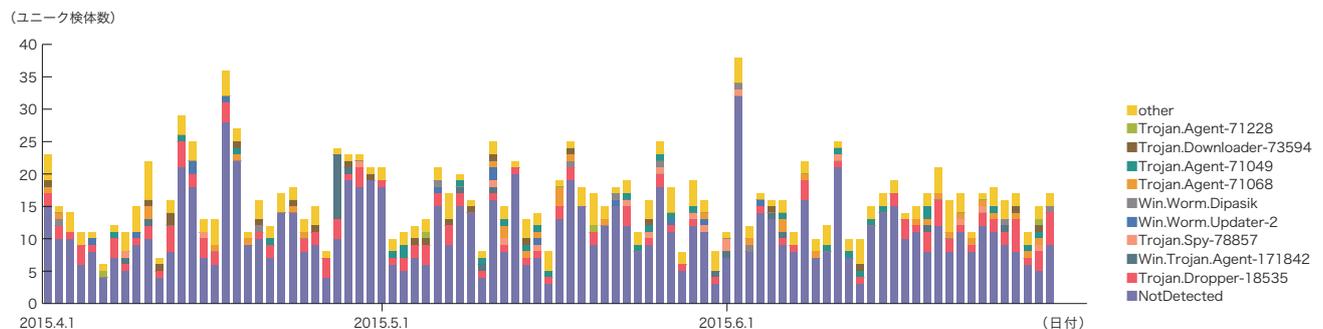


図-9 ユニーク検体数の推移 (Confickerを除く)

*37 Command & Controlサーバの略。多数のポットで構成されたポットネットに指令を与えるサーバ。

間中、米国に割り当てられたIPアドレスからの感染活動が増加したためです。Conficker Working Groupの観測記録^{*38}によると、2015年7月1日現在で、ユニークIPアドレスの総数は775,060とされています。2011年11月の約320万台と比較すると、約24%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。

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^{*39}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。

2015年4月から6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-10に、攻撃の推移を図-11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、中国63.3%、日本20.7%、米国7.6%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発生件数は前回に比べて大幅に減少しました。これは中国からの大規模な攻撃が減少したためですが、中国からの攻撃は依然として継続して発生しており、多くの割合を占めています。

この期間中、4月22日から4月23日にかけて中国の特定の攻撃元から特定の攻撃先に対する攻撃が発生していました。5月14日には中国の別の特定の攻撃元より特定の攻撃先に対する攻撃が発生しています。6月21日には米国やヨーロッパ各国の複数の攻撃元より特定の攻撃先に対する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。

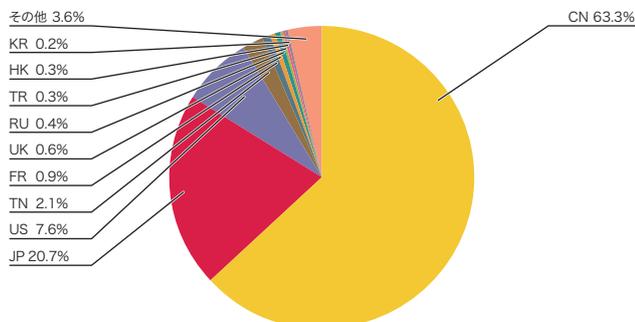


図-10 SQLインジェクション攻撃の発信元の分布

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。

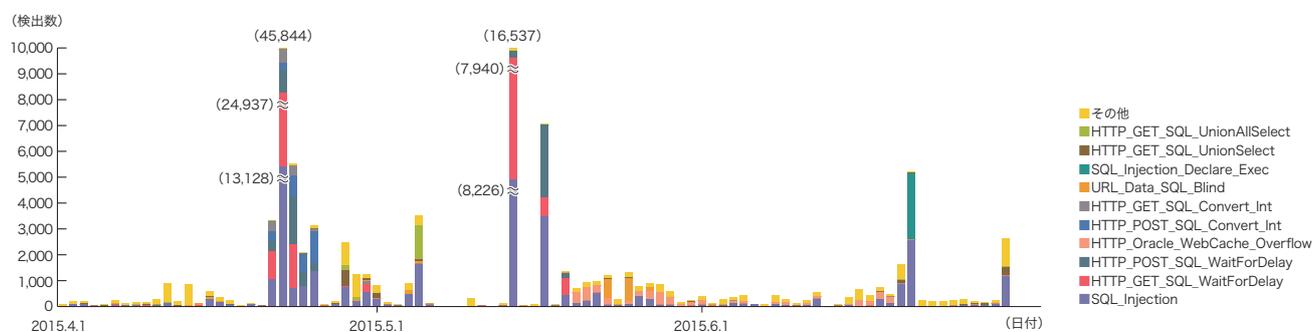


図-11 SQLインジェクション攻撃の推移(日別、攻撃種類別)

*38 Conficker Working Groupの観測記録 (<http://www.confickerworkinggroup.org/wiki/pmwiki.php/ANY/InfectionTracking/>)。

*39 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。

1.3.4 Webサイト改ざん

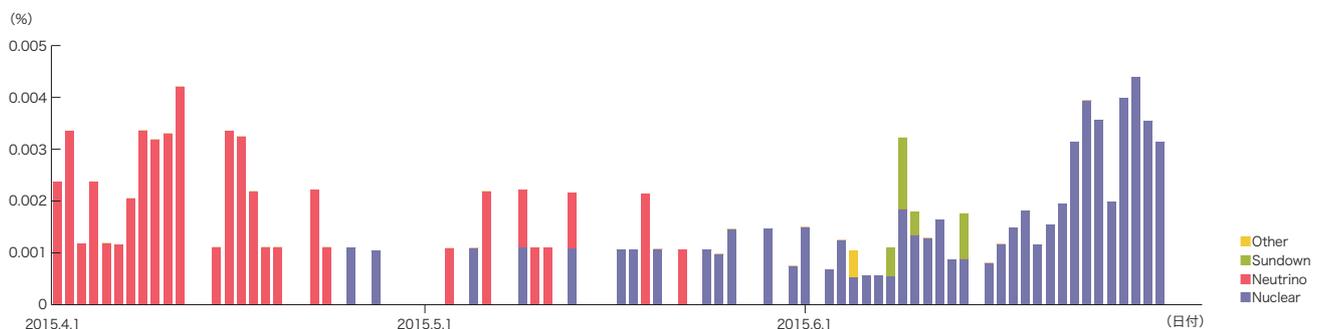
MITFのWebクローラ(クライアントハニーポット)によって調査したWebサイト改ざん状況を示します*40。

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。

2015年4月から6月の期間に検知されたドライブバイダウンロードは、2015年1月から3月の期間の2倍以上に増加しています。前半はNeutrinoが多く観測されましたが、後半はほとんどNuclearへとシフトしています(図-12)。また、6月中旬の一時期、Sundownという新たなExploit Kitを検知しました。Sundownは特に日本国内のユーザを対象とした攻撃に用いられており*41、MITFにおける観測時には、CVE-2015-0311やCVE-2015-0313などのFlashの脆弱性を悪用して最終的にKasidetなどのマルウェアのダウンロードが発生したことを確認しました。

2015年7月1日にWebクローラシステムの更新を行い、機能の追加や構成の変更を実施したところ、MITFでは2015年1月以来検知していなかったAnglerによる攻撃が観測されるようになりました。Anglerの攻撃は他の攻撃の合計よりも多く、7月1日以降は攻撃検知の総数が3倍近くに増加しています。このことから、6月30日以前の期間においても、Anglerが日本国内のユーザを対象とした攻撃に用いられていた可能性が極めて高いものと考えられます。Anglerによる攻撃の観測状況や、その内容については「1.4.2 猛威を振るうAngler Exploit Kit」も併せて参照ください。

全体として、ドライブバイダウンロードによる攻撃が比較的多く発生している状況です。2015年7月初旬にはExploit Kitで悪用可能な複数の0day脆弱性(CVE-2015-5119、CVE-2015-5122など)が公開され、前述のAnglerやNuclear、NeutrinoなどのExploit Kitが極めて短時間でそれらを悪用する機能を取り込んでいます。ブラウザ利用環境では、OSやブラウザ関連プラグインの脆弱性をよく把握し、更新の適用やEMETの有効化などの対策を徹底することを推奨します。また、Webサイト運営者はWebコンテンツの改ざん対策に加え、外部の第三者から提供されるマッシュアップコンテンツの健全性についても確認することが重要です。



※注 調査対象は日本国内の数十万サイト。近年の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。

図-12 Webサイト閲覧時のドライブバイダウンロード発生率(%)(Exploit Kit別)

*40 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。

*41 Sundownについては「Fast look at Sundown EK」(<http://malware.dontneedcoffee.com/2015/06/fast-look-at-sundown-ek.html>)にてコントロールパネルなどから得られた情報を踏まえてその機能や攻撃先について言及されている。

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、機械学習とセキュリティ、猛威を振るうAngler Exploit Kit、ID管理技術の実際の利用についての3つのテーマについて紹介します。

1.4.1 機械学習とセキュリティ

機械学習は古くから研究されている人工知能の一分野で、人が経験から学習していくように、既知のデータから自動的に出力が改善されていくアルゴリズムを研究・開発する分野です。インターネット上に大量のデータが蓄積され、また、クラウドなどで安価に大量の計算資源を利用できるようになると連動するようにして、機械学習の学術的成果やITシステムへの応用が話題に上ることが多くなりました。

従来、判断基準を明示できる問題ではそれをプログラム化することで自動化が進められてきました。機械学習の手法を用いれば、基準の明示が難しい問題でも人による判断を学習させ模倣させることができ、自動化の促進に役立つと期待されています。

セキュリティの分野では、ブラックリスト、ホワイトリストやシグネチャといった明示的な基準による自動判断を行ってきましたが、それだけでは判断が容易でない多数の対象に機械学習の手法が有効ではないかと期待されています。また、セキュリティ担

当者を育成することの難しさや、担当者が日常的に行う判断作業の負担が課題として認識される中で、判断作業を自動化あるいは支援するための技術として機械学習が注目されています。

本稿では、まず機械学習とそのセキュリティ分野への応用例を概観した後、いくつかの種類の特徴について機械学習を適用することで対策が可能かどうかを検討します。また、機械学習を組み込んだシステム自体のセキュリティリスクについて考えます。

■ 機械学習とは

はじめに機械学習の主要な方式について説明します(図-13)。機械学習の利用にあたっては、あらかじめ入力と期待される出力の組からなる訓練データを用意しておきます^{*42}。この訓練データを使って人による判断を事前に「学習」させておき、自動的な判断を行わせたいシステムに判断用アルゴリズムと学習結果を組み込んで利用します。

判断用アルゴリズムは一般に多数の内部パラメータを備えており、設定値に応じて出力が様々に変化するように設計されています。先述した「学習」とは、本質的には、訓練データに沿った適切な設定値を求める計算のことです。判断用アルゴリズムとその設定値の計算手順「学習アルゴリズム」をセットとして、ニューラルネットワーク、サポートベクタマシン(SVM)、決定木学習など様々な手法が開発、研究されています。

実際の問題に適用するにあたっては、対象となる問題とデータを分析して、それに合った機械学習アルゴリズムを選択し、そのアルゴリズムが学習・判断しやすいように入力の前処理(統計的処理、テキスト処理など)を実装する必要があります。

■ セキュリティ分野への応用

既に機械学習の手法を組み込んだセキュリティ製品も存在します。例えば、スパムメールの判定に使われるベイジアンフィルタは、単純ベイズ分類器という機械学習の手法を応用した技術です。他にも、通信ログを学習して攻撃やマルウェアなどの異常を検知する製品、通信パケットを学習してマルウェアの通信を検知する製品などがあります。マルウェア検知の分野では、マル

