1W-07

複数の評価サービスの統合による 短縮 URL の安全性提示手法の提案

藤根　麻羽†　小倉　加奈代†　ベッド　バハドゥール　ビスタ†　高田　豊雄†

岩手県立大学ソフトウェア情報学部†

1. はじめに

Twitter[1]や Facebook[2]を代表とするソーシャルネットワークサービス(以下，SNS)は 2000 年代後半に登場し，現在幅広い年代のユーザに利用されている．SNS の投稿サービスには，冗長性を排除する目的で文字数に制限が設けられているものがある．投稿時の文字数制限への対処法として，短縮 URL サービスが利用されるようになった．短縮 URL サービスとは，対象の URL と対になる簡素な URL を生成するサービスであり，goo.gl[3]や bit.ly[4]が代表例である．冗長性の排除の他，アクセス統計や QR コード生成に利用することが可能なため，企業の SNS でのキャンペーン活動の際に利用されている．だが，短縮 URL サービスを利用すると行先 URL のドメイン名が変換されるため，URL の難読化が起きる．これにより行先サイトが正規のサイトであるかどうかを判別することができず，Drive-by-Download 攻撃[5]やフィッシングサイトへの誘導に利用される被害が起こっている[6]．ユーザは，行先サイトが安全であるか否かを URL 毎に判断する必要が生じる．本稿では，短縮 URL を展開した行先 URL を複数の安全性評価サービスによって検査し，その結果を統計的手法により統合，3 段階の安全性評価結果として提示する手法を提案し，その有用性を評価する．

2. 関連研究

本章では，サイトの安全性評価手法及びサイトの評価と提示方法について記述する．

坂松[7]はサイトの安全性と重要度に応じたパスワード管理ツールに関する研究の中で，サイトの安全性を評価している．具体的には，(1)ウイルス検査，(2)ブラックリスト判定，(3)安全性の評価ツールを利用している．各項目の検査結果を単純加算したものを 3 段階の評価にあてはめ，ユーザに安全性評価結果として提示している．坂松の研究においてサイトの安全性をユーザに提示する点が本研究の目的と同じであるため，本研究では坂松の手法を一部採用する．

國分ら[8]は SNS における悪性 URL の分析と防御に関する研究の中で，URL の最終アクセス先サイトの安全性をオンデマンド検査し，評価情報を提示する手法を提案している．本研究と同様に，複数の評価ツールを用いた評価結果情報を提示しているが，その提示方法やユーザが必要とする情報の取捨については言及されていない．そこで，評価ツールごとに重み付け値を設定し，最終的な安全性評価を提示することで，ユーザはより Web サイトにアクセスするか否かの選択が容易になると考える．

3. 安全性提示手法の提案

本提案では，短縮 URL 展開後の行先サイト URL を検査対象として安全性を評価する．ユーザは該当サイトの安全性を検査した安全性評価結果をもとにアクセスするか否かを決定する．事前に決定した各評価ツール別の重み付け係数を行先サイト URL の検査結果に適用し，合算する．提案手法構成図を図 1 に示す．

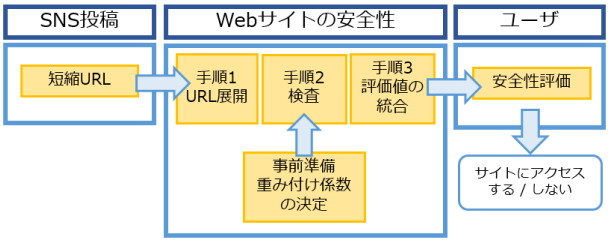

図 1．提案手法構成図

3.1 重み付け係数の決定(事前準備)

事前準備として，評価結果の統合を行う前に評価ツールの重み付け係数を決定する．使用する安全性評価サービスとして Google が提供する Google Safe Browsing やブラックリストのデータベースを保持する SPAMHAUS 等を複数利用する．これらの安全性評価サービスを予備調査したところ，検査結果にゆれが生じることが明らかになった．この理由として各評価ツールが独自のアルゴリズムや判断基準を設定し，Web サイトのクローリングを行なっていることが考えられる．評価結果の偏りを考慮する目的で複数の評価ツールを利用する．重み付け係数は線

Proposal Security Presentation Method of Shortened URL by Integrating Multiple Evaluation Services
†Iwate Prefectural University, Faculty of Software and Information Science

形回帰モデルを採用し，分析した値を各評価ツールに設定する．

3.2 URL 展開(手順 1)

対象となる短縮 URL を展開し，行先 Web サイトの URL を取得する．実際にはユーザが短縮 URL へのマウスオーバー動作によって Web サイトの安全性評価を開始する．

