平成 29 年度 修士論文

短縮URLサービスの実態調査

Understanding the Usage of URL Shortening Services in the Wild

指導教員　森 達哉 准教授

早稲田大学　基幹理工学部　情報理工学科

学籍番号 5116F080

星野 遼

提出日： 2018 年 1 月 30 日

概要

短縮 URL と呼ばれるサービスがある．URL を他人と共有するシチュエーションとして，ソーシャルネットワーキングサービスでの共有があるが，これらには文字数が制限されているものもあり，あまりにも長い URL は扱いにくい．また，メールやメッセージアプリケーションでも，ツールによっては長い URL が半端に改行されてしまい読みにくくなってしまうケースがある．短縮 URL サービスでは，長い URL を 14 文字〜24 文字程度にまで短縮することができ，上記の問題を解消する．しかし，短縮 URL サービスは適切に利用されない場合，いくつかの危険を伴うことがある．

本研究では，短縮 URL サービスでどのような URL が短縮の対象となっているかを調査した．そのために，様々な短縮 URL サービスの短縮 URL 及び対となる元の長い URL (Long URL) を収集するためのクローラーを作成した．また，そのクローラーを用いて Short URL と Long URL の収集を行った．最終的に約 1,600 万の短縮 URL を収集・分析し，悪用される危険のある URL の有無や，有害な URL の有無を調査した．その結果，未だにアクセスが有効な文書や画像などの共有 URL の短縮 URL の存在が明らかになった．そのような共有 URL にはプライバシーに関わる情報が含まれているものも存在した．また，悪性サイトの URL も数多く短縮されており，その中には短縮 URL サービス側で悪性であると判断されていないものも存在することが明らかになった．さらに，短縮 URL を多重にかけているような多重短縮 URL も存在し，その中にも悪性サイトの URL が存在することがわかった．多重短縮 URL は最大で 327 重もの深さをもつものも存在した．ただし多重短縮 URL の深さと悪性サイトの数に相関は見られなかった．また，多重短縮 URL の場合，短縮 URL サービス側が最終到達先の悪性 URL を検知できていない可能性があることが分かった．

このような結果を踏まえて，プライバシー情報漏洩の危険性のある利用を防ぐためにユーザが使用方法を改善したり，悪性 URL へのアクセスを回避するためにサービス側が検知精度あげると共に，ユーザが短縮 URL の危険性を理解し，注意して利用する必要がある．

目次

図目次

表目次

第 1 章　序論

本章では，本研究における背景，目的，主な貢献について示す．また，本論文の構成について示す．

1.1　はじめに

短縮 URL と呼ばれるサービスがある．メールやメッセージアプリ，SNS でのウェブサイトの共有の需要から，長く扱いにくい URL を短い URL に変換するサービスである．Google 社の `goo.gl` や Bitly 社の `bit.ly` など有名なもの，Twitter 社が自社の SNS 内で使用するための `t.co`，さらにマイナーなものまで含めると数十個の短縮 URL サービスが存在する [1, 2, 3].

しかし，短縮 URL はその利便性とは裏腹に，悪用される危険性もある．例えば，クラッカーが悪用する場合である．攻撃者は悪意を持って，一般ユーザを有害なサイトへ誘導しようとする．その際，有害なサイトの URL を短縮 URL サービスを用いて短縮することで，攻撃者が用意した本来の URL や IP を隠すことが可能である．また，正規のユーザが誤った使い方をすることで，プライバシー情報の漏洩につながる危険性もある．例えば，あるユーザが何らかの文書ファイルを，大学の友人や仕事の同僚と共有したいとする．その際，Google Drive や DropBox などのクラウドストレージサービスの共有リンクを使うことで簡単にファイルを共有することが可能である．共有リンクは，長いランダム文字列で構成されており，共有リンクを知る者しかファイルにアクセスすることができないという形で安全が担保される．ここで，もしユーザが長いランダム文字列で構成された URL に不便を感じ，短縮 URL サービスを用いて短縮してしまうとする．すると本来は推測不可能であるはずの共有リンクに，ブルートフォース攻撃などによってアクセスすることが可能である．現にブルートフォースすることで何らかの URL を短縮した短縮 URL を見つけることが可能である．

本研究では，このような短縮 URL サービスの悪用や誤用について調査することを目的に，様々な短縮 URL サービスの短縮 URL を収集した．具体的には，様々な短縮 URL サービスのクローラを作成し，約 3 ヶ月間短縮 URL を収集した．また，様々な手法によるクロールを試みて，約 1,600 万の短縮 URL を収集した．そして集めた短縮 URL を解析し，短縮 URL サービスの使用法の実態を調査した．具体的には，短縮 URL サービスで短縮された URL の FQDN,

URI スキーム， ファイルタイプ， 悪性 URL の有無， 多重短縮 URL について解析し，考察・対策を示した．

1.2 研究背景

1.2.1 短縮 URL サービス

1.1 節の冒頭にも述べた通り，短縮 URL サービスは長く扱いにくい URL を短縮するサービスである．例えば，Google の短縮 URL サービスで **`https://nsl.cs.waseda.ac.jp/achievements/awards/` (48 文字)** という URL を Google 社の `goo.gl` で短縮すると，**`https://goo.gl/65TeRC` (21 文字)** といった具合に文字数を約半分にまで減らすことができる．Google 社の goo.gl 以外にも `is.gd`, `v.gd`, `ow.ly`, `tiny.cc`, `ux.nu` など数多くの短縮 URL サービスが存在する．

1.2.2 短縮 URL の原理

短縮 URL にアクセスすると，最終的には短縮する前の長い URL (以下，LongURL) にアクセスすることができる．これにはリダイレクトが用いられており，主に HTTP リダイレクトが行われている．HTTP リダイレクトでは，HTTP のステータスコードによってリダイレクトの種類を伝え，HTTP ヘッダの Location ヘッダにリダイレクト先が書き込まれる．この手法でリダイレクトを行なっているか否かは，短縮 URL にアクセスした際の HTTP ヘッダを参照することで確認することができる．HTTP リダイレクトを行う短縮 URL の例を図 1.1 に示す．

リダイレクト手法には HTTP リダイレクトの他，meta タグを利用したリダイレクト手法や JavaScript の location.href を使ったリダイレクト手法も存在する．本研究で扱う短縮 URL サービスは全て HTTP リダイレクトを用いていたが，`tr.im` という短縮 URL の一部は JavaScript によるリダイレクトを使用していた．その例を図 1.2,1.3 に示す．

1.3 研究目的

本研究では，短縮 URL サービスの利用実態の調査を行う．具体的には一般ユーザがプライバシー情報の漏洩の危険性のある URL を短縮していないか，あるいは攻撃者が悪性なサイトの URL を短縮していないか，またそれらの悪性 URL を各サービスが検知しているかを確かめることを目的に，約 1,600 万の短縮 URL をクロール・収集した．また，いくつかのサービスを比較するために複数のサービスに対するクローラを作成した．そして集めた短縮 URL を解析し，短縮 URL サービスの利用実態を明らかにした．