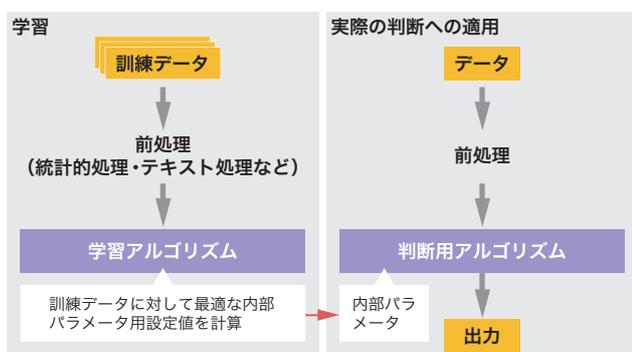


図-13 教師あり機械学習

*42 ここでは教師あり学習と呼ばれる方式について説明する。他に教師なし学習と呼ばれる方式もある。

ウェアのバイナリや振る舞いの特徴を学習させ、その結果を検出エンジンに組み込んだクライアント機向け製品が存在します。

個別の製品だけでなく、インシデント対応の現場に機械学習を応用している事例もあります。インドネシアのId-SIRTIIでは、IDSを多数導入している環境で、シグネチャでは対応していない攻撃についてそれぞれ機械学習による検知エンジンを開発し、自らの状況に合わせた攻撃検知システムを構築しています*43。また、セキュリティ担当者にとって情報収集の作業は欠かせません。他の企業や組織で発生したセキュリティ上の事件・事故の情報を日常的に収集し分析することもその1つです。オランダのNCSC-NLでは担当者の分析作業を支援するために、収集したニュース記事に対して機械学習を適用し、同一事象に関する記事のグループ化と、カテゴリ別のタグ付け(DDoS攻撃、情報漏えいなど)を自動で行うシステムを開発しています*44。

ここまで、機械学習の手法を組み込んだ個別のシステムについて見てきました。次に、現実の脅威として3つ、内部犯行、標的型攻撃、Webサービスへの攻撃を例にとり、それぞれ機械学習を適用することで有効な対策ができるかどうか検討します。

■ 脅威への対策

第一の例として、内部犯行の脅威に対して機械学習の適用が有効かどうかを検討します(図-14)。内部犯行の形態とし

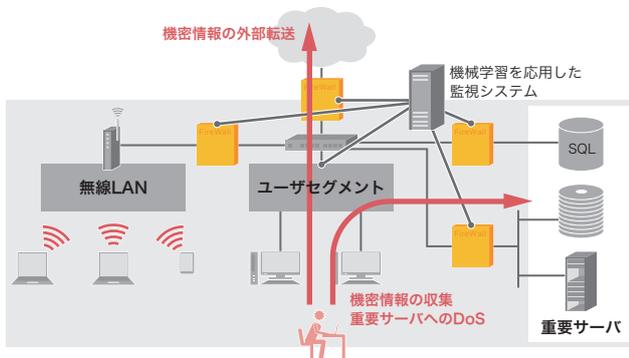


図-14 内部犯行対策への機械学習の適用

ては、機密データの持ち出しや重要なサーバへのDoSが考えられます。

まず、組織内ネットワークのトラフィックから機械学習により、正常時のユーザごとの通信プロファイルを学習させるとします。ここでは、通信の宛先、量、種類(メール、Web、ドキュメント転送など)といった情報を通信プロファイルと呼ぶことにします。通常の業務で発生する通信プロファイルをすべて列挙して、将来にわたり変化に逐一追従させる作業を人手で行うことは非常に困難ですが、機械学習ではこれを自動化できます。

その上で、正常なプロファイルから外れる通信を検知するシステムを構築し、トラフィックを監視させます。このようにすれば、内部者が重要サーバから通常では考えられない程大量のデータを取得したり、本来業務では利用しない重要サーバからデータを取得したりといった、機密情報の不正収集が疑われる行為を検知できます。社外への大量データ転送という、不正持ち出しが疑われる行為もまた、正常な通信プロファイルからの逸脱として検知できます。また、重要サーバにDoSをしかけようとして大量アクセスを行ったとしても、同様に検知できます。

第二の例として、標的型攻撃の脅威について検討します(図-15)。内部犯行対策の例と同様にユーザごとの通信プロファイルを学習させておきます。また、標的型攻撃マルウェアのバイナリ

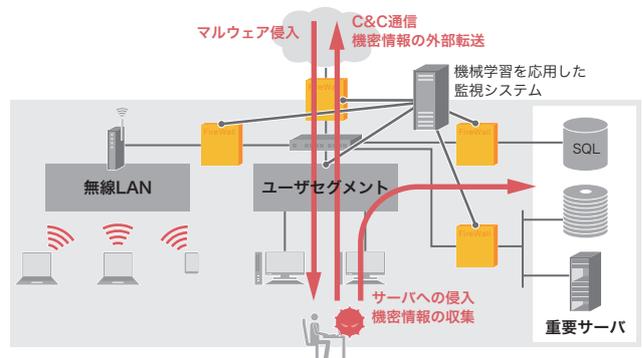


図-15 標的型攻撃対策への機械学習の適用

*43 Bisyron MASDUKI (Id-SIRTII), Muhammad SALAHUDDIEN (Id-SIRTII), "Implementation of Machine Learning Methods for Improving Detection Accuracy on Intrusion Detection System (IDS)" (<http://www.first.org/conference/2015/program#pimplementation-of-machine-learning-methods-for-improving-detection-accuracy-on-intrusion-detection-system-ids>).

*44 Edwin TUMP (NCSC-NL), "Machine Learning for Cyber Security Intelligence" (<http://www.first.org/conference/2015/program#pmachine-learning-for-cyber-security-intelligence>).

的特徴を学習させておきます。マルウェアの通信特性を学習させておくことも有用かもしれません。そして学習結果を元にトラフィックの監視と、メールやWebなどの外部との通信に含まれるバイナリの検査を行わせます。このようにすれば、標的型攻撃マルウェアが監視点を通過して侵入することを検知できます。仮に侵入を許しクライアントPCがマルウェアに感染したとしても、感染拡大、サーバへの侵入、機密データ収集などの組織内通信が正常な通信プロファイルからの逸脱として検知できます。更に、C&Cサーバとの通信や機密データの転送といった外部との通信も同様に検知できます。

最後の例として、外部向けに公開しているWebサービスの保護について検討します(図-16)。公開Webサービスでは、SQLインジェクション攻撃や管理インターフェースへの不正アクセスの問題が頻繁に取り上げられています。まず、機械学習を利用して通常時のURL形式、送受信データの内容や量、ページ遷移を学習させておき、これらから外れる通信を検知するシステムを作ります。通常と異なるURL形式や受信データ内容・量として検知されたWebリクエストは、SQLインジェクション攻撃の疑いがあります。サーバからの応答データ量が通常を大きく上回っている場合には、DBからデータを取得されている疑いがあります。また、管理用ログインページを経由しない管理ページへのアクセスは、管理インターフェースへの不正アクセスが疑われます。

以上のように、脅威に応じて適切な対象を学習、監視することにより、機械学習の手法が有効な対策として機能すると考えられます。

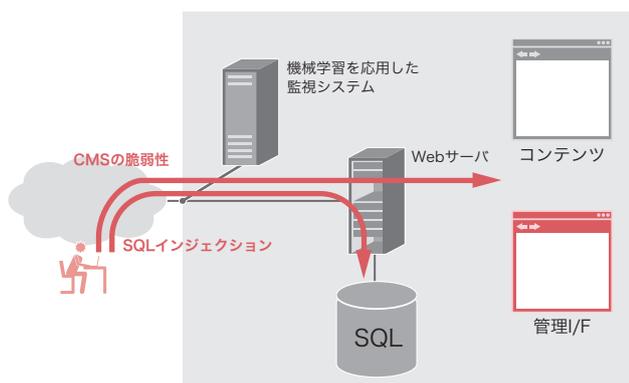


図-16 Webサービス保護への機械学習の適用

■ 機械学習システムのセキュリティリスク

機械学習の手法を導入することで新たなリスクは発生するでしょうか。

容易に考えられるのは、機械学習を組み込んだシステム自体への侵入です。システム内部を操作され、攻撃を見逃すように改ざん、停止される危険性が存在します。もっとも、これは機械学習に限らず、監視システム全般に存在するリスクと言えます。

また、設置環境から正常、異常の判別基準を継続的に学習していくシステムでは、別のリスクも存在します。ゆっくりとした変化で、徐々に本来の量や質に近づけていく攻撃をされた場合に、それを正常と判断して見逃す可能性が考えられます。

例えば、ファイルサーバから資料を1通取り出して外部に送信したとします。1通だけなら攻撃ではないと判断して、正当な外部送信として学習するかもしれません。次に別の資料を2通取り出して外部に送信すると、1通は正常なので、2通も正常と判断されて学習するかもしれません。これを繰り返して、最終的には大量の資料を外部送信できるかもしれません。

これもアナマリ検知の技術が登場したときに、既に指摘されていたことです。仮に、攻撃完了までに非常に長い期間を要するのであれば、攻撃のハードルはそれだけ高いと言えるでしょう。そのためには学習が緩やかである程望ましいということになりますが、その一方で正当な業務利用の形態の変化には素早く追従することが望まれます。このジレンマの解決には、両者の適切なバランスを見極めることが必要です。

機械学習の導入リスクに関する議論は、いまだ十分になされたとは言えない状況です。今後、新たな指摘が出てくる可能性もあり、引き続きその動向に注目する必要があります。

■ まとめ

セキュリティ上の事象の検知、分析、対応にかかる作業を自動化と迅速性で支援することは、セキュリティ関連技術の大きな目的の一つです。ブラックリストとホワイトリストに代表される、明確な異常と正常を列挙する方法論では、現実的に列挙が困難な、広大な中間域が残ります。この中間域を自動的に解消できる可能性を持った手法として、機械学習が期待されて

おり、今後も機械学習の適用例はますます増えていくと考えられます。

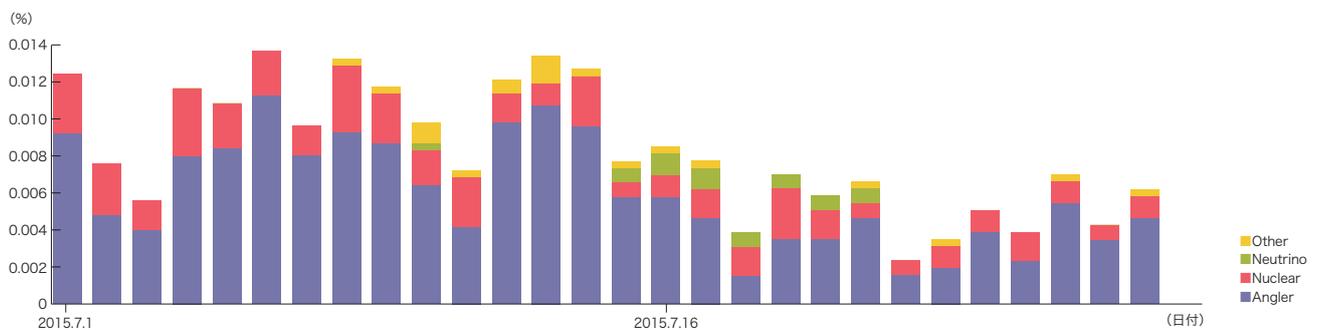
Exploit Kitについて、機能の概要やペイロード、攻撃環境の傾向などを紹介します。

1.4.2 猛威を振るうAngler Exploit Kit

IJでは、2015年7月1日に実施したMITFのWebクローラシステムの更新後、Anglerによる攻撃を多数検知するようになりました(図-17)^{*45}。本稿では、2015年7月現在、日本国内を対象としたドライブバイダウンロードに多用されているAngler

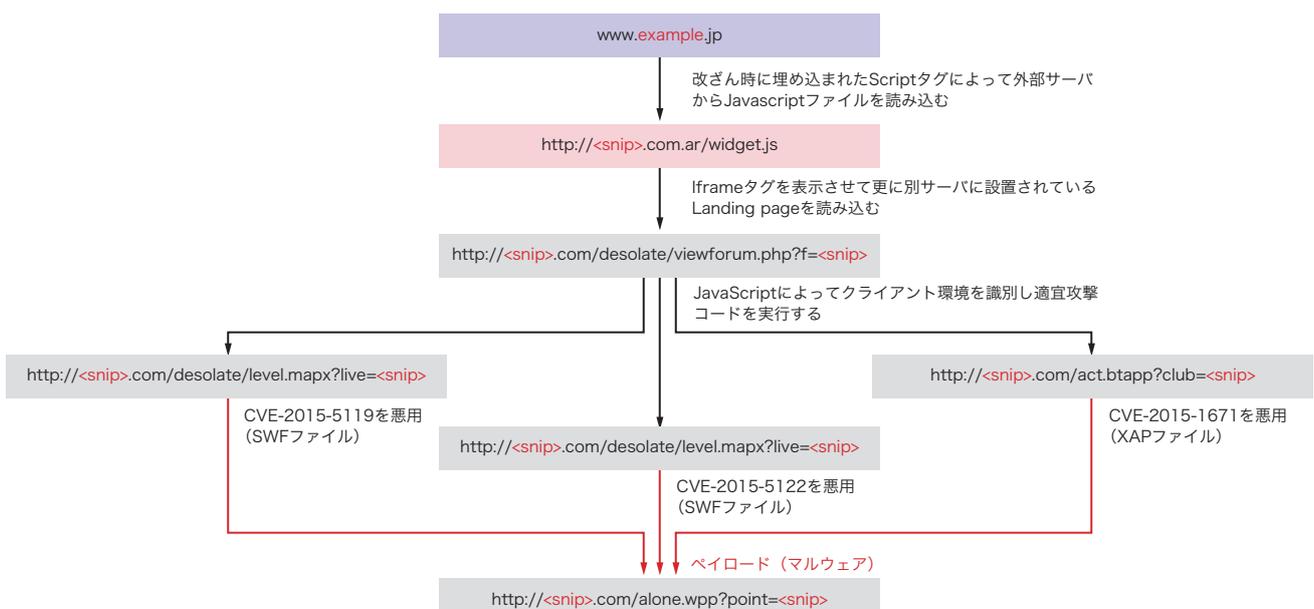
■ 攻撃の流れと隠蔽手法

本稿執筆時点(2015年7月下旬)に観測されているAnglerの典型的なURL遷移の例を示します(図-18)。この例では、改ざんされたWebサイト内のHTMLコンテンツに、外部のJavaScriptファイルを読み込むScriptタグが挿入されてお



