SNS における悪性 URL の分析と防御

Analysis of Malicious URLs on Social Networking Services and Protection

○國分 佑太朗[1]，中村 章人[1]

Yutaro KOKUBUN, Akihito NAKAMURA

1 会津大学 コンピュータ理工学部　School of Computer Science and Engineering, University of Aizu

Abstract　Cyber attack is one of the most serious threats facing many organizations in the Internet era. In addition to hardening computer and network systems, it is important to surely deliver security information to end users. This paper presents the results of analysis on malicious messages and URLs sent on a social networking service; Twitter. We also present a method and system for safe-browsing of URL links. URLs, and sometimes shortened URLs, on SNS may be utilized for malicious activities to redirect users to unexpected resources, e.g. phishing and malware, by obscuring the final destinations. The analysis results show that about 20% of messages contain at least one URL, 0.1% of messages contain malicious URLs. The users sent such messages and messages themselves have short lifetime on Twitter. That is, they are removed soon after malicious activities; 99% of them are disappeared within one month. The proposed system enables users to know how safe a particular Web resource might be before users dereference it. Our system retrieves and delivers safety information of the URLs on user's demand by a simple operation.

キーワード　情報セキュリティ，SNS，悪性 URL，短縮 URL，フィッシング，マルウェア

１．はじめに

ソーシャルネットワーキングサービス（Social Networking Service: SNS）は、スマートフォンと共に普及した便利なコミュニケーション手段である。特にTwitterはサービス開始から10年程度が経過し、世界中で30億人以上の利用者を持つ、もっとも普及したサービスの一つである。個人や企業、政府機関などが、不特定多数に向けてさまざまな情報発信に利用している。一方で、フィッシングやマルウェア感染などを目的としてSNSを悪用することが問題になっている。

SNSのメッセージ中にURLを掲載することで、受信者をWebリソースに誘導できる。Twitterではメッセージ文字数の制限から、短縮URLが用いられることが多い。短縮URLは、最終的なアクセス先が利用者から隠蔽されてしまうため、フィッシングやマルウェア感染を目的とした悪意あるサイトへの誘導など、セキュリティ上のリスクが高い。

本論文では、Twitterを対象とした、悪性URLに関する分析結果を示す。約200万件のメッセージを収集し、そこに含まれる約40万件のURLを分析した。

また、WebページやSNSメッセージなどに含まれるURLを安全に利用する手法とシステムについて述べる。最終アクセス先リソースの安全性をオンデマンドで検査し、ユーザに評価情報を提示する。本手法では、利用者はWebページやアプリケーション画面に表示されたURLリンクを通常通り手繰るだけで、追加の操作を必要としない。検査機能をプロキシとして実装することで、透過的なWebブラウジングを可能にしている。また、再帰的に複数回の短縮をされたURLに対して、複数リダイレクトを制限するWebブラウザの安全機能に関係なく安全性を検査できる。

本論文の構成は次のとおりである。2章では、Twitterメッセージを分析するために作成したデータセットについて述べる。3章では、短縮URLの概要とTwitterでの短縮URLの使用状況について述べる。4章では、Twitterメッセージに含まれる悪性URLに関する分析結果を示す。5章では、我々が提案するURLの安全性検査手法と悪性URLに対する防御策について述べる。6章で今後の課題を示し、7章で結論を述べる。

２．分析用データセット

本章では、Twitter を対象とした悪性 URL に関する分析を行うために作成したデータセットについて述べる。

２．１．Twitter メッセージの収集・保存方法

まず Twitter メッセージの収集方法について述べる。メッセージの取得には Twitter Streaming APIs [10]を用いた。この API を用いると、全メッセージからランダムサンプリングしたメッセージをほぼリアルタイムに取得できる。

この API は HTTP 上の REST スタイルの Web API で、応答に含まれる Twitter メッセージは JavaScript を利用した JSON 形式である。本研究では、収集したメッセージの格納に MongoDB [11]というドキュメン

ト型データベースシステムを用いた。MongoDB は JSON 形式のデータをそのまま格納でき、集約関数を含むクエリ言語を提供しているため、Twitter メッセージの格納と分析に適している。