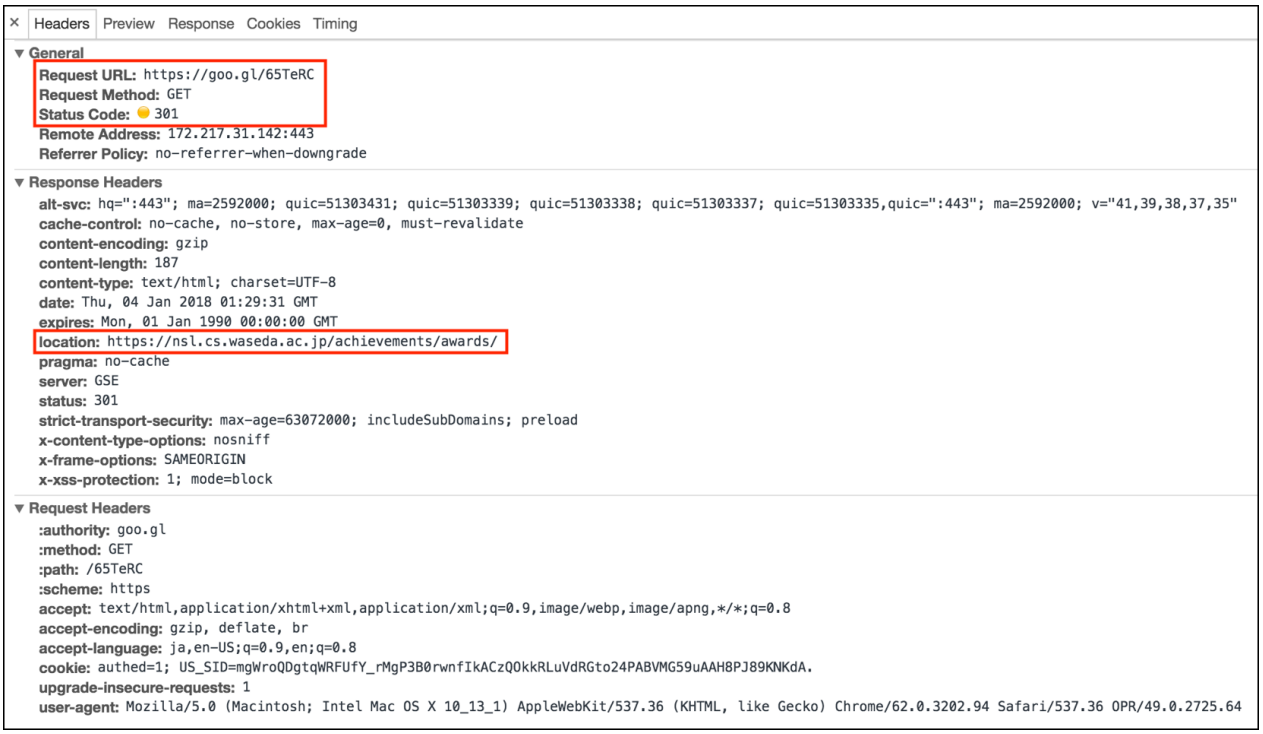

図 1.1　Opera ブラウザのデベロッパーツールで確認した短縮 URL アクセス時のリクエストヘッダとレスポンスヘッダ

1.4　主な貢献

本研究の貢献を以下に示す．

1. 様々な短縮 URL サービスに対するクローラを作成した．
2. 約 1600 万の短縮 URL を収集した．
3. 短縮 URL の LongURL の FQDN からクラウドサービスの共有 URL を調査し，プライバシー情報に関わる共有 URL の存在を発見し，ユーザによる危険な誤用を明らかにした
4. Twitter の直近のツイートから収集した短縮 URL にも悪性 URL が存在し，サービス側が悪性であると検知できていない可能性があることを示した

1.5　論文の構成

本論文の構成は以下の通りである．第二章で本研究に関する既存研究を述べる．第三章では短縮 URL の収集手法及び解析事項について説明する．第四章では収集した URL の解析結果を示す．第五章では四章の結果を踏まえた考察・対策を述べ，第六章でまとめる．

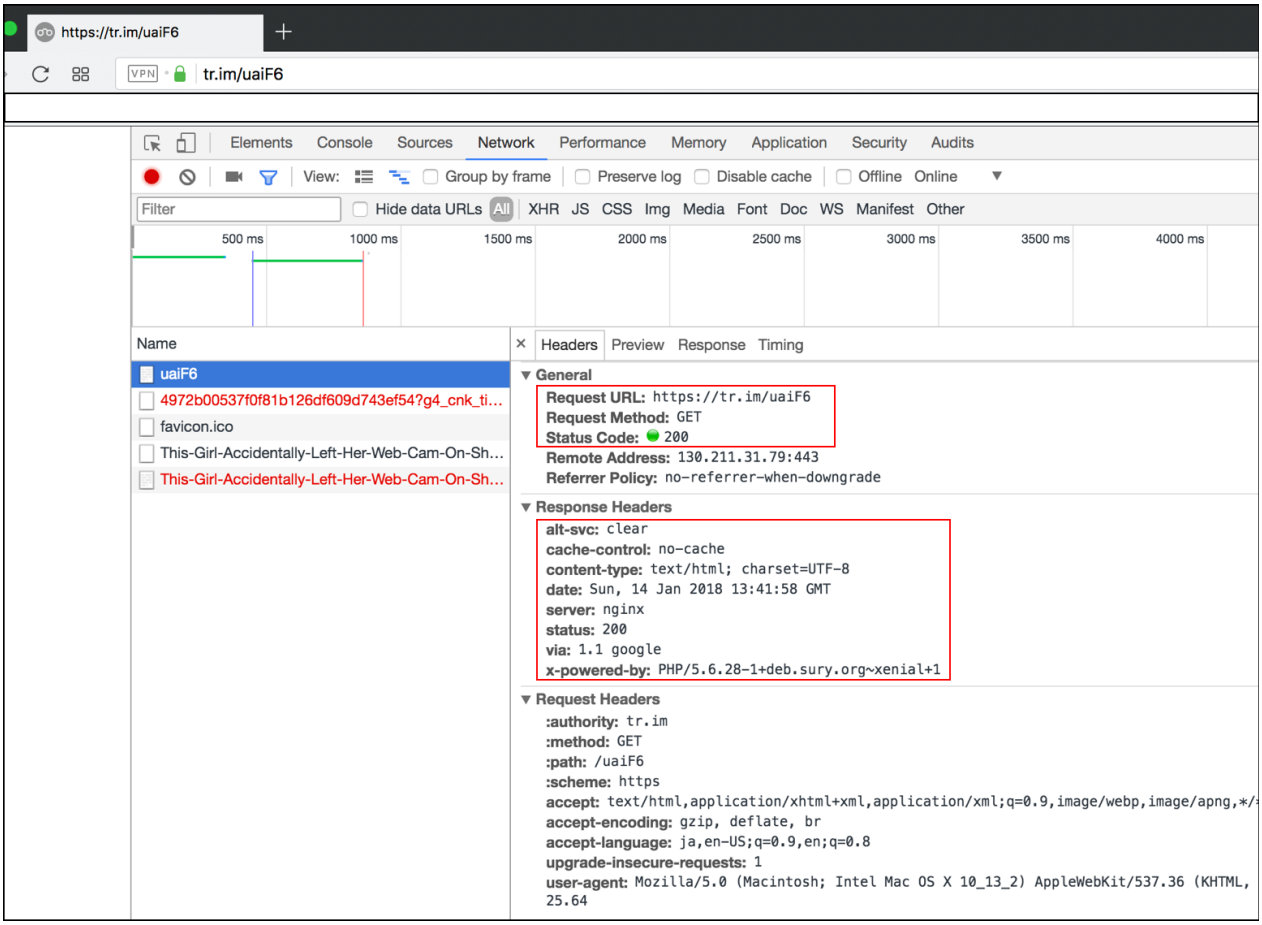

図 1.2　Opera ブラウザのデベロッパーツールで確認した tr.im へのアクセス時のリクエストヘッダとレスポンスヘッダ．goo.gl の時と異なり，status が 301 ではなく location ヘッダが存在しない．