※注 調査対象は日本国内の数十万サイト。近年の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。

図-17 2015年7月1日~28日のドライブバイダウンロード発生率(%) (Exploit Kit別)



※注 FQDN、Path共に頻繁に変更される。また、ディレクトリやパラメータ名や値には毎回異なる文字列が用いられるため、URLパターンのみでExploit Kitを識別することは困難な場合が多い。

図-18 Anglerの典型的なURL遷移の例(2015年7月)

*45 ドライブバイダウンロード攻撃全般の傾向については「1.3.4 Webサイト改ざん」参照。

■ ペイロード

Exploitが成功した際にダウンロードされるペイロードについては、2015年7月時点では以下の2種類のマルウェアを確認しています。

- CryptoWall 3.0
- Necurs

CryptoWall 3.0は、国内外で頻繁に悪用されているランサムウェアの一種で、クライアントPCに保存されているファイルを実行して暗号化し、復号キーと引き換えに金銭(ビットコイン)の支払

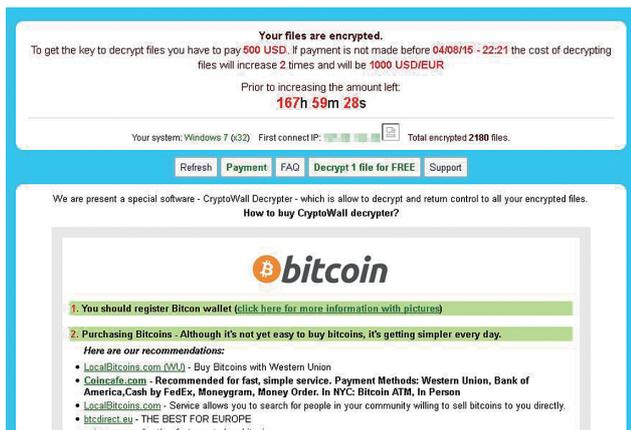


図-23 CryptoWall 3.0によって表示される脅迫メッセージ

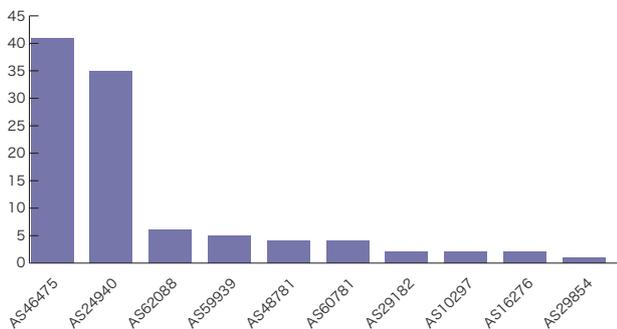


図-24 2015年7月1日～28日の期間に観測された AnglerのInfectorのIPアドレス数(ASごと)

いを要求します(図-23)。なお、CryptoWall 3.0は暗号化を行う前に攻撃者の管理サーバと鍵を共有するのですが、システムのProxy設定を自動取得する機能を備えていないため、Proxy経由でしか外部とのHTTP接続ができないような環境では暗号化は行われません^{*48}。一方、NecursはWindows Firewallや他のアンチウイルスソフトウェアなどを無効化したり、外部からの命令を受信して実行するRATのような機能を備えています^{*49}、多くの場合、更に別のマルウェアを導入させるためのダウンロードとして用いられます^{*49}。

■ Infectorの変遷

Landing pageやペイロードをホストするサーバ(Infector)は短い期間で使い捨てられています。2015年7月1日から28日の期間に、IJは102個のIPアドレスをAnglerのInfectorとして確認しましたが、個々のIPアドレスの平均生存期間(同じ攻撃に利用され続ける期間)は約1.3日でした。なお、同時期のNuclearのInfectorは約1.6日なので、これはAnglerに限った傾向ではなさそうです。ただし、AnglerのInfectorのIPアドレスを管理しているAS^{*50}には偏りがあり、前述の102個のInfectorは10個のASに集約され、更に全体の4分の3以上が上位2つのASに属しています(図-24)。Nuclearにも同様の傾向はありますが、Anglerに比べると多くのASに分散しているといえます(図-25)。また、AnglerがInfectorに利用するFQDN

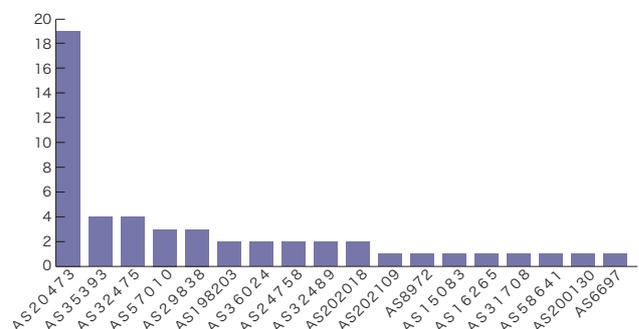


図-25 2015年7月1日～28日の期間に観測された NuclearのInfectorのIPアドレス数(ASごと)

*48 本稿執筆時点(2015年7月下旬)の状況。将来のバージョンでOSのProxy設定を自動取得する機能を備える可能性は否定できない。

*49 Necursの一般的な情報は「Trojan:Win32/Necurs」(<https://www.microsoft.com/security/portal/threat/encyclopedia/Entry.aspx?Name=Trojan:Win32/Necurs>)などで確認できる。

*50 自律システム(Autonomous System)。同一のルーティングポリシー配下にあるIPネットワークの集合のこと。一般的に通信事業者や大規模なホスティング事業者などは固有のASを形成している。

はIPアドレスよりも更に短時間で変更されます。IJの観測では、個々のFQDNのAレコードが30分で削除され、そのドメインが使い捨てられていることを確認しています。このため、IPアドレスやFQDNのブラックリストによる対策が効果を発揮しづらくなっています*51。

■ 対策

企業などの組織ネットワークにおけるドライブダウンロード対策としては、クライアント環境での適切なパッチ管理やプログラム実行可能領域の制限、EMETなどの脆弱性攻撃対策機能の有効化、Webブラウザプラグインの「Click to Play」の有効化などが挙げられます*52。HTTP Proxyを強制し、アクセスログを保管しておくことも、被害の緩和や早期発見に有用です。また、前述のFQDNの生存時間が短いという特徴を逆用して、例えばアクセスログから比較的アクセス数の少ないFQDNを抽出し、数十分後にそれらのAレコードの存在を確認するようなオペレーションを行えば、アクセスログの中からInfectorノードを抽出できる可能性があります。

Webサイト運営者、管理者の立場では、Webアプリケーションやコンテンツの厳格な管理が重要です*53。更に、前述のようにInfectorが一部のASに偏っている点を鑑み、サイトの運営環境(PaaSやIaaS、あるいはホスティング事業者など)についても、外部のブラックリストなどに登録されていないかなどの調査を行っておくことを推奨します。

1.4.3 ID管理技術～オンライン認証にパスワードを使わない方法へ～

あるID(識別子)を持つユーザがサーバにログインする際には、ユーザは秘密情報であるトークンを用いることでサーバは当該IDの属性情報を認証し、様々なリソースへの利用を認可することになります。このとき、IDに紐付けられ、そのIDを持つ属性情報や認可情報を保証する証明書としてクレデンシャルが流通します。このクレデンシャルは公開情報ですが、ユーザからはシステムの裏側で流通しているため直接ユーザが目にすることは少ないかもしれません。

前号*54ではこのようなID管理技術について解説を行いました。特にユーザ認証においては、パスワード認証に代表される"Something you know"だけではなく"Something you have"もしくは"Something you are"にカテゴライズされるトークンとの併用が進んでいることを取り上げました*55。ユーザは重要な通信に対し、多少利便性を損なってでも、強固な本人認証方式を選択できる時代が始まりつつあります。本稿ではこれまで広く利用されてきたパスワード認証に置き換わりつつあることを示すいくつかの技術動向を、実際のサービス事例と共に紹介します。

■ FIDO Alliance の動き

FIDO(Fast IDentity Online) Alliance*56は従来よりも利便性が高く、より安全なログイン方法に関するオープンな標準化の策

*51 IJで確認したCryptoWall 3.0の検体には、Sophos社のThreat Analysisの解析結果に記載されている、2014年12月以来継続して利用されていると推測されるC&Cサーバへの接続を試みるものがあった。「Troj/Ransom-BBD」(<https://www.sophos.com/en-us/threat-center/threat-analyses/viruses-and-spyware/Troj-Ransom-BBD/detailed-analysis.aspx>)。Exploit Kitのペイロードとして利用されるマルウェアのC&Cサーバのブラックリストを用いれば、ドライブダウンロードの被害を軽減できる可能性がある。

*52 クライアント環境におけるマルウェア感染対策については、本レポートのVol.21 (<http://www.ij.ad.jp/company/development/report/iir/021.html>)の「1.4.1 標的型攻撃で利用されるRAT「PlugX」」末尾で詳しく紹介している。

*53 Webサイト運営者、管理者の立場での改ざん対策については、本レポートのVol.24 (<http://www.ij.ad.jp/company/development/report/iir/024.html>)の「1.4.2 国内金融機関の認証情報などを窃取するマルウェア「vawtrak」」や、IJ Security Diary「BHEK2を悪用した国内改ざん事件の続報」(<https://sect.ij.ad.jp/d/2013/03/225209.html>)の末尾で詳しく紹介している。

*54 本レポートのVol.27 (<http://www.ij.ad.jp/company/development/report/iir/027.html>)の「1.4.2 ID管理技術～利便性と安全性の観点から～」を参照のこと。

*55 パスワードを入力する端末がキーロガーが仕込まれていたりマルウェアなどに感染するなど信頼できないケースや、別のサービスで利用されていたデータベースから使い回しパスワードが漏えいしてリスト型攻撃も頻発しており、パスワードだけを用了た認証方式の信頼は揺らいでいる。これに対する解決策としては、毎回異なるパスワードを用いるワンタイムパスワードや、従来のパスワード認証と併用して他の複数のトークンを用いる多要素認証も利用されるようになってきた。ワンタイムパスワードや多要素認証は本レポートのVol.26 (<http://www.ij.ad.jp/company/development/report/iir/026.html>)の「1.4.3 ID管理技術」を参照のこと。

*56 "About The FIDO Alliance"(<https://fidoalliance.org/about/overview/>)。

定を目指している非営利団体であり、2012年6月に設立されています。FIDO仕様では、オンライン認証はオフライン処理とオンライン処理の2つを組み合わせることで実現されています。従来から利用されているパスワード認証はオンライン処理のみで行われていると考えると理解しやすくなります。オンライン処理では、ユーザ(Prover)からサーバ(Verifier)にIDなどと共にパスワードが送信されています。このとき、パスワードはTLSなどの安全なチャネルを確保した上で送信する必要があり、別の安全なチャネルを利用しないケースではパスワードが簡単に第三者によって傍受されてしまいます。FIDO仕様では認証に関する通信をセキュアなチャネルを通して行わなくてもよいように、公開鍵暗号を用いたチャレンジ&レスポンスによる認証^{*57}をオンライン処理で行っています。一方でオフライン処理ではFIDO準拠デバイスがユーザをローカルで認証しており、ローカル認証に成功すると前述の公開鍵暗号認証で用いる秘密鍵の利用をユーザに許可する仕組みとなっています。オンライン処理の部分だけに注目すると、公開鍵暗号を用いたチャレンジ&レスポンスによる認証という単純なプロトコルになっているため、生体認証を含む様々なローカル認証をサポートすることが容易で、拡張性が高い点からFIDOへの注目と期待が集まっています。

