偽ショッピングサイトへ誘導する踏み台サイトの検出のための実態調査

道下 大悟[1]　小林 諭[2,a)]　山内 利宏[2,b)]

概要：近年，正規のショッピングサイトを模倣し金銭を詐取することを目的とした，偽ショッピングサイトによる被害が増加している．これまでに，Web 検索から偽ショッピングサイトへの誘導手口として，改ざんされた正規の Web サイト（以降，踏み台サイト）を経由することが明らかになっている．また，踏み台サイトではクローキングによる解析回避機能を有することが明らかになっている．このため，踏み台サイトの検出，および対処が実現できれば，偽ショッピングサイトに誘導される機会を抑制することができる．本稿では，踏み台サイトの検出のための有効な手法の実現を目的として，踏み台サイトの実態調査を行った結果を報告する．実態調査として，Web 検索に基づいて踏み台サイトを効率的に収集する検索条件を調査し，Top Level Domain と商品名に関するキーワードをクエリとして Google 検索を行うことが効率的な収集につながることを明らかにした．また，収集した踏み台サイトに対して，解析回避機能の起動条件を調査し，クローラやユーザのアクセス結果の違いを明らかにした．

キーワード：偽ショッピングサイト，踏み台サイト，解析回避機能

Investigation towards Detecting Springboard Websites for Fake Shopping Websites

Daigo Michishita[1]　Satoru Kobayashi[2,a)]　Toshihiro Yamauchi[2,b)]

Abstract: Recently, the number of victims of fake shopping websites that imitate legitimate shopping sites and aim to defraud people of money has been increasing. It has been shown that the fake shopping websites use springboard websites, which are defaced legitimate sites, as leading paths to themselves. In addition, springboard websites often have analytical evasion function. Therefore, we can remove the leading paths for fake shopping websites by detecting and addressing the springboard websites. In this paper, we collect and investigate the existing springboard websites for discussing a methodology for detecting them. We identified effective search terms for collecting springboard websites using search sites, and found that it is effective to use Google search with queries of TLD and product names. We also investigated the conditions for activating analytical evasion functions in the collected springboard websites, and clarified the differences in search results between crawlers and users.

Keywords: Fake shopping websites, Springboard websites, Analytical evasion function

1. はじめに

近年，正規のショッピングサイトを模倣し金銭を詐取することを目的とした，偽ショッピングサイトによる被害が増加している．例えば，日本サイバー犯罪対策センター（以降，JC3）が公表した「悪質なショッピングサイト等に

[1] 岡山大学 大学院環境生命自然科学研究科
Graduate School of Environmental, Life, Natural Science and Technology, Okayama University
[2] 岡山大学 学術研究院 環境生命自然科学学域
Faculty of Environmental, Life, Natural Science and Technology, Okayama University
[a)] sat@okayama-u.ac.jp
[b)] yamauchi@okayama-u.ac.jp

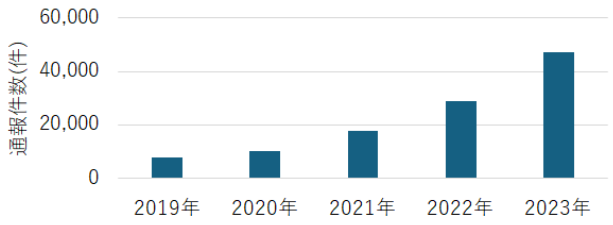

図 1 JC3 による悪質なショッピングサイト等の通報件数に関する統計情報

関する統計情報（2023 年）」[1] によると，セーファーインターネット協会から JC3 へ共有された悪質なショッピングサイト等の通報件数は，2023 年で 47,278 件となっており，2022 年の 28,818 件から約 1.6 倍に増加している（**図 1**）．こうした現状より，偽ショッピングサイトからユーザを保護するための対策が必要となっている．

これまでに，偽ショッピングサイトは改ざんされた正規の Web サイト（以降，踏み台サイト）を経由することが明らかになっている [1]．また，踏み台サイトは，クローキングによりユーザのみを偽ショッピングサイトへ誘導する機能（以降，解析回避機能）を有していることが明らかになっている [2]．これにより，ユーザのみが効率的に偽ショッピングサイトへ誘導され，被害が増加する要因となっている．