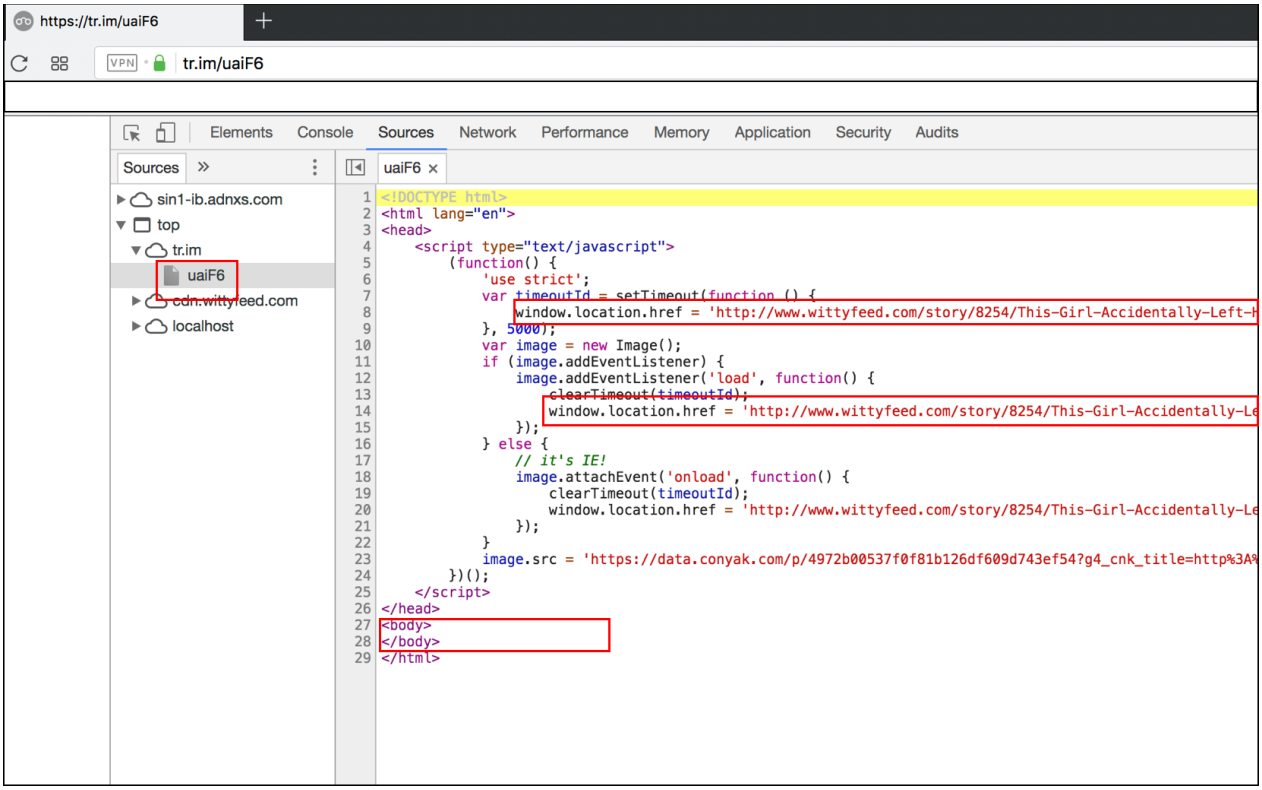

図 1.3　Opera ブラウザのデベロッパーツールで確認した tr.im の短縮 URL の HTML．script タグに location.href によってリダイレクトする処理が書かれていることがわかる．

第 2 章　関連研究

この章では，本研究との関連研究を示す．はじめに短縮 URL の危険性を指摘している研究を示す．次にメールに記載された短縮 URL を収集し解析を行った研究を示す．最後に特徴的な短縮 URL 収集手法を用いている研究を示す．

Neumann らは過去 7 年間のメールを蓄積し，そこに記載されている短縮 URL を収集することで，7 万個の短縮 URL を収集した．その中から，スパムへのリンクである短縮 URL を調査し，その割合を示した [1]．また，この研究で短縮 URL の脅威モデルを示しており，本研究でも示しているプライバシー情報を含む共有 URL の短縮や悪性 URL の短縮も指摘されている．

Georgiev らは，短縮 URL が作り出す URL が短いことに着目し，短縮 URL のランダム部分を総当たりすることにより 4,300 万個の短縮 URL を収集した [4]．その短縮 URL の中で Google ドキュメント等の共有 URL を抽出し，そのうち約 7% は書き込み可能なファイルであることを示している．我々の研究でも総当たりを利用した短縮 URL の収集を行なっている．また，我々の研究ではドキュメントだけでなくプライベートな写真などの漏洩の危険もあることを示している．

Maggi らは，短縮 URL のリダイレクト先の情報 (タイトル，ページサイズなど) を，アクセス前に取得できるブラウザアドオンを作成し，２年間ユーザに利用してもらうことで短縮 URL を収集した [5]．この手法の利点は，インターネット上でユーザが実際に出くわした短縮 URL を効率よく収集できることである．Georgiev らが総当たりで集めた URL は，現在では使われていないものが含まれているのに対し，Maggi らの手法であれば比較的に新しい短縮 URL を多く手に入れることができる．我々の研究では，Twitter のストリーム API を利用してリアルタイムにツイートを収集することでできる限りフレッシュな短縮 URL を集めた．

第 3 章　短縮 URL の収集手法と解析

本章では，短縮 URL の収集手法について述べる．はじめに収集手法について示す．次に収集した URL の解析項目を示す．最後に各短縮 URL サービスの特徴を述べる．なお，本研究では 3.1 節に示す手法のうち，API と Location ヘッダを用いた総当たり法と，Twitter のストリーム API を用いたクロールを行なっている．

3.1　短縮 URL のクロール手法

本研究では，2 種類のクロールを行っている．短縮 URL のパスのランダム部分を総当たりして存在する短縮 URL を収集する方法と，ウェブ上をクロールして短縮 URL を集める手法である．

3.1.1　総当たり法

短縮 URL のパスのランダム部分を総当たりして求める．文字数 1　5 文字程度までならこの手法で収集することができる．6 文字程度までなら確率的に収集することもできるが，7 文字以上だと一気に確率が下がってしまい収集が難しくなる．本研究では総当たり法の中でも以下の 3 種類の収集方法を試している．3.1.1 項の API は多量のリクエストが来ることを想定しているのに対し，Location ヘッダやプレビュー・クッションページを利用するような手法は，サービスに直接大量のリクエストを送ることになってしまうため，収集の際は適当な間隔をあけ，サービスに必要以上の負荷をかけないよう注意しなければならない．

Location ヘッダを用いる

短縮 URL と対になる元の LongURL を知るだけであれば，HTTP ヘッダの Location ヘッダを見るのが最もシンプルである．1.2.2 項でも述べたように，ほとんどの短縮 URL サービスは HTTP リダイレクトを利用している．そのため，短縮 URL に対して head リクエストを送りヘッダ情報を取得し Location ヘッダを確認することができればリダイレクト先を特定することができる．ただし，他のリダイレクト手法 (meta タグや JavaScript の location.href を使う方法)

でリダイレクトするサービスがあった場合は使うことができない．またデフォルトでクッションページにアクセスするような機能を持っている，あるいはリダイレクト先を危険な URL と判断した場合にクッションページを表示するような短縮 URL サービスの短縮 URL でも使うことができない．クッションページとは，短縮 URL にアクセスした際，最終的なアクセス先である LongURL の情報が記載されたページ (プレビューページ) を前もって表示し，その LongURL への遷移の許可をユーザに促すためページである．つまり，リダイレクトが起きずにクッションページが表示されるので HTTP ヘッダにはクッションページへのリクエストの情報しか入っていため LongURL を取得できないということである．

プレビュー・クッション・メンテナンスページを用いる

上述したような，クッション・プレビューページを採用しているサービスであれば，そこに表示される LongURL を取得することで URL を収集できる．もちろんクッション・プレビューページが無いサービスでは使用できない．例外として，クッションページの有無を設定できるようなサービス (ux.nu, v.gd 等) であれば，クッションページの表示を OFF にした後に，3.1.1 項の手法で収集することも可能である．

また，短縮 URL サービスによっては，メンテナンスページというページが用意されているものがある．メンテナンスページは，ある LongURL を短縮して生成された短縮 URL に対し，どの程度アクセスがあったか (短縮 URL のクリック数) やどの国からのアクセスがいくつあるか等の解析情報を参照することができるページである．メンテナンスページであれば，前述の解析情報も閲覧することが可能である．ただし，メンテナンスページの情報は JavaScript で処理していることが多いため，ブラウザ自動化などの技術を使う必要がある．メンテナンスページを利用するシチュエーションとしては，前述の手法が使えない場合や API は用意されていないがメンテナンスページが存在するような短縮 URL サービスで解析情報を得たい場合が考えられる (本研究ではそのようなサービスは見つかっていない)．また，API のリクエスト制限が厳しい場合にこちらの方法を代用することもできるだろう．