FIDO Allianceでは現在大きく分けて2つの仕様が策定されています。UAF(Universal Authentication Framework)標準は生体認証(バイオメトリクス認証)を、U2F(Universal 2nd Factor)標準は多要素認証^{*58}を取り扱っています。前者のUAF標準に

ついては、2015年5月、NTTドコモがFIDO Allianceへボードメンバーとして加入し^{*59}、同時にUAF標準を実装した2製品がFIDO準拠(FIDO Certified)として認定されています^{*60}。これらのスマートフォンは虹彩認証や指紋認証が搭載されており^{*61}、パスワードレスでのログインによる利便性が謳われています。また、多要素認証に関するU2F標準としてはGoogle Login Service^{*62}がFIDO準拠リストに挙げられています。2段階認証プロセスでのセキュリティキー^{*63}の1つとしてU2F標準のUSBデバイスが利用可能となっており、国内でも複数の代理店から販売されています。現在、Google Chromeバージョン40以降のブラウザでのみ利用可能ですが、Microsoft社も次期OSであるWindows10での対応を表明しており^{*64}、利用が拡大すると予想されます。また、F2F標準はNFCやBluetooth経由での利用について拡張されるとのアナウンスもありました^{*65}。更に6月には米国政府システムにおけるセキュリティ標準化仕様を策定しているNISTや英国内閣組織がFIDO Allianceへ加入しており^{*66}、政府機関での利用も検討されていると考えることができます。NISTからはInteragency Reportとしてスマートカードを利用した多要素認証に関するレポートのドラフトが7月に公開されました^{*67}。

■ ワンタイムパスワードの利用

オンラインゲームにおいて、仮想通貨や仮想アイテムを狙った不正ログインが相次いだことから、パスワード認証ではなく、スマートフォンなどの他のデバイスと連動したワンタイムパ

*57 サーバ(Verifier)からのチャレンジに対して、ユーザ(Prover)しか知りえない秘密鍵を用いて作成するデータ(例えばチャレンジに対するデジタル署名)をレスポンス(返答)することで実現される。

*58 複数の認証方式を組み合わせることで1回の認証を行うこと。例えば、通常のパスワードに加えて、ワンタイムパスワードの入力をもって認証を行うケースがそれに当たる。

*59 "FIDO Alliance Welcomes NTT DOCOMO, INC. to Board of Directors" (<https://fidoalliance.org/fido-alliance-welcomes-ntt-docomo-to-board/>)。

*60 "FIDO Certified" (<https://fidoalliance.org/certification/fido-certified/>)。

*61 NTT Docomo、「2015夏モデルの10機種を開発」(https://www.nttdocomo.co.jp/info/news_release/2015/05/13_00.html)。プレスリリースには4製品がFIDO準拠とあるが、FIDO CertifiedにリストされていないのはSAMSUNG製品であるためと思われる。

*62 <https://accounts.google.com/>

*63 Googleアカウントヘルプ、「2段階認証プロセスにセキュリティキーを使用する」(<https://support.google.com/accounts/answer/6103523?hl=ja>)。

*64 "Microsoft Announces FIDO Support Coming to Windows 10" (<http://blogs.windows.com/business/2015/02/13/microsoft-announces-fido-support-coming-to-windows-10/>)。

*65 "FIDO Alliance Equips U2F Protocol for Mobile and Wireless Applications" (<https://fidoalliance.org/fido-alliance-equips-u2f-for-mobile-and-wireless-applications/>)。Bluetooth SIG、「Bluetooth SIG and FIDO Alliance Deliver Two-factor Authentication Via Bluetooth Smart」(<http://www.bluetooth.com/Pages/Press-Releases-Detail.aspx?ItemID=233>)。

*66 "FIDO Alliance Announces Government Membership" (<https://fidoalliance.org/fido-alliance-announces-government-membership/>)。

*67 NIST IR 8055、「Derived Personal Identity Verification(PIV) Credentials(DPC) Proof of Concept Research」(http://csrc.nist.gov/publications/drafts/nistir-8055/nistir_8055_draft.pdf)。執筆時にはパブリックコメントの受付中である。

スワード利用への移行が推奨されました^{*68}。その際、ワンタイムパスワードの利用者に、ゲーム中で利用されるアイテムの増強などのボーナスを与えるといった、利用拡大のための独自の施策がとられています。更に不正ログインの被害が確認された場合には、できる限りのデータ救済措置が取られるというサービスも見受けられます。

また、オンラインバンキングにおいてもワンタイムパスワードの利用が広がっています。オンラインバンキングへのログインでは、従来は通常のパスワードもしくは暗証番号を用いた認証が行われることが一般的でした。ログイン中に振込などの重要なトランザクションが行われる際には、これまで乱数表が書かれたカードを書留などで安全に郵送した上で、トランザクションごとに異なる組み合わせの乱数(パスワード)を入力させる利用者認証を行っていました。この乱数表カードの利用には、もし攻撃者に内容を傍受されていた場合には、多数のトランザクションを行うことで徐々に乱数表の内容が漏れてしまう問題がありました。このため、乱数表に書かれたパスワードだけでなく、カード発行日を入力させるなどで、安全性を補う試みが行われています。しかしそこには限界があるため、一定時間ごとにパスワードが自動的に変更される専用デバイスへの切り替えが行われました。その後、入力インタフェースを持つ専用デバイスが利用されるようになりました。専用デバイスでは入力とトランザクションを結びつけることができ、例えば振込先の口座番号を入力させることでより安全な認証を行う方式へと変化しています。現在、このトランザクション認証は一部の銀行でしか対応していませんが、他銀行でも同じ専用デバイスが用いられているため、今後対応することが予想されます。

更に、専用デバイスやスマートフォンのアプリを用いたワンタイムパスワードではなく、デバイス不要のワンタイムパスワード

の利用も広がっています。前者・後者共に同じ用語が用いられていますが、概念が異なります。まず利用者はメールアドレスやショートメッセージを受け取れることが前提となります。通常のパスワードによるログインを行うと、短時間だけ利用可能なトークン(ワンタイムパスワード)がサーバでランダムに生成され、メールやショートメッセージを介して送信されます。ユーザがそれを受け取り、サーバに送信することで認証が完了する仕組みです。インターネットと携帯電話網の両方を利用するなど、ログインで利用したチャネルとワンタイムパスワードの受信に利用したチャネルを分けることで、両者を同時に傍受されない限りはより安全性を確保できるメリットがあります。また専用デバイスを利用しないため、比較的成本の安いシステム運用が可能となります。オリジナルパスワードを守るために、それぞれのスマートフォンなどのデバイスで、一時的に利用するためのパスワードを割り振るといった観点では派生パスワードと考えることもできます。

■ ソーシャルログイン

加えて、ソーシャルログインと呼ばれる手段を提供するサイトも増えつつあります。あるリソースを提供するサーバで独自にID・パスワード管理を行わず、SNSやポータルサイトをIdPとして利用する、つまりユーザ認証手段としてのみ利用して認証を行うものです。新しいサーバの利用時にユーザが既に利用しているサーバとの連携許可を与えることで、新しいサーバにユーザ登録を行うことなくログインすることが可能となる方式です。これは、サーバ・ユーザ双方で新しいパスワードを管理する必要がなくなるというメリットがあります。しかし、これまでである特定のSNS内など限られた範囲で利用していたIDで、新たなサーバとの連携を許可する際には、連携を行ったサーバ間での情報の流通に留意する必要があります。

*68 例えば、ガンホーゲームズ、セキュリティ対策(<http://www.gungho.jp/security/security.html>)、NEXONサポート、ワンタイムパスワード(<http://www.nexon.co.jp/support/security/otp-guide.aspx>)、GMOゲームボット、ワンタイムパスワードをつかってみよう。

情報収集などの目的で過大な情報を参照しようとするサーバが存在しますし、悪意のあるサーバと連携したために、SNS上で不要なメッセージが発出される事件なども発生しています。また、本来ならば異なるレムムの異なるIDでそれぞれ実施していた独立の活動が、連携したサーバ間でやり取りされる情報に基づいて、同一人物による行動として結び付けられてしまう可能性があります。このため、ログイン中に別のエンティティ（派生ID）として立ち振る舞うようIDを切り替える機能を活用したり、あるサーバの利用を終えた時に確実に連携を解除する手続きを行うなど、利用者側での慎重な対応が必要となる場合があります。

■ リスクベース認証

最後にリスクベース認証について紹介します。今年7月、主要サービスにおけるパスワード管理に関する実態調査^{*69}が総務省から発表されました。パスワード設定が可能な文字や文字数制限、パスワード管理時にハッシュ化しているかなどに加え、同一IPアドレスからのログイン試行に関連する項目についても調査が行われました。認証時に普段とは異なる振る舞いをする、例えばありえない国外からのログインなど、普段のユーザの利用環境との違いを検知し、IDをロック（一時的にIDを無効化）したり、普段のパスワード認証に加え他の認証を行う多要

素認証を実施したりするケースがしばしば見られるようになりました。ユーザに対して常に多要素認証を強要したりするなど一方的に利便性を損ねるのではなく、サーバがユーザの挙動の違いを自動的に検知した場合にのみ、今後起こりうる「リスク」を配慮し、より強固な認証方式の利用にシフトするという、利用者の利便性に配慮した中間的な認証方式も実際に利用されはじめています。

このように認証に関する状況は、利用者の利便性とリスクを考慮してトータルコストを最適化する方向にあります。認証を行う対象となるユーザの情報リテラシに応じて、トータルコストは変化しますから、その組織ごとの最適方式を見つける必要があるでしょう。またユーザは固定されておらず、常に増減していることから、状況に応じた見直しが求められます。

1.5 おわりに

このレポートは、IJ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、機械学習とセキュリティ、猛威を振るうAngler Exploit Kit、ID管理技術の実際の利用について紹介しました。IJ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）

IJ サービスオペレーション本部 セキュリティ情報統括室 室長。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IJグループの緊急対応チームIJ-SECTの代表として活動し、CSIRTの国際団体であるFIRSTに加盟。Telecom-ISAC Japan、日本シーサート協議会、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。

土屋 博英（1.2 インシデントサマリ）

土屋 博英、永尾 禎啓、鈴木 博志、梨和 久雄（1.3 インシデントサーベイ）

永尾 禎啓（1.4.1 機械学習とセキュリティ）

梨和 久雄、鈴木 博志（1.4.2 猛威を振るうAngler Exploit Kit）

須賀 祐治（1.4.3 ID管理技術 ～オンライン認証にパスワードを使わない方法へ～）

IJ サービスオペレーション本部 セキュリティ情報統括室

協力：

春山 敬宏、小林 稔、小林 直、加藤 雅彦、根岸 征史、桃井 康成、平松 弘行 IJ サービスオペレーション本部 セキュリティ情報統括室

*69 総務省、「ウェブサービスに関するID・パスワードの管理・運用実態調査結果」(http://www.soumu.go.jp/menu_news/s-news/01ryutsu03_02000099.html)。

ブロードバンドとモバイルのトラフィックを比較

2.1 概要

本レポートでは、毎年、IIJが運用しているブロードバンド接続サービスのトラフィックを分析して、その結果を報告しています*1 *2 *3 *4 *5 *6。今回も、利用者の1日のトラフィック量やポート別使用量などを基に、この1年間のトラフィック傾向の変化を報告します。また、今回からモバイルサービスのトラフィックについても同様にトラフィック傾向を報告します。

図-1は、IIJのブロードバンドサービス及びモバイルサービス全体について月平均トラフィック量の推移を示したグラフです。トラフィックのIN/OUTはISPから見た方向を表し、INは利用者からのアップロード、OUTは利用者へのダウンロードとなります。トラフィック量の数値は開示できないため、それぞれの最大値を1として正規化しています。

ブロードバンドに関しては、この1年のトラフィック量は、INは15%の増加、OUTは38%の増加となっています。この8年でみると、INが1.6倍、OUTが3.8倍と、アップロード量に比べダウンロード量が大きく伸びています。

