# IIR

Internet
Infrastructure
Review

Jun.2024
Vol.63

定期観測レポート

## 日々高度化するサイバー攻撃から お客様を保護するための取り組み

フォーカス・リサーチ(1)

## W3C標準化活動： RDF Dataset Canonicalization

フォーカス・リサーチ(2)

## IIJにおけるDRMの取り組み

IIJ
Internet Initiative Japan

# Internet Infrastructure Review

June 2024 Vol.63

## エグゼクティブサマリ

2022年11月30日、OpenAI社からChatGPTが発表され、その能力は世界中で大きな反響を呼びました。その後もChatGPTには様々な機能が追加され、他社からも生成AIが発表されるなど、世の中では生成AIブームが巻き起こっています。多くの組織においても付加価値向上や効率改善に活用され、社会への実装が進んでいます。

ChatGPTの発表から1年半後の2024年5月13日、OpenAI社は最新モデルであるGPT-4oをリリースしました。既に利用された方はその進化を自ら体験されていることと思いますが、評価記事や紹介動画を見るだけでも、変化のほどを垣間見ることができます。生成AIの技術開発のスピードには目を見張るばかりです。

その一方で、AIのネガティブな側面も多く指摘されるようになり、日本では今年に入ってからAIが詐欺に利用されているというニュースを目にすることが増えました。著名人に扮した画像や音声のディープフェイクが使われる事件も増えています。また、選挙やプロパガンダにAIが利用されるといったことは、以前から懸念されていました。今年は世界各国で重要な選挙が予定されており、警戒を強める必要があります。

そうしたなか、欧州連合(EU)でAI規制法案が5月21日に承認されました。2026年から本格適用されるもので、リスクの高さによって規制の強さを4段階に分け、リスクの高い違反に対しては高額の制裁金が課せられます。もっとも厳格に禁止されたAIは、特定の個人や団体に不利益をもたらすソーシャルスコアリングや犯罪行動予測などで、それに次いで厳しく規制すべき高リスクAIには、入学、採用、生体認証、インフラ運営が含まれるなど、幅広く要件と義務が定められました。また、生成AIで作成した画像や音声などは、AIによるものだと明示することが求められます。

EUのAI規制法は、これから世界のAI規制のスタンダードになりうるものです。事業にAIを利用する上で法的規制を理解・遵守するのは当然のこととして、高い倫理観を持つことが強く求められます。今後もAIが社会の隅々にまで普及するなかで、個々人がAIに対するリテラシーを身につけることがますます重要になっていくでしょう。

「IIR」は、IIJで研究・開発している幅広い技術を紹介しており、日々のサービス運用から得られる各種データをまとめた「定期観測レポート」と、特定テーマを掘り下げた「フォーカス・リサーチ」から構成されます。

1章の定期観測レポートは「メッセージング」です。電子メールはインターネット創生以来、長く利用されている重要なアプリケーションです。大量の宛先に向けて、非常に簡易にメッセージを送付することを可能にした電子メールの歴史は、メールシステムの管理者がabuse行為と戦ってきた歴史でもあります。今回は、近年見られるabuse行為の状況と、それに対するIIJの新しい取り組みを紹介します。送信ドメイン認証に関して大きな動きがあったこの1年の動向と課題についても説明します。

2章のフォーカス・リサーチでは「RDF Dataset Canonicalization」を取り上げます。RDF(Resource Description Framework)は、WEBで情報を表現するための枠組みで、W3Cにおいて標準化されています。筆者はW3Cにて、RDF Dataset Canonicalization(RDFで表現されたデータの正規化の仕組み)の標準化活動に携わっています。RDFの概要を説明した上で、正規化が必要となる背景・手順・課題など、標準化活動の状況について紹介します。

3章のフォーカス・リサーチは「動画配信におけるDRM(Digital Rights Management)」についてです。複製が容易なデジタルコンテンツを流通させる上で、コンテンツの権利を守るための技術は必要不可欠です。ここでは動画のDRMを解説していますが、今日のようにインターネットの動画配信が盛んになったのは、DRMによる貢献が大きいと言えます。エンドユーザが動画を楽しんでいる裏側でどのような処理が行われているのか、思いを馳せていただければと思います。

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、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てまいります。

**島上 純一**(しまがみ じゅんいち)

IIJ 常務取締役 CTO。インターネットに魅かれて、1996年9月にIIJ入社。IIJが主導したアジア域内ネットワークA-BoneやIIJのバックボーンネットワークの設計、構築に従事した後、IIJのネットワークサービスを統括。2015年よりCTOとしてネットワーク、クラウド、セキュリティなど技術全般を統括。2017年4月にテレコムサービス協会MVNO委員会の委員長に就任し、2023年5月に退任。2021年6月より同協会の副会長に就任。

# 日々高度化するサイバー攻撃から お客様を保護するための取り組み

## 1.1 電子メールが新たな時代へ

前回の報告[*1]から1年、2023年は電子メール関連業界が大きく動いた年となりました。

本稿の前半では、IIJが観測している最新の攻撃手法について報告し、新しく取り組み始めた対策内容を紹介します。後半では、この1年間で観測している送信ドメイン認証技術DMARC対応率の劇的変化について報告します。

電子メールは組織内外のコミュニケーション手段として重要なインフラでありながら、一度構築するとなかなか変更されないものです。しかし、世の中の攻撃手法やセキュリティ動向は絶えず変化しており、常に対策を講じ続けていく必要があります。この機会に電子メールインフラを見直してはいかがでしょうか。

## 1.2 お客様を脅威から保護する取り組み

### 1.2.1 abuse対応とは

メールサービスを悪用され、フィッシング(詐欺)メールを送信されるケースが後を絶ちません。

多くのISP(インターネットサービスプロバイダ)やホスティングメールサービスでは、電子メールを送信するとき、メールアカウントのユーザIDとパスワードの組み合わせ(クレデンシャル)を用いてSMTP認証し、認証に成功した場合のみ電子メールを送信できるのが一般的です。これは認証することで利用者を識別することと、第三者によるメールサービスの不正利用から守ることが目的です。

しかし、悪意のある者によって利用者のクレデンシャルが何らかの方法で盗まれ、メールサービスを用いてフィッシングメールの送信に利用される(アカウントの乗っ取り)行為が日常的に発生しています[*2]。これはIIJに限らず、他ISPや他社サービスでも発生しており、このようなインターネット上での迷惑行為や不正行為をまとめて、一般的に「abuse(アビューズ)行為」と呼びます。

### 1.2.2 abuse行為が行われると

悪意のある者によってメールサービスがフィッシングメールの送信に不正利用されると、どのようなことが起こるでしょうか。

近年は、単に迷惑な広告メールを送信するのではなく、攻撃対象となる宛先のユーザが利用しているWebサービス・アプリなどのIDやパスワードを盗み出す手段としてフィッシングメールを送信しています。彼らの最終目的はフィッシングメールで騙したユーザからIDやパスワードを盗み出し、銀行口座やクレジットカード番号を盗むことで、金銭的な利益を得ることです。近年はクラウドサービスの利用も進み、この動きがますます加速しています。

当然ながら、このようなフィッシングメールを送信する行為は多くのメールサービスの利用規約や約款で禁止されています。彼らは規約違反としてメールの送信が制限される前の、極めて短時間のうちに大量のフィッシングメールを送信することで、攻撃の成功率を上げようとしているのです。まさに「数打てば当たる」状況となっています。

＊1　IIR Vol.59(https://www.iij.ad.jp/dev/report/iir/059.html)。
＊2　ひと昔前はクレデンシャルの総当たり攻撃によってパスワードを発見されるケースもありましたが、この行為自体がabuse行為であり効率の悪い手法です。昨今は事前の認証試行をされることなく、1回目からフィッシングメールの送信に成功しているケースがほとんどですので、悪意のある者が何らかの方法で事前にクレデンシャルを入手していると考えるのが自然です。

### 1.2.3 abuse行為の問題点

このような状況を放置すると、攻撃対象となったユーザへの被害のみならず、メールサービスでも次のような悪影響が発生します。

- 悪意のある者によって、メールサービスから大量にフィッシングメールが送信されることで設備が過負荷になり、サービス障害の発生や、可用性の低下につながる(図-1(1)、(2))。
- フィッシングメールを送信されてしまった結果、宛先メールサーバやセキュリティベンダーなどで、メールサービスがフィッシングメールの送信元として登録され、他の正当な利用者のメールが宛先で迷惑メール判定され、受信拒否などされる(図-1(3)、(4))。
- あるセキュリティベンダーが、別のセキュリティベンダーの脅威情報を参照していることがあり、一度登録された脅威情報の解除までには時間が掛かるため、この影響が長引く(図-1(5)、(6)、(7))。

従って、IIJのメールサービス「IIJセキュアMXサービス」では、サービスの安定稼働や、他のお客様への悪影響回避のため、abuse行為が発生した場合は即座に調査し、メールサービス利用者のクレデンシャルを強制変更したり、該当の通信を遮断したりするなど、日夜設備の保護に努めています。

### 1.2.4 abuseと通信の秘密

日本では、電気通信事業者が取り扱う電気通信について知得などの行為をすることは、電気通信事業法第4条によって禁止されています[*3][*4]。しかし、abuse行為の発生が明示的に認知されており、これを放置するとサービス利用者が他人の権利を侵害する不法行為に加担してしまったり、自身が被害に遭ったりする強い蓋然性(がいぜんせい)がある場合に、それらの事態を回

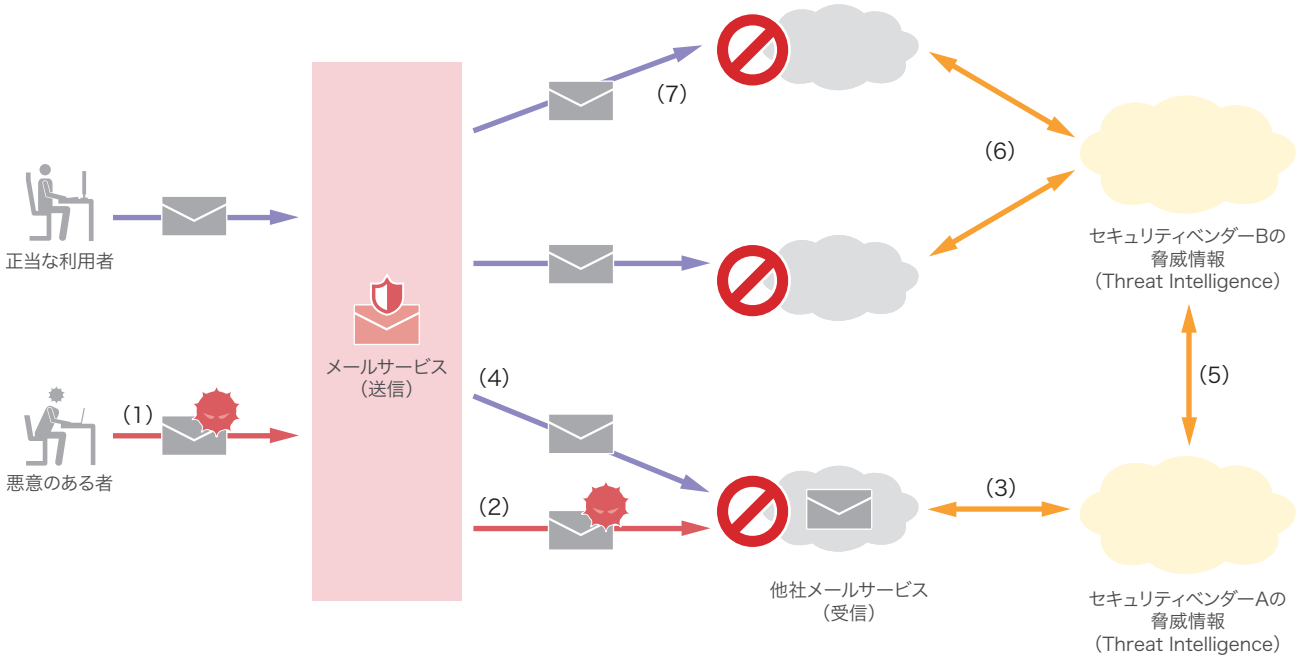



図-1 abuse行為が与えるメールサービスへの悪影響

＊3　この「通信の秘密」の概念は、もともと郵便物(信書)の通信を指すものでした。いつ、誰が、どの相手と、どのような内容のやりとりをしているかといった情報を第三者に知られない権利です。日本国内ではインターネットにおいても適用されますが、そのように整理される国は世界でも少数派で、諸外国では検閲が合法な国が圧倒的多数とされています。「ネットを監視も干渉もしない国は、日本を含むたった4カ国だけ」(谷脇康彦著「教養としてのインターネット論」(日経BP, 2023年)米国NPO調査(https://freedomhouse.org/sites/default/files/2022-10/FOTN2022Digital.pdf)。

