

商品画像検索による偽ショッピングサイト URL 収集方式の 提案と実態調査

吉井 優輝^{1,*} 川村 慎太郎¹ 中村 渚¹ 田中 秀一¹
安田 真悟¹ 井上 大介¹

概要: フィッシングサイトの1つとして、正規のショッピングサイトを模倣した偽ショッピングサイトの存在が挙げられる。このようなサイトは、購入者の金銭、クレジットカード情報や個人情報の搾取を目的としている。また正規ショッピングサイトなどの掲載商品画像を無断転載していることも確認されている。フィッシングサイトは、正規サイトに不正アクセスしリダイレクト用スクリプトを仕込み踏み台とすることで、正規サイトにアクセスしたユーザを誘導する。同時に Web 検索エンジンの検索結果を不正に操作する SEO ポイズニングにより、ユーザのフィッシングサイトへのアクセスがより増加しやすくなる。上記の背景から、NICT ではユーザに Web センサとしてタチコマ・セキュリティ・エージェントと呼称するアプリケーションを配布し、Web アクセスログを収集し、Web 媒介型攻撃の観測・分析を実施してきた。本取り組みは WarpDrive プロジェクトと呼称し、ユーザによる Web アクセスを主としたデータ収集を通じて受動的な Web 媒介型攻撃観測を可能とする。フィッシングサイトとははじめとする悪性サイトは出現や削除のスパンが短いことからより能動的な悪性サイト情報収集を実施することで、悪性サイトの傾向分析やより多くの URL 情報獲得が可能と考えられる。本研究では、多くの悪性サイト URL を能動的に収集し解析性能向上を目的とする。ショッピングサイトでの商品閲覧時はほぼ確実に商品画像を閲覧すること、多くの偽ショッピングサイトでは同じ商品画像の使いまわしが行われている。そこで、本稿では偽ショッピングサイトを対象に掲載された商品画像をシードとし、Google Cloud Vision API を用いた類似画像掲載サイトの URL 収集の実施、リダイレクト前・リダイレクト後の URL の収集の方式を提案する。実際に、偽ショッピングサイトと疑われるサイトの商品画像 426 枚を用いて実態調査を実施したため報告する。

キーワード: Web 媒介型攻撃観測, Web セキュリティ, サイバーセキュリティ, 悪性サイト

Proposal and Empirical Study on Fake Shopping Website URL Collection via Product Image Search

Masaki Yoshii^{1,*} Shintaro Kawamura¹ Nagisa Nakamura¹ Hidekazu Tanaka¹
Shingo Yasuda¹ Daisuke Inoue¹

Abstract: Among the various forms of phishing websites, one prominent example is fake shopping websites that imitate legitimate e-commerce platforms. These fraudulent sites are primarily designed to exploit consumers by stealing money, credit card details, and personal information. It has also been confirmed that such sites frequently engage in the unauthorized reproduction of product images originally published on legitimate shopping platforms. Furthermore, phishing websites often lure users by embedding redirect scripts into compromised legitimate sites, thereby turning them into stepping stones. In addition, by manipulating search engine rankings through SEO poisoning, the likelihood of unsuspecting users being directed to these malicious websites is further increased. To address these threats, the National Institute of Information and Communications Technology (NICT) has developed and distributed an application called the Tachikoma Security Agent, which functions as a Web sensor for users. By collecting web access logs, NICT has been able to observe and analyze Web-based attacks as part of an initiative known as the WarpDrive Project. This project primarily relies on passive data collection from user web activity, thereby enabling large-scale monitoring of Web-mediated cyber threats. However, because phishing and other malicious websites tend to emerge and disappear within short timeframes, a more active approach to collecting malicious site information is expected to enable more effective trend analysis and the acquisition of a wider range of URL data. In this study, we aim to actively collect a large number of malicious website URLs to enhance analytical performance. Since product images are almost invariably viewed during e-commerce browsing sessions, and because fake shopping websites often recycle identical product images, we propose a technique that uses these product images as seeds for malicious site detection. Specifically, the proposed method leverages the Google Cloud Vision API to identify and collect URLs of websites hosting visually similar images, including both pre-redirect and post-redirect URLs. Based on this approach, we conducted an empirical investigation using 426 product images obtained from suspected fake shopping websites, and we report the results in this paper.