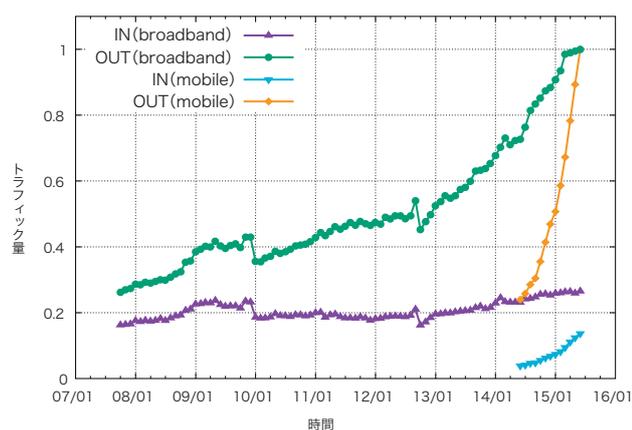


図-1 ブロードバンド及びモバイルの月間トラフィック量の推移

モバイルに関しては、この1年の数字しかありませんが、この1年で、INは3.5倍に、OUTは4.2倍にと、急速にトラフィック量が伸びてきています。ただし、総量ではまだブロードバンドよりひと桁以上少ない状況です。

2.2 データについて

ブロードバンドに関しては、今回も前回までと同様に、個人及び法人向けのブロードバンド接続サービスについて、ファイバーとDSLによるブロードバンド顧客を収容するルータで、Sampled NetFlowにより収集した調査データを利用しています。モバイルに関しては、個人及び法人向けのモバイルサービスについて、使用量についてはアクセスゲートウェイの課金用情報を、使用ポートについてはサービス収容ルータでのSampled NetFlowデータを利用しています。

トラフィックは平日と休日で傾向が異なるため、1週間分のトラフィックを解析しています。今回は、2015年6月1日から7日の1週間分のデータを使っています。ブロードバンドに関しては、前回解析した2014年5月26日から6月1日の1週間分と比較します。

ブロードバンドでは契約ごと、モバイルでは複数電話番号の契約があるので電話番号ごとの集計となっています。ブロードバンド各利用者の使用量は、利用者に割り当てられたIPアドレスと、観測されたIPアドレスを照合して求めています。また、NetFlowではパケットをサンプリングして統計情報を取得しています。サンプリングレートは、ルータの性能や負荷を考慮して、1/8192～1/16384に設定されています。観測された使用量に、サンプリングレートの逆数を掛けることで全体の使用量を推定しています。IIJの提供するブロードバンドサービスにはファイバー接続とDSL接続がありますが、今ではファイバー接続の利用がほとんどとなっていて、2015年には観測された

- *1 長健二郎。ブロードバンドトラフィックレポート：この一年でトラフィック量は着実に増加、HTTPSの利用が拡大。Internet Infrastructure Review. Vol.24. pp28-33. August 2014.
- *2 長健二郎。ブロードバンドトラフィックレポート：違法ダウンロード刑事罰化の影響は限定的。Internet Infrastructure Review. Vol.20. pp32-37. August 2013.
- *3 長健二郎。ブロードバンドトラフィックレポート：この1年間のトラフィック傾向について。Internet Infrastructure Review. Vol.16. pp33-37. August 2012.
- *4 長健二郎。ブロードバンドトラフィックレポート：マクロレベルな視点で見た、震災によるトラフィックへの影響。Internet Infrastructure Review. Vol.12. pp25-30. August 2011.
- *5 長健二郎。ブロードバンドトラフィックレポート：P2Pファイル共有からWebサービスへシフト傾向にあるトラフィック。Internet Infrastructure Review. Vol.8. pp25-30. August 2010.
- *6 長健二郎。ブロードバンドトラフィック：増大する一般ユーザのトラフィック。Internet Infrastructure Review. Vol.4. pp18-23. August 2009.

ユーザ数の95%はファイバー利用者で、ブロードバンドトラフィック量全体の98%を占めています。

2.3 利用者の1日の使用量

まずは、ブロードバンド及びモバイル利用者の1日の利用量をいくつかの切口から見ていきます。ここでの1日の利用量は各利用者の1週間分のデータの1日平均です。

図-2及び図-3は、ブロードバンドとモバイル利用者の1日の平均利用量の分布(確率密度関数)を示します。アップロード(IN)とダウンロード(OUT)に分け、利用者のトラフィック量をX軸に、その出現確率をY軸に示していて、2014年と2015年を比較しています。X軸はログスケールで、10KB(10^4)から100GB(10^{11})の範囲を示しています。一部の利用者はグラフの範囲外にありますが、概ね100GB(10^{11})までの範囲に分布しています。

ブロードバンドのINとOUTの各分布は、片対数グラフ上で正規分布となる、対数正規分布に近い形をしています。これはリニアなグラフで見ると、左端近くにピークがあり右へなだらかに減少するいわゆるロングテールな分布です。OUTの分布はINの分布より右にずれていて、ダウンロード量がアップロード量より、ひと桁以上大きくなっています。2014年と2015年で比較すると、INとOUT共に分布の山が右に少し移動して、

利用者全体のトラフィック量が増えていることが分かります。OUTの分布を見ると、分布のピークはここ数年間で着実に右に移動していますが、右端のヘビーユーザの使用量はあまり増えていないので、分布の対称性が崩れてきています。一方で、INの分布は右側の裾が広がっています。以前は、ここによりはっきりした山がINとOUT両方にあり、IN/OUT量が対称なヘビーユーザを示していました。そこで便宜上、大多数のIN/OUT非対称な分布を「クライアント型利用者」、右側の小数のIN/OUT対称なヘビーユーザの分布を「ピア型利用者」と呼んできました。今回もその慣習に従います。ここ数年で、ピア型利用者の山は小さくなりほとんど識別できなくなりました。これは、ヘビーユーザの割合が減少していることを示しています。

図-3のモバイルの場合、ブロードバンドに比べて利用量は大幅に少ないことが分かります。また、使用量に制限があるため、分布右側のヘビーユーザの割合が少なく、左右非対称な形になります。極端なヘビーユーザも存在しません。外出時のみの利用や、使用量の制限のため、各利用者の日ごとの利用量のばらつきはブロードバンドより大きくなります。そのため、1週間分のデータから1日平均を求めると、1日単位でみた場合より利用者間のばらつきは小さくなります。1日単位で同様の分布を描くと、分布の山が少し低くなり、その分両側の裾が持ち上がりますが、基本的な分布の形や最頻出値はほとんど変わりません。

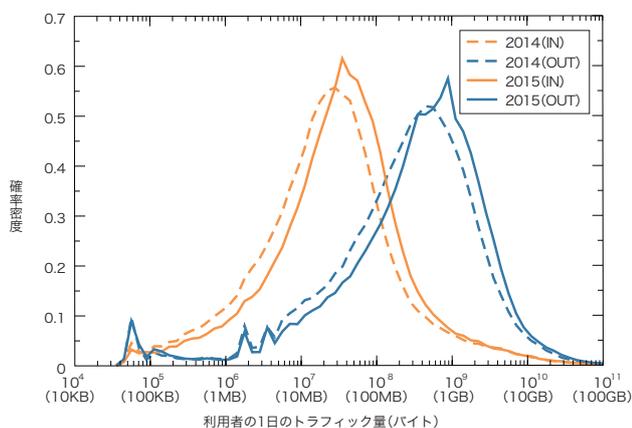


図-2 ブロードバンド利用者の1日のトラフィック量分布
2014年と2015年の比較

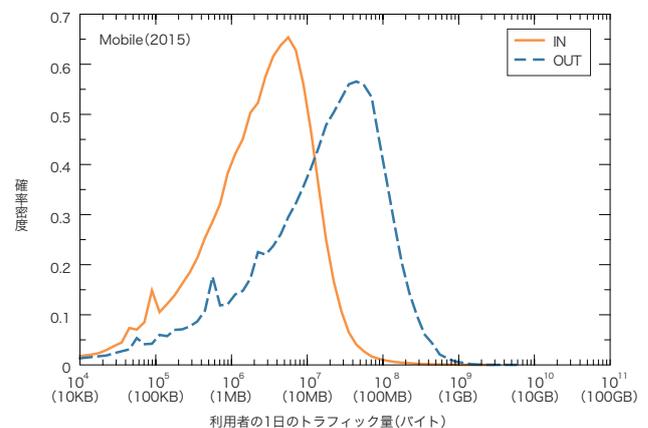


図-3 モバイル利用者の1日のトラフィック量分布2015年

表-1は、ブロードバンド利用者の1日のトラフィック量の平均値と、分布の山の頂点にある最頻出値の推移を示します。今回分布の山に対して頂点が少しずれているので、最頻出値は分布の山の中央に来るように補正しています。分布の最頻出値を2014年と2015年で比較すると、INでは28MBから40MBに、OUTでは447MBから708MBに増えていて、各利用者のトラフィック量が、特にダウンロード側で増えていることが分かります。一方、平均値はグラフ右側のヘビーユーザの使用量に引っ張られるので、2015年には、INの平均は467MB、OUTの平均は1621MBと、最頻出値よりかなり大きな値になります。2014年には、それぞれ437MBと1287MBでした。モバイルでは、表-2に示すように、極端なヘビーユーザが少ないため、平均と最頻出値にはそれ程差がありません。最頻出値は、

年	IN (MB/day)		OUT (MB/day)	
	平均値	最頻出値	平均値	最頻出値
2005	430	3.5	447	32
2007	433	4	712	66
2008	483	5	797	94
2009	556	6	971	114
2010	469	7	910	145
2011	432	8.5	1,001	223
2012	410	14	1,026	282
2013	397	18	1,038	355
2014	437	28	1,287	447
2015	467	40	1,621	708

表-1 ブロードバンド利用者の1日のトラフィック量の平均値と最頻出値の推移

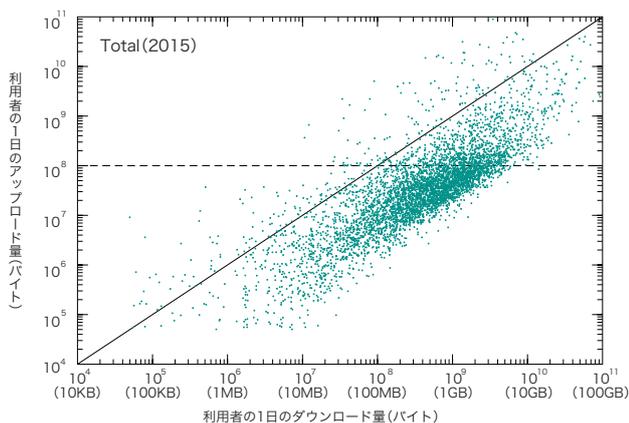


図-4 ブロードバンド利用者ごとのIN/OUT使用量

INで5.5MB、OUTで40MBで、平均値は、INで6.0MB、OUTで46.6MBです。

図-4及び図-5は、利用者ごとのIN/OUT使用量を5,000人をランダムに抽出してプロットしています。X軸はOUT(ダウンロード量)、Y軸はIN(アップロード量)で、共にログスケールです。利用者のIN/OUTが同量であれば対角線上にプロットされます。

対角線の下側に対角線に沿って広がるクラスは、ダウンロード量がひと桁多いクライアント型の一般ユーザです。ブロードバンドでは、以前は右上の対角線上あたりを中心に薄く広がるピア型のヘビーユーザのクラスがはっきり分かっていましたが、今では識別ができなくなっています。便宜上、クライアント型とピア型に分けましたが、実際には、クライアント型の一般ユーザでもSkypeなどのピア型のアプリケーションを利用し、また一方のピア型のヘビーユーザもWebなどのダウンロード型のアプリケーションを利用しているので、その境界はあいまいです。つまり、多くの利用者は両タイプのアプリケーションを異なる割合で使用しています。また、各利用者の使用量やIN/OUT比率にも大きなばらつきがあり、多様な利用形態が