＊4　郵便局員がはがきの表面を見て宛先のポストに投函することと同様に、インターネットではIPパケットのヘッダ、電子メールではSMTPプロトコルの通信内容を見なければ、相手に届けることができません。そのため、このような行為は「通信の秘密を侵害するが、正当業務行為である」と整理されます。

避するために電気通信に関与することは、緊急避難や正当業務行為として違法性が阻却され得ると整理されています。

また、IIJと利用者間の契約において、abuse行為に相当する行為は禁止されており、契約違反が明らかであるときは、IIJとしても契約当事者としての諸措置をもって対処することができます。

### 1.2.5 不正利用の準備行為の発見

ここ数年、前触れもなくフィッシングメールを送るのではなく、フィッシングメールが送られる数日前に数通、無害と思われるメールを送信する「お試し送信」を日々の運用業務の中で複数回発見しています。例えば、メールの件名に次のような情報が記載されています。

```
メールアドレス ; ログイン ID; パスワード ;SMTP サーバ名 ; ポート番号 ; 送った通数 ; 認証方式
(例)
iij-taro@example.jp;iij-taro;password;mail.securemx.jp;465;2;LOGIN
```

このようなメールの宛先には、探索行為の結果を収集していると考えられるメールアドレスが記載されており、数日すると実際にフィッシングメールの送信が行われるという因果関係が明らかになってきました(図-2)。

しかし、このような準備段階の時点では他人の権利を侵害している、または権利侵害の強い蓋然性(がいぜんせい)があるとは言い難く、必ずしもabuse行為とは言えません。そのため、このような準備段階を察知しつつも、実際にabuse行為が行われるまで、私たちは具体的な対処を行う根拠がない状況でした。サービスの品質を維持し、お客様を保護するための設備を運用する立場としては、とても歯がゆい状況です。

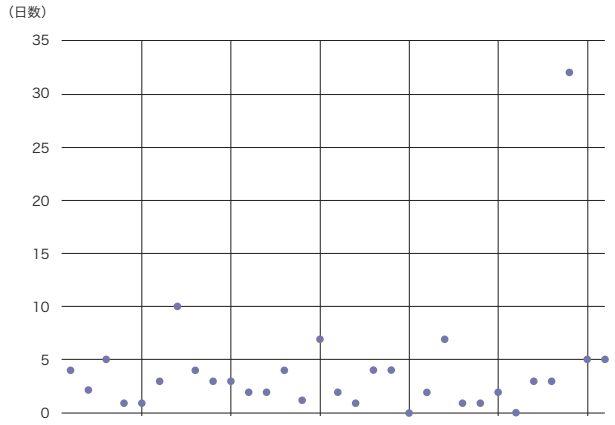



**図-2 特定期間中の探索行為から、実際にabuse行為が行われるまでの日数**

### 1.2.6 IIJの新しい取り組み

このままではお客様を保護できず、指をくわえて見ていることしかできません。そこで新しい枠組みを整備する必要があると考えました。IIJのサポート部門、法務部門を巻き込み、こうした不正利用のための準備行為を察知した場合、実際にフィッシングメールが送信される前に、必要な範囲で通信を制限する枠組みの設計に取り掛かりました。

フィッシングメールが送信される前に対策できれば、表-1のようなメリットがあります。

関係者と議論を重ねた結果、次の手順をもって契約上の行為として導入することとなりました。

- すべてのお客様に対して、事前に新たに取り組む内容を具体的に説明する。
- IIJセキュアMXサービスの約款にも具体的な規定を設ける変更を行うことで、お客様に十分に認知していただく。

**表-1 対策によるメリット**

| 視点 | メリット |
|---|---|
| メールサービス事業者(IIJ)視点 | ・フィッシングメールを送信される前に対策できるため、サービス障害を回避したり、メールの宛先に届かないような状況に陥るのを回避できる。 |
| 利用者視点 | ・クレデンシャルが漏えいしている状況を早期に察知できる。<br>・メールを盗み見られたり、情報流出する被害を抑えられる。 |

既存のお客様については運用管理担当者様宛へお知らせがメールで送付されていますのでご確認ください。該当の約款は2024年5月1日の改訂で盛り込まれています。IIJセキュアMXサービス個別規程第12条(不正利用などのおそれへの対処)をご参照ください。

ちなみに、実際に行われた迷惑行為の事後対処を「abuse対応」と呼ぶのに対して、不正利用の準備行為を事前に察知してお客様を保護するこの取り組みについて、社内では「ディフェンス対応」と名付けました。

### 1.2.7 IIJのディフェンス対応

従来は日々の運用業務の中で発見していた探索行為ですが、人力での実施には限界があります。そこで大規模なログの分析基盤として社内でも活用している「illumino[*5]」を用いて、不正利用の準備行為と思われる事象を、機械学習を用いて検出しています(図-3)[*6]。

実は最初からディフェンス対応のためにSplunkを活用していたのではなく、もともとはabuse対応を効率化するための調査用として利用を検討していました。ところが調査を進めていく段階で、こうした不正利用の準備行為を発見し、この検出にも機械学習が適用できるのではないかと考え精度を高めていった結果、かなり高い確率でこの準備行為を発見することに成功しました。

### 1.2.8 まとめ

他人の通信を媒介している電気通信事業者は、当該の法令によって事業遂行を厳しく規律されていることは、一般消費者にはあまり知られていません。しかし、インターネットは世界中と繋がっています。私たちが悪意のある者からお客様を保護するとき、その行為は常に「通信の秘密」の保護とも隣合わせであり、双方のバランスについて留意しながら日々業務にあたっています。

IIJでは、日々高度化するサイバー攻撃から、お客様を保護するための取り組みを続けてまいります。

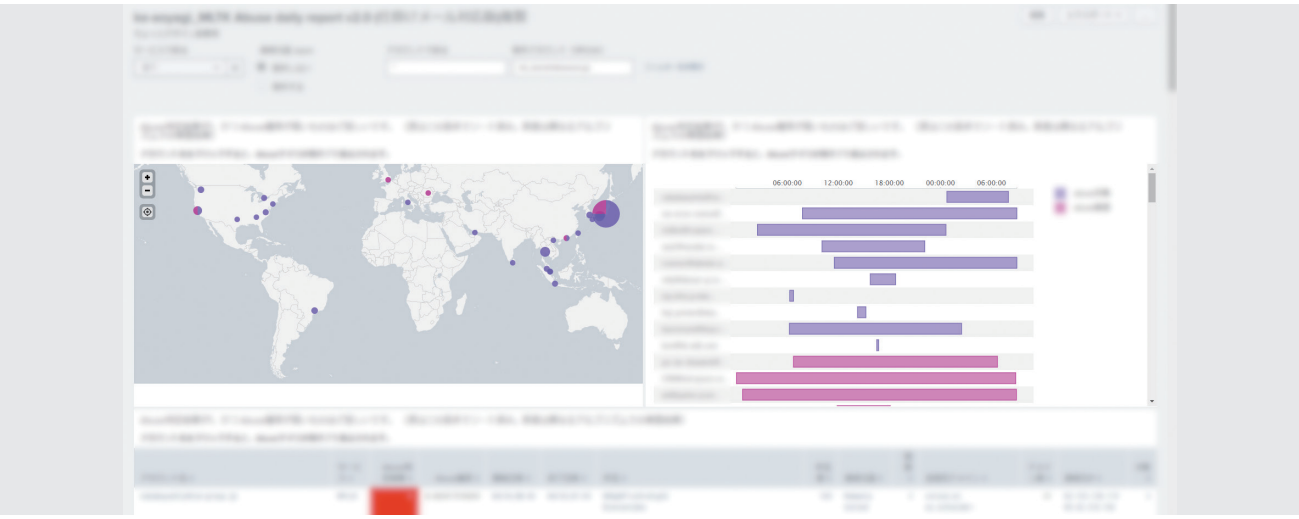


図-3 ディフェンス対応用Splunkダッシュボード

＊5　社内情報分析基盤「illumino」:Internet Infrastructure Review Vol.57(https://www.iij.ad.jp/dev/report/iir/057/03.html)。
＊6　Splunkの活用事例 本レポートのVol.48(https://www.iij.ad.jp/dev/report/iir/048/03.html)「フォーカス・リサーチ(2)Splunkによる日本語文章解析処理」。

## 1.3 送信ドメイン認証に対する大きな動き

### 1.3.1 金融業界でのDMARC対応要請とその動向

2023年2月、総務省・警察庁・経済産業省から、クレジットカード会社など金融機関に対して、なりすましメール対策としてDMARCポリシーの導入の要請がありました[*7]。

従前より、各クレジットカード会社などに対してはなりすましメールによる被害が増加しており、対策の必要性が叫ばれていましたが、この要請を機に、本格的な対策が始まったようです。図-4の金融業界ドメインのDMARC対応率の増加を見てもその様子が良く分かります[*8]。

DMARCポリシーを記載しているドメインは、2023年1月には20%程度しかありませんでしたが、1年後の2024年1月頃には、全体の80%まで増加する結果となりました。ただ、DMARCポリシーを導入したが、いまだにp=noneのままであるドメインも多く存在します。DMARC の効果を正しく得るためには、p=quarantine、p=rejectに変更が必要です。金融業界では2023年にこのような大きな動きがありましたが、日本のドメイン全体に目を向けると、まだまだ対応していない企業が多くある様子が窺えます(図-5)[*9]。

金融業界に続いて、その他の業界でもDMARC対応が進むよう、今後も動向に注目していきたいと思います。

### 1.3.2 GoogleとYahoo!(米国)の送信ドメイン認証非対応メール受信拒否ポリシー声明発表

2023年10月、GoogleとYahoo!(米国)が2024年2月より、送信ドメイン認証に対応していないメールは受信拒否する、という発表をしました。GoogleとYahoo!の両社は、日頃からかなりの量のspamやbulk mailに悩まされており、それらを受信拒否するために送信ドメイン認証していないメールを一律拒否することにしたということです。