偽ショッピングサイトによる被害への対策として，偽ショッピングサイトの検出，および通報 [3] に基づくサイトの閉鎖などの早期対処が試みられている．しかし，偽ショッピングサイトの早期対処が実現できたとしても，攻撃者は，踏み台サイトの遷移先を新規に作成した偽ショッピングサイトに変更することで，再度ユーザは Web 検索から新しい偽ショッピングサイトへ同じ踏み台サイトを経由して誘導されてしまう．

そこで，検出対象として，偽ショッピングサイト本体ではなく，その踏み台サイトに着目する．踏み台サイトの検出と対処を行うことで，Web 検索から偽ショッピングサイトへ誘導される機会を減らし，ユーザを保護することが期待できる．

本稿では，日本語が含まれる偽ショッピングサイトへ誘導する踏み台サイトを検出するための有効な手法の実現を目指し，踏み台サイトの実態調査の調査方法と結果を報告する．実態調査として，Web 検索に基づいた踏み台サイトの収集を行い，踏み台サイトを効率的に収集する方法を調査した．また，踏み台サイトの検出に利用しうる有用な特徴を明らかにするために，収集した踏み台サイトに対して，解析回避機能に関する分析を行った．

本稿の主な貢献は以下の通りである．

- Web 検索に基づき踏み台サイトのドメイン名を 135 件収集した．また，Web 検索において，検索サイトとして Google を用いること，TLD (Top Level Domain) と商品名に関するキーワードをクエリにすることが踏み台サイトを効率的に収集するために有効であることを明らかにした．
- 収集した踏み台サイトに対して，解析回避機能について分析した結果，日本国内での使用率が高い検索エンジンのクローラがアクセスした際は，インデックス登録のための商品説明を記載した Web ページを多く表示することを明らかにした．また，ユーザがアクセスした際は，機器の違いによるアクセス制御が行われていることを明らかにした．

2. 関連研究

2.1 偽ショッピングサイトへ誘導する踏み台サイトに関する研究

踏み台サイトは多くのユーザを偽ショッピングサイトへ誘導するための SEO (Search Engine Optimization) ポイズニングの機能を果たしていることが明らかになっている．また，ユーザのみを偽ショッピングサイトへ誘導する解析回避機能を有していることが明らかになっている．

まず，踏み台サイトの SEO ポイズニングの機能に関して，才納ら [4] は，ユーザの Web アクセスログを分析し，Web 検索結果に偽ショッピングサイトへ誘導する踏み台サイトがどの程度存在しているのかを調査した．調査の結果，商品に関する検索クエリ実行時は平均 21.4 位に，それ以外の検索クエリでは平均 29.1 位に踏み台サイトが表示されることが明らかになった．

次に，踏み台サイトの解析回避機能に関して，小寺ら [2] は，偽ショッピングサイトや踏み台サイトにおけるクローキングによる解析回避機能を調査した．調査の結果，収集された踏み台サイトの内，HTTP リクエストヘッダ内の Referer に検索エンジンを設定した場合のみ偽ショッピングサイトにリダイレクトさせる踏み台サイトが 99.8%を占めることを明らかにした．また，HTTP リクエストヘッダ内の User-Agent に Googlebot などのクローラの文字列を設定した場合，検索インデックスに掲載されるための商品説明ページを応答するケースがあることが明らかになった．

2.2 偽ショッピングサイトの検出に関する研究

偽ショッピングサイトの検出は，これまで多くの研究で取り組まれている．特に，高精度な検出の実現のために機械学習などを用いる手法が複数提案されている [5–10]．

堺ら [5] は，偽ショッピングサイトの自動検出システムの開発を行った．まず，新規に登録されたドメイン名のリストから HTML のソースコードの取得を試みる．次に，取得した HTML のソースコードを fast text でベクトル化を行い，事前に学習した LightGBM を用いて，偽ショッピングサイトであるかの判定を行う．検出精度を評価するため，偽ショッピングサイトと正規の Web サイトを 1,000 件ずつ判定した結果，偽陽性 4 件，偽陰性 26 件で，正解率