Keywords: Observation of Web-based Attacks, Web Security, Cyber Security, Malicious Websites

1. はじめに

Web 媒介型攻撃にはサイト閲覧するだけでマルウェアの

ダウンロードが実行される Drive-by download、正規サイトを模倣したフィッシングサイトが存在する。フィッシング

1 国立研究開発法人情報通信研究機構

National Institute of Information and Communications Technology
* y.masaki@nict.go.jp

サイトの1つには正規のショッピングサイトを模倣した偽ショッピングサイトの存在が挙げられる。このようなサイトは、購入者の金銭、クレジットカード情報や個人情報の搾取を目的としている。

フィッシングサイトの誘引には正規サイト（ショッピングサイトに限らず、教育機関、司法機関や観光サイトなど）に対する不正アクセスを実施し、改ざんやリダイレクト用スクリプトの仕込みを実施するパターンが存在する（以下踏み台サイト）。同時に Web 検索結果の不正操作する SEO ポイズニングにより、検索上位にフィッシングサイトへの踏み台サイトが出現しやすくなる。

上記の背景から国立研究開発法人情報通信研究機構では、Web 媒介型攻撃の実態把握と対策技術向上のためのユーザ参加型 Web 媒介型サイバー攻撃対策プロジェクト WarpDrive (Web-based Attack Response With Practical and Deployable Research InitiatiVE) [1] を推進している。WarpDrive では、PC 向けにタチコマ・セキュリティ・エージェント（以下、タチコマ SA）、Android スマートフォン向けにタチコマ・セキュリティ・エージェント・モバイル（以下、タチコマ・モバイル）を参加ユーザに無償配布している。参加ユーザの Web アクセスの観測・分析・警告などを実施している。偽ショッピングサイトを含むフィッシングサイトや悪性サイトの分析・観測が実施可能である [2][3][4]。

WarpDrive で観測する Web アクセスログは参加ユーザの Web アクセスに基づく。フィッシングサイトは踏み台サイトの修正、フィッシングサイト自体の URL の変更やサイト内容の変更・削除が頻繁に実施される。偽ショッピングサイトでは、17 日間継続的にアクセスを実施した場合、ドメイン変更が実施された割合は 13.7%存在していることが確認されている [5]。フィッシングサイトの増加や発生時期によって、その実態の変化が考えられることから、ユーザのアクセスログだけではなく、能動的なログ収集・解析が必要となる。ショッピングサイトの商品閲覧時には商品画像を閲覧すること、商品画像は偽ショッピングサイトで転載されているパターンが存在する。よって、商品画像を用いた Web 検索で商品画像を用いたサイト URL 収集が可能と考えられる。

本研究では、多くの悪性サイト URL を能動的に収集し解析性能向上を目的とする。本目的を満たすためには、以下の要件を設定する。

【要件 1】商品画像をシードとすることで、半能動的に偽ショッピングサイト URL の収集が可能となる方式。

【要件 2】実際に収集した URL から偽ショッピングサイトの実態調査が可能であること。

本研究における半能動とは、シードに基づき能動的に URL 収集と悪性判定を行う動作を指す。まず【要件 1】を満たす、偽ショッピングサイトと疑われるサイト URL 収

集方式を提案する。タチコマ SA およびタチコマ・モバイルが収集した偽ショッピングサイトに掲載されている商品画像をシードとして用いる。画像掲載サイト URL、リダイレクト先サイト URL を収集し、悪性判定、TLD 利用数、ドメイン利用数を評価する。また、実際に悪性と判定された URL を用いてサイトにアクセスし、特徴の分析を実施する。本稿では、前述により悪性 URL 収集がどの程度可能なのか定量的に性能評価を実施する。

本稿の構成は以下のとおりである。2 章で商品画像を用いた偽ショッピングサイト URL の収集方式を示す。3 章で提案方式の評価を実施する。4 章で偽ショッピングサイトと疑われるサイトの一部にアクセスを実施し、特徴を抽出した結果についてしめす。5 章で関連研究について示す。6 章で研究倫理について示す。7 章で本稿のまとめを示す。

2. 商品画像検索による偽ショッピングサイト URL 収集方式の提案

2.1 偽ショッピングサイトの商品画像について



図 1 偽ショッピングサイトと疑われるサイトの商品画像による検索結果例

Figure 1 Example of Search Results Based on Product Images from Suspected Fake Shopping Websites.

文献 [5] において、偽ショッピングサイトの商品画像の読み込み手法は以下 2 種類確認されている。

- 1) サイトの img タグの src 属性に正規のショッピングサイトの画像 URL を直接指定する場合。
- 2) サイトの img タグの src 属性に画像を代理で取得するための画像プロキシサーバの URL を指定する場合。

上記の手法を用いる理由として、攻撃者側の画像収集の工数を減らすこと、より多くの偽ショッピングサイトのミラーサイトを多く生成し、アクセスユーザのクレジットカード情報の収集や個人情報の収集、金銭の不正取得の確率を上げたい思惑があると考えられる。よって、実際に偽ショッピングサイトの商品画像を用いた画像検索を実施することで多くのミラーサイトや踏み台サイトが検索結果として出現すると考えられる。図 1 にタチコマ SA で収集され