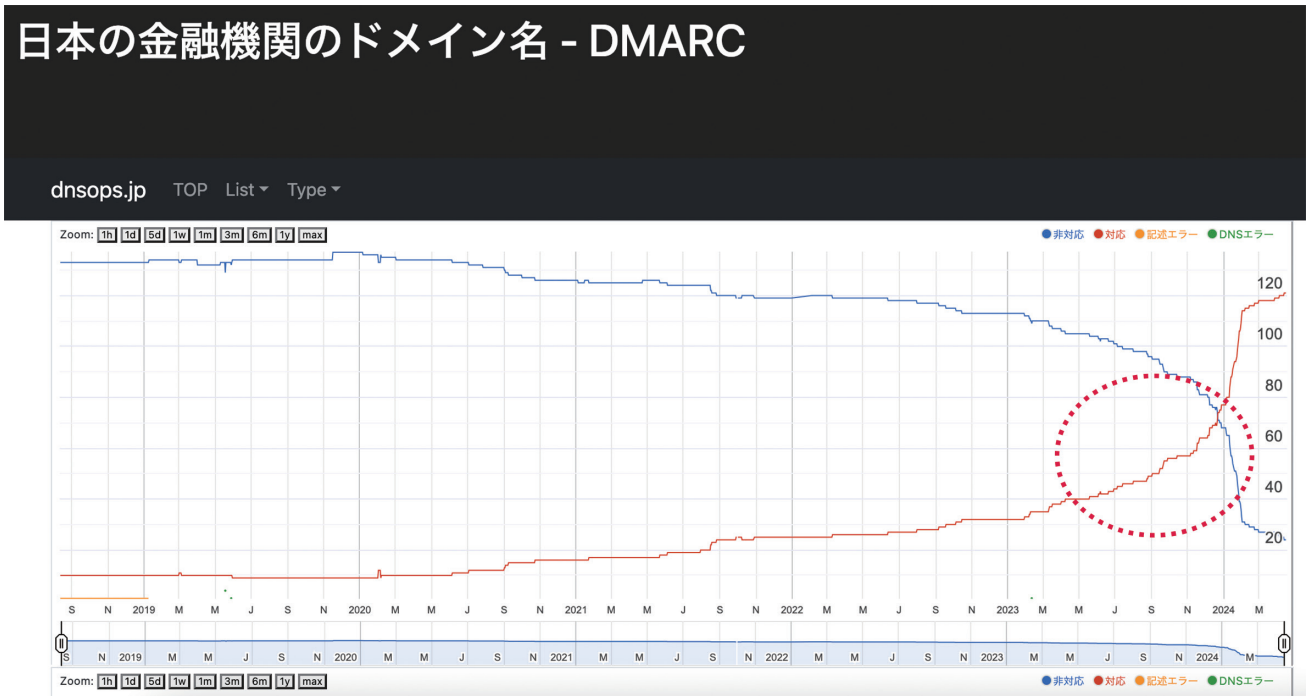



図-4 日本の金融機関ドメインのDMARC対応の様子

*7 総務省、「クレジットカード会社等に対するフィッシング対策強化の要請」(https://www.soumu.go.jp/menu_news/s-news/01kiban18_01000184.html)。
*8 日本の金融機関のドメイン名 - DMARC(https://stats.dnsops.jp/chart/jp-bank/dmarc)。
*9 日本のDNSSEC/SPF/DMARC 対応状況 - DMARC(https://stats.dnsops.jp/chart/all/dmarc)。

送信ドメイン認証技術は、2006年にSPFがRFC 4408、2007年にDKIMがRFC 4871、2014年にDMARCがRFC 7208としてそれぞれ公開されています。DMARCは2014年の公開から、9年経過した2023年でもなかなか普及しない状況が続いていました(図-5)[*10][*11]。

そのような状況下での、最大級のメール受信量を誇る世界最大手のGoogleとYahoo!の受信拒否ポリシー適用声明は、日本に限らず世界中に衝撃を与えました。メールの到達率が重要なサービス指標となるメール送信事業者は早急なDMARCへの対応を求められる形となりました。

発表の直後、2023年10月に開催されたグローバルな迷惑メール対策団体「M$^3$AAWG」の会議では、GoogleとYahoo!の担当者を招いて本アナウンスに対する質問会が緊急開催されました。各国から参加しているメール送信事業者を中心に、どの程度のコミットメントが求められるのか、本当に対応しないとメールを受信してもらえないのかなど、白熱した質問会となりました。

その後、2023年11月に開催された、日本でインターネットのセキュリティについて議論するワーキンググループ「JPAAWG」の会議では、日本国内でメール事業を展開している各事業者を中心に本件に関する議論もなされました。メール送信事業者に限らず、ドメインオーナーである各企業やメールボックス事業者ももちろん対応が必要であり、IIJでも法人、個人それぞれに展開している各サービスで対応を進めていました。

IIJの法人向けサービスでは既に各送信ドメイン認証(SPF、DKIM、DMARC)に対応する仕組みを準備していましたが、各機能の利用は利用者であるお客様による設定の変更や準備が必要であったため、お客様からの送信ドメイン認証に関する問い合わせの量は2023年末から激増する形となっていました。

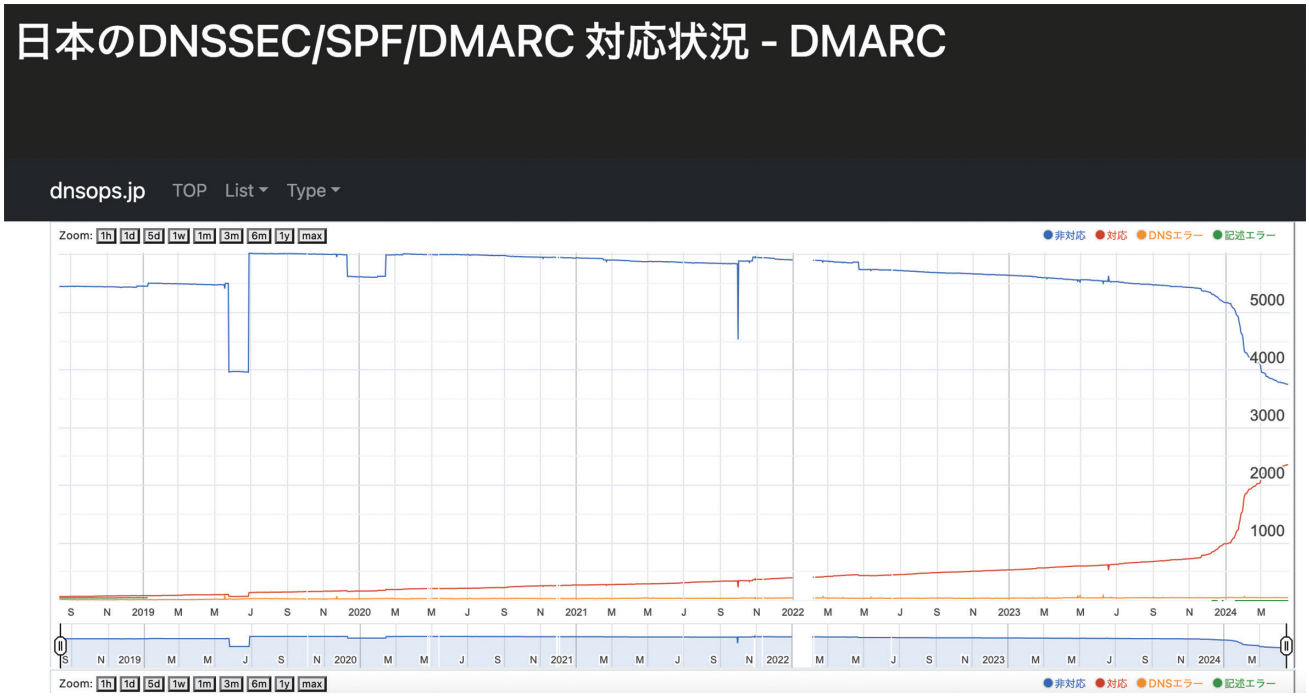



図-5 日本のドメインのDMARCの様子

*10 日本のDNSSEC/SPF/DMARC 対応状況 - DKIM(https://stats.dnsops.jp/chart/all/dkim)。
*11 日本のDNSSEC/SPF/DMARC 対応状況 - DMARC(https://stats.dnsops.jp/chart/all/dmarc)。

こうして各事業者や各企業、組織が対応した結果、IIJセキュアMXで受信しているメールの各送信ドメイン認証の対応率が激変しました(図-6)。

DKIM署名の割合が15%強増加し、DMARCレコードを記載した送信元ドメインの割合(DMARCパイチャートにおけるnone以外の割合)が42%から75%に30%以上増加しました。2022年には32%であったことを振り返ると、2023年はやはり、前述した事項を理由として、DMARCレコード対応をした事業者や企業が増えたと考えられます。

ただ、これらはあくまでも"DMARCレコードが存在していること"までしか確認しておらず、DMARCポリシーがp=none、quarantine、rejectのいずれかまでは見ていません。Googleは、今後も各ドメインオーナーには、DMARC集計レポートを確認しながらp=noneからquarantine、rejectに変更していく対応が求められると考えられます。

### 1.3.3 送信ドメイン認証技術が抱える問題点

近年、クラウドサービスの利活用が進み、企業がオンプレミス設備から直接インターネットへメールを送信すること自体が減少傾向にあります。そのような状況から一部では、SPFレコードの評価のみではそのメールの信頼性を担保するには値しないのではないか、という議論が出始めています。

また、複数のクラウドサービスからメールを送信するために、SPFレコードのDNSルックアップの上限である10回を超えてしまい、SPFレコード自体がerrorとなるドメインも、一部で確認されています。

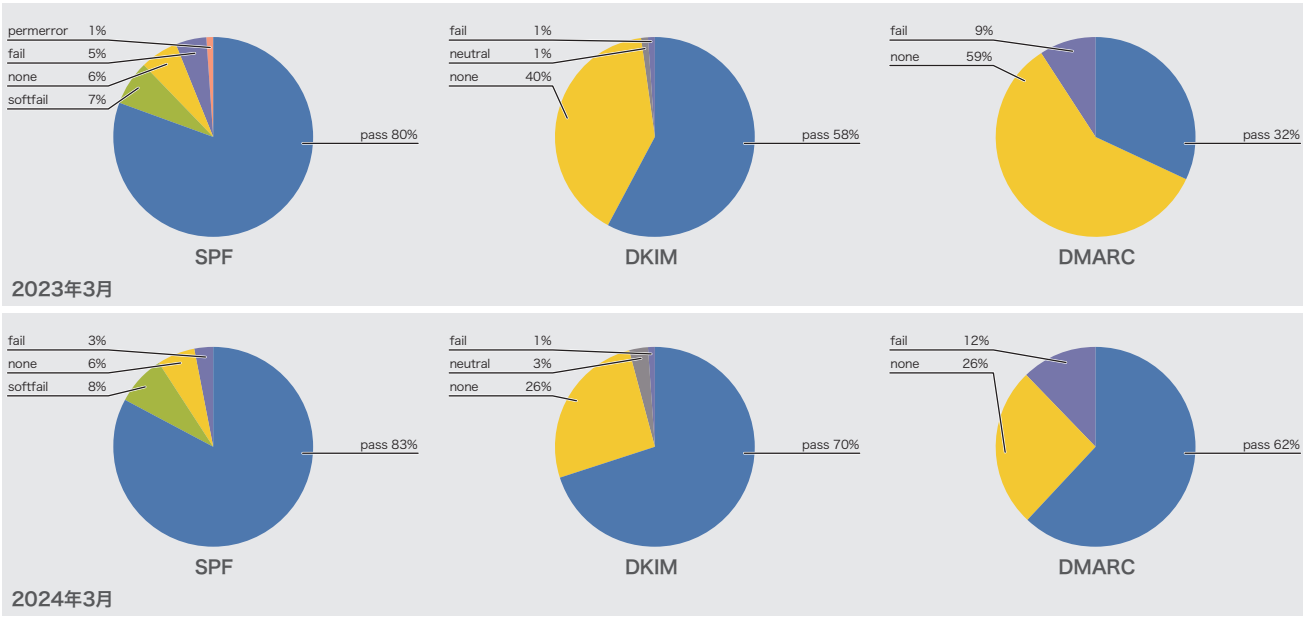



図-6 セキュアMXにおける受信メールに対する送信ドメイン認証対応割合(2023年・2024年比較)

