Web検索から偽ショッピングサイトへの誘導の実態調査

才納 明英[1,a)]　高田 一樹[2,7]　藤田 彬[3,7]　小出 駿[4]　金井 文宏[5]　秋山 満昭[6]　田辺 瑠偉[7]
吉岡 克成[7,8]　松本 勉[7,8]

概要：偽ショッピングサイトによるユーザの金銭や個人情報の窃取が深刻化しており，早急な対策が求められている．これまでに偽ショッピングサイトの検知に関する研究が行われているが，ユーザがどのようにして偽ショッピングサイトに誘導されるか調査した研究は少ない．本研究では，実ユーザの Web アクセスログを分析することで，Web 検索から偽ショッピングサイトへと到達する可能性のある状況についての実態調査を行った．具体的には，数百人規模のアクティブユーザの Web アクセスログから検索結果ページに表示された URL を抽出するとともに，Web クローラを用いて実際に表示された Web サイトにアクセスし，リダイレクト等の特徴により偽ショッピングサイトを判定した．分析の結果，検索結果ページの約 5% に偽ショッピングサイトに到達する踏み台サイトの URL が含まれていたことを確認し，身近な脅威であることを確認した．また，踏み台サイトは検索結果の 20 位前後に出現する場合が多いことを確認した．さらに，ユーザが検索を起点として実際に偽ショッピングサイトに到達した事例を 26 件確認した．偽ショッピングサイトへの到達前後の行動からは，商品やその情報を得ようとして様々な Web 検索を試み，過去に訪問したことがないものを含めて様々なサイトへアクセスする振る舞いが伺えた．

キーワード：偽ショッピングサイト，踏み台サイト，Web 検索

Investigation on Fake Shopping Sites via Web Searches

Akihide Saino[1,a)]　Kazuki Takada[2,7]　Akira Fujita[3,7]　Takashi Koibe[4]　Fumihiro Kanei[5]
Mitsuaki Akiyama[6]　Rui Tanabe[7]　Katsunari Yoshioka[7,8]　Tsutomu Matsumoto[7,8]

Abstract: Theft of users' money and personal information by fake shopping websites is becoming increasingly serious that countermeasures are required. Although there have been studies that focus on detecting fake shopping sites, few studies have investigated how users are redirected to these websites. In this study, by analysing the web access logs of real users, we investigate how users are redirected to fake shopping sites from web searches. We first extracted URLs displayed on search result pages using Web access logs of several hundred active users. We then detected fake shopping sites by using a web crawler to access to these websites and by finding features such as redirects. As a result, approximately 5% of the search result pages contained the URL of a springboard site that redirected to a fake shopping site. In many cases, springboard sites appear around the rank of 20 in each web search results. Furthermore, we confirmed 26 cases in which users were actually redirected to the fake shopping site. The user behavior before reaching fake shopping sites and after reaching the sites shows that the user was trying to obtain the product and its information and accessed various sites including those they had never visited in the past.

Keywords: Fake shopping website, Springboard website, Web search

[1] 横浜国立大学大学院環境情報学府
Graduate School of Environment and Information Sciences, Yokohama National University
[2] 株式会社セキュアブレイン
SecureBrain Corporation
[3] 情報通信研究機構
National Institute of Information and Communications Technology
[4] NTT セキュリティ・ジャパン株式会社
NTT Security (Japan) KK
[5] NTT コミュニケーションズ株式会社

1. はじめに

近年，インターネットショッピングの普及に伴い，正規のショッピングサイトを装いユーザの個人情報や金銭を窃取する偽のショッピングサイトが増加傾向にあり，対策が求められている．実際に，偽ショッピングサイトで商品を購入すると代金を支払ったにも関わらず商品が届かない場合や，商品が届いたとしても粗悪品や偽物が届いた被害が知られている．日本サイバー犯罪対策センターによると，2022 年に偽ショッピングサイトを含む悪質なショッピングサイトについて 28,818 件の通報が寄せられている [20]．

これまでに，偽ショッピングサイトの検知や実態解明に関する研究 [8], [18], [19] が活発に行われてきた．しかし，ユーザがどのようにして偽ショッピングサイトに誘導されるか調査した研究は少ない．攻撃者は，検索結果での表示順位を不正に操作する手法であるブラックハット SEO を行い，予め用意したウェブサイトや改竄した正規のウェブサイト（以降，踏み台サイトとする）を検索結果に表示させる [2]．攻撃者が踏み台サイトを検索結果に表示させる目的として，偽ショッピングサイトの URL を検索エンジンのクローラから隠してブロックリストへの追加を回避することなどが考えられる．その後，ユーザが踏み台サイトへアクセスすると，Location ヘッダや JavaScript により自動的に偽ショッピングサイトへとリダイレクトされる．このため，本研究では攻撃の起点となる Web 検索に着目してユーザの検索行動の実態調査を行う．

以降では，実ユーザの Web アクセスログを用いて (1)Web 検索結果に偽ショッピングサイトへリダイレクトされるリスクがどの程度存在するか調査するとともに，(2) どの程度のユーザが検索結果に表示された踏み台サイトにアクセスするか調査することで，偽ショッピングサイトへの対策を講じることを目指す．具体的には，検索エンジンの検索結果を表示したページ（Search Engine Result Page：以降，SERP とする）を分析し，SERP に表示された URL にクローラを用いて実際にアクセスを行い，リダイレクト等の特徴を観測することで偽ショッピングサイトを判定する．また，偽ショッピングサイトへ到達したユーザの Web アクセスログを抽出してその検索行動を分析する．