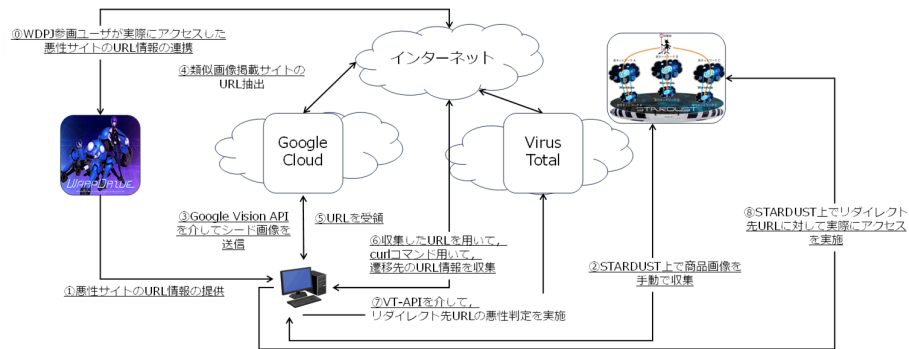


図 2 商品画像を用いた偽ショッピングサイト URL 収集方式

Figure 2 A Method for Collecting URLs Using Product Images from Fake Shopping Websites.

た偽ショッピングサイトと疑われるサイトの商品画像で検索した結果例を示す。商品画像やサイト名について研究倫理に配慮しモザイクを施している。図 1 の検索結果では、まったく同じ画像が用いられたサイトが 70 件以上出現していることが分かる。また出現しているサイトの URL の TLD は「.sn」(セネガル)や「.cl」(チリ)などが用いられている。実際にアクセスするとドメインが全く異なるサイトにリダイレクトすることが見受けられた。検索結果として出現したサイトは、正規サイトが改ざんされ、偽ショッピングサイトを含む悪性サイトにリダイレクトすることが見受けられた。

よって、画像検索によって出現したサイトの URL を収集することで踏み台サイトを含むリダイレクトが発生するサイト URL を抽出できると考えられる。

2.2 提案方式

図 2 に偽ショッピングサイトの商品画像を用いた偽ショッピングサイト URL 収集方式の概要を示す。以下に方式順序を示す。

- ⑩ シードとなる偽ショッピングサイトと疑いのあるサイトの商品画像は WarpDrive 参画ユーザが実際にアクセスしたサイトから手動で収集することを前提とすることから本手番は⑩と設定する。
- ① タチコマ SA が具備する BlockList より、偽ショッピングサイトの疑いがある URL 情報の提供を受ける。
- ② STARDUST 基盤 [6] で実際に URL にアクセスを実施し、商品画像を手動で収集する。収集した画像をシード画像と呼称する。
- ③ Google Cloud Vision API [7] (以下、Google Vision API) を介してシード画像を Google Cloud に送信する。
- ④ Google Vision API の機能である“ウェブエンティティとページ” [8] を用いて類似画像掲載サイトの URL を抽出する。
- ⑤ Google Cloud より URL 抽出結果を受領する。
- ⑥ 収集した URL を用いて、curl コマンドを用いて、リダイレクト先の URL を受領する。

- ⑦ リダイレクト先の URL を Virus Total [9] API (以下、VT-API) を介して Virus Total が保有するアンチウイルスエンジン (以下、AV エンジン) でスキャンを実施し、悪性と判定した AV エンジン数 (以下、スコア数) を取得する。
- ⑧ STARDUST 基盤上で悪性判定された URL にアクセスし、目視で特徴を抽出する。

また先行研究では、リダイレクトには検索エンジンサイトの Referer ヘッダが付与されていないと遷移しないパターンが確認されている [5]。よって Selenium [10] などを用いて人の Web 操作を模倣する手法も考えられるが、検索エンジンに対する負荷を鑑みて、本方式では curl コマンドによりリダイレクト先の URL を収集する。遷移条件は踏み台サイトにより変わることが考えられるため、実際に手動で URL を収集する手法と curl で収集する方法で性能差があると考えられることから 4 章で評価する。

③~⑦はそれぞれ個別で機能を Python3.8.10 で実装した。

STARDUST 基盤, Google Vision API, VT-API の概要と提案方式採用理由については 2.3 節から 2.5 節に別途示す。

2.3 STARDUST 基盤

STARDUST 基盤 [6] は人間的な行動のサイバー攻撃の観測を実現するサイバー攻撃誘引基盤として NICT が研究開発を実施している。単純な偽ショッピングサイトの商品画像収集には Windows Sandbox などを選択肢としてあげられる。ただし、悪質なプログラムが実行されるおそれがある。また、Windows Sandbox を用いたサイバー攻撃の事例も、確認されている [11]。偽ショッピングサイトの商品画像収集、実際に悪性プログラムが実行された際の解析用データの収集が可能となることから STARDUST 基盤を採用した。

2.4 Google Cloud Vision API

Google 検索エンジンによる実際の画像での画像検索は、数十件単位で URL の検出が可能となるが、機械的に検索を実施することは Google 社の規約違反となる [12]。また、

専用の Google 検索エンジン内の画像検索と紐づく API も提供されていない。よって Google Vision API [7] を用いることで上記課題を解決する。

Google Vision API は Google Cloud Platform が提供する画像認識サービスである。機械学習を利用し、画像から物体検出や顔検出、類似画像や完全一致画像が掲載されているサイトの URL を抽出することが可能となる。ただし、API の仕様上、10 件前後の URL 数のみとなる。