２．２．分析用データセットの概要

本研究で作成した分析用データセットの概要を表 1 に示す。

2017 年 5 月に 11 日間を要して約 206 万個のメッセージを収集した［表 1 (1) (2)］。それらのメッセージの内、URL を含むものは約 66 万個で、写真と動画を除くと、URL を含むメッセージは約 45 万個であった［表 1 (3) (4)］。

これらのメッセージから URL を抽出し、重複を取り除いた結果、約 42 万個のユニークな URL を得た［表 1 (5)］。

表 1: Twitter 分析用データセットの概要

(1)	メッセージ収集期間 （メッセージの発信期間）	2017 年 5 月 17 日 ～ 2017 年 5 月 28 日 （11 日間）
(2)	メッセージ数	2,063,507
(3)	URL を含むメッセージ数	655,690 （(1)の 32%）
(4)	URL を含むメッセージ数 （写真・動画を除く）	445,658 （(1)の 22%）
(5)	メッセージに現れる ユニークな URL 数 （(4)のメッセージに現れる URL から重複を除いた URL の数）	419,733 （(4)の 94%）

３．短縮 URL

ここでは、短縮 URL の概要および Twitter での短縮 URL の利用状況について述べる。

３．１．短縮 URL の概要

短縮 URL とは、ある URL（元 URL）に対してその長さを短くした URL である。短い URL は、SNS で投稿記事の長さが制限されている場合に文字数の消費を抑える、Web ページの見栄えを損ねないなどの利点がある。

表 2: 短縮 URL の例

元URL	https://en.wikipedia.org/wiki/Japan
短縮URL 例1	http://bit.ly/1LXn5Yl
短縮URL 例2	https://goo.gl/J0HsVM

元 URL から短縮 URL を生成し、短縮 URL から元 URL への逆変換を行う Web サービスを URL 短縮サービスという。短縮 URL は、URL 短縮サービスのドメイン名と、元 URL に対応する一意なキーとで構成される。例えば表 2 の例 1 では、bit.ly がドメイン名、1LXn5Yl がキーである。同一の元 URL に対して、生成される短縮 URL はサービスごとに異なる（例 1 と例 2 は同一の元 URL から生成）。短縮 URL から元 URL への逆変換は、その短縮を行ったサービスだけが行える。

表 2 の例に示したように、短縮 URL から元 URL を推測することは困難である。この性質を悪用すると、フィッシングやマルウェア感染を目的とした悪性リソースへ利用者を誘導する攻撃が可能になる[12,13]。

３．２．短縮 URL の利用状況

表 1 (5) の URL には短縮 URL が含まれている。これらの URL の内、100 回以上使用されている短縮 URL サービスの短縮 URL だけを数え上げると 約 9 万 8 千個である［表 3 (6)］。すなわち、少なくとも 23%以上の URL が短縮 URL であった。先に述べたとおり、利用者は、これらの短縮 URL を見ただけでは最終的なアクセス先のドメイン名などはわからない。

同一の元 URL に対して複数の短縮 URL が存在しえる。このため、短縮 URL を伸長した上で重複を除去すると、収集したメッセージに現れるユニークな URL は約 24 万個であった［表 3 (7)］。これは、約 42 万個の URL が指示する Web リソースの数を意味する。

短縮 URL の生成に使用された URL 短縮サービスと、それらによって生成された URL 数を表 4 に示す。

表 3: 短縮 URL の利用状況

(6)	短縮 URL の数 （100 回以上使用されている短縮 URL サービスの URL のみ）	97,829 （(5)の 23%）
(7)	短縮 URL 伸長後の ユニークな URL 数 （Web リソース数）	239,643 （(5)の 57%）

表 4: よく利用される URL 短縮サービス（上位 10）

ドメイン名	URL 数
bit.ly	23,449
fb.me	21,425
youtu.be	11,821
goo.gl	9,339
ift.tt	6,900
dlvr.it	5,960
ow.ly	3,235
twcm.me	1,328
cas.st	1,267
nico.ms	1,145

４．悪性 URL の分析

本章では、Twitter を対象とした悪性 URL の分析結果について述べる。分析対象は表 1 に示したとおり、約 206 万個のメッセージと約 42 万個の URL である。