これを回避するために、一部の事業者ではCNAMEレコードを用いてincludeされているレコードを1レコードに展開するようなサービスを提供している状況です。本来SPFレコードのincludeに制限がある理由は、includeが多いSPFレコードがDNS lookup amplifierとなり得る可能性を鑑みて制限されているものです[*12]。

クラウドサービスの中には、膨大な数のIPアドレスレンジが設定されているSPFレコードが指定されているものもあり、そのようなクラウドサービスから自身のドメイン名を用いたメールを送信する際は、DKIM署名を実施することで、SPFによる規格の制限を回避し、送信ドメイン認証のDKIMでpassさせることができます。

一方で、DKIMも万能ではなく、DKIM replay attackの対策や、DKIM署名鍵の有効期限を適切に管理することが必要です[*13]。

DMARCについても、転送メールやメーリングリストなど、DKIM署名後にヘッダ情報が書き換わってしまう従来からあるメールの仕組みが原因で、DKIM署名の検証に失敗する事象への対応が各事業者間で根深く残っています。

このDKIMの検証に失敗する状況を回避するためにヘッダ情報などが書き換わったあとで再署名するARCという仕組みがありますが、DKIM同様、どの署名ドメインを信頼するべきかは、受信側の判断に委ねられています。

送信ドメイン認証に関する課題はまだ多くありますが、IIJとして情報収集や発信ならびにIETF規格策定への参加等、尽力していく所存です[*14]。今回、GoogleとYahoo!(米国)の発表を発端としたSPF、DKIM、DMARCへ対応するというのはあくまでもスタートであり、それぞれ継続的な対応が求められるということに気をつけなければなりません。

執筆者:

1.1 電子メールが新たな時代へ、1.2 お客様を脅威から保護する取り組み
**古賀 勇**(こが いさむ)
IIJ ネットワーク本部 アプリケーションサービス部 メールサービス運営課 課長。
2007年IIJ入社。メールサービスの運用業務に従事し、現場でメールに関する動向を調査。お客様のメールボックスを守るため、最新の攻撃手法や、迷惑メールのトレンド、対策情報などを発信・公演。$M^3$AAWG、WIDE Project、openSUSEなどで幅広く活動中。

1.3 送信ドメイン認証に対する大きな動き
**今村 侑輔**(いまむら ゆうすけ)
IIJ ネットワーク本部 アプリケーションサービス部 運用技術課 リードエンジニア。
2015年IIJ入社。メールサービスの運用業務に従事。IIJ Europeでの就業経験を活かし、日々グローバルに活躍中。

*12 Datatracker IETF、11. Security Considerations, 11.1. Processing Limits(https://datatracker.ietf.org/doc/html/rfc7208#section-11.1)。
*13 IETF、DKIM Replay Problem Statement(https://www.ietf.org/archive/id/draft-ietf-dkim-replay-problem-00.html)。
*14 Datatracker IETF、The Authenticated Received Chain (ARC) Protocol(https://datatracker.ietf.org/doc/html/rfc8617)。

# W3C標準化活動:RDF Dataset Canonicalization

## 2.1 はじめに

本稿では、筆者がWorld Wide Web Consortium(W3C)で標準化に協力し、2024年5月にW3C勧告となったばかりのRDF Dataset Canonicalization[*1]について紹介します。RDF Dataset Canonicalizationは、Resource Description Framework(RDF)で表現されたデータを正規化する仕組みです。以降では、RDFがどのようなものであるか、そしてRDFの正規化とはどんな処理であるか、更にどういった場面でそれが必要とされるのかを説明します。また、W3Cにおける標準化活動の経緯や、具体的な正規化の手順についても説明します。

## 2.2 RDFとは

RDFはWeb上の情報(リソース)を記述するためのフレームワークとして、W3Cで標準化されたものです。RDFを使うことで、異なるデータベースやアプリケーションの間でデータを簡単に連携できるようになります。そのため、生命科学、薬学、図書館などの分野で広く使われています。1999年に初版がW3C勧告となり、その後2004年にRDF 1.1[*2]が勧告となりました。本稿執筆時(2024年5月)はRDF 1.2[*3]の標準化が行われている最中です。

RDFは、情報を「主語」「述語」「目的語」の3つの要素で表現します。この3つの組はRDFトリプルと呼ばれます。例として、日本の様々なコンテンツを検索できるジャパンサーチ[*4]から取得した、枕草子に関するRDFトリプルを以下に示します。

- 主語: <https://jpsearch.go.jp/data/bibnl-20853658>
- 述語:<http://www.w3.org/2000/01/rdf-schema#label>
- 目的語: "枕草子:対訳"

RDFトリプルは一般的な文と同様に、「主語の述語は目的語である」と読むことができます。このRDFトリプルが表しているのは「...bibnl-20853658の...labelは枕草子:対訳である」という文であるといえます。ここで、主語<https://jpsearch.go.jp/data/bibnl-20853658>は、ジャパンサーチがある図書に割り当てた識別子です。述語<http://www.w3.org/2000/01/rdf-schema#label>はW3CのRDFスキーマ[*5]で定められた用語で、この後に来る目的語"枕草子:対訳"が、主語のラベル(簡単な説明文)であることを意味しています。つまりこのRDFトリプルは、識別子<https://jpsearch.go.jp/data/bibnl-20853658>を持つ情報に"枕草子:対訳"というラベルが付くことを表しています。

このようにRDFトリプルでは<https://jpsearch.go.jp/data/bibnl-20853658>のようなURL[*6]で多くの情報を表現します[*7]。URLを使うのは、データの作成者が表現しようとしている情報を正確に伝えるためです。もし主語と述語がURLを使わず、単に20853658、labelと表されていた場合、20853658がどこで定められた識別子なのか、またlabelという述語の意味が何であるか、読み手が正しく理解するのは難しくなってしまうことでしょう。

RDFトリプルは、図-1のように2つのノード(丸や四角で囲まれた情報)を矢印で繋いだ図として描かれることもあります。図-1では読みやすさを考慮して、URLの一部https://jpsearch.go.jp/data/をdata:という省略形に書き換えています。同様にhttp://www.w3.org/2000/01/rdf-schema#はrdfs:で置き換えています。以降でもこのような短縮表記を使います。

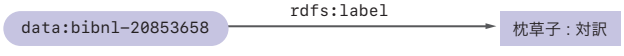



**図-1 枕草子に関するRDFトリプルの例**

＊1 Dave Longley, Gregg Kellogg, Dan Yamamoto: RDF Dataset Canonicalization. W3C Recommendation, 2024/05/21.(https://www.w3.org/TR/rdf-canon/)。
＊2 Richard Cyganiak, David Wood, Markus Lanthaler: RDF 1.1 Concepts and Abstract Syntax. W3C Recommendation, 2014/02/25.(https://www.w3.org/TR/rdf11- concepts/)。
＊3 Olaf Hartig, Pierre-Antoine Champin, Gregg Kellogg, Andy Seaborne: RDF 1.2 Concepts and Abstract Syntax. W3C Working Draft, 2024/05/02(https://www.w3.org/TR/2024/WD-rdf12-concepts-20240502/)。
＊4 ジャパンサーチ(https://jpsearch.go.jp/)。
＊5 Dan Brickley, R.V. Guha: RDF Schema 1.1. W3C Recommendation, 2014/02/25(https://www.w3.org/TR/rdf11-schema/)。
＊6 正確には、URLを一般化したInternationalized Resource Identifier(IRI)が使われます。
＊7 この例では目的語がURLではなくただの文字列で表現されていますが、目的語にURLを取るトリプルも一般的です。

RDFトリプルを集めたものはRDFグラフと呼ばれます。先程と同様、枕草子に関するRDFトリプルをジャパンサーチから追加で取得すると、図-2のようなRDFグラフを作ることができます。

このRDFグラフでは、data:bibnl-20853658のラベルが「枕草子:対訳」であることに加えて、その制作に関わった人が清少納言であることと、訳編に関わった人が守屋新助であることが表されています。

RDFトリプルを集めたものがRDFグラフでしたが、RDFグラフを集めたものはRDFデータセットと呼ばれます。本稿のテーマであるRDF Dataset Canonicalizationはその名のとおり、RDFデータセットを正規化する方法です。ただし本稿では説明を簡単にするため、RDFグラフとRDFデータセットを区別せずに扱います。

## 2.3 RDFの空白ノード

図-2の例にはおもむろに_:b152539105や_:b152573899という奇妙な名前のノードが登場しました。これらは空白ノード(blank node)と呼ばれる、識別子(URL)を持たない特殊なノードです。巨大なRDFグラフを作る際など、すべてのノードにURLを付けるのはときに煩雑な作業になります。そこで、他のグラフとは繋がることのない中間的なノードなどには、URLをもたない空白ノードが使われることがあります。