2.5 Virus Total API

VT-API [9] は Virus Total に対してファイルアップロード、URL スキャン、スキャン結果の取得がプログラマ的に可能となる API である。Virus Total は複数のセキュリティベンダの AV エンジンを用いて URL スキャンやマルウェアの解析を実施し、誤判定（悪性ではないサイトを悪性と判定する、悪性サイトを悪性ではないと判定するなど）を減らせることから採用した。本研究では、悪性判定 “malicious” を出力した AV エンジン数をスコアとして用い、関連研究 [13] を参考にスコア 3 以上が判定された URL を悪性と判断する。

3. 提案方式評価

3.1 評価項目

本提案方式は悪性 URL 収集を目的としていることから、収集した URL とその URL の悪性判定などを評価する必要がある。よって、以下の観点で実施する。前提条件として、タチコマ SA が収集した偽ショッピングサイトと疑われる 1 サイトから、商品画像を画像の重複がないよう 426 枚収集し、提案方式に利用している。重複の判断は、画像のデータサイズが同一かつ目視で重複を判断している。

[評価 1] 収集 URL 数

Google Vision API による収集 URL 数と、その URL を利用したリダイレクト先 URL の手動収集数、curl を用いて Google の Referrer、Yahoo の Referrer および Bing の Referrer を設定した際のリダイレクト先 URL の自動収集数を評価する。Google、Yahoo および Bing それぞれで評価する理由は、Referrer の種類によりリダイレクト条件が変わることが先行研究で実施されているためである [5]。リダイレクト先の URL 収集において、Google Vision API による収集 URL = リダイレクト先 URL であること、リダイレクト先が存在しない場合はカウントしないものとする。

[評価 2] 踏み台サイト TLD 評価

踏み台サイトとして利用されているサイトの TLD 数を評価し、国別および gTLD の利用数を定量的に分析する。Google の Referrer を設定した curl コマンド実施結果を用いる。

[評価 3] 踏み台サイト頻出ドメイン評価

企業名やブランド名、商用ドメイン数を定量的に評価する。どのようなサイトが踏み台サイトとして利用されている

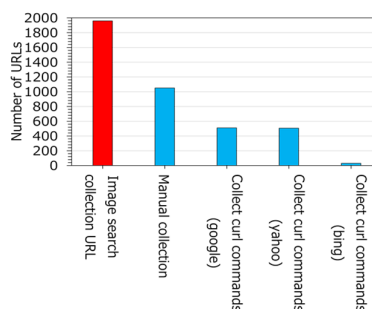


図 3 収集 URL 数の評価結果

Figure 3 Evaluation Results of Collected URL Count.

るか分析する。Google の Referrer を設定した curl コマンド実施結果を用いる。

[評価 4] VT-API によるスキャン結果

リダイレクト先 URL の AV スコア数が 3 以上のものを検出した URL 件数を評価する。比較対象として、AV スコア 1 以上の場合、2 以上の場合についても列挙する。手動 URL 収集の場合と Google の Referrer を設定し、curl を実行した際の収集 URL 数両方を評価する。

[評価 5] VT-API によるスキャン実施 URL の TLD 評価

VT-API によるスキャンを実施した URL の TLD 数について評価を実施し、傾向を定量的に示す。

3.2 [評価 1] 収集 URL 数

図 3 に収集 URL 数を示す。まず、Google Vision API による収集 URL 数は 1959 件となった。手動で 1959 件の URL にアクセスし、リダイレクトしたサイトの URL 数は 1050 件となった。Google の Referrer を設定し、curl コマンドを実行した場合、収集した URL 数は 509 件であった。Yahoo の Referrer を設定し、curl コマンドを実行した場合、収集した URL 数は 506 件であった。Bing の Referrer を設定し、curl コマンドを実行した場合、収集した URL 数は 28 件であった。

1959 件の URL 収集数について、収集画像数 426 枚と平均値をとった場合、1 枚あたり約 5 件の URL が収集できることが分かる。

手動でリダイレクト先 URL を収集した場合と curl コマンド実施した場合を比較し、前者のほうが URL 数は 1050 件と多い傾向にある。Referrer 設定のみではアクセスを拒否するような設定が施された踏み台サイトが存在するためと考えられる。2.2 節に示したとおり、Selenium による自動化により時間的コストの削減は可能となるが、規約上利用はできない。よって、curl による時間的コスト削減とリダイレクト先 URL の収集数はトレードオフの関係と言える。

curl コマンド実施時の Referrer の種別により、収集可能な URL 数に差分があることが分かった。Google > Yahoo > Bing という大小関係であった。日本国内ユーザを誘引する

ことを目的としていることから、日本国内ユーザの利用数が多い Google や Yahoo を対象としている攻撃者が多いと考えられる。

3.3 [評価 2] 踏み台サイト TLD 評価および [評価 3] 踏み台サイト頻出ドメイン評価

図 4 に踏み台サイトで用いられている TLD 数のうち上

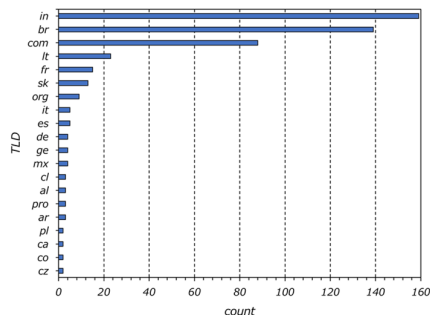


図 4 踏み台サイト TLD 数評価

Figure 4 Evaluation of the Number of TLDs in

Redirector Sites.