４．１．URL の悪性判定

まず、Twitter メッセージに現れる URL の悪性判定結果について述べる。

URL で指示された Web ページなどのリソースが悪性のものであるかどうかを判定することをここでは URL の悪性判定と呼び、悪性のリソースを指示する URL を悪性 URL と呼ぶ。悪性の Web リソースには、フィッシングやマルウェア感染の危険性がある[12,13]。本研究では、URL の悪性判定に Google Safe Browsing APIs [7,8]を用いた。この API は、要求した URL が悪性であるかどうかと、悪性である場合はその脅威種別を応答する。脅威種別には、malware、social engineering、unwanted software の 3 種類がある。

短縮 URL を伸長して得た約 24 万個のユニーク URL［表 3 (7)］に対して悪性判定を行った結果、621 個の悪性 URL を検出した［表 5 (8)］。これらの悪性 URL を含んでいたメッセージ数を数えると、2,050 個であった［表 5 (9)］。これは、収集した全メッセージ数に対して約 0.1%、URL を含むメッセージ数に対して約 0.5%の割合である。

悪性リソースの生存期間は短いと言われている。そこで、メッセージ収集後に約 1 か月経過してから、悪性 URL を含むメッセージが生存しているかどうかを調べた。すると、そのようなメッセージは 99%以上が削除されていた［表 5 (10)］。

以上の結果から、悪性 URL を含むメッセージは 1 千個に 1 個程度の割合で存在し、そのほとんどは短期間のうちに削除されていることがわかる。これは証拠隠滅を図るためと推測できる。

表 5：URL の悪性判定結果

(8)	悪性 URL の数 （判定方法：Google Safe Browsing API、 判定日：2017 年 6 月 26 日）	621 （(5)の 0.1%、 (7)の 0.3%）
(9)	悪性 URL を含むメッセージ数	2,050 （(1)の 0.1%、 (3)の 0.5%）
(10)	悪性 URL を含むメッセージの内、生存期間が 1 か月以上のメッセージ数 （調査日：2017 年 6 月 26 日）	15 （(9)の 0.7%）

４．２．メッセージの分析

悪性 URL を含むメッセージ（本文）に関する分析結果を示す。

まず、メッセージの言語コードを調べた。悪性 URL を含む 2,050 個のメッセージには 24 種類の言語コードが使用されており、言語判定不能（URL のみのメッセージ）のもわずかにあった。メッセージ数の多い順に 10 位までを表 6 に示す。英語が最も多く、欧州の言語が上位を占めている。日本語のメッセージは存在しなかった。

表 6：悪性 URL を含むメッセージに使用されている言語（メッセージ数の上位 10）

言語コード	日本語名	メッセージ数
en	英語	1,262
de	ドイツ語	260
nl	オランダ語	177
fr	フランス語	158
es	スペイン語	44
pt	ポルトガル語	28
it	イタリア語	16
da	デンマーク語	15
es	エストニア語	12
In, ht	インドネシア語，ハイチ語	11

メッセージに使用された単語の出現頻度を調べると、図 1 に示すように、卑猥な単語がよく使用されていた。これらの単語の検索結果に悪性 URL を含むメッセージを表示させ、それらのメッセージから悪性リソースへ誘導することを狙っていると推測できる。

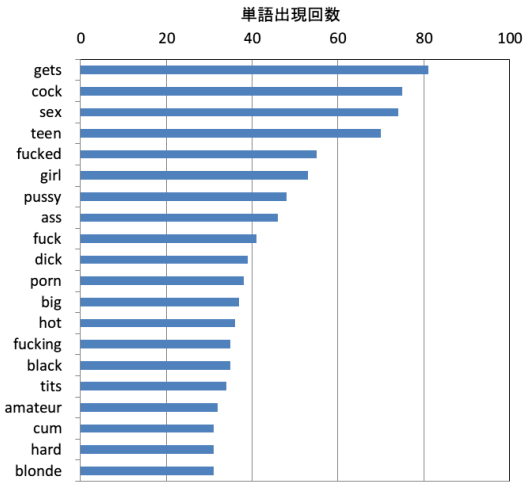

図 1：悪性 URL を含むメッセージの単語出現頻度（英語のみ上位 20 単語）

４．３．ユーザの分析

悪性 URL を含むメッセージの送信者に関する分析結果を述べる。

まず、これらのメッセージの送信アカウント数は 1,942 個であった［表 7 (11)］。同一人物が複数のアカウントを利用している場合と、複数の人物が同一アカウントを共有している場合とが考えられるので、このアカウント数が攻撃者数と一致するとは限らない。