3.3 検査(手順 2)

取得した URL を評価ツールによりフィッシングサイト等の悪性サイトである疑いがないかを検査する．異常を検知した場合は 0，正常と判断した場合は 1 として，評価ツールごとに検査結果を収集する．事前に決定していた重み付け係数を各調査結果に乗算し補正後の評価結果とする．

3.4 評価値の統合(手順 3)

手順 2 で得られた補正後の評価結果を合算し，合算した値を高・中・低の 3 段階評価に当てはめ，最終的な評価結果としてユーザに提示する．

4. 評価

採用する評価ツールの重み付け係数の妥当性，及びユーザに提示する 3 段階評価の妥当性の 2 つの観点から評価する．Google Safe Browsing 等のサイトの安全性評価ツールを 66 件使用した．本稿では悪性サイトをフィッシングやマルウェア配布サイトなどの悪意のある行為を行うサイトと定義する．良性サイトを ALEXA 及び dmodztool.com から，悪性サイトを Malware Domain List 及び Phishtank からサンプルとして取得し，Web サイトの安全性評価の妥当性を検証する．評価には良性及び悪性サイトの URL それぞれ 500 件ずつ，合計 1,000 件取得した．これらのサイト URL を評価ツールによって検査した結果を次節以降の評価に利用する．分析用データセットとして，Twitter Streaming API より実際の Twitter 投稿を収集し，展開した URL を使用する．

4.1 重み付け係数の有効性

章 3.1 で述べた評価ツールの重み付け係数処理に関して，良性及び悪性サイト群を用い，真偽の判断が行えるかを検証する．統計的手法として線形回帰分析を採用する． 前述の 1,000 件の良性及び悪性 URL をランダムに並べかえ，そのうち 700 件を訓練用データとして，300 件をテストデータとして用いる．訓練用データに線形回帰分析を繰り返し適用することで適切な重み付け値を得る．訓練用データを評価ツールにより検査し，1,000 回の試行回数で分析を行った．重み付け係数は学習率を 0.001 に初期値として設定し，最適化アルゴリズムには Adam 法を用いた．評価の結果，テストデータに対して正答率 98.84%が得られた．

4.2 3 段階評価の妥当性

章 3.4 で述べた安全性評価値について，閾値の妥当性を評価対象とする．閾値の設定評価には悪性サイト群と分析用データセットを用い，対象となる悪性 URL が SNS 投稿に含まれていると仮定し，Web サイトの安全性を評価する．

5. おわりに

本稿では，複数の評価サービスの統合による短縮 URL の安全性提示手法を提案した．今後は評価ツールの重み付けについて訓練用データ及びテストデータの件数を増やし分析，評価するとともに，安全性提示時のユーザインタフェースの検討や被験者による実証実験により本提案手法の有用性を検証する．また，本研究では SNS 投稿に含まれる URL を対象として安全性を評価したが，今後は，投稿メッセージ本体および投稿者の信頼性等を本提案と併せることでより精度の高い安全性判断指標の確立を目指す．

謝辞

本研究は JSPS 科研費 16K01025 の助成を受けたものである．

参考文献

[1] Twitter, available from <https://twitter.com/> (accessed 2017-12-24).
[2] Facebook, available from < https://www.facebook.com/> (accessed 2017-12-24).
[3] Google URL Shortener, available from <https://goo.gl/> (accessed 2017-02-01).
[4] Bitly | URL Shortener and Link Management Platform, available from <https://bitly.com/> (accessed 2017-04-30).
[5] Trojanized Propaganda App Uses Twitter to Infect, Spy on Terrorist Sympathizers | McAfee Blogs, available from <https://securingtomorrow.mcafee.com/mcafee-labs/trojanized-propaganda-app-uses-twitter-to-infect-spy-on-terrorist-sympathizers/> (accessed 2017-05-28).
[6] Apple Credentials | McAfee Blogs, available from<https://securingtomorrow.mcafee.com/mcafee-labs/active-ios-smishing-campaign-stealing-apple-credentials/> (accessed 2017-05-28).
[7] 坂松春香: サイトの安全性と重要度に応じたパスワード管理ツールの検討, 岩手県立大学 2015 年度博士前期課程(ソフトウェア情報学)論文, 2015.
[8] 國分佑太朗, 中村章人: SNS における悪性 URL の分析と防御, 社会情報学会(SSI)2017, オンライン入手先 <http://gmshattori.komazawa-u.ac.jp/ssi2017/wp-content/uploads/2017/03/30.pdf> (参照 2017-12-24).