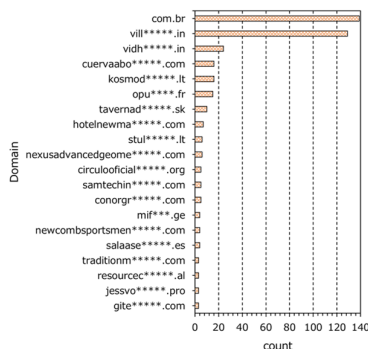


図 5 企業、ブランドおよび商用ドメイン数

Figure 5 Number of Corporate, Brand, and Commercial Domains.

位 20 件を示す。一番多かった TLD は “.in (インド)” で、159 件であった。二番目は “.br(ブラジル)” で、139 件であった。三番目は “.com (gTLD)” で、88 件であった。

図 5 に企業名やブランド名、商用ドメインの件数のうち上位 20 件を示す。研究倫理観点から一部マスク (*) を施す。一番多かったドメインは “.com.br(ブラジル商用ドメイン)” で 139 件であった。2 番目は “.vill****.in (インド観光サイト)” で 129 件であった。3 番目は “.vidh****.in (法律または教育関連サイトなど)” で 24 件であった。

ホテルサイトを含む観光業を展開するサイトの改ざんについては新型コロナ禍以降の各国のインバウンド需要による影響が考えられる。観光・ホテルサイトは多くのユーザの個人情報などが入手できること、セキュリティ対策が不十分であることが標的となっていることが一部要因として考えられる [14]。

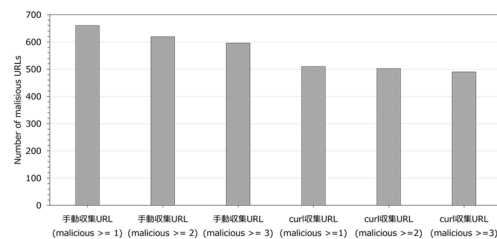


図 6 VT スキャンによる悪性判定結果

Figure 6 Maliciousness Detection Results by VT.

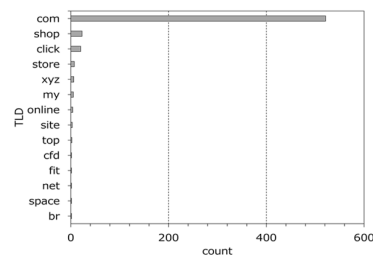


図 7 手動収集時の悪性判定 URL の TLD 数評価

Figure 7 Evaluation of TLD Count in Maliciously Identified URLs Collected Manually.

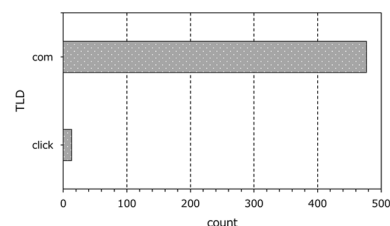


図 8 curl コマンド収集時の悪性判定 URL の TLD 数評価

Figure 8 Evaluation of TLD Count in Maliciously Identified URLs Collected via curl Commands.

3.4 [評価 4] VT-API によるスキャン結果

図 6 に VT スキャン結果について示す。手動で収集したリダイレクト先 URL のうち、VT が “.malicious” を 1 以上判定した URL 数は 669 件であった。2 以上の URL 数は 619 件であった。3 以上の URL 数は 596 件であった。次に curl コマンドかつ google.com で収集したリダイレクト先 URL のうち、VT が “.malicious” を 1 以上判定した URL 数は 596 件であった。2 以上の URL 数は 509 件であった。3 以上の URL 数は 490 件であった。

上記の結果から、手動で収集した場合、取得 URL 数 1050 件のうち 596 件 (約 57%) が “.malicious” 3 以上の URL が占めている結果となった。curl で取得した場合、取得 URL 数 596 件のうち 490 件 (約 82%) が “.malicious” 3 以上の URL が占めている結果となった。curl コマンドは踏み台サイト内の JavaScript 内に記述されたリダイレクト先 URL 収集を実施しているが、手動収集の場合、遷移先の URL を収

集している。よって、手動動作による遷移条件を多く含めた形でリダイレクトを発生させることが可能なため、より多くの悪性 URL を収集可能だとわかる。ただし、踏み台サイトが修繕された場合、正規サイトのトップページに遷移する場合が存在することから、正規サイトの収集数も増加し、VT スキャン結果に悪性以外で含まれる形となることが考えられる。また、curl コマンドによる“malicious”判定のリダイレクト先 URL 数が約 82%であった。リダイレクトには改ざんサイト内の JavaScript に対してリダイレクト先 URL を記述し、ユーザを誘引する傾向の存在を改めて観測できる結果となった。