メッセージの収集から約 1 か月経過した後にこれらのアカウントを調査してみると、99%以上が削除され、23 アカウントだけが存続していた［表 7 (12)］。このことから、攻撃者は悪性 URL を頒布する目的で Twitter アカウントを作り、実際にメッセージの送信を行い、目的を達成した後ですぐにアカウントを削除するという一連の行動をしていることが推測できる。

次に、アカウントの作成時期を調べた。すでに削除されたアカウントについては情報を取得できず、メッセージ収集から約 1 か月後に存続していた 23 アカウントだけを調査した。すると、過去 3 年以内に作成されたアカウントが約 61%、その内 2017 年になってから（すなわち過去半年以内に）作成されたアカウントが約 30%あった［表 7 (13,14)］。

これらの比較的新しいアカウントからのメッセージ送信数を調べてみると、50%のアカウントで 1 万個以上のメッセージが送信されていた［表 7 (15)］。2017 年 4 月 6 日に作成されたあるアカウントからは、同年 6 月 26 日までの 3 か月弱の間に約 5 万 1 千個のメッセージが送信されていた。

これらの 23 アカウントの中には、メッセージだけでなく、プロフィールに悪性 URL を記載しているものも 3 個あった［表 7 (16)］。SNS では、魅力的なプロフィール写真を利用して他者のメッセージにコメントし、コメントされた利用者がプロフィールを調べに来るのを待ち構えて悪性リソースに誘導する攻撃方法が知られている。

表 7: 悪性 URL を含むメッセージの送信アカウント

(11)	アカウント数	1,942
(12)	メッセージ送信後、1 か月以上存続したアカウント数（調査日: 2017 年 6 月 26 日）	23（(11)の 1.2%）
(13)	過去 3 年以内に作成されたアカウント数	14（(12)の 61%）
(14)	2017 年になってから作成されたアカウント数	7（(12)の 30%）
(15)	過去 3 年以内に作成されたアカウントで、1 万個以上のメッセージを送信しているアカウント	7（(13)の 50%）
(16)	プロフィールに悪性 URL を記載しているアカウント数	3

５．URL のオンデマンド安全性検査による防御

本章では、我々が提案するURLの安全性検査手法とシステム[12,13]について述べる。

５．１．URL 安全性検査の手順とシステム構成

本システムは、URLの安全性検査を要求するクライアントと、安全性検査を実施するサーバ、外部サービスとで構成される［図 2］。

以下に、URL安全性検査の手順を示す。

URL 安全性検査の手順:

（１）URL安全性検査クライアントCは、短縮URL（URL_S）をサーバに送る。

（２）URL安全性検査サーバSは、URL_Sのトップレベルドメインに対応する外部URL短縮サービスを呼び出し、伸長結果の元URL（URL_L）を得る。URL_Sが多重に短縮されている場合は、伸張処理を繰り返す。キャッシュに検査結果が存在する場合は、その検査結果をCに送り、以下の処理を省略する。