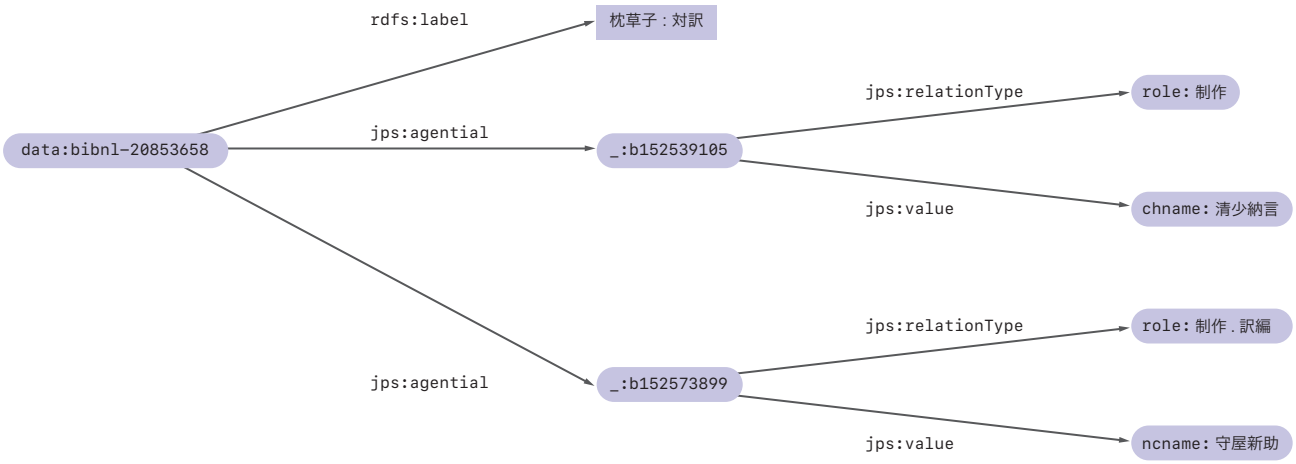



図-2 枕草子に関するRDFグラフ

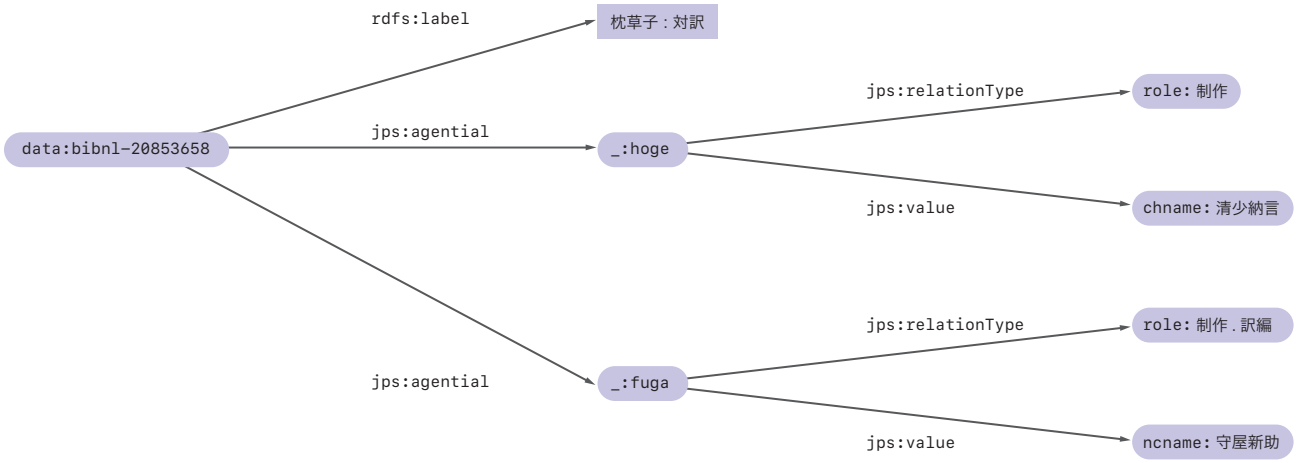



図-3 枕草子に関するRDFグラフのもう1つの例

空白ノードに付けられた名前は、あくまで一時的な名前です。同じRDFグラフでも、扱うシステムや環境によって、空白ノードの名前は変わってしまうことがあります。例えば上の_:b152539105と_:b152573899をそれぞれ_:hogeと_:fugaに置き換えた図-3のRDFグラフは、置き換え前のRDFグラフと同じものとして（正確には同型なグラフとして）扱われます。

そのおかげで、RDFグラフの作成者は空白ノードの名前付けに頭を悩ませる必要がなくなるという利点があります。また、RDFグラフをデータ化するときにそれらを省略できるというメリットもあります。例えば、図-3のRDFグラフはJSON-LD[*8]という仕様を使うと次のように表せます。ここでは空白ノードの名前を意識する必要がなくなっています。

```
{
  "@context": { ... },
  "@id": "data:bibnl-20853658",
  "rdfs:label": "枕草子 : 対訳",
  "jps:agential": [
    {
      "jps:relationType": "role:制作",
      "jps:value": "chname:清少納言"
    },
    {
      "jps:relationType": "role:制作.訳編",
      "jps:value": "ncname:守屋新助"
    }
  ]
}
```

## 2.4 Canonicalization（正規化）

便利な空白ノードですが、決まった名前を持たないという特徴が問題を生むこともあります。例えば、2つのRDFグラフが同じグラフであるか確認したい場合や、グラフとグラフの差分を知りたい場合、またあるRDFグラフに更新があったか知りたい場合などに、この空白ノードの扱いが問題となってきます。また、RDFグラフに作成者のデジタル署名を付ける場合、署名を付けたときの空白ノードの名前と、それを後に検証するときの空白ノードの名前が一致していなければ検証に失敗してしまいますが、それは空白ノードの性質上保証されません。

そこで、システムや環境によらず、名前を持たない空白ノードに決まった名前を付ける方法が必要となりました。それが今回のテーマであるRDF Dataset Canonicalizationです。例えば図-2と図-3に登場した2つのRDFグラフは、Canonicalizationを行うことでどちらも同じ図-4のグラフに変換されます。

Canonicalization後の空白ノードには_:c14n0、_:c14n1という新しい名前が与えられます[*9]。これらは、元々空白ノードに付けられていた_:b152539105や_:hogeといった値には左右されず、グラフに現れるURLや文字列と、グラフの構造に基づいて、決まった方法で計算されます。従って元の空白ノー

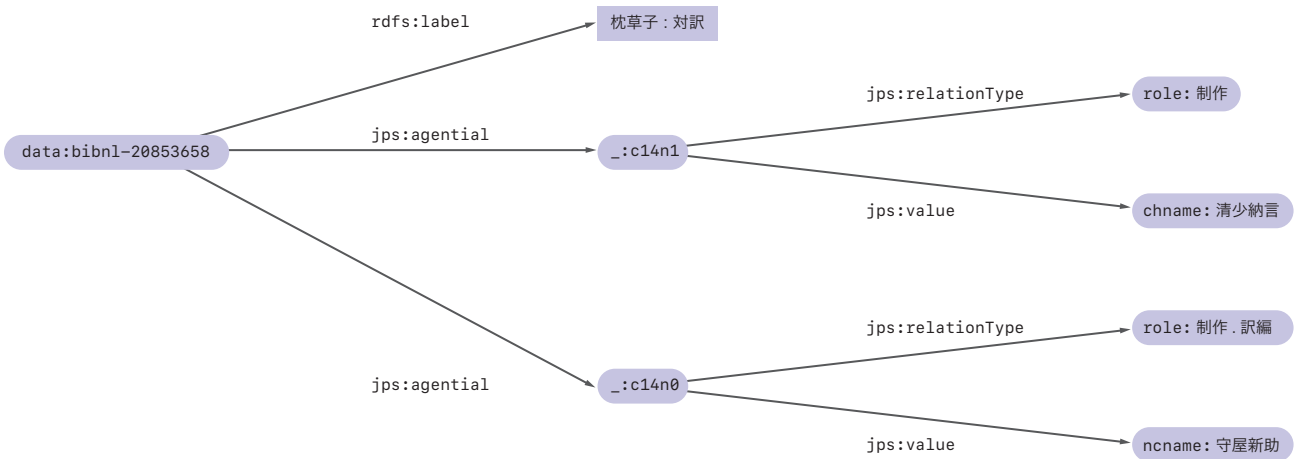



図-4 Canonicalizationを行ったRDFグラフの例

*8 Gregg Kellogg, Pierre-Antoine Champin, Dave Longley: JSON-LD 1.1. W3C Recommendation, 2020/07/16.（https://www.w3.org/TR/json-ld11/）。
*9 "c14n"は"canonicalization"の省略形です。

ドにどのような名前が付けられていたとしても同じグラフを得ることができるわけです。

空白ノードの名前が確定したところで、後はこれをCanonical N-Quadsと呼ばれる形式[*10]で出力すれば、図-5のような正規化されたデータを得ることができます。正規化後のデータを使うことで、RDFグラフの差分計算や更新の確認、デジタル署名やハッシュ値の計算が簡単に実現できます。

本レポートのVol.52(https://www.iij.ad.jp/dev/report/iir/052.html)[*11]で取り上げたW3C Verifiable Credentialというデジタル証明書は、デジタル署名のついたRDFデータセットです。デジタル署名をつける前に、このRDF Dataset Canonicalizationを使って空白ノードの名前を正規化することで、署名するデータと検証するデータが同じになることを保証しています。

## 2.5 標準化活動

標準化には得てして長い時間がかかります。RDF Dataset Canonicalizationの標準化にも10年以上もの長い歳月が費やされました。議論自体は早い段階から始まっていたものの、標準化の必要性や最適な方法についてのコンセンサスが長らく得られなかったことが一因です[*12]。

まず2009年から2010年にかけてW3CでCanonicalizationの仕様化に関する議論が始まりました。2012年には、Digital Bazaar社のDave LongleyとManu SpornyがUniversal RDF Graph Normalization Algorithm(URGNA2012)を提案しました。更にその3年後には、改訂版であるUniversal RDF Dataset Normalization Algorithm(URDNA2015)が提案され、今回標準化された仕様のベースとなりました。

その後、W3C内でVerifiable Credentialの議論が活発になり、2021年にはRDFデータにデジタル署名を付けるためのLinked Data Signatures Working Groupの立ち上げが提案されました。しかし、署名プロセス全体の標準化に関しては合意形成に至らず、中止となりました。その代替案として、RDFの正規化に焦点を当てたRDF Dataset Canonicalization and Hash Working Group(RCH WG)[*13]が提案され、2022年7月に承認されました。

こうしてようやく、RDF CanonicalizationをW3C勧告とするための作業が始まりました。そして2024年5月21日、標準化活動のゴールであるW3C勧告へと至ることができました[*14]。

筆者はWG Chairからの招待を受けて、2022年8月に招聘専門家(Invited Expert)としてRCH WGに参加し、同11月からはEditorとして協力しました。きっかけは、筆者らのVerifiable Credentialsに関する国際会議での発表内容[*15]がWG co-chairの目に止まったことでした。

RCH WGの活動は、GitHub上での議論と隔週の電話会議が中心です。GitHub上では問題提起や修正案の投稿が行われ、電話

```
<https://jpsearch.go.jp/data/bibnl-20853658> <https://jpsearch.go.jp/term/property#agential> _:c14n0 .
<https://jpsearch.go.jp/data/bibnl-20853658> <https://jpsearch.go.jp/term/property#agential> _:c14n1 .
<https://jpsearch.go.jp/data/bibnl-20853658> <rdfs:label>"枕草子 : 対訳".
_:c14n0 <https://jpsearch.go.jp/term/property#relationType> <https://jpsearch.go.jp/term/role/制作 . 訳編>.
_:c14n0 <https://jpsearch.go.jp/term/property#value> <https://jpsearch.go.jp/entity/ncname/守屋新助>.
_:c14n1 <https://jpsearch.go.jp/term/property#relationType> <https://jpsearch.go.jp/term/role/制作>.
_:c14n1 <https://jpsearch.go.jp/term/property#value> <https://jpsearch.go.jp/entity/chname/清少納言>.
```

**図-5 正規化された結果**

*10 一般にN-Quads形式のデータはソートされている必要はなく、また区切りとしての空白文字や改行文字の個数に制限はないですが、ここでは辞書順にソートした上で、区切り文字は1つになるよう制限を加えたものを用います。これはCanonical N-Quadsと呼ばれます。
*11 Internet Infrastructure Review vol.52「2. フォーカス・リサーチ(1) Verifiable CredentialとBBS+署名」(https://www.iij.ad.jp/dev/report/iir/052/02.html)。Vol.52で言及していた「LD Canonicalization」が、本稿で紹介するRDF Dataset Canonicalizationの旧称です。
*12 Phil ArcherによるメールHttps://lists.w3.org/Archives/Public/semantic-web/2024May/0030.html)。
*13 W3C RDF Dataset Canonicalization and Hash Working Group(https://www.w3.org/groups/wg/rch/)。
*14 RDF Dataset Canonicalization and Hash Working Group Charter(https://w3c.github.io/rch-wg-charter/)。
*15 Dan Yamamoto, Yuji Suga, Kazue Sako: Formalising Linked-Data based Verifiable Credentials for Selective Disclosure. 2022 IEEE European Symposium on Security andPrivacy Workshops (EuroS&PW)(https://doi.org/10.1109/EuroSPW55150.2022.00013)。

会議では課題の解決に向けた議論やメンバーの合意形成が行われます。その結果がGitHub上での編集活動を通じて仕様文書に反映されていきます。筆者にとっては初めての標準化活動であり、有識者の議論に付いていくのもやっとでしたが、文面の提案やPull Requestのレビュー、リファレンス実装の提供など、できることから貢献に努めました。

W3Cの仕様は、読者がその内容を正しく実装できることが重要な要件です。本稿執筆時(2024年5月)、RDF Dataset Canonicalizationには9つのオープンソース実装が寄せられており、開発言語もC++、Elixir、Java、JavaScript、Ruby、Rust、TypeScriptと多岐にわたっています[*16]。筆者もRustによるオープンソース実装を提供しました[*17]。