専用のセキュリティエージェントを用いて 2023 年 5 月 31 日から 2023 年 6 月 23 日までの 24 日間のユーザの Web アクセスログを収集した．Web 検索結果に含まれる URL にアクセスしたところ，3,489 件の踏み台サイト URL と 3,448 件の偽ショッピングサイト URL を検出した．また，観測期間中にユーザの検索により得られた全 SERP の約 5%（1,767/36,948）に偽ショッピングサイトに誘導する踏み台サイトの URL が含まれており，検索を行ったユーザの約 35%（302/851）は踏み台サイトを含む SERP を調査期間内に 1 回以上表示し，約 6%（51/851）は平均で 1 日 1 回以上表示していた．このように日々の Web 検索に偽ショッピングサイトにリダイレクトされるリスクが存在し，身近な脅威であることを確認した．踏み台サイトは検索結果の 10 位以内に出現する場合は比較的少ないものの 20 位前後に出現する場合が多く，商品情報の転載元としてメルカリなどの正規サイトが利用されていることを確認した．また，ユーザが実際に偽ショッピングサイトへアクセスした 26 件の事例では全て検索が起点であり，その偽ショッピングサイトへの初めてのアクセスであった．偽ショッピングサイトへとアクセスする前後のユーザの行動を見ると，商品に関して検索クエリを変えながら検索を行ったり，複数の検索結果にアクセスするなど，商品やその情報を得ようとするユーザの心理が伺えた．更に，偽ショッピングサイトの URL ブロックリストを作成し，Web アクセスログ内のユーザへの配信シミュレーションを行ったところ，実際にユーザ保護に役立つことを確認した．

2. 関連研究

偽ショッピングサイトを含む悪意のあるウェブサイトやソーシャルエンジニアリング攻撃に着目した研究が活発に行われている．具体的には，URL やドメイン名や IP アドレスの文字列特徴を用いて悪意のあるウェブサイトやドメイン名を検出する手法 [3], [4], [9], [16]，HTML などのサイト内コンテンツを用いて悪意のあるウェブサイトを検出する手法 [5]，機械学習を用いてフィッシング攻撃を検出する手法 [10], [12]，ウェブサイト内の HTML 要素を意図的に選択することで多段階のソーシャルエンジニアリング攻撃を自動的に収集及び検出する手法 [15]，ユーザのアクセスを木構造でモデル化し悪意のあるリダイレクトを検出する手法 [7] などが提案されている．

同様に，偽ショッピングサイトに着目した研究が行われている．論文 [8], [18] では，機械学習を用いて高い精度で偽ショッピングサイトを検出する手法が提案されている．また，論文 [19] では 99.8%の踏み台サイトが解析回避機能を有することや一定数のユーザが偽ショッピングサイトに到達している可能性があることを明らかにした．

一方で，実ユーザのログに基づいた研究が行われている．論文 [6] では，数百人規模のユーザからコンピュータの構成とユーザの行動に関するデータを数年間収集し，ユーザへのアンケートを基にユーザのセキュリティ意識と実際の

NTT Communications Corporation
[6] NTT 社会情報研究所
NTT Social Informatics Laboratories
[7] 横浜国立大学先端科学高等研究院
Institute of Advanced Sciences, Yokohama National University
[8] 横浜国立大学大学院環境情報研究院
Faculty of Environment and Information Sciences, Yokohama National University
[a)] saino-akihide-tb@ynu.jp

コンピュータの状態の関係について調査した．論文 [14] では，1000 人規模のユーザの Web アクセスログを分析した結果，悪性 URL への到達経路はブックマークが多いことが報告されている．論文 [13] では，2 万人程度のユーザから HTTP アクセスのログを 3 か月収集し，機械学習を用いてユーザが悪意のあるコンテンツに到達する直前に到達を予測するシステムが提案されている．

このように様々な研究が行われているが，ユーザの検索行動に着目して偽ショッピングサイトへ誘導されてしまう状況について調査した研究は我々の知る限り存在しない．

3. データセット

本研究では，Web 媒介型攻撃からユーザを守ることを目的とした WarpDrive プロジェクトで公開されているセキュリティエージェント [17] を利用して，2 種類のログ（SERP ログと Web サイトアクセスログ）を収集した．当セキュリティエージェントは PC 版の Google Chrome に拡張機能としてインストールして利用することができ，ユーザは規約に同意することで無料で利用することができる．

SERP ログ: SERP ログは，ユーザが行った検索に関する情報を取得したログである．本研究での検索とは, Google（google.com, google.co.jp）, Yahoo!JAPAN（yahoo.co.jp）, Bing（bing.com）のいずれかの検索エンジンを用いた検索のみを対象とし，Baidu や DuckDuckGo などの他の検索エンジンを用いた場合や，“google.co.uk” や “yahoo.com” などの別ドメイン名の検索エンジンを用いた場合は調査から除外した．日本での検索エンジンのシェアは Google・Yahoo!・Bing の合計で 99%を超えており [1]，別の検索エンジンを用いるユーザは殆ど存在しないと考えられる．また，Google・Yahoo!・Bing を用いる場合であっても，日本以外の国をサービスの対象としている URL で検索を行うユーザも殆ど存在しないと考えられる．セキュリティエージェントは，ユーザが “google.com”・“google.co.jp”・“yahoo.co.jp”・“bing.com” のいずれかの検索エンジンを用いて検索を行った場合に JavaScript 実行後の HTML ソースを収集する．収集した SERP ログには以下が含まれる．