API を用いる

短縮 URL サービスによっては，API が用意されているものもある [6, 7]．この API を使うことで，独自の短縮 URL を持たないようなサービスでも，プログラム的に短縮 URL を生成することが可能である．その中には，短縮 URL を渡すとリダイレクト先の LongURL を返す (この工程を expand と呼ぶ)API も存在するため，これを利用することで収集ができる．また，メンテナンスページの項で述べたような解析情報を取得できるような API が用意されている短縮 URL サービスもあり，最も簡単かつ多くの情報の収集が期待できる．ただし，API にはリクエスト制限が付いていることが多く，同一ユーザ (同一 IP, 同一アカウント) から収集するのは効率が悪い場合もある．アクセス元を Tor などで匿名化したり複数プロキシを切り替えながらアクセスすることで緩和できるが，匿名化に利用される IP や有名なプロキシの IP からのリクエ

ストを弾くサービスもあるので注意する必要がある．

3.1.2 ウェブ上のクロール

こちらは存在するかわからない短縮 URL を求める総当たり法に対し，ウェブ上に確実に存在する (した) 短縮 URL を収集する．総当たり法に比べて多少テクニカルになる代わりに，ユーザが実際に遭遇した可能性の高い短縮 URL を収集することができる．本研究では主に Twitter のストリーム API を用いて Twitter の投稿から収集している．

検索エンジン

検索エンジンの検索演算子を使うことで，詳細な条件で検索することが可能である [8]．例えば，`site:` という検索演算子を使い `site:nsl.cs.waseda.ac.jp` のように検索すると，`nsl.cs.waseda.ac.jp` のドメインを持つサイトだけを抽出することが可能である．また，これを利用して `goo.gl` や `bit.ly` のページを検索することで短縮 URL を収集するのが検索エンジンを用いた手法である．この手法は検索エンジンの API を使う方法とブラウザを自動化して検索エンジンから直接検索し，その結果をスクレイピングする方法がある．しかしどちらの手法も仕様上検索結果全てを見ることができず，収集効率は悪いため，本研究では後述の方法を使っている．

サービス内検索

この方法では，短縮 URL を収集したい SNS 等の投稿サービスを決め，そのサービスのサービス内検索機能を用いて収集する．例えば，Google+ や Twitter などでこの手法を扱うことができる．本研究では Twitter を対象に短縮 URL を収集したが，Twitter の場合は Twitter のストリーム API を使った方が効率が良く，主として API を用いた方法で収集を行なった．

API

SNS 等の投稿サービスによっては，投稿を取得する API が用意されているものがある．例えば Twitter であればタイムラインを取得するためのストリーム API がある．この API を使うことで，リアルタイムにツイートを収集することが可能であり，短時間で多くの短縮 URL を収集することが可能である．本研究では Twitter のストリーム API を用いて `t.co` の短縮 URL を収集した．

3.2 解析内容

3.2.1 短縮されている URL のドメインの分析

短縮されている URL のドメインを分析することで，どのようなサービスの URL が短縮されているかを把握することができる．つまり，ユーザが短縮 URL をどのような目的で使用しているかを知ることができる．ここでクラウドサービス等の共有 URL が数多く確認されれば，共有 URL によるプライバシー情報漏洩の可能性も高いと言える．さらに，もしも悪性 URL のドメインが発見されれば，攻撃者による悪用の危険性が高いことが示せる．

3.2.2 短縮されている URL のファイルタイプの分析

短縮 URL が，動画サイトの URL やブログの投稿などの共有のみに使われているなら，LongURL はパスのみ (`https://nsl.cs.waseda.ac.jp/` 等) になるか，html や php などのファイルを指す．逆に，png や pdf 等の画像や文書ファイルを指しているものがあれば，その中にプライバシー情報に関わるものがある可能性がある．特に jpg などの場合は exif 情報などに個人情報が含まれている可能性も考えられる．ただし，通常 jpg ファイルや png ファイルなどはサイトに埋め込むためにも利用され，pdf ファイルは電子マニュアル等として公開されているものも多い．そのため，特定のファイルタイプの存在自体がプライバシー情報の漏洩を示すわけではない．

3.2.3 短縮されている URL の URI スキームの分析

URI スキームの分析を行う理由は，過去に URI スキームの一つである data スキームと短縮 URL を組み合わせた攻撃が存在したためである [9]．URI スキームとは，例えば `https://nsl.cs.waseda.ac.jp` という URL であれば https の部分を指す．URI スキームには，指定したメールアドレスを宛先としてメーラーを起動するための mailto スキームや，指定したデータを Web ページに埋め込むことができる data スキームなどが存在する．過去に存在した攻撃では，正規サイトに偽装した悪性サイトの URL 表示部分に，dataURI スキームを利用して正規のサイトの https の URL を記載することで，一見正規サイトの URL が表示されているかのように見せかけていた．このような偽装は短縮 URL に限らず存在するものである [10]．詳細は参考文献の URL を参照してほしい．　現在でも dataURI スキーム等の URI スキームを短縮しているものが存在しているなら，URI スキームの悪用による攻撃にも注意すべきであると言える．

3.2.4 短縮されている悪性 URL の割合の分析

短縮されている悪性 URL を調査する．本論文では攻撃者の悪用による短縮 URL の危険性を示している．この分析で悪性 URL の短縮が高い割合で確認できてしまう場合，今後も攻撃者の悪用の危険性があると言える．悪性 URL の判定には，ブラックリストと VirusTotal と呼ばれる悪性 URL やファイルの検査サービスを利用する．

3.2.5 多重に短縮された URL の分析

短縮 URL を多重にかけているものが存在する．例えば，`http://goo.gl/XhEB` にアクセスすると `http://bit.ly/cFwXZR` にリダイレクトが発生し，再び `http://goo.gl/VL1Q` へリダイレクトするというような挙動をする．このように何度もリダイレクトを行う挙動は多段リダイレクトなどと呼ばれる．我々はこのような短縮 URL を多重短縮 URL と呼び，短縮の回数を深さと表現する．一方で，多段リダイレクトは攻撃者がドライブバイダウンロード攻撃を行う際，ユーザを入り口サイトから誘導する際に使われることがある．そこで我々は，多重短縮 URL が攻撃のために悪用されているのではないかという仮説を立て，実際に多重短縮 URL の LongURL に悪性 URL が存在するかを調査した．

3.3 各短縮 URL の特徴

本研究では，Google 社の `goo.gl`, Bitly 社の `bit.ly`, TinyURL 社の `tinyurl.com`, HootSuite 社の `ow.ly`, MEMSET Hosting 社の `is.gd`,Twitter 社の `t.co` の合計 6 つの短縮 URL サービスを扱う．注目した短縮 URL サービスの特徴を以下の表 3.1 に簡単にまとめる．

Service	Owner	API	Maintenance URL
goo.gl	Google	✓	✓
bit.ly	Bitly	✓	✓
tinyurl.com	TinyURL	✗	✗
ow.ly	HootSuite	✓	✗
is.gd	MEMSET Hosting	Only create	✗
t.co	Twitter	✗	✗

表 3.1 各短縮 URL サービスの特徴

`goo.gl` は，Google のサービスであり知名度が高く，API やメンテナンスページも充実している．メンテナンスページは解析情報をクライアント側のスクリプトで処理しているので，これらの情報をプログラムで取得するにはブラウザ自動化や，必要に応じてヘッドレスブラウザ