## 2.6 Canonicalizationの手順

RDF Dataset Canonicalization仕様が定めるアルゴリズムはRDF Canonicalization algorithm version 1.0、通称RDFC-1.0と命名されています。本稿ではRDFC-1.0の概要を説明します。

RDFC-1.0は、入力されたRDFグラフ内の空白ノードにラベル付けをする正規化(canonicalize)と、正規化されたRDFグラフをCanonical N-Quads形式の正規化されたデータとして出力する直列化(seriarize)の2つのステップからなります。

第1の正規化のステップでは、まずグラフ内の空白ノード1つ1つについて1次ハッシュ(first degree hash)と呼ばれる値を計算していきます。これは空白ノードの周辺の情報をハッシュ関数と呼ばれる特殊な関数に入れて、長さの決まったハッシュ値と呼ばれるデータを得るものです。直観的には、空白ノードの周辺の情報を使って、その空白ノードに名前を付ける操作に相当します。

空白ノードに付けられた1次ハッシュの値がすべて異なっていれば、後はそれを辞書順[*18]に並べ直すことで、空白ノードに順番を付けることができます。この順番に従って_:c14n0, _:c14n1, _:c14n2, ...という要領でラベル付けをすれば、めでたく正規化の処理は完了です。

図-6の例を使って、具体的にこの流れを説明します。このRDFグラフは4つのノードを含み、そのうち2つ(:pと:u)がURLを持つ通常のノードで、残りの2つ(_:e0と_:e1)が空白ノードです。

ここで空白ノード_:e0を含むRDFトリプルだけを抜き出し、Canonical N-Quads形式で表すと以下のようになります。

```
:p :q _:e0 .
_:e0 :s :u .
```

これが_:e0の周辺情報に対応します。ここで空白ノードに付けられている「仮の」名前e0をaで置き換え、次の文字列を得ます。

```
:p :q _:a .
_:a :s :u .
```

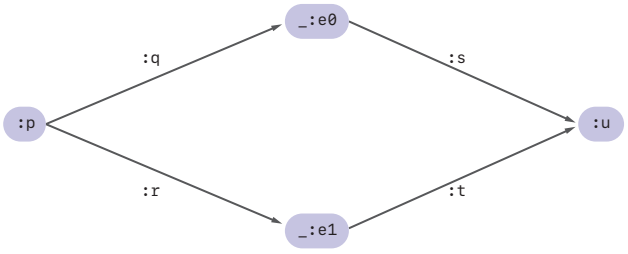



**図-6 2つの空白ノードを含むRDFグラフの例**

*16 Gregg Kellogg: RDF Dataset Canonicalization and Hash 1.0 Processor Conformance.(https://w3c.github.io/rdf-canon/reports/)。
*17 zkp-ld/rdf-canon(https://github.com/zkp-ld/rdf-canon)。
*18 正確にはUnicode Codepointの順序に従って整列させます。

これをハッシュ関数に入力し、得られたビット列を16進数で表現した21d1dd5ba21f3dee9d76c0c00c260fa6f5d5d65315099e553026f4828d0dc77aが、空白ノード_:e0の1次ハッシュ値になります。この1次ハッシュ値には_:e0の周辺情報が埋め込まれており、_:e0と他の空白ノードを区別するために使うことができます。

同様に_:e1を含むRDFトリプルを抜き出すと

```
:p :r _:e1 .
_:e1 :t :u .
```

のようになり、先程と同様にe1をaで置き換えて得られた

```
:p :r _:a .
_:a :t :u .
```

のハッシュ値6fa0b9bdb376852b5743ff39ca4cbf7ea14d34966b2828478fbf222e7c764473が_:e1の1次ハッシュ値になります。

これらを辞書順で並べると、先頭が2で始まる_:e0の1次ハッシュ値の方が、先頭が6で始まる_:e1の1次ハッシュ値よりも辞書順で先に来ることが分かります。この結果、e0とe1の間に順序を定めることができました。あとはこの順序にしたがって、_:e0に_:c14n0、_:e1に_:c14n1という正規化識別子を与えれば正規化は完了です。

正規化前の空白ノードに付けられていた名前が何であっても、正規化後の結果が変わらないことが重要です。実際、図-6の例で、_:e0を_:hogeで、_:e1を_:fugaでそれぞれ置き換えても、それらの1次ハッシュ値が変わらないことが確認できます。1次ハッシュ値の計算の途中、空白ノードの名前をみなaに置き換えましたが、そのおかげで計算結果は元々付けられていた空白ノードの名前に依存しなくなるわけです。

図-6の例のようにそこまで複雑ではないRDFグラフは、1次ハッシュの計算のみで正規化ができ、話は比較的簡単です。しかしRDFグラフによっては、異なる空白ノードに同じ1次ハッシュが割り当たってしまうことがあります。例えば図-7のようなグラフでは、空白ノードの周辺の状況がまったく同じような空白ノードが存在し、これらには同じ1次ハッシュが割り当たってしまいます。

実際、_:e0と_:e1を見てみると、これらはどちらも主語:pから述語:qを経て到達する目的語になっており、更にどちらも述語:pを介して空白ノードを目的語とする主語になっていることが分かります。この結果、これらの1次ハッシュ値はまったく同じ値になってしまいます。

そこでRDFC-1.0では、同じ1次ハッシュが割り当てられた空白ノードが存在する場合に限り、次なる識別手段としてn次ハッシュ(n-degree hash)を計算することになっています。n次ハッシュの計算方法はRDFC-1.0の中でも複雑な処理になっているため、本稿での説明は割愛します。興味のある方は仕様をご覧ください。

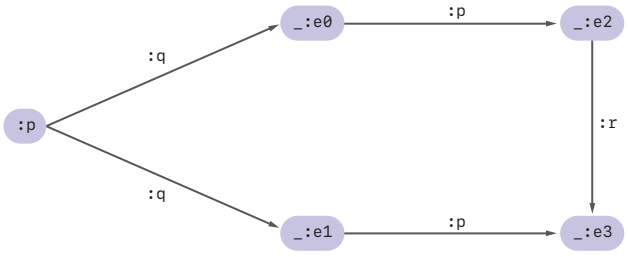



図-7 より複雑なRDFグラフの例

## 2.7 Canonicalizationの課題と対策

これまで見てきたように、RDF Dataset Canonicalizationの本質は空白ノードに順序付けをして正規化された名前を得ることにあります。従って、空白ノードを一切含まないRDFグラフのCanonicalizationでは、1次ハッシュやn次ハッシュの計算は必要なく、RDFグラフをN-Quadsとして表現した上でソートするだけの単純な処理(直列化)だけで済みます。

RDF Dataset Canonicalizationが必要以上に複雑な処理をしているとの誤解もありますが、Canonicaliationの複雑さは、入力されるRDFグラフに含まれる空白ノードの数と、空白ノードを含むグラフがどのような構造を持つかに依存して決まるものであって、実用上多くの場合、それらは単純な1次ハッシュ値計算のみで高速に終わることがほとんどです。

とはいえ、多くの空白ノードを含む特殊な構造を持つようなRDFグラフではn次ハッシュ値の計算に非常に長い時間を要するものも存在します[*19]。そこでRDFC-1.0の実装ではn次ハッシュの計算回数に上限を設け、上限を超えた場合にはエラーとして途中で終了させることが必須とされています。

また、RDFグラフがパーソナルデータや機密事項を含むような場合、正規化の結果からそれらが部分的に推測される可能性に注意が必要です。正規化の計算はグラフ内のデータに基づいて行われるため、正規化の結果作られた_:c14n0などの名前には、グラフ内の情報が部分的に含まれてしまいます。

例えばAliceの配偶者がBobであることを意味する、図-8のようなRDFグラフがあったとします。そしてこのグラフの正規化を行い、デジタル署名をつけて図-9のように保存したとします。

あるとき、何らかの理由でAliceは配偶者の名前を隠したまま、自身が結婚をしているという事実のみを表現したいと思ったとします。Verifiable Credentialの選択的開示という手法を使うと、デジタル署名の有効性を保ったまま配偶者の名前を隠し、図-10のような検証可能なRDFグラフを作ることができます。

しかしこのグラフを見た人が、Aliceの配偶者はBobかCharlieのどちらかであるということまでは知っていたとします。そして***として隠されている部分にBobとCharlieの名前を

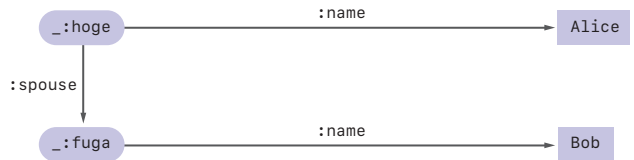



図-8 AliceとBobに関するRDFグラフ

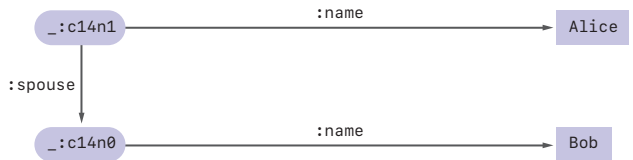



図-9 図-8の正規化後

*19 仕様ではこれらをPoison Datasetと呼んでいます。RDFグラフのCanonicalizationはグラフ同型問題という難しい問題と同じ難しさを持つことが分かっているため、入力によって非常に長い計算時間を要するのは避けがたい課題です。

入れて再び正規化をしたところ、図-11のように、正規化ラベルが異なる2つの結果が得られたとします。これらをAliceの公開したグラフと比較すれば、Aliceの配偶者はBobであったと特定することが可能になります。

これは特殊な状況を想定したものですが、Verifiable CredentialsでRDF Dataset Canonicalizationを使う場合には注意すべき性質になります。そこで、W3C Verifiable Credentials Data Integrity[*20]という、Verifiable Credentialsのセキュリティやプライバシーを保護するための仕様の中で、こうした問題の回避について議論がなされています。

## 2.8 おわりに

以上、筆者がW3Cで標準化に協力し、W3C勧告となったRDF Dataset Canonicalizationについて、仕様と標準化活動の概要を紹介しました。RDF Dataset Canonicalizationを使うことで、RDFグラフの差分計算や更新の確認、更にはハッシュ値の計算やデジタル署名の生成も容易になります。これによりデータ管理が効率化され、RDFグラフに偽造耐性や真正性を持たせることも可能となります。私自身もこの仕様の一利用者として、Verifiable Credentialsやそのアプリケーションの研究開発に活用しています。本稿が皆様にも興味や関心をもっていただくきっかけとなれば幸いです。

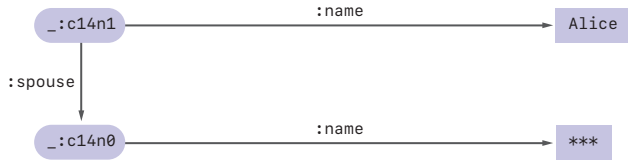



図-10 Bobの名前が隠されたRDFグラフ


図-11 2つの異なる結果

執筆者:
**山本 暖**（やまもと だん）
IIJ セキュリティ本部 セキュリティ情報統括室 シニアエンジニア。
2021年より現職。デジタルアイデンティティと情報セキュリティに関わる調査・研究活動に従事。

*20 Manu Sporny, Dave Longley, Greg Bernstein, Dmitri Zagidulin, Sebastian Crane: Verifiable Credential Data Integrity 1.0. W3C Candidate Recommendation Draft, 2024/04/28.(https://www.w3.org/TR/2024/CRD-vc-data-integrity-20240428/)。