- ユーザ ID：ユーザを一意に識別するための識別子
- URL：SERP 自体の URL
- 日時：ユーザが検索を実行した日時
- HTML ソース：検索結果に表示されるタイトル・スニペット・URL・表示順位などの情報が含まれる

Web サイトアクセスログ: Web サイトアクセスログは，Web サイトにアクセスする際にブラウザからサーバに対して行ったリクエスト及びレスポンスに関する情報を取得したログであり，以下の情報が含まれる．

- ユーザ ID：ユーザを一意に識別するための識別子
- URL：セキュリティエージェントが収集する URL．アクセスしたページだけでなく，読み込まれるリソースもこれに含まれる（例 ‘.js’,‘.css’）．
- リソースタイプ：読み込まれるリソースの種類．main_frame（タブにロードされたトップレベルのドキュメント）や sub_frame（<iframe>要素または<frame>要素にロードされたドキュメント），script（<script>要素によって実行される，またはワーカーで実行されるようにロードされるコード）など．
- タブ ID：リソースが読み込まれたブラウザタブを識別する ID
- 元タブ ID：生成元のタブ ID．新しくタブが生成されたときに，そのタブで読み込まれるリソースタイプが main_frame のログにのみ設定される．
- リファラ：リクエストを発生させる原因となった参照元の URL
- ページ遷移タイプ：ページ遷移の種類．link（ユーザがリンクをクリックした）や form_submit（ユーザがフォームを送信した），auto_bookmark（ユーザがブックマークをクリックした）など．
- ページ遷移修飾子：ページ遷移に関する追加情報．client_redirect（ページ上の JavaScript またはメタリフレッシュに起因するリダイレクト）や server_redirect（サーバから送信された 3XX HTTP ステータスコードに起因するリダイレクト），forward_back（ユーザが「進む」または「戻る」ボタンをクリックした場合）など．
- アクセス日時：ユーザが URL にアクセスした日時

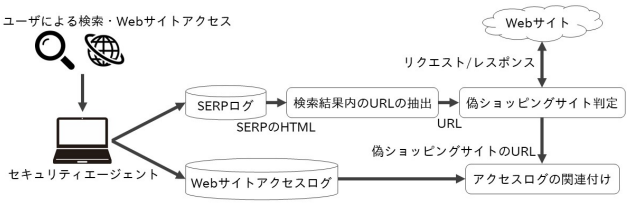

図 1 実態調査の流れ
Fig. 1 Flow of Survey

4. 実態調査の流れ

本研究における調査では，前述の 2 種類のログを入力として，Web 検索結果や検索行動を分析するためのデータの抽出を行う．調査の流れを**図 1** に示す．はじめに，偽ショッピングサイトへリダイレクトする踏み台サイトの可能性のある URL を収集するため，(1) 検索結果に含まれる URL を抽出する．また，(2)Web クローラを用いて抽出した URL にアクセスしてリダイレクトなどの特徴を用いて偽ショッピングサイトであるか判定する．続いて，偽ショッピングサイトにアクセスしたユーザの検索行動を調

査するため，(3) 偽ショッピングサイトの URL を用いて Web サイトアクセスログから偽ショッピングサイトにアクセスしたユーザを抽出して検索行動に関連付けを行う．

4.1 検索結果に含まれる URL の抽出

前章で述べたように，セキュリティエージェントはユーザが特定のドメイン名の検索エンジンで検索を行った場合に HTML ソースを収集する．ただし，一部の SERP では動的に検索結果が読み込まれて表示される場合がある．これらの検索結果については HTML ソースから抽出することができないため，分析の対象外とした．検索エンジンによって HTML の構成は大きく異なるため，Google・Yahoo! JAPAN・Bing の 3 つの検索エンジンの HTML ソースに対して異なる処理を行った．なお，画像検索や動画検索などは分析の対象外とし，通常の検索結果のみを分析の対象とした．検索エンジン側の都合で HTML の構成が突然変化することがあるが，出来る限り調査に影響を与えないように留意し，適宜処理を修正した．

検索を行ったユーザは，SERP に表示された URL にアクセスした後に元の SERP に戻ることがある．セキュリティエージェントの仕組み上，最初に SERP を表示したときと同一の SERP に戻ったときにも SERP ログが収集される．同一ユーザが行った同一の検索について重複したデータをそのまま調査に用いることをさけるため，同一ユーザから得られた SERP について，検索時刻から 5 分が経過するまでに検索クエリと URL と表示順位が完全に一致する SERP が存在した場合にはそれらを同一の検索行動とみなし，先に検索されたもののみを調査対象の SERP とした．