が 98.5%であった．

また，Bitaab ら [6] は，偽ショッピングサイトを含む不正な電子商取引の Web サイト（以降，FCW）の検出器 Beyond Phish を開発した．この研究ではクラウドソーシングを通じて収集した FCW から特徴を分析・抽出し，ニューラルネットワークによる偽ショッピングサイトの判定を行う．検出精度を評価するために，自動収集した FCW の未経験のデータで判定を行った結果，誤検知率が 1.34%で検知率が 98.34%であった．

2.3 課題

2.1 節より，踏み台サイトはユーザのみを偽ショッピングサイトへ効率的に誘導させるために重要な役割を果たしていることがわかる．しかし，2.2 節で示したように，偽ショッピングサイトの検出に関する研究は進んでいるものの，踏み台サイトの検出技術については十分な議論がされていない．このため，踏み台サイトの検出を実現することを目的として，踏み台サイトの実態調査を行う．

3. 踏み台サイトの収集

3.1 Web 検索に基づいた踏み台サイトの収集

踏み台サイトの検出アプローチについて議論するためには，十分な数の踏み台サイトを収集し，共通する性質を調査・検討する必要がある．また，ユーザが被害を受ける可能性が高い事例を優先的に調査することが課題である．

先行研究 [4] より，ユーザが偽ショッピングサイトに誘導される流れをまとめたものを**図 2** に示す．ユーザはまず，検索サイトにて Web 検索を行うことで検索結果を得る．次に，検索結果に表示される踏み台サイトにアクセスすることで，リダイレクトが発生し，偽ショッピングサイトへ遷移される．

そこで，本稿では，ユーザが被害を受ける可能性が高い踏み台サイトを優先的に収集するために，Web 検索を起点として，踏み台サイトの収集を行う方法とその調査結果を報告する．まず，調査に有用な検索サイトを把握するために，検索エンジンの違いによる踏み台サイトの収集結果の比較を行う．次に，効率的な踏み台サイトの収集を行うために，踏み台サイトの収集数が多いキーワードと TLD の組み合わせを調査する．

また，調査過程で踏み台サイトは機能しており，リダイレクトが発生するが，遷移先の偽ショッピングサイトが閉鎖しているケースがあった．これについて，個別の調査を行った結果を報告する．

3.2 検索エンジンの違いによる踏み台サイトの収集結果

検索サイトが採用する検索エンジンの違いに着目して，踏み台サイトの収集結果の比較を行う．調査方法として，異なる検索エンジンを用いる 2 つの検索サイトそれぞれについて，以下の手順で調査を行った．

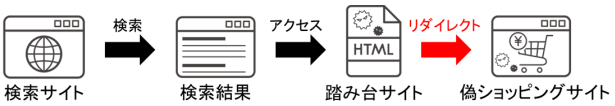

図 2 偽ショッピングサイトにアクセスするまでの流れ

表 1 検索エンジンの違いによる踏み台サイトの収集結果

	Google	Bing
踏み台サイトを 1 件以上発見したクエリ	23	0
踏み台サイト数	43	0
踏み台サイトのドメイン名数	28	0

(1) ユーザの商品検索を模倣したキーワードをクエリとして検索を行う．

(2) 検索結果の上位にアクセスを行い，踏み台サイトや偽ショッピングサイトであるかの判定を行う．

検索サイトとして，異なる検索エンジンを用いており，日本国内での使用率が高い Google と Bing を使用した [11]．ユーザの商品検索を模倣したキーワードとして，“格安” や “オンラインショッピング” といったショッピングサイトに関連する単語を使用した．また，商品に関するカテゴリ分けを行い “ファッション” や “家電” などのカテゴリ名，カテゴリ名に関連する会社名，および会社名に関連する商品名を使用した．

検索結果については，最大上位 50 位までを調査対象とする．これは，先行研究 [4] において，ユーザの Web アクセスログを分析した結果，検索結果上位 50 位を超えると踏み台サイト数が急激に減少することが明らかになっているためである．

偽ショッピングサイトの判断基準として，以下の条件に該当する Web サイトを，踏み台サイトを経由する偽ショッピングサイトの疑いがある Web サイトとして収集する．