# IIJにおけるDRMの取り組み

## 3.1 はじめに

DRMとは、Digital Rights Managementのことで、デジタルコンテンツの利用制限、複製などの著作権を制御する技術です。日本語ではデジタル著作権管理と言います。本来は、文章や画像といった静的コンテンツの他、音楽、ゲームなど、デジタルコンテンツ全般に応用できる技術ですが、本稿では筆者が携わっている動画配信プラットフォームである「IIJ Media Sphereサービス」の観点から、動画配信におけるDRMについて解説します。

## 3.2 DRM概要

音楽や映像を含めコンテンツのデジタル化が90年代に急速に進み、90年代後半にはインターネットの普及も伴って、その流通が一変しました。街のレンタルビデオショップがレンタルDVDショップになり、今ではオンラインの配信サービスで、好きな場所で好きなときに、好きな作品を手軽に視聴できます。

このように、コンテンツのデジタル化は、私たちの生活に多大な恩恵をもたらしましたが、一方で、劣化なく簡単にコピーできるため、海賊版の流通が容易に想像されました。違法コンテンツの流通は、コンテンツの権利者だけでなく、制作現場や販売、配信など様々な立場の企業の利益を損ないます。こうした状況が蔓延した場合、業界自体が衰退し、最終的には魅力あるコンテンツも生まれなくなるでしょう。結果的に、消費者である私たちが楽しめなくなってしまいます。

このような背景で、コンテンツ自体だけでなく業界全体の発展を守るためにも考えられた技術がDRMです。DRMを使うことで、コンテンツの適切な利用や、複製などの制御が期待されます。

1つ注意しなければならないのは、DRMは完璧な技術ではなく、本質的に視聴環境を制限するという点です。悪意がないエンドユーザでも、環境によって視聴できないケースがどうしても出てきます。また、不正利用に関しても過去にいくつかの事例が報告されています。

しかしながら、現在のインターネットを利用したコンテンツ配信では、広く普及している安定した技術でもあり、今後もその必要性は高まると考えられています。

## 3.3 IIJでのDRMサービスの変遷

読者の方々が意識しているかどうかはともかく、コンテンツ保護を目的として利用されているDRMは、そこまで新しい技術ではありません。この記事を執筆するにあたって、IIJの古参の社員にも尋ねたところ、既に90年代後半には、DRMの必要性を感じ、オンライン動画配信に関するニュースメディア「Streaming Media[*1]」が主催する展示会などで情報収集を行っていたとのことでした。

IIJが具体的なサービスとしてリリースしたのは、2008年のFlash VideoにおけるDRM[*2]です。その後、PlayReadyに対応し、2015年にはオープンスタンダードであるMarlin DRMを採用したサービスもリリース[*3]しています。

現在のIIJ Media Sphereサービスでは、DRMシステムとして、AppleのFairPlay Streaming、GoogleのWidevine、MicrosoftのPlayReadyに対応しています。現在、この3種類のDRMに対応することで、かなり網羅的に動画の再生環境をカバーできると考えています。

今回IIJでDRM機能を開発し、IIJ Media Sphereの機能としてサービス提供することで、動画プレイヤー環境を含め、コンテンツのパッケージングも意識することなくDRMによるコンテンツ保護機能を容易に利用できるようになってい

＊1　Streaming Media(https://www.streamingmedia.com/)。
＊2　IIJ、Flash Video配信ソリューションにDRM機能を追加(https://www.iij.ad.jp/news/pressrelease/2008/pdf/FlashDRM.pdf)。
＊3　IIJ、「IIJ DRMサービス/ExpressPlay®」を提供開始(https://www.iij.ad.jp/news/pressrelease/2015/0126.html)。

ます。またIIJとしても外部のDRMプロバイダを利用しないことでコストが見積もりやすくなりました。加えて、自社開発及び自社設備での運用は、サービスの持続可能性においても一層の信頼を提供できるとも考えています。動作環境の長期間の保証は、ソースコードも含めたプログラムもそうですが、基盤となるプラットフォームの保守も大事だからです。そして、今後もサポートを含め、デバイスについての情報提供、レポーティング、分析などの機能を実装し、様々な付加価値を創造していければとも思います。

なお、IIJでは、Widevineに関して、私を含め複数人の社員が認定プログラムに合格しており、Certified Widevine Implementation Partnerとしての資格を取得しています。

## 3.4 DRMの機能

現在のDRMは、コンテンツを暗号化することによってコンテンツを保護し、適切な環境でのみ利用させるというのが基本的なコンセプトです。前述のとおり、DRMは日本語ではデジタル著作権管理と訳されますが、利用させる際において、具体的に様々な管理をすることができます。これらの機能は、コンテンツ事業者や、配信事業者が自身のビジネスを考慮して、ポリシーとして利用する機能です。

最も想像しやすい機能は、動画の再生許可です。期待する条件でのみコンテンツの再生を管理することができます。また、同時再生デバイスの制限といった制御も馴染みがあるかもしれません。

加えて、コンテンツのクオリティに関しても制御できます。例えば、オーディオのみの再生だけ許可することや、SD画質(480p)、HD画質(1080p)、UHD画質(4k〜)、それぞれの再生環境を管理することもできます。

HDCPについても同様です。近年、テレビやスマートフォンをはじめとする各種デバイスを購入する際、または様々な配信サービスの案内などで、HDCPという単語を見かけたことがあるかもしれません。HDCPも、DRMを構成する技術の1つです。HDCP は、High-Bandwidth Digital Content Protection を表し、2000年にインテルによって開発された暗号化技術で、不正コピーを防ぐことを目的とした著作権保護技術です。HDCPはHDMIなどのデジタルインタフェースの暗号化に用いられており、映像伝送するときに用いられます。映像の出力側と共に、ディスプレイなどの入力側もHDCPに対応していないと、コンテンツの再生ができなかったり、画質が制限されたりする可能性があります。分かりやすい例だと、アナログデバイスへの出力を制限したいケースなどがあります。現在普及しているHDCP 2.2は、そのリリースから10年以上経過しているため、あまり心配する必要はないと思いますが、HDMIなどを使って映像伝送したいケースで、コンテンツの再生に問題がある場合、各種デバイスがHDCPに対応しているか確認することをお勧めします。

最後にデバイスのセキュリティレベルについて紹介します。例えば、Widevineでは、デバイスごとにL1、L2、L3といった3つのセキュリティレベルを規定しています。Widevineが最もセキュアなデバイスと認定しているレベルがL1で、TEE(Trusted Execution Environment)と呼ばれるハードウェアでの復号、動画再生ができるデバイスを対象としています。L3はTEEの搭載がなく、ソフトウェアでの復号、動画再生を行うデバイスです。これらを元に、高画質コンテンツは、L1デバイスのみで再生を許可し、L3デバイスでは、低画質のみ再生できるといった制御ができます。

Androidスマートフォンであれば、「DRM Info」といったアプリで、自身のスマートフォンのセキュリティレベルなどを確認できますので、興味がある方は試してみてください。

## 3.5 DRMの仕組み

以下では、DRMの仕組みについて、暗号化の過程と、復号の過程に分けて説明します。

### 3.5.1 コンテンツの暗号化

前述のとおり、DRMは、コンテンツを暗号化することが基本的なコンセプトとなります。そのため、一般的に動画のパッケージングの過程で、DRMプロバイダの提供するKeyサーバとやり取りを行い、暗号化が行われます。しかし、世の中には様々なDRMシステムが存在し、またその方式に関しても様々なやり方があります。DRMシステムごとにパッケージャーが個別の対応をしなくても良いように、業界ではDASH Industry Forum（DASH-IF）により開発されたCPIX（Content Protection Information Exchange）という規格が普及しています。もともとはMPEG-DASH用に開発されたものですが、現在ではHLSにも対応しています。CPIXを利用する利点は以下のとおりです。

1. インターオペラビリティ
   CPIXは、異なるDRMシステム間での相互運用性を可能にします。これにより、コンテンツプロバイダや配信プラットフォームは、複数のDRMプロバイダを利用することができ、様々なデバイスやプラットフォームでのコンテンツの配信が容易になります。
2. ワークフローの簡素化
   CPIXに準拠することで、コンテンツプロバイダや配信プラットフォームは、複数のDRMシステムに対して統一された鍵及び暗号化フォーマットを使用できます。これにより、複雑な鍵管理や暗号化手順が簡素化され、効率的なワークフローが実現できます。
3. セキュリティの向上
   CPIXは、セキュリティの観点からも利点があります。共通の暗号化フォーマットと鍵マッピングを使用することで、コンテンツの漏えいや不正コピーを防ぐための一貫したセキュリティポリシーを適用しやすくなります。
4. オープンな標準規格
   CPIXはオープンな標準規格であり、業界全体で採用されています。これにより、ベンダーロックインを回避できます[*4]。

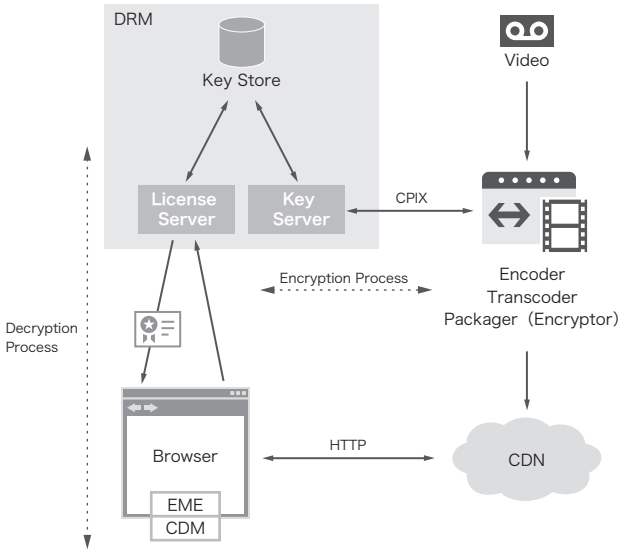



図-1 暗号化と復号の過程

*4 DASH Industry Forum、DASH-IF Implementation Guidelines: Content Protection Information Exchange Format（https://dashif.org/docs/CPIX2.2/Cpix.pdf）。

一例を挙げます。

```
  ’
    <cpix:CPIX id="cpixsample" xmlns:cpix="urn:dashif:org:cpix"
xmlns:pskc="urn:ietf:params:xml:ns:keyprov:pskc">
      <cpix:ContentKeyList>
        <cpix:ContentKey kid="f269e534-c4f1-4721-9d62-26dc7ed241bd"
explicitIV="8mnlNMTx...">
          <cpix:Data>
            <pskc:Secret>
             <pskc:PlainValue>vfMB2...</pskc:PlainValue>
            </pskc:Secret>
          </cpix:Data>
      </cpix:ContentKey>
      </cpix:ContentKeyList>
      <cpix:DRMSystemList>
        <cpix:DRMSystem kid="f269e534-c4f1-4721-9d62-26dc7ed241bd"
systemId="edef8ba9-79d6-4ace-a3c8-27dcd51d21ed">
          <cpix:PSSH>AAAAP3Bzc...</cpix:PSSH>
        </cpix:DRMSystem>
        <cpix:DRMSystem kid="f269e534-c4f1-4721-9d62-26dc7ed241bd"
systemId="9a04f079-9840-4286-ab92-e65be0885f95">
          <cpix:PSSH>AAADBnBzc...</cpix:PSSH>

<cpix:ContentProtectionData>PG1zcHI6cHJvIHhtbG...</cpix:ContentProt
ectionData>
        </cpix:DRMSystem>
        <cpix:DRMSystem kid="f269e534-c4f1-4721-9d62-26dc7ed241bd"
systemId="94ce86fb-07ff-4f43-adb8-93d2fa968ca2">
          <cpix:URIExtXKey>c2tkOi...</cpix:URIExtXKey>
        </cpix:DRMSystem>
      </cpix:DRMSystemList>
    </cpix:CPIX>
  `
```