4.2 偽ショッピングサイトの判定

Web ブラウザ自動化ツールである Puppeteer [11] を用いて検索結果の URL に実際にアクセスを試みた．コストの低減のため，google.com などの明らかに安全だと思われるドメイン名をもつ URL は分析の対象外とした．なお，この処理により除外したドメイン名は 801 件である．先行研究において，踏み台サイトの多くに解析回避機能があり，HTTP リクエストヘッダに検索エンジンの Referer を設定せずにアクセスした場合には偽ショッピングサイトへリダイレクトされないことが明らかにされている [19]．そのため，HTTP リクエストヘッダに検索エンジンの Referer を設定した．加えて，ユーザのアクセスを模倣するために HTTP リクエストヘッダに Accept・Accept-Encoding・Accept-Language・Connection・Cache-Control・User-Agent の各項目を追加した．アクセス後は，Location ヘッダや JavaScript に起因するリダイレクトについて調査し，到達した Web サイトの HTML ソースと読み込まれたリソース（JavaScript や画像ファイルなど）の URL を取得した．

検索結果に表示された踏み台サイトへアクセスすると，偽ショッピングサイトへリダイレクトされる．偽ショッピングサイトで表示される画像は正規サイトから直接読み込まれている場合があるが，正規サイトでこの挙動が発生することは少ないと考えられる．そのため，SERP に表示された検索結果の URL について，以下の条件を全て満たした場合に偽ショッピングサイトと，偽ショッピングサイトへリダイレクトするサイトを踏み台サイトと判定した．また，商品情報の転載元となっている正規サイトを特定した．

条件 1 検索結果の URL にアクセスすると，リダイレクトによって別サイトへのページ遷移が発生した

条件 2 最終的に到達したウェブサイトのドメイン名が正規サイトのドメイン名ではない

条件 3 正規サイトの画像を読み込んでいる

条件 3 のみを満たさない場合には，Web サイトに掲載されている商品のタイトルや画像，説明文などの情報を基に偽ショッピングサイトであるかを判定した．判定の過程で転載元となっている正規サイトを新たに特定した場合には，その正規サイトの情報も条件 2 及び条件 3 の条件に加えて利用することとした．5 章の実態調査では 21 種類の正規サイトを転載元とした．偽ショッピングサイトには数十万の商品が掲載されている場合があったが，同一ドメイン下に存在していることから，商品詳細ページの URL が異なっていてもドメイン名が同じであれば同じ偽ショッピングサイトを構成するページであると考えられる．実際に偽ショッピングサイトの URL のクエリパラメータ部を削除してドメイン名のみでアクセスを行ったところ，偽ショッピングサイトのトップページが表示された．そのため，ある URL を偽ショッピングサイトであると判定した場合に，その URL と同じドメインをもつ他の URL も偽ショッピングサイトであると判定した．

4.3 Web 検索と偽ショッピングサイト誘導の関連付け

偽ショッピングサイトへのアクセスを基軸として，その前後で行われた Web 検索や Web サイトへのアクセスとの関連付けを以下の (1) から (5) までの手順により行った．

(1) Web サイトアクセスログの中から，リソースタイプが `main_frame` であり，ドメイン名が判定した偽ショッピングサイトのドメイン名と一致した場合に，偽ショッピングサイトへのアクセスとした（あるユーザが偽ショッピングサイトに到達したことを示す）．

(2) 偽ショッピングサイトへのアクセスからユーザ ID，タブ ID，元タブ ID を特定し，Web サイトアクセスログからユーザ ID と一致する全 Web アクセスを抽出した．

(3) 抽出した Web サイトアクセスログから偽ショッピングサイトへのアクセスとタブ ID が一致し，アクセス日時が偽ショッピングサイトへのアクセスよりも古いものを対象として，最もアクセス日時が偽ショッピン

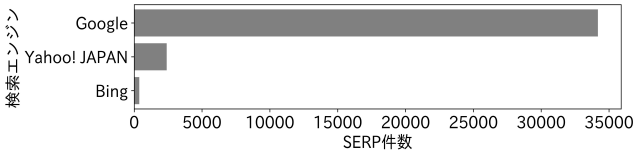

図 2 検索エンジン別の SERP 数
Fig. 2 Number of SERPs by search engine

グサイトへのアクセスと近いアクセスを関連付ける．偽ショッピングサイトへのアクセスの元タブ ID とが一致した Web アクセスについても同様の処理を行う．

(4) 関連付けを行ったアクセスのうち，ドメイン名が検索エンジンのドメイン名と一致し，URL に検索の際に用いられるクエリパラメータが存在する場合にユーザが検索を行ったとみなした．また，ユーザ ID，アクセス時刻，検索クエリ（URL のクエリパラメータから取得できる）を用いて SERP ログとの関連付けを行った．また，先程と同様の手順で偽ショッピングサイトへとアクセスするに至った検索行動の起点を探索した．

(5) 抽出した Web サイトアクセスログから偽ショッピングサイトへのアクセスとタブ ID が一致し，アクセス日時が偽ショッピングサイトへのアクセスよりも新しいものを対象として，最もアクセス日時が偽ショッピングサイトへのアクセスと近いアクセスを関連付ける．偽ショッピングサイトへのアクセスの元タブ ID とが一致した Web アクセスについても同様の処理を行う．

5. 実態調査

5.1 データセットの統計分析

今回の調査は 2023 年 5 月 31 日から 6 月 23 日までの 24 日間とした．なお，調査期間中に 1 日あたり 500 人程度のアクティブユーザが存在し，1 日あたり 250 人程度のユーザが Web 検索を 1 回以上行っていた．また，ユニークなアクティブユーザの合計は 1,161 人，Web 検索を行ったユニークなユーザの合計は 851 人であった．