の技術を使う必要がある．Google API Console に Google アカウントでログインすることで使用できる API を使うことで，短縮 URL とペアの LongURL はもちろん，解析情報も取得することが可能である．また，一つのアカウントで複数の API キーを取得することが可能だが，1 ユーザ (1 IP) につき，1 秒 1 リクエストという制限があるため，最大でも 1 日 86,400 回しか使うことができない．そこで，1 ユーザの定義が 1 IP であることを利用し，プロキシサーバ等を介してリクエストを送り，動的にプロキシを切り替えながらクロールを行なった．

bit.ly も goo.gl と同様，API やメンテナンスページが充実しており実用的にも非常に使いやすい．API のリクエスト制限は 1 分間に 100 リクエスト，1 時間に 1,000 リクエストとなっている．bit.ly の短縮 URL の末尾に"+"を付加することでメンテナンスページにアクセスできる．

tinyurl.com もよく使われる短縮 URL の一つだが，API やメンテナンスページといった機能が存在しない．その代わり．preview.tinyurl.com/ 短縮 URL という URL にアクセスすることで，事前に短縮前の LongURL を確認することができる．goo.gl は tor やプロキシサーバを用いて IP を切り替えることで効率的に収集できたが，逆に tinyurl.com では tor を使うと収集できなくなるため注意が必要である．

ow.ly は使用するために会員登録が必要だが，API やドキュメントが充実しており使いやすい．会員登録が必須である分，実用的な使いやすさは goo.gl や bit.ly に劣る．API の制限は 1 秒 20 リクエスト，1 日 100,000 リクエストと多い．

is.gd はデフォルトでクッションページを表示される機能を持つ．クッションページの有無はユーザが変更することが可能である．また，API は LongURL を短縮するものしかないため，収集は Location ヘッダから行う．

t.co は，他の短縮 URL サービスと違い，Twitter 社が自社の SNS サービス内でのみ使用する短縮 URL である．ユーザが投稿したメッセージに URL が含まれている場合，自動的に t.co に短縮される．

第 4 章　結果

本章では，収集した短縮 URL の分析結果を示す．まず収集した短縮 URL の内訳を示した後，短縮されている FQDN や URI スキーム，悪性 URL の有無，多重に短縮された URL についての分析結果を述べる．

4.1　収集した短縮 URL

収集した短縮 URL の内約を表 4.1 に示す．

短縮 URL サービス	手法	収集数
goo.gl	総当たり (API)	13,705,868
bit.ly	総当たり (API)	860,534
	検索エンジン	30,829
	サービス内検索 (googlePlus)	83
tinyurl.com	総当たり (プレビューページ)	476,193
ow.ly	総当たり (HEAD リクエスト)	383,830
is.gd	総当たり (HEAD)	224,867
t.co	総当たり (HEAD)	317
	ツイッター API	444,349
	サービス内検索 (twitter)	877

表 4.1　収集した URL の内約

これらの結果はユニークな短縮 URL の個数となっており，一つの短縮 URL に一つの LongURL が存在する．t.co は，10,119,340 個の短縮 URL を収集したが，LongURL と対応付けられているものは 444,394 個だけであるため，その数値を記載している．

`goo.gl` は API を使うことで詳細な情報も取得することができるため API で収集を行なった． 13,705,868 回リクエストを送り， データを収集した． そのうち現在でも使われているものは 10,905,075 個であり， ステータスが REMOVED である削除済みのものが 2,800,793 個で

あった．短縮 URL サービスはしばしばその URL の永続性が懸念されるが，goo.gl では，一度短縮された URL は半永久的に存在し続けるとドキュメントで示されている．削除されるのは悪意のある URL であると判断された場合や，ユーザのプライバシーを侵害する可能性がある時だとも書かれている．しかしステータスには REMOVED 以外に MALWARE，PHISHING というものも存在する．そのステータスの明確な違いは分かっていない．

bit.ly は文字数 5 の空間を 1,601,200 リクエスト総当たりした． そのうち実在し収集に成功した短縮 URL は 860,534 個だった． bit.ly も API が存在するため今回は複数の API を用いて収集した．前述した通り API にはリクエスト制限が存在する上に， 1 アカウント 1 API しか与えられないため， 短期間でさらに大量に収集したい場合は head リクエストを用いる方法が良い．

tinyurl.com は HEAD リクエストを利用してもプレビューページを利用しても速度に変化がなかったためプレビューページを利用して収集した．

ow.ly, is.gd はそれぞれ 50,792 , 224,967 の URL を収集した． 数が少ないのは収集期間が短いためである．

t.co は他の短縮 URL とは違い，4 文字以下での存在数が非常に少なかったため，総当たりでの収集数は少ないものとなっている．t.co の短縮 URL のパスのランダム部分の長さはほとんどが 10 文字であった．

サービス内検索を用いた手法での収集は， googlePlus と Twitter 上で行なった． しかし，googlePlus は最近の投稿しか取得できないからか収集できる総数が少なく，Twitter も収集に時間がかかり効率が悪い．Twitter 上のツイートからの収集であれば，Twitter のストリーミング API を使う方法により比較的効率よく収集できる．ただし，Twitter のストリーミング API は新 API 公開と共に廃止される予定なので今後扱う場合には注意が必要である．

4.2 短縮された URL の FQDN

本研究では，短縮 URL の元となる LongURL の FQDN を調査した．FQDN の内約を見ることでおおよその用途を把握することができる．表 A.1 に goo.gl の短縮 URL における LongURL の FQDN を示す．上位の FQDN はほとんどが Facebook, Youtube, Google と有名サービスであることが見て取れる．また，1.1 節で述べたように google ドキュメントの URL も含まれており，その中には閲覧可能なものも存在する．そして，表には goo.gl や bit.ly のような短縮 URL のドメインも存在する．このドメインが存在する理由の一つは，短縮 URL が多重に短縮されているものがあるためである．さらにもう一つは，google のサービスである google map や photo は，共有 URL をなるべく短い URL にするような機能を持っており，その際の共有 URL のドメインは goo.gl のドメインになる．ただし，ランダム文字列部分の長さは 1 7 文字程度となっており，推測することは不可能な長さである．しかしながら，このよう

な共有 URL をさらに短縮しているものが見つかった．
ow.ly, is.gd の結果も同様の結果となった．こちらのデータは付録 A.1 に掲載する．

FQDN	個数	FQDN	個数
www.youtube.com	466,441	jumppath.com	60,934
apps.facebook.com	403,247	feedproxy.google.com	58,954
nnm.ru	372,029	www.sgrab.ru	54,756
twitter.com	285,154	www.es-noticia.com	52,817
www.facebook.com	253,874	www.flickr.com	47,616
www.google.com	140,182	www.trouw.nl	42,641
www.extremaduraaldia.com	101,298	maps.google.com	41,815
goo.gl	94,085	www.extremaduraaldia.tv	40,629
www.jolatefri.com	79,980	www.orkut.com.br	37,903
bit.ly	69,731	docs.google.com	37,223

表 4.2　goo.gl の longURL における上位 20FQDN

収集した bit.ly の短縮 URL の LongURL の FQDN の上位 20 個を表 4.3 に示す．上位サイトは goo.gl と同様，twitter や google など日本でも有名なサイトが多いことが見て取れる．

FQDN	個数	FQDN	個数
twitter.com	33,783	my.saynow.com	7,336
friends.myspace.com	24,585	www.saynow.com	5,942
www.youtube.com	21,927	myloc.me	4,577
lolquiz.com	21,015	mp3.3gp.fm	4,231
www.facebook.com	20,279	twitrss.dyndns.org	4,204
fun140.com	20,240	www1.tour.ne.jp	3,938
feedproxy.google.com	17,656	rover.ebay.com	3,740
justsignal.com	16,464	www.etsy.com	3,651
news.google.com	9,924	www.booking.com	3,290
www.google.com	8,109	blip.fm:80	3,087

表 4.3　bit.ly の longURL における上位 20FQDN

収集した t.co の短縮 URL の LongURL の FQDN の上位 20 個を表 4.4 に示す．t.co のデータは goo.gl や bit.ly と違い，ツイート上から収集したものであるため，実際にユーザの目に触れた可能性の高い短縮 URL が集まっている．故に，このデータに悪性 URL やプライバシー情報の漏洩の恐れがある場合は特に危険だと言える．表を見ると，総数のほとんどを

twitter.com のドメインが占めていることがわかる．また，短縮 URL のドメインが多いことも見て取れる．このことからも，1.1 節で述べたように SNS サービスで短縮 URL が用いられることがわかる．さらに，goo.gl や bit.ly に比べて文書等の共有 URL は少なかった．bit.ly と t.co の docs.google.com の数を比較すると，bit.ly では約 86 万個の短縮 URL のうちの 275 個 (0.03%) の LongURL が docs.google.com であったが，t.co は約 44 万個のうち 20 個 (0.003%) と 10 分の 1 程の数しか存在しなかった．