（３）URL_Lの安全性検査を外部の検査サービス（設定により複数）に委譲し、検査結果（$R(URL_S)$）を作成する。

（４）$R(URL_S)$をキャッシュに保存する。

（５）$R(URL_S)$をクライアントに送る。

（６）Cは、サーバから検査結果$R(URL_S)$を受け取り、それを別の画面（ウィンドウやタブ、またはポップアップ）に表示する。

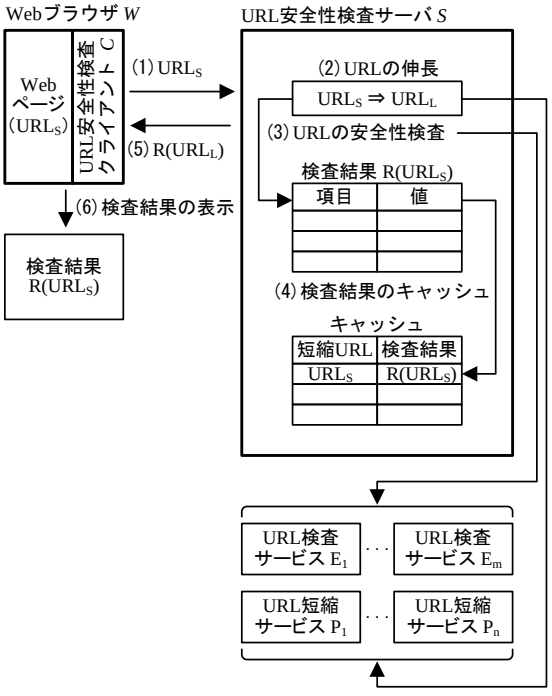

図 2: URL安全性検査の手順とシステム構成

５．２．ユーザインタフェース

我々が開発したシステムのユーザインタフェース（UI）について説明する。UIについて重要視した要件は、利用者のユーザビリティを損ねないことである。利用者に要求する操作が難しかったり煩雑であったりしては、この安全性検査機能が利用されない。これからアクセスしようとするWebサイトなどの安全性を確かめず、すぐにそのリソースにアクセスしてしまう。本システムでは、URLの入力やそのための画面遷移など、不要な操作ステップを極力排除し、１ステップまたは0ステップで検査結果を提示できるようにした。0ステップとは、明示的な検査操作を一切必要としないという意味である。

利用者の操作手順（実装方法 1）：

（１）Webブラウザ上で、短縮URLの上でマウス右クリックする。

（２）コンテキストメニューが表示されるので、短縮URL安全性検査機能を選択する。メニューには『短縮URL「実際の短縮URL」を検査』と表示される。[図 3]

（３）別のWebページ（またはポップアップ）が開き、評価情報が表示される。これを参考に、元のページで短縮URLのリソースにアクセスするかどうか判断する。

図 3：URL 安全性検査のユーザインタフェース 1

既存のURL検査システム[5,6]では、利用者はまずURL安全性検査サービスが提供するWebページを開き、対象のURLをタイピングやコピー＆ペーストでHTMLフォームに入力しなければならない。これに対し、我々のシステムでは、利用者は画面遷移やURLの入力を必要とせず、画面上のURLに対する直接操作ができる。このことから、既存システムに比べてユーザビリティが高いといえる。

利用者の操作手順（実装方法 2）：

（１）Webブラウザ上で、短縮URLの上でマウスをURLの上に乗せる。安全なURLの場合は画面に変化がないが、危険なURLの場合にはURLの背景色が変化し（赤色）、クリックする前に警告を与える。

（２）安全なURLをクリックした場合はそのリソースにアクセスする。警告にもかかわらず危険なURLをクリックすると、別のWebページ（またはポップアップ）が開き、評価情報が表示される。

図 4：URL 安全性検査のユーザインタフェース 2
（上の URL にマウスオーバーしている）

５．３．検査結果情報

URL安全性検査サーバがクライアントに返す検査結果は以下の項目から構成される。

表 8：URL 安全性検査結果の内容

短縮 URL 情報: 短縮 URL、元 URL、作成日時、多重短縮の場合に限り伸長過程（短縮 URL と元 URL のリスト）
元 URL 情報: 脅威の種別（マルウェア、フィッシングなど）、MIME タイプ、文字コード、タイトルなどのコンテンツ情報、サムネール画像
レピュテーション情報: リファラ、アクセス数（伸長回数）、アクセス元地域

５．４．システムの実装

提案手法に基づくシステムの基本的な機能は実装済みである。クライアント側はChromeブラウザを対象とし、Chrome Extensions [9]を使用してJavaScriptで開発した。他の主要ブラウザへの対応は今後進めていく。サーバ側はRuby on Railsを使用してRuby言語で開発した。

サーバのURL安全性検査機能はRESTスタイルのWeb APIからの呼び出しが可能で、異なる実装形態のクライアントやまったく別のアプリケーションの実装にも利用できるようにした。

外部のURL短縮サービスではBitly、Google、HootSuite、TinyURLに対応し、URL検査サービスにはGoogle Safe Browsing APIs (v4)[7, 8]を用いた。

６．今後の課題

本章では、悪性URLの頒布状況の分析および防御策に関する今後の研究課題について述べる。

多様なSNSおよび頒布経路の分析

本研究では、Twitterを対象に悪性URL頒布の分析を行った。Twitterでは、メッセージが140文字に限定されているため、コンテキスト情報が少ないので、利用者はURLで指示されたリソースをすぐに見てしまうという行動パターンが推測される。