調査期間内に 36,948 件の SERP から 513,553 件の URL を収集した．調査対象とした Google, Yahoo! JAPAN, Bing の各検索エンジン別の SERP 数を**図 2** に示す．Google での検索が全体の約 93%を占めるのに対し，Yahoo! JAPAN での検索は約 7%，Bing での検索は約 1%であった．これは，Google Chrome のデフォルトの検索エンジンが Google であることが影響していると考えられる．

ユーザは SERP に表示される検索結果の件数を選択することができ，Google では最大 100 件の検索結果を 1 つの SERP として表示させることができる．但し，ユーザが x 件の検索結果を表示する設定にしていたとしても，検索結果が x 件未満であれば表示される検索結果はその値となる．調査対象の SERP において，1 つの SERP に表示された検索結果の数を**図 3** に示す．SERP に表示される検索結果の件数は 10 件前後が最も多く，検索結果を 10 件ずつ表示するデフォルトの設定のまま検索エンジンを利用しているユーザが大半であることが推測される．一方で，50 件以上の検索結果を表示するユーザも見受けられた．

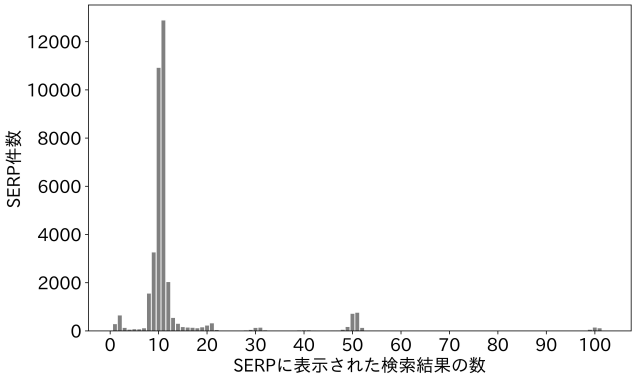

図 3 SERP に表示された検索結果の数
Fig. 3 Number of search results displayed in one SERP

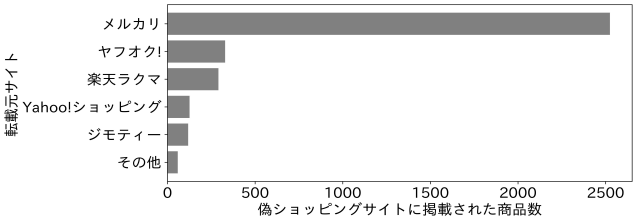

図 4 偽ショッピングサイトの商品情報の転載元サイト
Fig. 4 The site from which the product information on the fake shopping site is reprinted

前述の手法による判定の結果，3,489 件を踏み台サイトの URL，3,448 件を偽ショッピングサイトの URL と判定した．踏み台サイトと判定された URL と偽ショッピングサイトと判定された URL についてそれぞれドメイン名で集約すると，踏み台サイトは 1,219 件，偽ショッピングサイトは 2,650 件であった．偽ショッピングサイトの商品情報の転載元となっている正規サイトについての調査結果を**図 4** に示す．商品情報の転載元はメルカリが 7 割以上と突出して多く，ヤフオク!や楽天ラクマなども転載元として用いられていた．これらは日本で有名なフリマサイトであり利用者も多いため，掲載される商品数の転載元として利用されていると考えられる．

検索を行ったユーザの約 35%（302/851）は踏み台サイトを含む SERP を調査期間内に 1 回以上表示し，約 6%（51/851）のユーザは踏み台サイトを含む SERP を平均で 1 日 1 回以上表示していた．また，全体の SERP の約 5%（1,767/36,948）に踏み台サイトの URL が含まれていた．踏み台サイトの URL を含む SERP を表示した回数をユーザ毎に見ると，9 割以上のユーザが 10 回未満であったが，最大で 224 回表示したユーザが存在した．表示した回数にはばらつきが見られるが，これはユーザ毎に検索回数も大

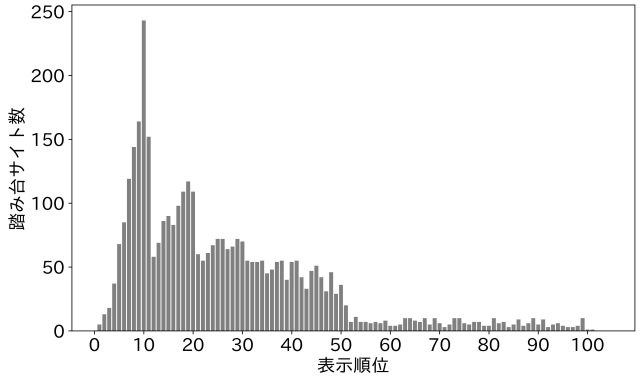

図 5 検索結果における踏み台サイトの表示順位
Fig. 5 Display position of stepping-stone sites in search results