- 検索結果に表示されている URL のドメイン名と異なるドメイン名の URL にリダイレクトされる．
- ショッピングサイトとしてのコンテンツを表示する．

また，先行研究 [4,12] を参考に，以下の条件から総合的に偽ショッピングサイトであるかの判断を行う．

- 商品の割引が過大である．
- 会社情報が記載されているが，住所や会社情報が実際には存在しないものである．
- 正規サイトの画像を読み込んでいる．

さらに，アクセス結果を踏み台サイトを経由する偽ショッピングサイトと判断した場合，検索結果に表示されている URL を踏み台サイトの URL として収集する．調査は 2024 年 8 月 12 日から 2024 年 8 月 14 日に行い，ショッピングサイトに関連する単語 12 件，カテゴリ名 22 件，会社名 23 件，商品名 115 件（合計 172 件）をキーワードとした．

調査結果を**表 1** に示す．調査の結果，Google を用いた検索では踏み台サイトを 1 件以上発見できたクエリが 23

件あり，合計 43 件の踏み台サイトを収集した．一方で，Bing を用いた検索では踏み台サイトを発見できなかった．これは，Google と比較して Bing は，日本語以外の他言語の Web ページや関連する商品の公式 Web ページを検索上位に優先的に表示する傾向であったためであると推察する．このことから，ユーザが被害を受ける可能性が高い踏み台サイトを収集するためには，検索サイトとして Google を使用することが有用であることが分かった．

Google を利用して踏み台サイトを発見した際のキーワードはすべて商品名であった．また，特定のカテゴリではなく，幅広いカテゴリの商品名で踏み台サイト，および偽ショッピングサイトを発見した．このことから，偽ショッピングサイトは特定のカテゴリの商品に限定して存在するわけではなく，幅広い商品をターゲットに存在していることが分かった．また，Web 検索による踏み台サイトの収集においては，幅広い商品名による検索が網羅的な踏み台サイトの収集につながると推察する．

3.3 キーワードと TLD を組み合わせた踏み台サイトの収集結果

3.2 節で収集した踏み台サイトの TLD の結果を**表 2** に示す．この内，最も多く確認した TLD は “.com” であった．また，先行研究 [2] でも収集した 2,996 件の踏み台サイトのドメイン名の内，最も多く確認された TLD は “.com” で 1,638 件あった．このことから，踏み台サイトで使用されている TLD には傾向があり，踏み台サイトの TLD を指定することで，より効率的にユーザが被害を受ける可能性が高い偽ショッピングサイトを収集できる可能性がある．

そこで，踏み台サイトの収集数が多いキーワードと TLD を組み合わせて以下の手順で調査を行った．

(1) 検索コマンドを用いて TLD を指定し，ユーザの商品検索を模倣したキーワードと合わせて検索を行う．
(2) 検索結果にアクセスを行い，踏み台サイトや偽ショッピングサイトであるかの判定を行う．

検索サイトは，3.2 節の結果に基づき収集効率が良い Google を用いる．また，TLD を指定するために “site:(TLD)（キーワード）” をクエリとして検索を行う．TLD として，先行研究 [2] で明らかになった踏み台サイトで使用頻度が高い 4 つの TLD (.com, .org, .ru, .net)，および日本語を含んだ偽ショッピングサイトを対象とするため “.jp” を使用する．なお，アクセスを行う件数，および踏み台サイトと偽ショッピングサイトの判定条件は 3.2 節と同じである．調査は 2024 年 8 月 15 日から 2024 年 8 月 18 日に行い，ショッピングサイトに関連する単語 3 件，カテゴリ名 3 件，会社名 6 件，商品名 30 件（合計 42 件）をキーワードとした．

調査結果を**表 3** に示す．調査の結果，踏み台サイトを 199 件，踏み台サイトのドメイン名を 76 件収集した．この結果より，検索コマンドで TLD を指定することで，指定した TLD の踏み台サイトを複数収集することができることが分かった．また，踏み台サイトのドメイン名数に着目した際，地理的制限なしに登録することができる gTLD (Generic Top Level Domain) に属する TLD (.com, .org, .net) を指定することで，踏み台サイトを多く収集することができた．このことから，踏み台サイトの TLD の種類と数には傾向があり，踏み台サイトに使用されている傾向が強い TLD を指定することで，効率的な収集ができることが分かった．