Twitter以外のSNSや、最近の写真・動画を中心とするSNSについても同様の分析を行い、相互に比較することで新たな知見が得られる可能性がある。また、SNS以外に、一般公開されていない個人間のメールやメッセージが頒布経路になっている可能性がある。スマートフォンで普及しているLINEもその一つである。

複数アカウントの結託の分析

本研究では、複数のアカウント同士の関係は分析できていない。Twitterではアカウント間にフォロー関係があり、悪性URLの頒布に悪用されている可能性がある。

日本国内の頒布状況の分析

本研究で作成したデータセットには日本語のメッセージが含まれていない。国内での悪性URLの頒布状況を分析するために、現在、日本語のTwitterメッセージのみを収集し分析を始めたところである。

また、国内最大の掲示板である2チャンネルを対象にして同様の分析を行う予定である。

安全性検査結果に基づく利用者行動の抑制

ブラウザの警告に対する利用者行動の大規模な調査研究[1]によれば、2種類のブラウザ（Mozilla FirefoxとGoogle Chrome）でそれぞれ9.1%および18.0%の利用者がフィッシングサイトの警告を無視していた。また、マルウェアの警告に関してはそれぞれ7.2%と23.2%の利用者が無視していた。この問題は、技術的な対策だけでは解決できない。セキュリティに対する人の心理的・行動的特性に関する研究[2,3,4]が必要であろう。

７．結論

本論文では、SNSメッセージの悪用方法に着目し、Twitterメッセージに含まれる悪性URLの分析結果を示した。32%のメッセージにURLが含まれており、そのうちの0.5%のメッセージがURLを用いた悪性リソースへの誘導に利用されている。

このような悪性のメッセージおよびそれらを送信したユーザアカウントは、短期間のうちに削除される。約1か月の間に、99%以上のメッセージおよびユーザアカウントが削除されていた。これは、証拠隠滅を図るためと推測できる。

また、悪性メッセージは、半数以上が英語で記述され、卑猥な単語が多用されている。検索結果からメッセージに到達させ、そこに含まれるURLから悪性リソースへ誘導する狙いと思われる。

悪性URLのリスクに対抗する手段として、本論文ではオンデマンドでリンク先リソースの安全性を検査するしくみを示した。一般利用者のユーザビリティを考慮して、透過的に安全性検査を実行できる点が特徴である。提案手法はクライアント・サーバ構成のシステムを実装済みで、Google Chromeなどの一般的なブラウザで動作する。

参考文献

[1] Akhawe, D., Felt, A.P. (2013): “Alice in Warningland: A Large-Scale Field Study of Browser Security Warning Effectiveness”, 22nd USENIX Security Symposium, 2013.

[2] 西岡大, 齊藤義仰, 村山優子: “専門知識のないユーザを対象とした情報セキュリティ技術に関する安心感の構造”, 情報処理学会論文誌, Vol.54, No.9, 2013, pp.2197-2207.

[3] 寺田剛陽, 津田宏, 片山佳則, 鳥居悟: “IT被害に遭いやすい心理的・行動的特性に関する調査”, マルチメディア、情報処理学会 分散協調とモバイルシンポジウム2014論文集, 2014, pp.1498-1505.

[4] 寺田剛陽, 鳥居悟, 安野智子, 瀧澤弘和, 新真知: “リスク認知に基づく標的型メール対策の検討”, 情報処理学会研究報告セキュリティ心理学とトラスト, Vol.2013-SPT-5, No.9, 2013.

[5] ExpandURL, http://www.expandurl.net/

[6] 短縮URLチェッカー, http://x-1.jp/

[7] Google 透明性レポート セーフブラウジングのサイトステータス, https://www.google.com/transparencyreport/safebrowsing/diagnostic/

[8] Google Safe Browsing APIs (v4), https://developers.google.com/safe-browsing/v4/

[9] Google Chrome Extensions, https://developer.chrome.com/extensions

[10] Twitter Streaming APIs, https://dev.twitter.com/streaming/overview

[11] MongoDB, https://www.mongodb.com/

[12] 中村章人, 松尾卓朗, 林 隆史: 短縮URLの安全な利用に向けて, 2016年社会情報学会（SSI）学会大会, 2016.

[13] 中村章人, 松尾卓朗, 西川直登: 短縮URLのオンデマンド安全性検査, 情報処理学会 第79回全国大会, 2017.