年	IN (MB/day)		OUT (MB/day)	
	平均値	最頻出値	平均値	最頻出値
2015	6.0	5.5	46.6	40

表-2 モバイル利用者の1日のトラフィック量の平均値と最頻出値

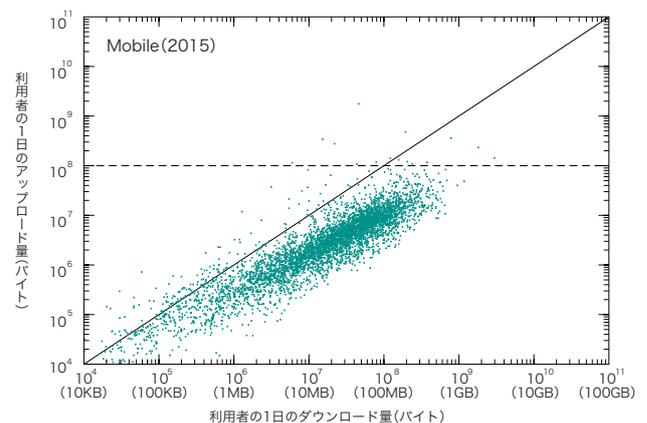


図-5 モバイル利用者ごとのIN/OUT使用量

存在することが窺えます。ここでは、2014年と比較しても、ほとんど違いは確認できません。

モバイルでも、OUTがひと桁多い傾向は同じですが、ブロードバンドに比べて利用量は少なく、IN/OUTのばらつきも小さくなっています。

図-6及び図-7は、利用者の1日のトラフィック量を相補累積度分布にしたものです。これは、使用量がX軸の値より多い利用者の、全体に対する割合をY軸に、ログ・ログスケールで示したもので、ヘビーユーザの分布を見るのに有効です。グラフの右側が直線的に下がっていて、べき分布に近いロングテールな分布であることが分かります。ヘビーユーザは統計的に分布していて、決して一部の特殊な利用者ではないと言えます。モバイルでも、ヘビーユーザはべき分布していますが、ヘビーユーザの割合が少なく、ヘビーユーザにおいてもOUTの利用量がINより数倍大きくなっています。

利用者間のトラフィック使用量の偏りをみると、使用量には大きな偏りがあり、結果として全体は一部利用者のトラフィックで占められています。例えば、ブロードバンド上位10%の利用者がOUTの65%、INの90%を占めています。更に、上位1%の利用

者がOUTの29%、INの63%を占めています。ここ数年のヘビーユーザ割合の減少に伴い、僅かながら偏りは減ってきています。

モバイルでは、上位10%の利用者がOUTの49%、INの50%を、上位1%の利用者がOUTの13%、INの19%を占めています。ここからも、モバイル利用者のヘビーユーザ割合が少ないことが分かります。

2.4 ポート別使用量

次に、トラフィックの内訳をポート別の使用量から見ていきます。最近では、ポート番号からアプリケーションを特定することは困難です。P2P系アプリケーションには、双方が動的ポートを使うものが多く、また、多くのクライアント・サーバ型アプリケーションが、ファイアウォールを回避するため、HTTPが使う80番ポートを利用します。大雑把に分けると、双方が1024番以上の動的ポートを使っていればP2P系のアプリケーションの可能性が高く、片方が1024番未満のいわゆるウェルノウンポートを使っていれば、クライアント・サーバ型のアプリケーションの可能性が高いと言えます。そこで、TCPとUDPで、ソースとデスティネーションのポート番号の小さい方を取り、ポート番号別の使用量を見てみます。

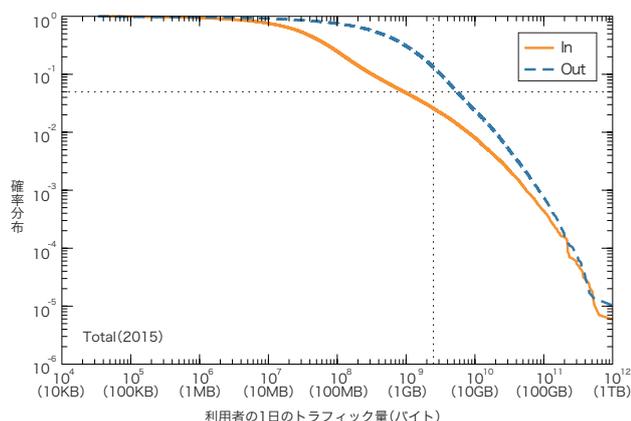


図-6 ブロードバンド利用者の1日のトラフィック量の相補累積度分布

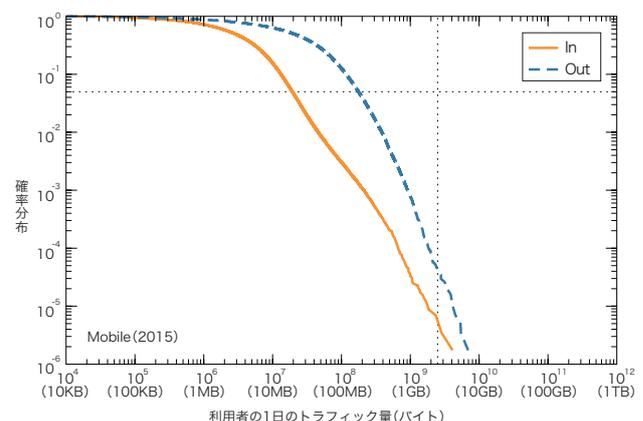


図-7 モバイル利用者の1日のトラフィック量の相補累積度分布

また、全体トラフィックは、ピア型のヘビーユーザのトラフィックに支配されているので、クライアント型の一般利用者の動向を見るために、少し荒っぽいですが、1日のアップロード量が100MB未満のユーザを抜き出して、これをクライアント型利用者として扱います。これは、図-4では、IN=100MBにある水平線の下側の利用者にあたります。

表-3はブロードバンド利用者のポート使用割合を、全体とクライアント型利用者について、2014年と2015年で比較したものです。

2015年の全体トラフィックの81%はTCPです。HTTPの80番ポートの割合が、2014年の45%から38%に減って、HTTPSの443番ポートの割合が、9%から23%に増えています。減少傾向のTCPの動的ポートは、2014年の24%から2015年には18%にまで減りました。動的ポートでの個別のポート番号の割合は僅かですが、Flash Playerが利用する1935番が最大で総量の約2%ありますが、後は0.5%未満となっています。TCP以外のトラフィックでは、UDPでもHTTPSの443番ポートのトラ

フィックがあり、GoogleのQUICプロトコルだと思われます。他はほとんどVPN関連です。

一方、クライアント型利用者に限ると、2014年には75%を占めていた80番ポートが、2015年には53%へと22ポイント減少しました。一方で、2番目に多いHTTPSの443番ポートが、2014年の14%から35%へと21ポイント増えていて、HTTPの一部がHTTPSに移行したと考えられます。また、動的ポートの割合は、7%から5%に減少しています。

表-4はモバイル利用者のポート使用割合で、全体的にブロードバンドのクライアント型利用者の数字に近い値となっています。

HTTPSの利用拡大については、2013年6月に米国家安全保障局(NSA)の通信傍受プログラムの存在が問題になって以降、暗号化通信を行うHTTPSを常時使用するサービスが増えてきているためです。2015年のHTTPSを利用するトラフィック量について事業者別内訳を調べると、その約8割はGoogle社関連で、同社の積極的なHTTPS採用の取り組みが窺えると同時に、YouTubeのトラフィック量がHTTPSの利用量を押し上げていると思われます。

protocol port	2014		2015	
	total (%)	client type	total (%)	client type
TCP	80.2	97.4	80.8	94.7
(< 1024)	56.3	90.1	63.3	89.9
80(http)	44.9	74.8	37.9	53.2
443(https)	9.3	13.8	23.3	35.1
81	0.0	0.0	0.5	0.7
182	0.0	0.0	0.4	0.3
554(rtmp)	0.4	0.3	0.2	0.1
22(ssh)	0.3	0.0	0.2	0.0
(>= 1024)	23.8	7.3	17.5	4.8
1935(rtmp)	2.5	4.0	1.8	2.4
8080	0.4	0.2	0.3	0.1
7144(peercast)	0.3	0.0	0.2	0.0
UDP	12.5	1.8	11.4	2.6
443(https)	0.0	0.0	0.9	0.9
1701(12tp)	0.0	0.0	0.2	0.3
ESP	6.9	0.7	7.4	2.6
IP-ENCAP	0.2	0.0	0.2	0.0
GRE	0.2	0.0	0.2	0.0
ICMP	0.0	0.0	0.0	0.0

表-3 ブロードバンド利用者のポート別使用量

protocol port	2015
	total (%)
TCP	93.8
80(http)	52.5
443(https)	37.4
81	0.5
993(imaps)	0.5
1935(rtmp)	0.5
UDP	5.2
1701(12tp)	1.8
443(https)	1.0
4500(nat traversal)	0.3
80(http)	0.2
53(dns)	0.1
ESP	0.7
GRE	0.3
ICMP	0.0

表-4 モバイル利用者のポート別使用量

図-8は、ブロードバンド全体トラフィックにおけるTCPポート利用の週間推移を、2014年と2015年で比較したものです。ここでは、TCPのポート利用を80番、443番、その他のウェルノウンポート、動的ポートの4つに分けてそれぞれの推移を示しており、ピーク時の総トラフィック量を1として正規化して表しています。前回までは443番ポートはその他のウェルノウンポートに含んでいましたが、その利用割合が増えてきたため、今回から443番ポートも別にプロットしています。2014年と比較すると、全体でも443番ポートの割合が更に増え、動的ポートの利用が減少している傾向が確認できます。全体のピークは21:00から翌1:00、土日には昼間のトラフィックが増加していて、家庭での利用時間を反映しています。

図-9のモバイルでは、80番ポートと443番ポートの推移を示し、朝から夜中までトラフィックの高い状態が続きます。平日には、朝の通勤時間、昼休み、夕方から夜中にかけての3つのピークがあり、ブロードバンドとは利用時間の違いがあることが分かります。

2.5 まとめ

この1年間のブロードバンドトラフィックの傾向として、前回報告したHTTPSの利用がこの1年で大きく拡大したことが挙げられます。その他の全体のトレンドには大きな変化はありませんが、この1年間でダウンロード量は38%、アップロード量も15%増加しました。その前の1年間では、27%と13%だったので、この1年で増加率も増えています。

今回から報告に加えたモバイルトラフィックは、ブロードバンドに比べるとまだ量的には少ないですが、この1年間で大きく伸びてきています。モバイル利用者では、利用帯域や使用量の制約などからブロードバンドに比べてヘビーユーザの割合が少なくなっています。また、そのポート別の利用割合はブロードバンドのライトユーザに近くなっていて、マルチデバイス対応サービスの普及に伴って、モバイルでもブロードバンドと同じようなサービスを使っていることが窺えます。その一方で、利用時間では平日の通勤時間帯や昼休みの利用が目立つなど、ブロードバンドとの違いもあります。

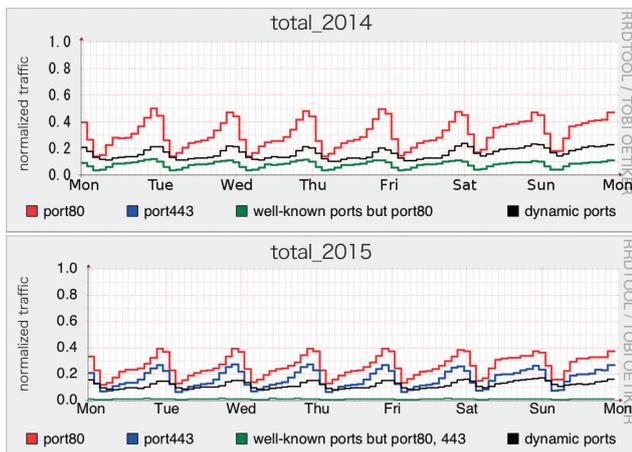


図-8 ブロードバンド利用者のTCPポート利用の週間推移
2014年(上)と2015年(下)

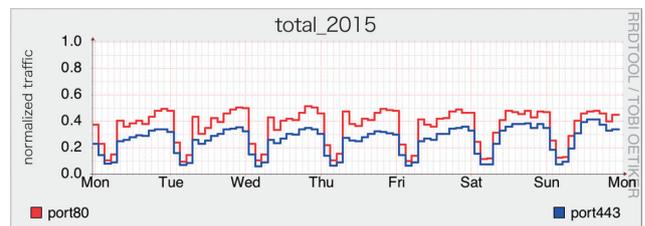


図-9 モバイル利用者のTCPポート利用の週間推移 2015年



執筆者：
長 健二郎 (ちょう けんじろう)
株式会社IJ インベーションインスティテュート 技術研究所所長。

ソーシャル・ビッグデータ

3.1 ビッグデータの現状

「ビッグデータ」という言葉は「クラウド・コンピューティング」などと同じように非常に概念的な用語です。定義が曖昧模糊としていて意味するところが時間の経過と共に変化していくので、非常に感覚的に使われることが多いです。それ故に長く生き残っていく用語だと思えます。しかし、流行りすたりのあるパスワードとしての「ビッグデータ」は、2015年の現在、賞味期限切れを迎えつつあるように感じられます。この印象は、イギリスのシンクタンクであるガートナーが毎年発表しているハイプサイクルでも、2014年には「ビッグデータ」は幻滅期に突入していることから、それなりに裏付けのある話でしょう(図-1)。