FQDN	個数	FQDN	個数
twitter.com	381,621	ow.ly	800
bit.ly	6,248	lin.ee	550
goo.gl	2,904	tinyurl.com	386
dlvr.it	2,697	v.ht	379
ift.tt	2,464	www.instagram.com	313
cards.twitter.com	1,318	www.youtube.com	295
bnent.jp	1,096	amzn.to	278
youtu.be	891	www.bigo.tv	247
buff.ly	842	dld.bz	237
goat.app.link	839	www.liveme.com	216

表 4.4　t,co の longURL における上位 20FQDN

4.3　短縮された URL の遷移先のファイル

短縮 URL の LongURL が示すファイルの拡張子を表 4.5 に示す．html や php などの Web サーバのコンテンツの他，png や jpg，pdf などのコンテンツも見られる．ここに記載していないものでは rar, や zip などの圧縮ファイルや wmv などの音声ファイルも見られた．

4.4　短縮された URL の URI スキーム

短縮された URL の URI スキームについて調査した．　結果を表 4.6 に示す．goo.gl は http などのプロトコルは許しているが，data: や mailto: などの短縮は許していないため，4 種類のみとなっているが，bit.ly や tinyurl.com，is.gd は多くの URI スキームが見て取れる．中には 3.2.3 項で示したような dataURI スキームも見られた．

goo.gl		bit.ly	
ファイルタイプ	個数	ファイルタイプ	個数
html	1,892,606	html	138,703
php	1,102,618	php	102,509
jpg	275,537	cfm	29,129
aspx	245,839	aspx	23,835
htm	244,170	asp	15,161
asp	140,118	htm	11,883
jsp	97,506	jpg	5,166
shtml	82,755	shtml	4,069
pdf	33,777	jsp	1,805
png	33,642	cgi	1,226

表 4.5　goo.gl の longURL における上位 20 のファイルタイプ

サービス	発見された URI スキーム
goo.gl	http, https, ftp, rtsp
bit.ly	http, https, file, ftp, chrome, feed, res, mms, itms, market
tinyurl.com	http, https, file, ftp, res, mms, mailto, news, about, aim, rtsp, data, ms-help, telnet, view-source, ed2k, callto, hxxp, gopher, view-source, itms
ow.ly	http, https
is.gd	http, https, ftp, mailto, feed, file, chrome, mms, rtsp, irc, ed2k, webcal, teamspeak， xmpp, aim, telnet, gtalk
t.co	http, https

表 4.6　短縮された URL に使われていた URI スキーム

4.5　悪性 URL の短縮 URL

我々は，収集した短縮 URL の中に悪性 URL を短縮しているものがないか調査を行なった．悪性 URL を判定するためにはフリーのブラックリストと VirusTotal を使用した．VirusTotal はファイル名やファイルのハッシュ値を入力するとそのファイルがマルウェアであるかどうかの判定結果を返すサービスである．VirusTotal の機能には，Web サイトの URL を入力することでそのサイトが悪性であるか否かを判定できるサービスもあり，今回はこの機能を利用した．また，VirusTotal は API も提供しており，無料の API は 1 分間に 4 リクエストの制限で使う

ことが可能である．今回はこの API を複数取得して利用したが，それでもリクエストの制限が厳しいため，収集した URL 全てに対してスキャン及び検査結果の取得を行うことは困難であった．そこで，本研究ではフリーのブラックリストを用い，集めた URL から悪性ドメインを持つものを抽出し，抽出したドメインを VirusTotal で検査した．ブラックリストは，spamhaus や phishtank などのブラックリストが載っている malwaredomains.com のドメインリスト，malwaredomainlist.com のドメインリスト，Blacklists UT1 の adult, phishing, malware に分類されるドメインリスト，openphish.com のフィッシングサイトのドメインリストを使用している [11, 12, 13, 14]． 分析結果を表 4.7 に示す．

サービス	Malware & Phishing	VirusTotal	Including Adult	VirusTotal (Including Adult)
goo.gl	41	18	2,062	141
bit.ly	8	3	399	19
tinyurl.com	1	0	110	3
is.gd	1	1	130	5
ow.ly	1	0	125	7
t.co	2	0	93	3

表 4.7　短縮 URL の元の LongURL のうちブラックリストに含まれていた数．悪性サイト及びフィッシングサイトの結果と，それに加えてアダルトコンテンツの配信サイトも考慮した際の数，VirusTotal での検査後の数を示している．

4.6　多重の短縮 URL の LongURL

4.2 節でも示したように，多重に短縮されている短縮 URL が存在する．本研究では，この短縮 URL の最終的な到達先を調査した．母数の多い goo.gl,bit.ly, tinyurl.com において，多重短縮 URL の数とその深さ，さらに多重に短縮された URL を VirusTotal にかけた結果を表 4.8 に示す．

bit.ly では，多重に短縮された URL の元の LongURL だけで 18 個の悪性 URL が発見された．goo.gl では，別々の短縮 URL サービスで相互の短縮 URL が短縮されており無限にリダイレクトが続くというものが存在した．例えば，goo.gl/NUFc が tinyurl.com/qxe123 を短縮しており，tinyurl.com/qxe123 が goo.gl/NUFc を短縮しているという具合である．これは短縮 URL サービスのカスタム URL サービスを利用したものと考えられる．このような挙動をリダイレクトループと呼び，実装ミスや悪用目的で発生することがあるが，短縮 URL で実現されている意図はわからない．実ブラウザでアクセスした場合，リダイレクトループを検知して途中でアクセスが停止する．

また，多重短縮 URL の深さの最も深いものは 327 重という深さだった．

サービス	goo.gl	bit.ly	tinyurl	深さ最大	Blacklist+VirusTotal
goo.gl	77,065	67,257	1,040	327	2
bit.ly	9	501	1,411	3	0
tinyurl	0	0	9,673	27	0

表 4.8　多重に短縮された URL

第 5 章　議論

本章では，本研究における背景，目的，主な貢献について示す．また，本論文の構成について示す．

5.1　結果に対する考察

FQDN の調査では，短縮 URL のおおよその利用方法を把握することができた．1.1 節で示したように，Google ドキュメントやクラウドストレージの共有 URL を短縮するような用途も多く見て取れた．その文書や写真の中から無作為に選んだものを閲覧してみたところ，稀に家族写真のようなプライベートな写真にアクセスできてしまうことが分かった．表 4.3 で示した FQDN の `blip.fm` という URL を Virus Total でスキャンしたところ，約 6 年前に一部のアンチウィルスソフトが攻撃をブロックしたことがあるというコメントが存在した．現在でも一部の解析サイトでは suspicious という扱いになっている．悪性 URL が短縮されたことがある例が存在するというだけでなく，そのドメインを持つサイトが今回収集した短縮 URL の中では上位 20 位に入るというのは，一つの大きな成果であると考える．

表 4.4 では `t.co` の LongURL の FQDN の分析結果を示した．この FQDN からは，SNS で短縮 URL が多く利用されていることが見て取れた．また，投稿自体を誰でもできる性質から，プライベートな文書や写真などの共有はほとんどされていないことが分かった．

FQDN の調査の結果，3.2.1 項で述べたように，短縮 URL サービスにて短縮すべきではない共有 URL の短縮が見つかったため，プライバシー情報の漏洩の可能性が高いことが分かった．また，悪性 URL のドメインも多く見られたため，攻撃者による悪用の危険性も高いことが分かった．他に多重に短縮された短縮 URL が発見されたが，多重短縮 URL については後述する．

4.4 節では， URI スキームから利用用途を調査した．http や https 以外の URI スキームを短縮している例も存在した．短縮 URL サービスは URL の形式ではない文字列や URL スキームをサポートしているものも多いため，それ自体は自然である．URI スキームに対応していないサービスもあり，例えば ux.nu では mailto スキームは「危険な URL」と判断され，短縮することができない．しかし，mailto スキームは直接メールを送るわけではなく，指定のアドレスを補完した状態でメーラーを開く機能であるため，攻撃用とのものとは考えにくい．これらの