収集した踏み台サイトにおいて “.co.jp” に属する URL が確認された．“.co.jp” は通常取得の場合，日本で登記している企業が審査に基づいて登録可能 [13] であり正規のウェブサイトである可能性が高い．このことから，正規のウェブサイトが改ざんされ踏み台サイトとして利用されている可能性が高いと推察する．

表 2 3.2 節で収集した踏み台サイトの TLD（Google）

順位	TLD	踏み台サイトの件数
1	.com	10
2	.jp	2
2	.sk	2
2	.me	2
2	.cz	2

表 3 TLD とキーワードを組み合わせた場合の踏み台サイトの収集結果

	.com	.org	.ru	.net	.jp
踏み台サイトを 1 件以上発見したクエリ	16	24	23	21	15
踏み台サイト数	44	63	34	36	22
踏み台サイト数ドメイン名数	30	20	8	16	2

3.4 偽ショッピングサイトが閉鎖された踏み台サイトの調査

3.2 節や 3.3 節において，検索結果に表示された URL にアクセスを行った際，検索結果に表示された URL と異なるドメイン名の URL にリダイレクトが発生し，遷移先でサーバエラーレスポンスを表示するケースがあった．これは，偽ショッピングサイトは閉鎖されたものの，踏み台サイトは機能し続けているものである可能性がある．そこで，以下の条件に該当するサイトを偽ショッピングサイトが閉鎖後も機能している踏み台サイトとして収集した．

- 検索結果のタイトルがショッピングサイトに関連する．
- 検索結果に表示されている URL のドメイン名と異なるドメイン名の URL にリダイレクトさせる．
- 検索結果に表示されている URL のドメイン名直下にアクセスすると，検索時のタイトルに関連しないコンテンツが確認できる．

表 4 偽ショッピングサイト閉鎖後も機能している踏み台サイト数

	件数
1 件以上発見したキーワード数	84
踏み台サイトの数	193
踏み台サイトのドメイン名数	103

- 検索結果に表示されている URL のドメイン名を site 内検索すると，統一性のない商品のページタイトルが確認できる．

収集は 3.3 節で行った踏み台サイトの収集と同条件で行った．収集結果を**表 4** に示す．収集の結果，210 件のクエリで検索を行った際，偽ショッピングサイト閉鎖後も機能している踏み台サイトを 1 件以上確認できたクエリが 84 件あり，合計 193 件収集した．また，偽ショッピングサイト閉鎖後も機能している踏み台サイトのドメイン名数は 103 件収集できた．これは 3.3 節で収集した偽ショッピングサイトが機能している踏み台サイトのドメイン名数 76 件を上回っている．

また，偽ショッピングサイト閉鎖後も機能している踏み台サイトのドメイン名に対して，site 内検索を実行することで，閉鎖していない偽ショッピングサイトへ誘導している URL が検索結果の一部として得られるかを調査した．

調査の結果，52 件のドメイン名について，site 内検索の結果に閉鎖していない偽ショッピングサイトへの踏み台サイトが含まれることを確認した．

このことから，Web 検索において偽ショッピングサイト閉鎖後も機能している踏み台サイトのドメイン名は多数存在し，ほかの偽ショッピングサイトへの誘導に利用される可能性があり，潜在的な脅威がある．

4. 踏み台サイトの分析

4.1 踏み台サイトの解析回避機能に関する分析

踏み台サイトを検出する際の有用な特徴を特定するために，踏み台サイトが有する解析回避機能の現状を調査する．

先行研究 [2] では，踏み台サイトが有する解析回避機能の起動条件として，HTTP リクエストヘッダ内の User-Agent と Referer があることを明らかにしている．また，踏み台サイトをアクセスする対象として，ユーザのほかに Web サイトの情報を収集するクローラも考えられる．

そこで，本稿では，HTTP リクエストヘッダにおける User-Agent と Referer を変更することで，クローラとユーザによるアクセスを模倣し，踏み台サイトの挙動を調査した．なお，本調査では，ブラウザの自動化ツールである Selenium [14] を用いた．

4.2 クローラによるアクセス時の踏み台サイトの挙動分析

クローラによる踏み台サイトへのアクセス時の挙動を調査するために，User-Agent を検索エンジンのクローラに変更する．またクローラは検出した URL に直接アクセスを行うため，Referer は与えない．