もっとも日本国内に限定したハイプサイクルでは流行期の最後に位置するので、パスワードとしての「ビッグデータ」は実に微妙な立ち位置にあるのかもしれません(図-2)。

■ 日本におけるビッグデータの動向

ともあれ、ここ数年間の技術トレンドを象徴するパスワードであり続けた「ビッグデータ」も今ではやや古びてきて、今は「ビッグデータ」より少し具体化した他のパスワード、例えば「Internet of Things(IoT)」などが喧伝されることが多いように感じます。

「Internet of Things(IoT)」は、RFIDの国際標準化に貢献したKevin Ashtonが1999年に提唱した用語で、様々なオブジェクトにIDを付与することによりすべてをネットワーク化する、すなわち「モノのインターネット」を構築する概念です。非常によく似た考えとして「Cyber-Physical System(CPS)」というコンセプトも知られており、こちらの方は「物理的な実体を制御するエレメントが情報を共有し協調して機能するシステム」と定義されています。いずれも「物理世界とサイバー世界を融合する」ところが共通する概念です*1。

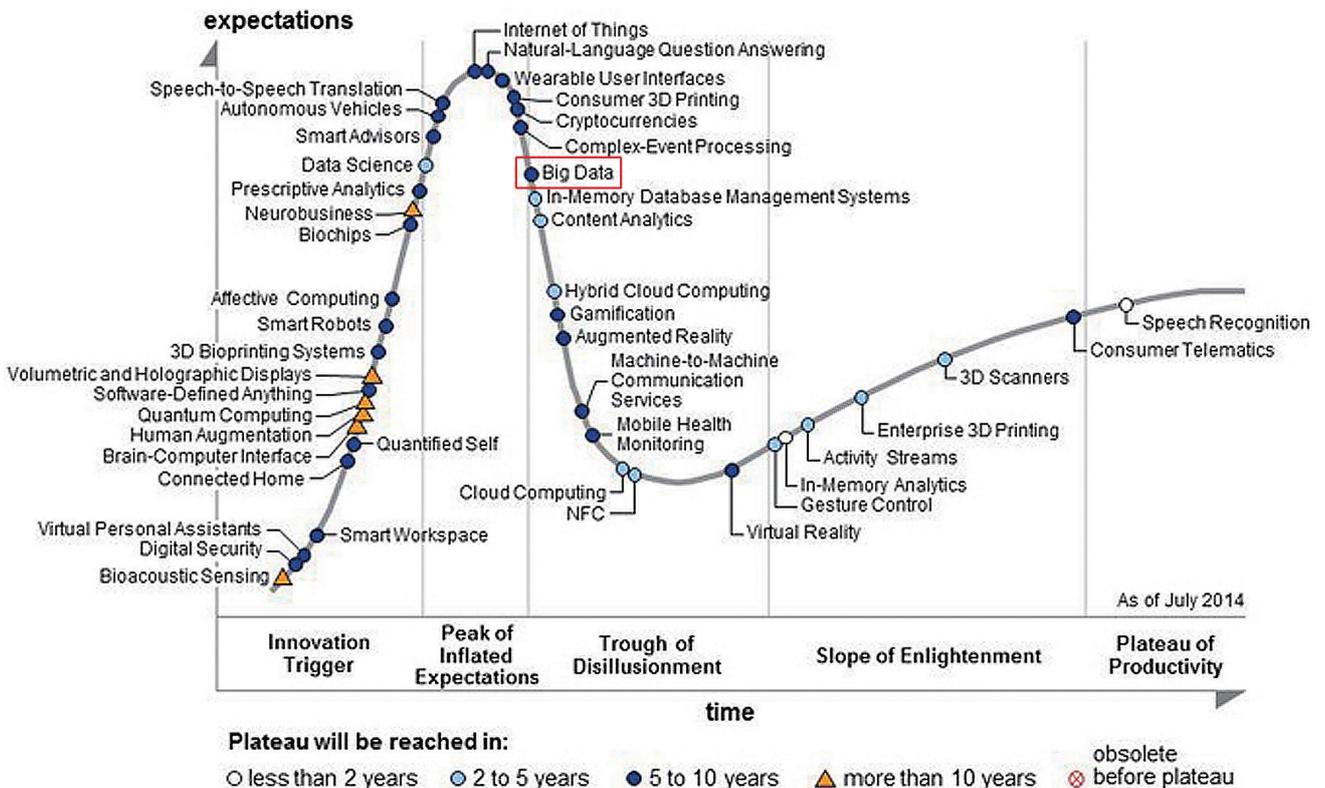


図-1 ハイプサイクル2014版

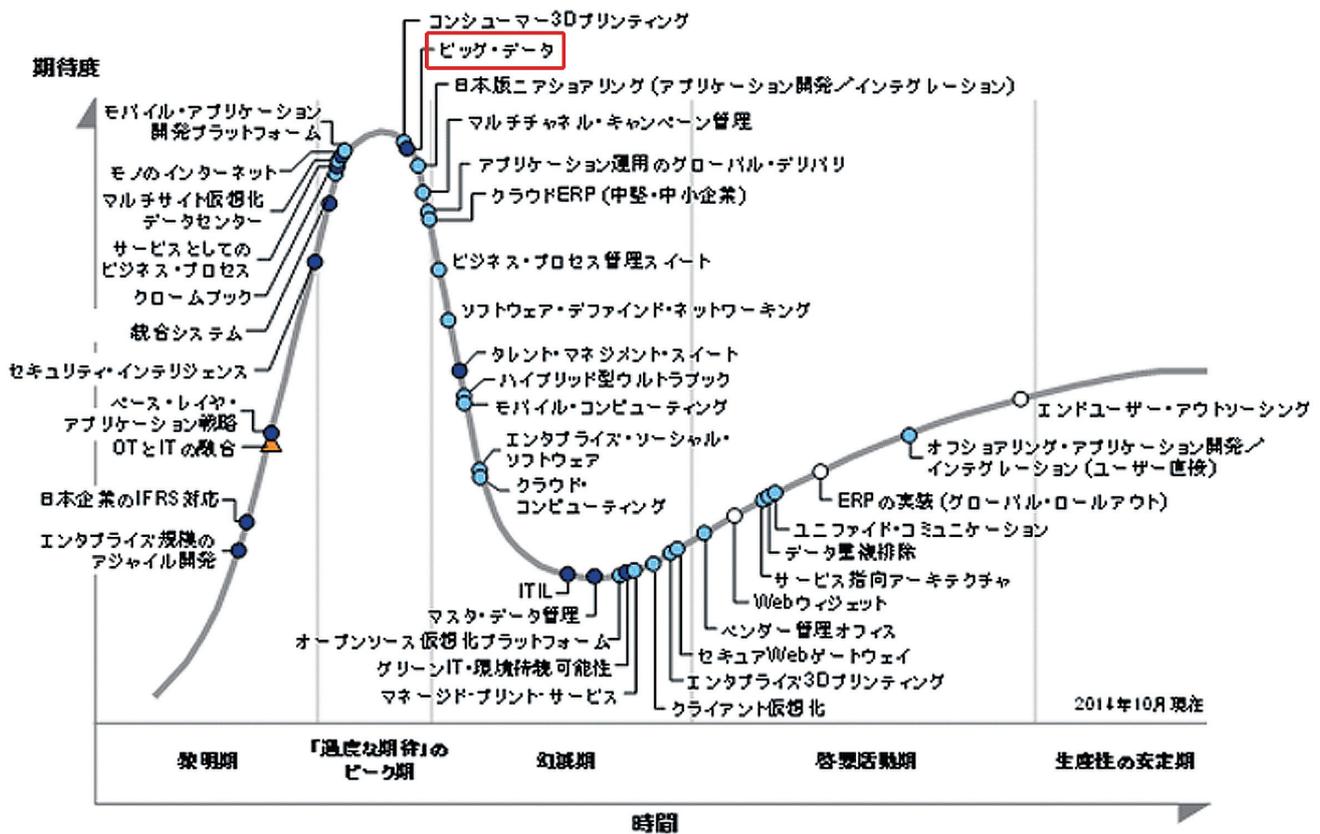
*1 「サイバーフィジカルシステムとIoT(モノのインターネット)実世界と情報を結びつける」岩野 和生、高島 洋典、情報管理 2-15 vol.57 no.11 (https://www.jstage.jst.go.jp/browse/johokanri/57/11/_contents/-char/ja/)。

今日、宅配便の荷物のタグや非接触型ICカードである定期券、あるいはスマートフォンなどには既にIDが付与されており、その移動などを追跡することが可能になっています。したがって「IDごとに生成され続ける位置情報」は文字どおりのビッグデータであり、これを分析することにより、様々な効率化・最適化を図ることができます。これがIoTやCPSが目指す「明るい未来」だとされています。例えば「スマートグリッド」や「スマートシティ」といった提案を目にしたことのある方は多いのではないのでしょうか？

すなわちIoTやCPSは、本質的に社会システムの革新・刷新を目指すスケールの非常に大きなパラダイムです。その研究開発は

社会生活を営む多くの人々に直接的な影響を与える可能性があります。例えば一般市民のプライバシーの問題などが挙げられます*2。筆者のようなビッグデータの分析手法に関心のある研究者にとって、社会的側面に配慮しなければならないという意味で、IoTやCPSに関わる研究は難しい課題を伴う研究対象であると考えます。

このような社会的問題が顕在化してくるのも、ビッグデータに関する研究開発が進展している証左だと思えますが、特にその研究の方向性を熟慮しなければならない踊り場状態にあるのが「2015年のビッグデータ研究の現状」ではないかと考えます。



出典：ガートナー（2014年10月）

図-2 日本におけるテクノロジーのハイブ・サイクル：2014年

*2 NICT（2013年11月25日、プレスリリース）、「大規模複合施設におけるICT技術の利用実証実験を大阪ステーションシティで実施」（<http://www.nict.go.jp/press/2013/11/25-1.html>）。日経コンピュータ（2014年3月11日、清嶋 直樹）、「JR大阪駅ビルの「顔識別」実証実験、プライバシー侵害の懸念から延期」（<http://itpro.nikkeibp.co.jp/article/NEWS/20140311/542723/>）。日経コンピュータ（2014年11月7日、清嶋 直樹）、「JR大阪駅ビルの監視カメラを使った顔識別実験、範囲限定で再開へ」（<http://itpro.nikkeibp.co.jp/atcl/news/14/110701801/>）。

実際に、IoTやCPSが指向する「実世界とサイバー空間を結びつける」アプローチは、「事実をデータ化する」という意味において、ビッグデータ分析に新たな可能性を提供してくれると思います。例えば「ある人間がある時間にある場所にいた」や「あるプラントのある箇所がある時間にある温度に達していた」といった事実を示すデータが分析に活用できるのであれば、少なくともその事実については(論理的に揺らぎのない)確定情報として扱えるので、分析結果の精度向上に寄与するであろうと推測できます。また、そのような確かなデータに基づいて分析・予測しなければならない社会的な需要もあります。例えば「原子力発電所の運転状態監視」であるとか「犯罪者の行動追跡」といった緊急性が高く、確定性が重視される分析では「事実を示すデータ」の価値は絶大です。

3.2 ソーシャル・ビッグデータ

IoTによりもたらされる「事実を示すデータ」を、ここでは便宜的に「ファクト・ビッグデータ」と呼ぶことにします。それとは対照的な特性を持つデータ、例えば「暮らしを豊かにする…」とか「あなたのお好みの…」といった個人の感性や嗜好性を考慮する分析で使用するデータは「ソーシャル・ビッグデータ」と呼べるのではないかと思います。TwitterやFacebookといったのソーシャル・メディアで流通するメッセージなどがその典型的な事例と言えるでしょう。

一般に「雑多な情報の寄せ集め」と理解されがちなソーシャル・ビッグデータですが、Amazonのチーフ・サイエンティストであるAndreas Weigendによれば、ソーシャル・ビッグデータは急激に増大しており、2009年1年間に生成されたデータ総量は、有史以来2008年までに生成されたデータ総量を超えており、これが顧客の商品情報の探索や収集に革命的な変革をもたらしたと報告しています*3。