URI スキームは，一般ユーザが素直に HTML 等に埋め込むとスパムの標的となりやすい (web 上のメールアドレスの収集で標的になる) ため短縮 URL を用いたのではないかと考える．ただし，表 4.6 の `tinyurl.com` で使われている dataURI スキームには注意が必要である．過去に dataURI スキームを短縮した短縮 URL を用いて正規サイトに偽装する攻撃が存在した [9]. dataURI スキームを用いたアクセス先の偽装は，短縮 URL に限らず存在する [10]．本研究内で目にした dataURI スキームはいずれも bmp 画像を定義していた．dataURI で画像を定義することでブラウザでの描画を高速化できることが知られており，その場合は画像データを直接記述する必要がある．そのために長くなった URI を短縮するために短縮 URL サービスを用いたのではないかと考える．

URI スキームの調査の結果，過去に悪用された dataURI スキームなど多くの URI スキームの利用が見られたため，攻撃者の利用にも注意する必要があると言える．あるいは，必要がなければサービス側は http や http 以外の URI スキームは短縮を拒否しても良いと考える．

4.5 節では悪性 URL の短縮 URL について調査した．いずれのサービスでも悪性 URL を短縮した短縮 URL が存在したため，過去に悪性 URL の短縮に使われていたと言える．Google の API では短縮 URL のステータスとして MALWARE であるか，あるいは PHISHING であるか等を確認することができるが，今回 VirusTotal で悪性と判断されたもので Google では悪性では無いと判断されているものも存在した．これは Web サイトの検査サービスによって悪性の基準が異なることによるものであると考えられる．Google は Google Safe Browsing という悪性サイトの検査サービスを公開していることもあり信頼できるが，それでも他のサービスで悪性と判断しているものをスルーしていることがあるということは念頭に置くべきである．また，短縮 URL サービスではアダルトサイトなどの URL の短縮にも多く使われていることがわかった．

Twitter の投稿から収集した約 44 万の t.co の短縮 URL にもブラックリストと VirusTotal にて悪性であると判断される URL が存在することが明らかになった．この 44 万は Twitter のストリーム API で収集した URL の一部であるため，実際にはより多くの悪性 URL が存在していたと考えられる．発見された 3 つの悪性 URL のうち，1 つは Twitter が悪性 URL を警告する際に使われる URL である `https://twitter.com/safety/unsafe_link_warning` のクエリにも存在しており，Twitter 側で悪性であることを検知していた．しかし残りの 2 つの悪性 URL は警告で使われる URL が存在していなかったため，Twitter 側で検知していなかったのではないかと考える．このように，多くのユーザが存在しているサービスでも，サービス側で検知されていない可能性のある悪性 URL が存在することが明らかになった．

悪性 URL の調査では，実際に悪性 URL が存在し，なおかつサービスによって悪性と判断しているものとそうでないものが実際に存在することが明らかになった．サービス側で網羅できない可能性がある以上，ユーザ側も注意する必要がある．

4.6 節では，多重に短縮された短縮 URL の分析結果を示した．Twitter の t.co のように自社サービス内で特別な短縮 URL を利用しているものが多重になることは考えられるが，一般の

ユーザが扱える goo.gl や bit.ly でも多重に短縮された URL が存在した．我々は 3.2.5 項にて，仮説として，リダイレクトを多段に含むことができることからドライブバイダウンロード攻撃などの導線として使われている可能性を示した．結果として，悪性 URL は存在したものの，多重短縮 URL の方が通常の短縮 URL に比べて悪性 URL が多いという結果は得られなかった．また，今回の調査では最大で 327 重という明らかに普通ではない多重 URL が存在した．しかし，最終的に到達するサイトを VirusTotal で検査したが，悪性とは判断されなかった．そのため, 多重短縮 URL の深さと悪性サイトの数には相関がないことが明らかになった. ただし，その最終的に到達するサイトには現在アクセスできずウェブアーカイブにも残っていないことから，過去に攻撃に利用されていた可能性は考えられる．正常な利用法での仮説としては，Twitter などの投稿サービスで独自の短縮 URL サービスを持たない者が Google や Bitly の API を使い，そのユーザが貼り付けた短縮 URL がさらに短縮されただけであるということが考えられる．しかしその場合は 2 重や 3 重止まりであり，不自然に多重になっているものの存在に疑問が残る．

多重短縮 URL の調査の過程で, 多重短縮 URL が悪性だった場合, 短縮 URL サービスは最終到達先の悪性サイトを検知できていない可能性があることが分かった．例えば, goo.gl にて短縮された http://tinyurl.com/42hmhr という短縮 URL は, 最終的に http://nobrain.dk/ というフィッシングサイトへリダイレクトされる．しかし，goo.gl では http://tinyurl.com/42hmhr が悪性であるとは判断されていなかった．また，tinyurl.com でも悪性検知されていないため，クッションページなどの設定がされていない場合直接フィッシングサイトに飛ばされてしまう．当初の仮説とは少し異なるが，有名な短縮 URL サービス (及び短縮 URL を利用する SNS 等) を使う前に，悪性検知の甘い別の短縮 URL でラップすることで，短縮 URL サービスの悪性検知を回避できる可能性があることが分かった．

5.2 制限

全てのドキュメントやクラウドサービスの共有 URL がプライバシーの漏洩に繋がるものであるかはわからない．今回いくつかの共有 URL において，プライバシーに関わると思われる URL を確認できたが，中には公開情報と思われる文書なども多く存在するが，その全てを確認することはできていない．

また，本研究で分析した URL は総数からすればほんの一部である．goo.gl において 1,000 万以上の URL を集めたが，まだ多くの短縮 URL が存在する．各サービスが表に示していない実態を調査するには，継続的な収集および解析が必要である．

短縮 URL サービスは他にも多くのものが存在するが，その全てを網羅できていない．マイナーなものは一般ユーザが使うことは少なく，攻撃者も一般ユーザに信頼されやすい有名な短縮サービスを利用する可能性が高いという考えからか，既存研究でもマイナーな短縮 URL をア

クティブに調査しているものは存在しない．しかし，有名なサービスはその分悪性 URL の検知等の対策を取っており，悪性サイトの URL を登録してもすぐに悪性判定され対策されてしまうだろう．その点マイナーで検知が緩い短縮サービスを使えば検知されにくいと考える攻撃者が存在する可能性もあるだろう．そのため，私はマイナーな短縮 URL サービスを今後調査する価値があると考える．

5.3 対策

主に 3 つの対策が考えられる．

1 つ目は，ユーザが短縮 URL を正しく活用するということである．本研究で，ユーザのプライバシー情報 (画像や文書) にアクセスできる恐れのある短縮 URL が未だに存在していることが明らかになった．長いことで安全性が担保される共有 URL を短縮することは控えるべきである．どうしても短縮する必要のある場合は，きちんとアクセス権限を設定した共有 URL を用いるべきである．

2 つ目は，短縮 URL サービス側が悪性 URL をなるべくリアルタイムに，また定期的に検査し，悪性 URL を除去することである．本研究にて，他のブラックリストや検査サービスで悪性であると判断されていても，短縮 URL サービス側では悪性であると判断されていない URL の存在が明らかになった．また，別の短縮 URL サービスで短縮された悪性 URL をさらに別の短縮 URL サービスで短縮した場合，悪性と検知されない場合があることも明らかになった．そのため，短縮 URL サービスではより悪性検知を正しく行うようにすべきである．また，URL スキームに関しては https や http 以外の data スキームや mailto スキームは短縮の対象から外しても良いのではないかと考える．

3 つ目は，ユーザが短縮 URL をクリックする際に十分注意することである．前述したように URL によっては，あるサービスでは悪性と判断されていないが，他の短縮サービスや悪性検査サービスでは悪性であると判断されている場合がある．ユーザはツールやブラウザ拡張を使うことで，短縮 URL の遷移先 URL を事前に確認することができる．

5.4 今後の課題

5.4.1 より多くの短縮 URL サービスの調査

本研究では 6 つの短縮 URL サービスについて調査したが，他にも多くのサービスが存在する． 本研究では，本稿に示した 6 つの短縮 URL サービス以外に，ux.nu, tiny.cc, v.gd, bit.do, nsfw.in, t.cn, dwz.in, tr.im, clc.li という短縮 URL に対しては HEAD リクエストやクッションページを利用して短縮 URL の expand を行えることを確認している．誤用や悪用の危険がないか調査するためには，このような短縮 URL サービスも調査すべきである．