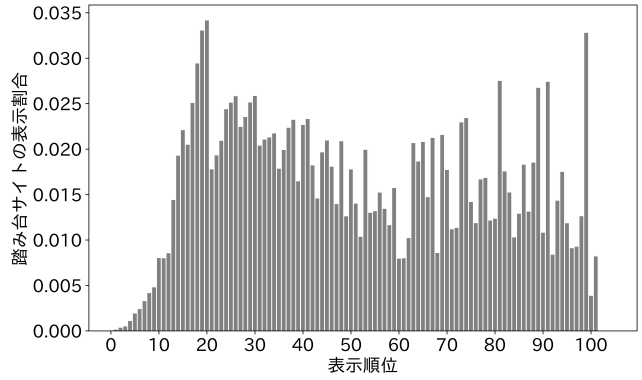

図 6 各順位別の踏み台サイトの表示割合
Fig. 6 Percentage of springboard sites by each ranking

きく異なることも影響しているものと考えられる．

5.2 ユーザの Web 検索の実態調査

踏み台サイトの URL を含む SERP が表示された際のユニークな検索クエリ数は 1,572 件であった．これは，調査期間に用いられた検索クエリ 30,375 件の約 5%にあたる．偽ショッピングサイトへとリダイレクトされる検索クエリの特徴について調査するため，手動で 1,572 件の検索クエリを確認した．その結果，約 34%はネットショッピングで購入可能な商品に関すると思われる検索クエリであり，その内半数以上が具体的な商品名や製品番号を含む検索クエリであった．また，会社名と思われる文字列を含む検索クエリが約 6%存在した．偽ショッピングサイトの商品情報はフリマサイトやショッピングサイトから転載されているため，出品・販売時に入力された商品名や製品番号及び会社名が転載され，検索結果に表示される状況になっていると考えられる．一方，それ以外の検索クエリには一見商品に関係しない文字列が含まれていた．例えば，「来週」という検索クエリで検索を行った場合には来週入荷予定との文言が説明に含まれた商品が表示されていた．商品に関すると思われる検索クエリによる踏み台サイトの平均表示順位は 21.4 位であり，それ以外の検索クエリによる踏み台サイトの平均表示順位は 29.1 位であった．このため，商品に関すると思われる検索クエリを用いた検索はユーザにとっての危険性も高くなると考えられる．

SERP には，何件目以降の検索結果をレスポンスとして返すかを示すパラメータが含まれている．また，このパラメータが含まれない場合は 1 件目以降の検索結果をレスポンスとして返す．検索結果の何件目（順位）に踏み台サイトが存在したかについて調査した結果を**図 5** に示す．この結果から，10 位付近がピークであり，20 位付近にも小さなピークが見られるが，50 位を超えると急激に減少する．これは，SERP に表示される検索結果の件数が 10 件や 50 件までに設定されていることも一因であると考えられる．そのため，x 位に表示された踏み台サイトの数を x 位以降を表示した SERP の数で除することで，踏み台サイトの表示割合を各順位別に調査した．調査の結果を**図 6** に示す．各順位における踏み台サイトの表示割合の平均は 1.5%であった．9 位までの踏み台サイトの表示割合は 0.5%未満と少ないが，20 位前後では約 3.0%となった．以上から，偽ショッピングサイトは検索結果の 20 位前後に表示される場合が多く，それ以降も概ね 1〜2%を推移する結果が得られ，一定のリスクが存在していることを確認した．50 位以降で割合にばらつきが生じているのは，50 位以降の検索結果を閲覧した事例が少ないことが影響していることが理由として考えられる．攻撃者は，ブラックハット SEO を行い踏み台サイトを上位の結果に表示させようとしているが，ユーザが最も閲覧すると考えられる 10 位以内に踏み台サイトを表示させることができている割合は少ない．

5.3 ユーザの検索行動の実態調査

偽ショッピングサイトへリダイレクトされたユーザの検索行動の調査結果を示す．はじめに，偽ショッピングサイトへリダイレクトされたアクセスは 26 件であった．また，偽ショッピングサイトへリダイレクトされたアクセスの内，SERP ログを発見できたのは 21 件であった．ユーザがアクセスした踏み台サイトの平均表示順位は 16.3 位であった．ユーザが 10 位以内の踏み台サイトへアクセスした事例は 6 件であり，20 位以内の踏み台サイトへアクセスした事例はこの 6 件を含めて 17 件であった．偽ショッピングサイトが SERP に表示された際の検索クエリにおいて，偽ショッピングサイトの商品情報の転載元のサイトが SERP に表示されている事例が 3 件存在した．これらの事例では転載元サイトが偽ショッピングサイトより上位に表示されていたが，転載元サイトへアクセスした事例は確認されなかった．一方で，SERP ログを発見できなかった 5 件についても，いずれも検索を起点としたアクセスであることを確認した．SERP ログから発見できなかった理由として，セキュリティエージェントが SERP ログを正しく収集できなかったことが考えられ，今後の課題とする．

表 1 Web 検索から偽ショッピングサイトへ誘導された事例
Table 1 Example of a behavior redirected to a fake shopping site via web search.