Weigendは、ソーシャル・メディアのユーザが、相互にメーカーサイトや自身やその友人の感想、あるいは購入した商品の周辺について明確な情報をポストする事例が増えており、メーカーが広報宣伝の一環として提供する情報より役立つ傾向にあるこ

と、また一部のメーカーは、ユーザの率直なデータ提供を奨励し、有益なデータには褒賞する系統立った方法を用意していることを指摘しています。

これらの事実を背景に、今日のオンライン・マーケティングの世界が、ユーザ間の共同作業により第三者にとって有益で明確なデータを作成するモデルに移行(e-businessからme-businessへの移行)しつつあることをWeigendは「ソーシャルデータ革命」と呼んでいます。ソーシャル・ビッグデータの分析は「顧客の期待が何を引き起こすのか?またそのような期待に直面する企業には何ができるのか?」との問いに対する答えを提示できるのかもしれませんが。インターネットで流通するビッグデータがますます増加していくことは自明ですが、今日のIoTの普及やソーシャルメディアの台頭が、流通するビッグデータの特性の二局化を促す状況を加速させているように考えています。

3.2.1 ソーシャル・ビッグデータの性質

ビッグデータ分析という観点で、ソーシャル・ビッグデータをファクト・ビッグデータとの対比で考えると、「機械的な生成が困難」「含意性が多元的」「意味的な曖昧さが許容される」「推測・憶測が許容される(誤りが許容される)」といった性質を持っていることが挙げられます。少々感覚的な表現を使えば、ファクト・ビッグデータが「機械が生成する硬いデータ」であるのに対し、ソーシャル・ビッグデータが「人間が生成する柔らかいデータ」と言えるのではないかと思います。

通常、データ生成に人間が介在すると、そこから得たデータの分析はその分難しくなります。例えば、大きな公園において一番過ごしやすいポイントを調査することを考えてみてください。「過ごしやすい」を「適正な気温」と仮定し、公園内に温度センサーのついたデバイスを1万個配置し、24時間データを収集し続けるシステムを構築して蓄積されたファクト・ビッグデータを分析する場合は非常に明快な結果が得られます。デバイスは公園内にくまなく設置できますし、設置さえすれば確実にその場所の気温データを得ることができます。1万個もあれ

*3 The Social Data Revolution(s) Andreas Weigend MAY 20, 2009 (<https://hbr.org/2009/05/the-social-data-revolution.html>).

ば、そのうちの幾つかは故障するでしょうが、故障が確認できれば直ちに交換すれば良いのです。人間が「過ごしやすい」と感じる気温の地点をくまなく調べ上げることができるでしょう。

これに対して、調査員を1万人雇って「各ポイントの温度を調べてください」と依頼するのがソーシャル・ビッグデータでの調査・分析の方法です。生真面目な調査員であれば温度計を持参し「ポイントXXはYY度でした」と温度センサー付きデバイスに迫る正確な報告をしてくれるかもしれませんが、温度計など持たずに「池の周辺は寒いです」と非常に曖昧な報告をする調査員もいるでしょう。華氏の温度計を見て「72度です」と答えたり、ポイントまで出向かずに適当に嘘の報告をしたり、中には「森林地帯は過ごしやすいです」と見当違いな報告をしてくる調査員もいるかもしれません。これでは公園内の「適正な温度」のポイントを見つけるのは大変です。

ですが、調査員にお願いするのが「過ごしやすいポイントを探してください」という依頼だとしたら、調査結果はガラッと変わってくるでしょう。調査員は、まず過ごしやすいポイントに向かって移動するでしょう。その途上で知り合った他の調査員と意見交換したり、一緒にポイントを探し始めたり、意見が一致するポイントが見つければ「ここが一番いいよね」といったやり取りをするのかもしれませんが。その結果、時間の経過と共に調査員が公園内の幾つかのポイントに集まるようになることが期待できます。そのタイミングを見計らって個々の調査員にヒアリングを行います。「今どこにいるのか?」「そこは過ごしやすいのか?」「そう思う理由は?」といった質問をすると、公園内の過ごしやすいポイントについて第三者が共感できる情報が得られるわけです。

では、どちらの調査結果が適切でしょうか?それは調査目的によって変わってくるでしょう。公園の管理や環境保全の担当者であればファクト・ビッグデータだけを必要とするのではないかと思います。しかし、公園内の売店のオーナーであれば、むしろソーシャル・ビッグデータによる調査結果を欲するかもしれ

ません。この事例での両者の決定的な違いは、ファクト・ビッグデータは調査方法が一様で結果が均質な事実に基づく客観的なデータで構成されているのに対し、ソーシャル・ビッグデータは調査員(人間)の主観に基づくデータの集積物だということです。両者は共にビッグデータですからエラーデータに対処する「信頼性の考慮」が必要ですが、更にソーシャル・ビッグデータに関しては「データは正しいのか?誤りなのか?どの程度信用するか?」といった信用性も考慮した分析手法が必要になります。

3.3 ビッグデータとしてのWikipedia

ソーシャル・ビッグデータと言えばTwitterで流通しているメッセージをイメージされる方が多いと思うのですが、筆者はWikipediaに着目しています。

最もポピュラーな電子辞書サービスであるWikipediaの概要をここで改めて紹介する必要はないと思いますが、ビッグデータとしてもWikipediaは注目すべき存在です。Wikipedia日本語版の「Wikipedia:統計」ページ^{*4}、あるいはWikipedia英語版の"Wikipedia:Statistics"^{*5}を見ると最新のWikipediaの記事総数などが掲載されています。

2015年7月20日時点でのWikipediaの総記事数は4,920,887件で総項目数(周辺情報も含める)36,748,410ページ、日本語に限定しても総記事数974,894件、総項目数2,785,007ページにもなります。このデータはCC-BY-SA 3.0^{*6}というライセンス条件に基づいて改変、複製などの2次の利用をすることもできます。一般的なオープンソースソフトウェアと同様に使える「ライセンス的に最も安全なビッグデータ」と言えるかもしれません。

Wikipediaのデータは"Wikimedia Download"^{*7}というページから辿ることができます。Wikipediaでは、常時データバックアップのタスクが動いており各言語で順次バックアップを行っています。筆者の知見では概ね1ヵ月単位で新たなバックアップが作成されます。

*4 <https://ja.wikipedia.org/wiki/Wikipedia:%E7%B5%B1%E8%A8%88>

*5 <https://en.wikipedia.org/wiki/Wikipedia:Statistics>

*6 「クリエイティブ・コモンズ」(<https://ja.wikipedia.org/wiki/%E3%82%AF%E3%83%AA%E3%82%A8%E3%82%A4%E3%83%86%E3%82%A3%E3%83%96%E3%83%BB%E3%82%B3%E3%83%A2%E3%83%B3%E3%82%BA>)。

*7 <http://dumps.wikimedia.org/>

このWikipediaデータは既に様々なところで活用されています。著名なところと言えば、例えばDBpedia^{*8}、これはWikipediaデータからLinked Open Data (LOD)を生成しデータベース化するプロジェクトです。このデータベースは自然言語処理やテキストマイニングなどに活用されるようです。

3.3.1 ソーシャル・ビッグデータとしてのWikipedia

ではWikiシステムを使った人海戦術で電子辞書としてのコンテンツを維持しているWikipediaはソーシャル・ビッグデータでしょうか？「辞書としての有用性を追求し、ライターが客観的な事実に基づく記述をするよう促すため、システムと運用の両面で数々の工夫を施しているWikipediaの記事データはファクト・ビッグデータに分類すべき」というのが筆者の意見です。もちろん、Wikiシステムを使って人手により記述・修正されているデータですので、機械的に生成されるデータに比べて、記述の誤りや事実誤認、記事相互での矛盾など、データとしての完全性に問題はありますが、その反面、事実が確認不能な事柄や定説が定まらない事柄なども網羅できる「間口の広いファクト・ビッグデータ」とも理解できます。

もっともWikipediaにはソーシャル・ビッグデータとしての顔もあります。2013年から公開されるようになった“Page view statistics for Wikimedia projects”^{*10}がそれに該当します。これはWikimediaプロジェクトの全ページについて、1時間単位でページビューカウントを集計したデータで、ここから辿ると2008年(正確には2007年12月)以降のページビュー情報を入手することができます。

筆者の研究グループでは2013年6月にこのデータに基づくランキングサービス^{*11}を立ち上げて以来、ランキングの変動の面白さ、すなわち「世間で話題になっているトピックがランキング上位にマークされる」振る舞いに注目していました(図-3)。

3.3.2 Tobias Preisの研究

ウォーリック大学の金融行動学の准教授であるTobias Preisによれば、インターネットを利用する個人の情報探索行動がこのような振る舞いを発生させていると説明しています。例えば、マスメディアなどで大々的に報道されているトピックに接すると、日常的にインターネットを利用する個人はサー



図-3 Wikipediaランキング

*8 <http://wiki.dbpedia.org/>
 *9 <http://ja.dbpedia.org/>
 *10 <http://dumps.wikimedia.org/other/pagecounts-raw/>
 *11 <http://www.gryfon.ii-j.co.jp/ranking/>

チエンジンで検索したり、Wikipediaを閲覧して、トピックに関する情報を収集する行動をとります。この行動の痕跡は検索エンジンのクエリデータやWikipediaのページビューとして記録されると推測されます。この現象に注目したPreisは2010年の論文"Complex dynamics of our economic life on different scales : insights from search engine query data"^{*12}で、サーチエンジンのクエリデータと株式市場の変動が相関していることを突き止めました。

続く論文"Quantifying Trading Behavior in Financial Markets Using Google Trends"^{*13}では、Google Trendsから得られるクエリデータに含まれる金融関連の98の用語の検索量の増大が金融市場の大きな損失に先行する傾向があることを示唆しました。

更に論文"Quantifying Wikipedia Usage Patterns Before Stock Market Moves"^{*14}では、Google Trendsでの分析の知見を使ってWikipediaの閲覧回数が株式市場の大規模な変動と相関することを発見しています。

これらの論文でのPreisの結論では、検索エンジンのクエリデータやWikipediaのページビューといったオンラインデータから、意思決定を迫られている人々の情報収集活動について新たな知見が得られる可能性を指摘しています。例えば、株式市場の大暴落といった事象は個々の投資家の意思決定の結果ですが、オンラインデータに注目していれば、その兆候を早期に発見することができるわけです。これはソーシャル・ビッグデータを活用した典型的なソリューションの一例でしょう。

3.4 まとめ

本稿ではソーシャル・ビッグデータとその性質を中心に議論してきました。Andreas Weigendが主張する「ソーシャルデータ革命」は、Eコマースといった特定の分野では、商品購入における人間行動に関する膨大なデータが蓄積され始めており、そのデータの分析を前提とした新たなアプローチの必要性が主張されています。

ソーシャル・ビッグデータが「人間の行動に関するデータ」であることから、その分析方法の開発は社会科学や行動科学の領域の研究となるでしょう。しかしながら、人間の行動に関してこれだけ広範囲かつ詳細なデータを入手できるような状況がこれまでなかったことを考えると、価値ある知識を抽出できる有効な手法を見出すには相応の時間が必要でしょう。筆者が知る限りでは、マイクロブログやSNSなどから得たデータを元に個人の行動を追跡、分析するアプローチは「なんらかの分析結果は得られるものの、そこから有意な知見を見出すことが難しい」という意味においてあまり成功していないと考えます。

一方、インターネット利用者の行動をマクロな現象と捉え、複雑系の分析手法を応用するTobias Preisのアプローチの方がむしろ有意な知見が得られやすいのではないかと考えます。この分析にはクエリデータやWikipediaのページビューなどが用いられていますが、Preisの論文は、独立した個人による集団行動の予兆を捉えることにより、短期的な予測が可能であることを示しているように思います。また、この分析によって得られた変化の著しい現象をより詳細に分析するには、ソーシャルメディアから得られたデータを使って個人の行動を追跡することにより、要因などの知見を得ることができるのではないかと、考えています。



執筆者：
藤田 昭人 (ふじた あぎと)

株式会社IIイノベーションインスティテュート (II-II) 企画開発センター チーフアーキテクト。2008年II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。

*12 http://www.tobiaspreis.de/publications/prs_ptrsa_2010.pdf

*13 <http://www.nature.com/srep/2013/130425/srep01684/full/srep01684.html>

*14 <http://www.nature.com/srep/2013/130508/srep01801/full/srep01801.html>



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。

©2008-2015 Internet Initiative Japan Inc. All rights reserved.
IIJ-MKTG019bA-1508GR-10000PR

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/>