5.4.2 リアルな攻撃の発見

既存研究でも，短縮 URL からセンシティブな情報にアクセスするという攻撃が指摘されているが，その攻撃が本当に存在するかどうかは確認されていない．例えば，あえてドキュメントの共有 URL を短縮しておいて，そのドキュメントへのアクセスが存在するかどうかを調べる方法や，カスタム URL 機能があるサービスを利用してあえてわかりやすい文字列を URL として利用することで囮の短縮 URL を作り調査をする方法が考えられる．我々の研究では「短縮 URL ハニーポット」と呼称し，今後の研究での調査を検討している．ハニーポットとは，あえて攻撃しやすいサーバ (囮のサーバ) を用意し，攻撃者の攻撃を待ち受けて攻撃者がどのような行動をするか監視する技術である．短縮 URL ハニーポットは従来のハニーポットの考え方を短縮 URL サービスを対象に扱えるよう取り入れたものである．短縮 URL ハニーポットではハニートークンと呼ばれる技術を使うことも効果的であると考えている．ハニートークンとは，囮となるデータ (今回ならドキュメント) にあえて記載する「自作の囮サービスの URL」と「囮ユーザの ID, パスワード」等の情報，あるいはその情報を使用して攻撃者の動向を監視する技術のことである．もしもこの囮サービスに囮ユーザでのアクセスが確認できれば，それはハニートークンを盗み見た攻撃者がアクセスしたということに他ならないため，攻撃者の行動を追跡することが可能だ．

第 6 章　まとめ

本研究では，短縮 URL サービスの誤用や悪用の実態を明らかにするため，短縮 URL サービスでどのような URL が短縮の対象となっているかを調査した．そのために，6 つの短縮 URL サービスの短縮 URL(Short URL) 及び対となる元の長い URL(Long URL) を収集するためのクローラーを作成し，収集を行った．最終的に約 1,600 万の短縮 URL を収集・分析し，悪用される危険のある URL の有無や，有害な URL の有無を調査した．

その結果，未だにアクセスが有効な文書や画像などの共有 URL へのアクセスが短縮 URL 経由で可能であることが明らかになった．そのような共有 URL にはプライバシーに関わる情報が含まれているものも存在した．また，悪性サイトの URL も数多く短縮されていることを示した．その中には短縮 URL サービス側で悪性であると判断されていないものも存在することが明らかになった．さらに，短縮 URL を多重にかけているような多重短縮 URL について掘り下げて調査を行った． 多重短縮 URL は攻撃者の用意した悪性サイトへのリダイレクトとして扱われているのではないかという仮説の元調査を行った．結果，多重短縮 URL の最終的な到達先 URL のいくつかは悪性サイトであると判断されたが，不自然に多重に短縮されているにも関わらずクリーンなサイトも存在したため，仮説を決定づけるまでには至らなかった．多重短縮 URL は最大で 327 重もの深さをもつものも存在した．多重短縮 URL の分析の過程で，悪性 URL を短縮 URL サービスで短縮した後，さらに別の短縮 URL サービスにかけた場合，悪性 URL を検知しない場合があることが明らかになった．

主に 3 つの対策が考えられ，1 つ目は，ユーザが短縮 URL を正しく活用するということである．長いことで安全性が担保される共有 URL を短縮することは控えるべきである．2 つ目は，短縮 URL サービス側が悪性 URL をなるべきリアルタイムに，また定期的に検査し，悪性 URL を除去することである．3 つ目は，ユーザが短縮 URL をクリックする際に十分注意することである．

今後の課題として，本研究で指摘したような総当たりにより本来は知り得ない共有 URL へアクセスできる短縮 URL を特定するという攻撃の存在を調査することをあげた．その手法として短縮 URL ハニーポット，短縮 URL ハニートークンという手法を考えた．この攻撃の存在確認は既存研究では成されていないため，今後進めていく必要があると考える．

参考文献

[1] Er Neumann, Johannes Barnickel, and Ulrike Meyer. Security and privacy implications of url shortening services. In *Proceedings of the Workshop on Web 2.0 Security and Privacy*, 2010.

[2] List of url Shorteners. https://bit.do/list-of-url-shorteners.php.

[3] List of 230 Free URL Shorteners Services. https://www.techmaish.com/list-of-230-free-url-shorteners-services/.

[4] Vitaly Shmatikov Martin Georgiev. Gone in six characters: Short urls considered harmful for cloud service. 2010.

[5] Federico Maggi, Alessandro Frossi, Stefano Zanero, Gianluca Stringhini, Brett Stone-Gross, Christopher Kruegel, and Giovanni Vigna. Two years of short urls internet measurement: Security threats and countermeasures. In *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13, pp. 861–872, New York, NY, USA, 2013. ACM.

[6] Bitly api documentation and resources. https://dev.bitly.com.

[7] URL Shortener API - Google Developers. https://developers.google.com/url-shortener/v1/getting_started.

[8] ウェブ検索の精度を高める. https://support.google.com/websearch/answer/2466433.

[9] Data uri scheme を用いた悪性短縮 url の例. http://d.hatena.ne.jp/Kango/20170206/1486351285.

[10] Data uri scheme を用いたアクセス先の偽装例. http://gigazine.net/news/20170112-gmail-embed-phishing-attack/.

[11] Dns-bh – malware domain blocklist by riskanalytics. http://www.malwaredomains.com/?page_id=66.

[12] Blacklists UT1. https://www.netnanny.com/blog/where-to-get-a-good-internet-blacklist/.

[13] malwaredomainlist.com. http://www.malwaredomainlist.com/hostslist/hosts.txt.

[14] openphish.com. http://openphish.com/feed.txt.

付録 A　その他の短縮 URL の情報

A.1　その他の短縮 URL の FQDN

FQDN	個数	FQDN	個数
groups.google.com	29,444	www.compaq.com	7,001
story.news.yahoo.com	22,226	www.geocities.com	6,479
cgi.ebay.com	21,222	www.upi.com	6,177
www.amazon.com	18,335	maps.yahoo.com	6,064
tinyurl.com	13,728	photos.groups.yahoo.com	4,689
www.google.com	13,551	groups.msn.com	4,562
www.reuters.com	12,707	support.microsoft.com	4,487
www.mapquest.com	11,866	www.microsoft.com	3,905
www.nytimes.com	7,815	e.ccialerts.com	3,655
www.jpost.com	7,118	news.bbc.co.uk	3,622

表 A.1　tinyurl.com の longURL における上位 20FQDN

FQDN	個数	FQDN	個数
www.youtube.com	6,323	www.last.fm	1,945
flickr.com	5,577	maplink.uol.com.br	1,936
github.com	4,461	www.amazon.com	1,840
g1.globo.com	3,372	www.nytimes.com	1,672
www.peabirus.com.br	2,891	maps.google.com	1,629
www.orkut.com	2,590	info.abril.com.br	1,548
euro2,008.phewse.com	2,556	youtube.com	1,383
twitter.com	2,229	www.palmnet.me.uk	1,324
en.wikipedia.org	2,210	www.google.com	1,221
earthquake.usgs.gov	2,125	www.estadao.com.br	1,188

表 A.2　is.gd の longURL における上位 20FQDN

FQDN	個数	FQDN	個数
snipurl.com	20,865	us.rd.yahoo.com	3,086
www.youtube.com	8,113	www.amazon.com	3,058
movielanka.net	7,089	www.creatingwebsuccess.com	2,735
twitter.com	6,152	www.mailchimp.com	2,663
sn.im	5,147	mashable.com	2,559
www.pettalez.com	2,478	www.pettalez.com	2,478
www.facebook.com	3,970	www.nytimes.com	2,396
ow.li	3,914	news.bbc.co.uk	2,247
alexking.org	3,300	www.flickr.com	2,004
www.fixya.com	3,233	news.google.com	1,847

表 A.3　ow.ly の longURL における上位 20FQDN.snipurl や sn.im は別の短縮 URL サービスである．snipurl は virustotal では悪性判定される．