3.5 [評価 5] VT-API によるスキャン実施 URL の TLD 評価

図 7 に手動で収集したリダイレクト先 URL のうち、“malicious”判定 3 以上の URL に含まれる TLD 種別とその数を示す。1 番利用数の多い TLD は“com (gTLD)”であり 521 件であった。2 番目に利用数の多い TLD は，“shop”で、23 件であった。3 番目に利用数の多い TLD は，“click”で、20 件あった。とくに 2 番目、3 番目の TLD は偽ショッピングサイトと疑われるサイトで利用数が多いことが調査されている [5][15]。また全部で 14 種類の TLD が利用されていることが分かった。

図 8 に curl コマンドで収集したリダイレクト先 URL のうち、“malicious”判定 3 以上の URL に含まれる TLD 種別とその数を示す。1 番目に利用数の多い TLD は，“com (gTLD)”であり、477 件であった。2 番目に利用数の多い TLD は，“click”であり、13 件であった。また、curl コマンド収集では TLD 種別は上記 2 種類のみであった。

上位に出現する TLD “com”や“shop”などは比較的安価もしくは無料でドメインを取得できることから、攻撃者側の利用数が多いと考えられる。curl コマンド収集による悪性判定 URL は“com”および“click”のみであった。“com”利用数が多い点については前述と同様の理由と考えられる。“click”については、マウスクリックによるページ遷移であることを攻撃者側が把握しやすいように利用しているものと考えられる。上記の傾向は、検索サイトが“bing.com”、“yahoo.co.jp”の場合でも同一である。

よって、TLD 種別によりリダイレクト条件を攻撃者側の意図によって変えていることが考えられるが、他 TLD の利用が利用されていない部分と同時に今後の分析課題とする。

4. 偽ショッピングサイトと疑われるサイトへのアクセスと特徴抽出および比較

4.1 偽ショッピングサイトの特徴

偽ショッピングサイトの特徴は多岐にわたる。日本サイバー犯罪対策センター（以下、JC3）[16] の調査では、以下の特徴が含まれていることが分かっている。

[特徴 1] 価格が安い。

表 1 偽ショッピングサイト特徴抽出結果

Table 1 Feature Extraction Results of Fake Shopping Websites.

①	ドメインのトップレベルドメイン（TLD）が「.xyz」「.top」「.buzz」「.cfd」「.store」「.click」「.casa」「.live」などが含まれている。
②	URL に意味不明な文字列やランダムな英数字が多用されており、サイトの内容と一致しない。
③	ブラウザのタブに表示されるタイトルと、実際のサイト名が一致しない。
④	会社概要欄の会社住所が存在しないもしくは別の組織の住所。
⑤	購入ページの支払い方法と利用規約や FAQ などに記載の方法に齟齬がある。
⑥	ほぼすべての商品が割引となっている。
⑦	法人名の欄に個人名（フルネームなど）が記載されている。
⑧	サイト全体の日本語に表記ゆれや日本語のおかしい箇所が存在する。
⑨	個人情報保護方針にて別の会社名や既存の会社名の記載が存在する。
⑩	商品注文ページの商品小計に対して消費税額が 0 円である。
⑪	購入者のレビューコメントに実際の商品の趣旨とは異なるレビューがされている。
⑫	同一の見目のサイトがほかにも複数存在している。
⑬	会社方針に「ハッセル払い戻しポリシーなし」という文言の記載がある。
⑭	会社方針に「世界中の無料税」という文言の記載がある。
⑮	商品画像が他大手ショッピングサイトにも存在する。
⑯	品薄などの文言の記載がある。
⑰	設立年月日 2013 年 8 月 18 日 という文言が含まれている。
⑱	「日用品雑貨類等の小売業」という文言が含まれている。
⑲	「◆創業 2004 年 12 月 6 日」という文言が含まれている。
⑳	株式会社 <商品のカテゴリ名><「商店」,「オンラインストア」,「専門店」など店を示す文字>が含まれている。