	アクセス日時	アクセス先	リファラ	ページ遷移タイプ	ページ遷移修飾子
(1)	2023/06/15 22:06:16	Google 検索（1 位〜10 位）	google	form_submit	—
(2)	2023/06/15 22:06:22	ショッピングサイト X（2 位）	google	link	—
(3)	2023/06/15 22:08:21	Google 検索（1 位〜10 位）	Web サイト X	form_submit	forward_back
(4)	2023/06/15 22:08:37	Google 検索（11 位〜20 位）	google	link	—
(5)	2023/06/15 22:08:56	踏み台サイト A（17 位）	google	link	—
(6)	2023/06/15 22:08:58	偽ショッピングサイト B	踏み台サイト A	link	client_redirect
(7)	2023/06/15 22:09:39	Google 検索（11 位〜20 位）	google	link	forward_back
(8)	2023/06/15 22:09:43	Web サイト Y (16 位)	google	link	—
(9)	2023/06/15 22:11:56	Google 検索（11 位〜20 位）	google	link	forward_back
(10)	2023/06/15 22:12:30	Google によるリダイレクト	google	—	—
(11)	2023/06/15 22:12:30	ショッピングサイト Z（順位不明）	google	link	server_redirect

これら 26 件のアクセスは全て Web 検索が起点であったため，検索は偽ショッピングサイトへと到達する際の入口となることが確認された．偽ショッピングサイトにアクセスした後に商品の購入を試みた形跡は確認されなかった．ユーザが偽ショッピングサイトへとアクセスする前後では，商品を探すために殆ど同じ検索クエリであるが一部の単語を少しずつ変更するなどして繰り返し検索を行う行動が 19 件，同一の検索クエリから複数の検索結果にアクセスする行動が 17 件，どちらにも当てはまらない検索行動が 2 件確認された．それぞれの事例について，誘導された偽ショッピングサイトにユーザがアクセスしたのは調査期間内で初めてであった．このような検索行動には，いずれも商品についての情報を得たいと切望するユーザの心理が反映されているものと推察され，普段は利用しない偽ショッピングサイトへのアクセスに至ったと考えられる．

5.4 偽ショッピングサイトへ到達した検索行動の事例

偽ショッピングサイトへと到達した検索行動の事例を**表 1** に示す．Web アクセスログの URL には適宜マスキングを行ってアクセス先として示した．また，Web サイトへのアクセスは青字として検索結果（黒字）と区別した．一連の検索行動において，Google 検索に用いたクエリはゲーミング PC の製品番号と思われる文字列で，5 回の検索全てにおいて同一であった．また，(1) から (7) までのアクセスのタブ ID は同一であった．(8) 及び (9) の 2 件のアクセスのタブ ID は同一であり ((1) から (7) までのアクセスのタブ ID とは異なる)，元タブ ID は (1) から (7) までのアクセスのタブ ID であった．これは，(8) 及び (9) のアクセスが行われたタブは (1) から (7) までのアクセスが行われたタブから新しく開かれたタブであることを示す．以上の結果から，この事例は同一の検索クエリから複数の検索結果にアクセスした検索行動である．

6. 考察

6.1 研究倫理

本研究で用いた SERP ログと Web サイトアクセスログは，セキュリティエージェントから得られたものである．ユーザはセキュリティエージェントをインストールする際に利用規約に同意する必要があり，Web サイトアクセスデータ及び HTML データを研究目的で収集・分析し，研究成果を発表することに同意を得ている．また，データを管理および利用する際には URL のクエリパラメータなどからユーザの個人情報が外部に漏洩しないように十分に留意し，論文中では検索クエリや URL などからユーザが特定されることのないよう適宜マスキングを行った．

6.2 ユーザ保護に向けた検討

ユーザが危険な Web サイトにアクセスする際に注意喚起を行う保護手法は広く利用されており，実際に Google Chrome には，ユーザが危険な Web サイトにアクセスしたりマルウェアをダウンロードしようとした際に警告画面を表示する機能が導入されている．偽ショッピングサイトについても同様に，アクセスしようとするユーザに対しても注意喚起を行うことで，偽ショッピングサイトによる被害を低減できる可能性がある．そのため，偽ショッピングサイトの URL リストに基づくユーザへの注意喚起のシミュレーションを行った．具体的には，5 章で検出した偽ショッピングサイトについて，検出した翌日にブロックリストとして配信した場合のユーザ保護効果を検証した．この結果，24 日間の観測期間中に 6 人のユーザによる 6 件の Web アクセスをブロックできることを確認した．数百人という小規模のユーザ集合の検索行動を 24 日間という短期間に分析したことで得られる簡易的なブロックリストであっても実際にユーザ保護の効果が確認でき，ユーザ数や分析期間を拡大すればさらに高い保護効果が期待できる．

また，偽ショッピングサイトへ到達してしまったユーザに対して事後的な通知を行うことで，偽ショッピングサイトの存在や過去の検索行動との関係をユーザ自身に知らせることで今後の検索行動の改善を狙うことを検討している．さらに実際に偽ショッピングサイトへと到達した際のユーザの心理状態や状況などについてアンケート調査することでより詳細な検索行動の把握と効果的な対策を検討する．

7. まとめ

数百人規模のアクティブユーザの Web アクセスログから SERP を分析するとともに，Web クローラを用いて表示された Web サイトにアクセスし，リダイレクト等の特徴により偽ショッピングサイトを判定することで，ユーザが Web 検索を起点として偽ショッピングサイトへと誘導される可能性のある状況について調査した．この結果，Web 検索が実際に偽ショッピングサイトへ到達する際の起点となっていることを確認した．ユーザを偽ショッピングサイトの脅威から保護するための取り組みが必要であり，対策の検討や通知実験の実施を今後の課題とする．

謝辞 本研究の一部は，国立研究開発法人情報通信研究機構 (NICT) の委託研究「Web 媒介型攻撃対策技術の実用化に向けた研究開発」の支援により得られた．本研究の一部は JSPS 科研費 JP21H03444 の助成を受けて行われた．

参考文献

[1] : Desktop Search Engine Market Share Japan — Statcounter Global Stats, StatCounter (online), available from ⟨https://gs.statcounter.com/search-engine-market-share/desktop/japan/⟩ (accessed 2023-08-15).