CPIXはXMLで表現されます。上記の例で、いくつかのポイントを解説します。

ContentKeyエレメントは、コンテンツの暗号化に必要な情報を表現しています。これはRFC 6030 Portable Symmetric Key Container(PSKC)を元に拡張されている項目です[*5]。

DRMSystemエレメントは、各DRMシステムに特有の情報を表現しています。systemIdがそれぞれのDRMシステムを識別しているidで、DASH-IFにより決められています[*6]。

上記の例では、edef8ba9-79d6-4ace-a3c8-27dcd51d21edがWidevineを表し、9a04f079-9840-4286-ab92-e65be0885f95がMicrosoft PlayReadyを表し、94ce86fb-07ff-4f43-adb8-93d2fa968ca2がApple FairPlayを表しています。

PSSHエレメントは、動画コンテナの1つである、mp4のPSSH(Protection System Specific Header)Boxに利用されるデータを表現しています。このBoxは、デジタルコンテンツの暗号化とデジタル著作権管理(DRM)システムに関連する情報を格納するために使用されるMP4ボックスの一種です。PSSHボックスには、暗号化キー、使用されている暗号化方式、及びその他のDRMシステムに関する情報が含まれます。

*5 RFC Editor、RFC 6030 Portable Symmetric Key Container(PSKC), OCTOBER 2010(https://www.rfc-editor.org/info/rfc6030)。
*6 DASH Industry Forum、Content Protection(https://dashif.org/identifiers/content_protection/)。

URIExtXKeyエレメントは、DRMシステムがApple FairPlayで利用されていることから想像できるように、HLS playlistの中で利用される、EXT-X-KEYに影響する項目になっています。

コンテンツのパッケージングの過程で、これらの情報を使い、晴れてコンテンツが暗号化されます。

なお、AWSで利用できるCPIXを元に拡張された、SPEKE(Secure Packager and Encoder Key Exchange)といった規格も非常に普及しています。こちらもオープンな仕様ですので、誰でも仕組みを理解できます[*7]。

### 3.5.2 コンテンツの復号

さて、ここまででコンテンツの暗号化がなされました。私たちが普段ブラウザなどで何気なく動画を視聴する場合、DRMの存在に気付かないケースがほとんどだと思います。しかし実際には、暗号化されたコンテンツを再生するために複雑な処理が発生しています。それらについて簡単に説明します。

コンテンツの復号において重要なのは、EME(Encrypted Media Extensions)と呼ばれる、Webブラウザとデジタル著作権管理(DRM)ソフトウェアの間のコミュニケーションチャンネルを提供するW3C仕様と、その中のCDM(Content

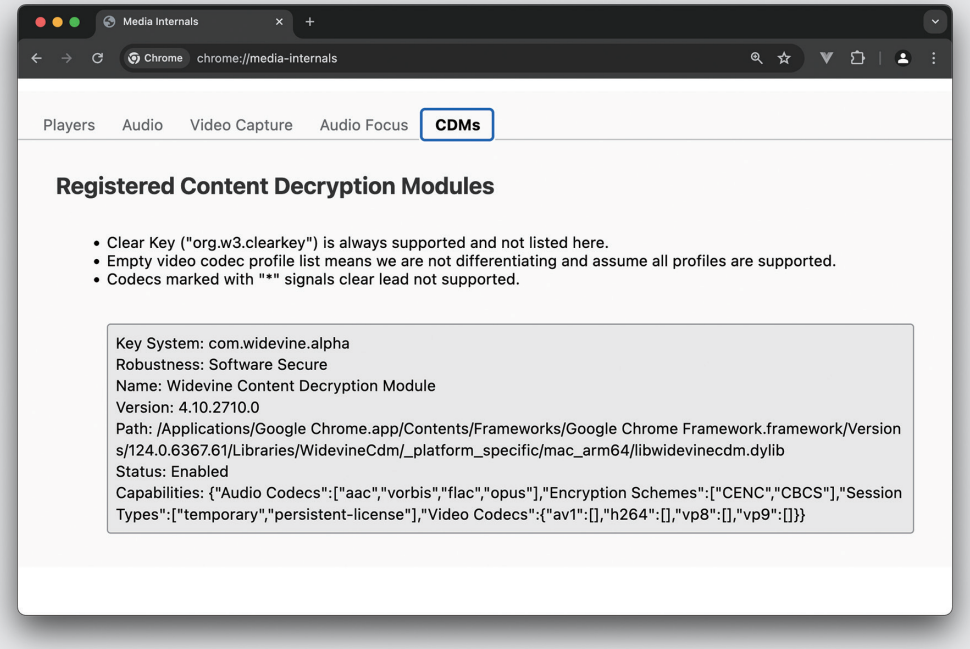



図-2 Chromeのmedia-internals

*7　AWS、SPEKE API 仕様(https://docs.aws.amazon.com/ja_jp/speke/latest/documentation/speke-api-specification.html)。

Decryption Module)と呼ばれるDRMベンダーが提供している復号モジュールです。

CDMはクローズドソースで、例えばWidevineは、Google Chromeの他、Firefoxにも、プラグインという形でCDMを提供しています(図-2、図-3)。しかし、Apple FairPlayのCDMはSafariをはじめとしたAppleプロダクトにのみ提供されているため、Google Chromeなどでは利用できません。

CDMは、EMEの仕様策定の過程でオープンなWebを支持する方面から強い反対がありましたが、結果として、CDM自体はあくまで復号のためのConfidentialityとIntegrityに責任を持ち、動画プレイヤーをはじめとするアプリケーションが通信を司るという形で、仕様が策定されました。EMEとCDMがこのような形で策定されなければ、DRMプロバイダも統一的な対応が取れなかったと思いますし、配信サービスごとに独自のアプリケーションを使わなければいけなかったかもしれません。

Google Chromeではアドレスバーに、「chrome://media-internals/」と入力することで、CDMの詳細情報を見ることができます。興味がある方は覗いてみてください。

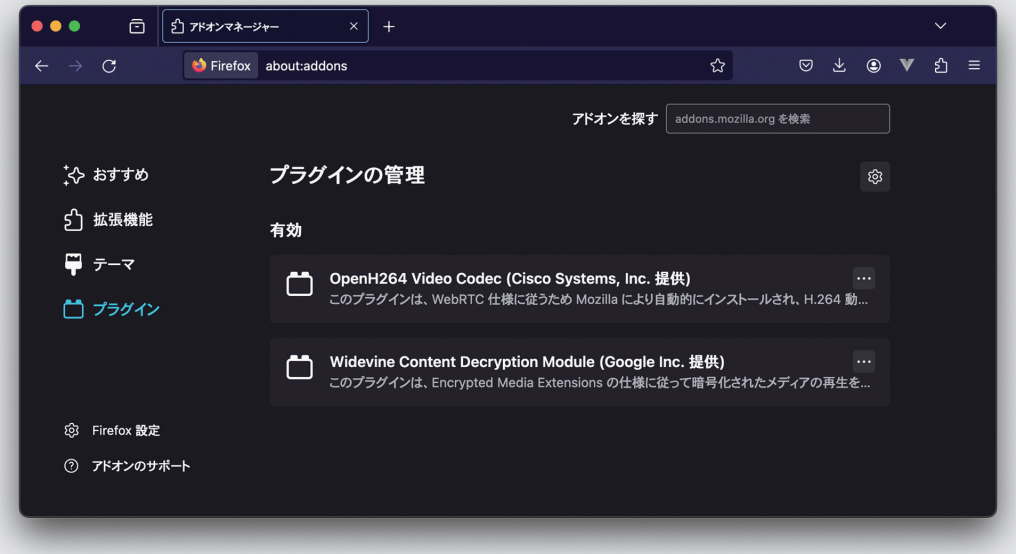
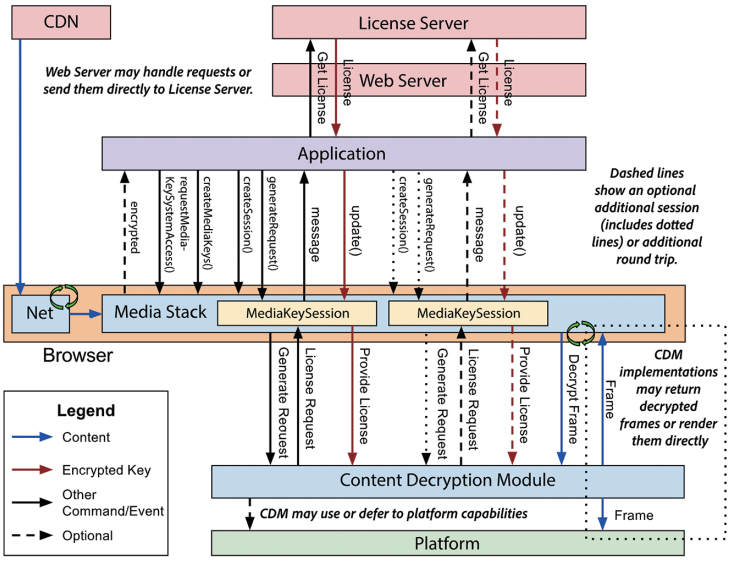



図-3 Firefoxのプラグイン

図-4は、EME及びCDMによる復号までの手順を表しています。動画プレイヤーはEMEを使ってCDMとやりとりを行って初めてコンテンツの再生ができます。手順は以下のとおりです。

1. 動画プレイヤーは、ブラウザからのencryptedイベントにより、コンテンツが暗号化されていることを知り、特定のCDMを利用するために、requestMediaKeySystemAccess()メソッドを呼びます。
2. createMediaKeys()メソッドにより、キーの初期化を行います。
3. キーの有効な期間を制御するためのセッションをcreateSession()メソッドで作ります。
4. Licenseを要求するための指示を、generateRequest()メソッドでCDMに行います。Licenseは、復号キーや、DRMの機能を使うための情報を含んでいる最も重要なデータです。
5. CDMはLicenseリクエストを生成し返します。動画プレイヤーはmessageイベントでLicenseリクエストを知ります。
6. 動画プレイヤーは、LicenseリクエストをLicenseサーバへ送信します。
7. Licenseサーバから返されるLienseを受け取ります。
8. 返ってきたLicenseを、update()メソッドを使ってCDMへ転送します。

DRMのポリシーによって、返されたLicenseを判断してCDMが復号を行い、動画の再生が行われます。




出典：W3C、Encrypted Media Extensions（https://www.w3.org/TR/encrypted-media/）。

**図-4 EME及びCDMによる復号までの手順**

## 3.6 おわりに

今回はIIJ Media Sphereサービスでの開発を通して、動画配信におけるDRMについて解説しました。近年では配信サービスを介して動画コンテンツを楽しんでいる方も少なくないと思います。その裏で実はDRMという技術が使われていること、またその仕組みがおぼろげながらイメージできるように説明してみました。限られた誌面ではありますが、お読みいただきありがとうございました。

執筆者:
**黒石 光雄**（くろいし みつお）
IIJ ネットワーク本部コンテンツ配信サービス部配信開発課 課長代理。
2002年IIJ入社以来、様々なサービス開発に従事。好きな言葉は「論よりコード」でEmacsのアップデートが楽しみ。

June 2024 Vol.63

IIJ
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。



本冊子の情報は2024年6月時点のものです。


IIJ-MKTG019-0063

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@iij.ad.jp　URL：https://www.iij.ad.jp