[特徴 2] 支払方法が銀行振込に限定されている。

[特徴 3] 不自然な日本語。

[特徴 4] URL のドメイン名に“xyz”や“.top”などの TLD が用いられている。

警察庁の調査 [17] では上記に加え、以下の特徴が含まれていることを調査している。

[特徴 5] 「品薄」等の表示により商品購入を急がせる。

[特徴 6] 会社概要に実在しない住所が記載されている。

トレンドマイクロ社 [18] の調査では、さらに以下の特徴が含まれていることを長際している。

[特徴 7] 設立年月日 2013 年 8 月 18 日 という文言が含まれている。

[特徴 8] 株式会社 <商品のカテゴリー名><「商店」,「オンラインストア」,「専門店」など店を示す文字>が含まれている。

[特徴 9] 「ハッセル払い戻しポリシーなし」という文言が含まれている。

[特徴 10] 「日用品雑貨類等の小売業」という文言が含まれている。

[特徴 11] 「◆創業 2004 年 12 月 6 日」という文言が含まれている。

上記, JC3, 警察庁およびトレンドマイクロ社の調査を踏まえつつ, タチコマ SA が収集した偽ショッピングサイトと疑われるサイトより, 特徴の確認および特徴を目視で抽出する。2025 年 5 月 30 日から 2025 年 7 月 10 日にかけて偽ショッピングサイトと疑われる 15 サイトを対象に特徴について抽出した。以下表 1 に結果を示す。今回特徴として改めて観測できた箇所は, ⑨, ⑩, ⑪, ⑭となる。⑨については文中に某総合電機メーカーの社名を確認したためである。⑩については購入ページにて, 商品小計と税込み額とは同一金額であったが, 消費税額欄の金額が 0 円であったためである。⑪については, 模型商品の購入者レビュー欄にて, 「生地がしっかりしている」といった商品種別と比較し, 趣旨の違うコメントがなされていたためである。⑭については, 趣旨や意味が読み取れないことに由来し, ⑬が用いられているサイトとほぼ同時に用いられている。

4.2 節では 3 章で収集した “malicious” 判定サイトに対して STARDUST 基盤上で実際にアクセスを実施し, 表 1 で示す特徴を含むか実際に確認を実施する。

4.2 偽ショッピングサイトへのアクセス

3.5 節で示した curl 収集かつ “malicious” 判定 3 以上の

表 2 アクセス実施 URL

Table 2 Accessed URLs.

アクセス実施 URL	
A	https://cfag**.o*****.za.com/index.php?main_page=product_info&products_id=10912
B	https://bqtdsc**.a*****.sa.com/index.php?main_page=product_info&products_id=29624
C	https://wazuq**.b*****.za.com/index.php?main_page=product_info&products_id=14704
D	https://ge*.m*****.click/index.php?main_page=product_info&products_id=9034
E	https://ubys**.r*****.za.com/index.php?main_page=product_info&products_id=11941
F	https://dfprv**.t*****.za.com/index.php?main_page=product_info&products_id=21320
G	https://best****.click/index.php?main_page=product_info&products_id=3585
H	https://owqd**.r*****.za.com/index.php?main_page=product_info&products_id=10827
I	https://mxl**.r*****.sa.com/index.php?main_page=product_info&products_id=8038
J	https://kxbi**.t*****.za.com/index.php?main_page=product_info&products_id=23566

URL からランダムで 10 件 URL を抽出し, STARDUST 基盤上の Windows 端末から実際にアクセスを実施した。表 2 に対象 URL を示す。研究倫理観点から一部マスク (*) を施す。アクセスは 2025 年 7 月 22 日に実施した。表 3 に表 2 に対する表 1 の特徴の有無について示す。

表 3 特徴有無結果

Table 3 Results of Feature Presence Evaluation.

	A	B	C	D	E	F	G	H	I	J
①				✓			✓			
②	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
③	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
④	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
⑤	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
⑥	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
⑦										
⑧	✓					✓	✓		✓	
⑨										
⑩				✓			✓			
⑪	✓		✓					✓	✓	✓
⑫			✓							
⑬										
⑭										
⑮	✓		✓			✓		✓	✓	✓
⑯							✓			
⑰				✓			✓			
⑱										
⑲										
⑳							✓			

表3の結果より、サイトAからJに共通する特徴は②から⑥までの特徴であることが分かった。次に⑦と⑨、⑬と⑭、⑱と⑲について、チェックはつかなかった。サイトDとGについて、①と⑰が他のサイトと比較してチェックがついた。①については、“click”のTLDが利用されていたためである。⑰については同一の文言が見受けられたためである。

“com”と“click”が付与されたサイトによって特徴が異なる点が存在することから、TLDによって構築された偽ショッピングサイトの様式や目的が異なることも考えられる。

5. 関連研究

Sakaiらは、偽ショッピングサイトを自動で検出し、実運用に耐えうる処理速度および精度を両立したシステムを提案し、JC3で実用化している[19]。文献[19]では、Webクローラによるデータ収集とHTMLコードに着目した機械学習によって、偽ショッピングサイトURLを検出するシステム開発が目的とされている。Webクローリングサブシステム(WCS)および機械学習サブシステム(MLS)を構築し、新規に公開されたWebサイトを対象にクロール、偽ショッピングサイトかどうかの判定を自動化した。MLSはFastTextとLightGBMが用いられている。判定精度は98.5%と高い結果となっている。またJC3で運用されているという実績が存在する。

本稿と比較し、能動性および検知率の観点では、より高く機械学習による精度の高い検知が可能な方式と考えられる。本稿の方式では、ユーザ利用を想定したとき、実際の商品画像を用いて偽ショッピングサイトと疑われるサイトのURL収集と、悪性判定を同時に実施が可能となる点において有用と考えられる。