[2] Center, J. C. C.: Revealed Threat of Fake Store, Japan Cybercrime Control Center (online), available from ⟨https://www.jc3.or.jp/threats/upload/threats_JC3_APWG_Revealed_Threat_of_Fake_Store.pdf⟩ (accessed 2023-07-23).

[3] Chiba, D., Tobe, K., Mori, T. and Goto, S.: Detecting Malicious Websites by Learning IP Address Features, *2012 IEEE/IPSJ 12th International Symposium on Applications and the Internet*, pp. 29–39 (2012).

[4] Chiramdasu, R., Srivastava, G., Bhattacharya, S., Reddy, P. K. and Reddy Gadekallu, T.: Malicious URL Detection using Logistic Regression, *2021 IEEE International Conference on Omni-Layer Intelligent Systems (COINS)*, pp. 1–6 (2021).

[5] Desai, A., Jatakia, J., Naik, R. and Raul, N.: Malicious web content detection using machine leaning, *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, pp. 1432–1436 (2017).

[6] Forget, A., Pearman, S., Thomas, J., Acquisti, A., Christin, N., Cranor, L. F., Egelman, S., Harbach, M. and Telang, R.: Do or Do Not, There Is No Try: User Engagement May Not Improve Security Outcomes, *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*, Denver, CO, USENIX Association, pp. 97–111 (2016).

[7] Hesham, M., Ruben, T., Zhi-Li, Z., Sabyasachi, S. and Antonio, N.: Detecting malicious HTTP redirections using trees of user browsing activity, *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications* (2014).

[8] Khoo, E., Zainal, A., Ariffin, N., Kassim, M. N., Maarof, M. A. and Bakhtiari, M.: Fraudulent e-Commerce Website Detection Model Using HTML, Text and Image Features, *Proceedings of the 11th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2019)* (Abraham, A., Jabbar, M. A., Tiwari, S. and Jesus, I. M. S., eds.), Springer International Publishing, pp. 177–186 (2021).

[9] Kidmose, E., Stevanovic, M. and Pedersen, J. M.: Detection of Malicious domains through lexical analysis, *2018 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*, pp. 1–5 (2018).

[10] Pratiwi, M. E., Lorosae, T. A. and Wibowo, F. W.: Phishing Site Detection Analysis Using Artificial Neural Network, *Journal of Physics: Conference Series*, Vol. 1140, No. 1, p. 012048 (2018).

[11] Puppeteer: Puppeteer, Puppeteer (online), available from ⟨https://pptr.dev/⟩ (accessed 2023-07-24).

[12] Saha, I., Sarma, D., Chakma, R. J., Alam, M. N., Sultana, A. and Hossain, S.: Phishing Attacks Detection using Deep Learning Approach, *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, pp. 1180–1185 (2020).

[13] Sharif, M., Urakawa, J., Christin, N., Kubota, A. and Yamada, A.: Predicting Impending Exposure to Malicious Content from User Behavior, CCS '18, New York, NY, USA, Association for Computing Machinery, p. 1487–1501 (2018).

[14] Takahashi, T., Kruegel, C., Vigna, G., Yoshioka, K. and Inoue, D.: Tracing and Analyzing Web Access Paths Based on {User-Side} Data Collection: How Do Users Reach Malicious {URLs}?, *23rd International Symposium on Research in Attacks, Intrusions and Defenses (RAID 2020)*, pp. 93–106 (2020).

[15] Takashi, K., Daiki, C. and Mitsuaki, A.: To Get Lost is to Learn the Way: Automatically Collecting Multi-step Social Engineering Attacks on the Web, *Proceedings of the 15th ACM Asia Conference on Computer and Communications Security* (2020).

[16] Verma, R. and Das, A.: What's in a URL: Fast Feature Extraction and Malicious URL Detection, *Proceedings of the 3rd ACM on International Workshop on Security And Privacy Analytics*, IWSPA '17, Association for Computing Machinery, p. 55–63 (2017).

[17] WarpDrive: 国立研究開発法人 情報通信研究機構 (online), available from ⟨https://warpdrive-project.jp/⟩ (accessed 2023-07-22).

[18] 堺啓介，竹重耕介，加藤一樹，栗原直樹，大野克己，橋本正樹：fastText と LightGBM を用いた偽ショッピングサイト自動検出システムの開発，コンピュータセキュリティシンポジウム 2022 論文集，pp. 887–894 (2022).

[19] 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， Vol. 62, No. 9, pp. 1523–1535 (2021).

[20] 日本サイバー犯罪対策センター：悪質なショッピングサイト等に関する統計情報（2022 年），日本サイバー犯罪対策センター（オンライン），入手先 ⟨https://www.jc3.or.jp/threats/topics/article-