6. 研究倫理

本研究では、商品画像を偽ショッピングサイトと疑われるサイトから収集した。対象サイトに対する負荷を考慮し手動で実施した。curlコマンドによるURL収集でも対象サイトに対する負荷を考慮し、1サイト1度のみコマンドを実施した。取得データはリダイレクトURL以外収集を実施していない。

7. おわりに

本稿では、商品画像をシードとし、半能動的に偽ショッピングサイトと疑われるサイトの入り口となる踏み台URLとリダイレクト先URLを収集する方式を提案した。上記を実現するために、[要件1]と[要件2]を設定した。1959件の画像検索URLとVT-APIによる“malicious”判定3以上のリダイレクトURLを490件取得することができた。よって、[要件1]を満たした。また、実態についても

定量的に調査することが可能であることを示し、[要件2]を一部充足したと考えられる。

今後、より精度の高いリダイレクトURL取得方法の提案と、認知科学的な側面からデータ分析を実施する。

参考文献

- [1] “WarpDrive”. <https://warpdrive-project.jp/>. (参照 2025-07-16).
- [2] Hasegawa, M. Saino, A., Fujita, A. Tanabe, R. Ganan, C. H. van Eten, M. Gañán, C. H. and Yoshioka, K. POSTER: Do You Sell This? Utilizing Product Searches to Find SEO-driven Fake Shopping Sites. Network and Distributed System Security Symposium (NDSS2025), 2025.
- [3] Miyashita, D. Kobayashi and S. Yamauchi, T. Investigation Towards Detecting Landing Websites for Fake Japanese Shopping Websites. 13th International Conference on Emerging Internet, Data and Web Technologies (EIDWT2025), Lecture Notes on Data Engineering and Communications Technologies, 2025, vol. 243, p. 107-119.
- [4] Yamauchi, T. Orito, R. Ebisu and Sato, M. Detection Unintended Redirects to Malicious Websites on Android Devices Based on URL-Switching Interval. IEEE Access, 2024, vol. 12, p. 153285-153294.
- [5] 小寺 博和. 小出 駿. 千葉 大紀. 青木 一史. 秋山 満昭. 偽ショッピングサイトによる攻撃手法の実態解明. 情報処理学会論文誌. 2021, vol. 62, no. 9, p. 1523-1535.
- [6] 金谷 延幸. 鈴木 悦子. 中村 大典. 梅村 勇貴. 佐藤 茂. 攻撃者の行動を観測するためのサイバー攻撃誘引基盤 STARDUST. 情報通信研究機構, 2024, vol. 70, no. 2, p. 15-27.
- [7] “Cloud Vision API”. <https://cloud.google.com/vision?hl=ja>, (2025-07-17 参照).
- [8] “ウェブエンティティとページを検出する”. <https://cloud.google.com/vision/docs/detecting-web?hl=ja> (2025-07-17 参照).
- [9] “Virus Total”. <https://www.virustotal.com/gui/home/url> (2025-07-17 参照).
- [10] “The Selenium Browser Automation Project”. <https://www.selenium.dev/documentation/> (2025-08-19 参照).
- [11] “MirrorFace によるサイバー攻撃について (注意喚起)”. https://www.nisc.go.jp/pdf/news/press/20250108_MirrorFace.pdf (2025-07-17 参照).
- [12] “Google Privacy & Terms”. <https://policies.google.com/terms?hl=en-US> (2025-08-19 参照).
- [13] Namrud, Z. Kpodjedo, S. and Talhi, C. AndroVul: a repository for Android security vulnerabilities. 2019, CASCON '19: Proceedings of the 29th Annual International Conference on Computer Science and Software Engineering, p. 64-71.
- [14] Florido-Benitez, L.: The role of cybersecurity as a preventive measure in digital tourism and travel. a systematic literature review, Computers, 2025, Springer, vol. 28, no. 28.
- [15] “【セキュリティレポート】2024 年上半期フィッシングサイトのドメインを独自に分析 低価格または無料で購入できるドメイン「.xyz」を悪用した攻撃が増加”. <https://www.daj.jp/webtopics/1168/> (2025-07-22 参照).
- [16] “偽ショッピングサイトに注意”. <https://www.jc3.or.jp/threats/topics/article-462.html> (2025-07-22 参照).
- [17] “偽ショッピングサイト・詐欺サイト対策”. <https://www.npa.go.jp/bureau/cyber/countermeasures/fake-shop.html> (2025-07-22 参照).
- [18] “日本向け偽ショッピングサイトを運用する犯罪アクターのグルーピング”. https://www.trendmicro.com/ja_jp/research/23/j-fake-shopping-sites.html (2025-07-22 参照).
- [19] Sakai, K. Takeshige, K. Kato, K. Kurihara, N. Ono, K. and Hashimoto, M. An Automatic Detection System for Fake Japanese Shopping Sites Using fastText and LightGBM. 2023, IEEE Access, vol. 11, p. 111389-111401.