表 5 クローラを模倣するための User-Agent の文字列

通番	User-Agent
1	`Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)`
2	`Mozilla/5.0 (compatible; bingbot/2.0; +http://www.bing.com/bingbot.htm)`
3	`Mozilla/5.0 (compatible; Yahoo! Slurp; http://help.yahoo.com/help/us/ysearch/slurp)`
4	`Mozilla/5.0 (compatible; YandexBot/3.0; +http://yandex.com/bots)`
5	`Mozilla/5.0 (compatible; Baiduspider/2.0; +http://www.baidu.com/search/spider.html)`

表 6 クローラを模倣してアクセスした結果（件数）

表 5 の通番（User-Agent）	1	2	3	4	5
踏み台サイトで商品説明表示	111	111	117	18	16
踏み台サイトでエラー表示	24	24	18	117	119

検索エンジンのクローラとして，日本国内での使用率が高い Google，Bing，および Yahoo!と，世界での使用率が高い YANDEX と Baidu を対象とした [11]．また，上記の方針に基づいて設定した 5 つの User-Agent の文字列（**表 5**）を，Selenium の ChromeOptions クラスにおける “`--user-agent`” の引数に指定することで，User-Agent を変更した．

調査は 2024 年 8 月 14 日から 8 月 19 日に行い，3.2 節，3.3 節，および 3.4 節で収集した 435 件の踏み台サイトのうち，ドメイン名が重複しない 135 件の踏み台サイトを調査した．

調査結果を**表 6** に表す．表 6 において，表頭の番号は表 5 の通番に対応する User-Agent をアクセスに用いたことを示す．

表 6 から，クローラを模倣してアクセスした結果，踏み台サイトとは異なるドメイン名の URL へのリダイレクトが行われず，商品説明を記載した Web ページが表示されるケース，および踏み台サイトとは異なるドメイン名の URL へのリダイレクトが行われずクライアントエラーやサーバエラーを返すケースがあったことが分かる．

また，表 6 より，日本国内で使用率が高い Google，Bing，および Yahoo!（表 6 の通番 1，通番 2，および通番 3）のクローラの User-Agent を設定した場合，海外で使用率が高い YANDEX，Baidu（表 6 の通番 4，および通番 5）のクローラの User-Agent を設定した場合よりも，商品説明を記載した Web ページを表示する割合が多いことが分かる．これは，収集した偽ショッピングサイトが日本語を含んだものを対象としており，検索エンジンへのインデックス登録を意図しているものと推察する．

4.3 ユーザによるアクセス時の踏み台サイトの挙動分析

ユーザによる踏み台サイトへのアクセス時の挙動を調査するために，User-Agent については，PC とモバイル端末を対象とする．これは，近年のインターネットショッピングでは，携帯電話・タブレットでの購入頻度と PC での購入頻度が高いためである [15]．また，Referer については PC とモバイル端末で使用率が高い検索エンジンを対象とする．これは，ユーザが Web 検索に基づいてショッピングサイトへアクセスを行うためである．

PC の User-Agent として，Windows の User-Agent を用いる．また，モバイル端末の User-Agent として，Android と iOS の User-Agent を用いる．Referer として，設定しない場合と PC とモバイル端末で使用率が高い 4 つ (Google, Bing，Yahoo!，DuckDuckGo) を対象とする [11]．

上記方針に基づいて設定した User-Agent と Referer の設定方法を**表 7**，および**表 8** に示す．User-Agent の変更方法は 4.2 と同様である．Referer の変更方法として，まず Referer に設定する検索エンジンを用いている検索サイトにアクセスを行う．次に，検索サイト上で任意のページにリダイレクトする JavaScript コードを実行することで Referer を設定した状態でのアクセスを行った．調査は 2024 年 8 月 14 日から 19 日に行い，4.2 節と同じ 135 件の踏み台サイトを対象とした．

調査結果を**表 9** に示す．表 9 において，表頭の 1 行目の番号は表 7 の通番に対応する User-Agent をアクセスに用いたことを示す．また，表頭の 2 行目の番号は表 8 の通番に対応する Referer をアクセスに用いたことを示す．

表 9 から，ユーザを模倣してアクセスした結果，4.2 節のアクセス結果に加えて，踏み台サイトのドメイン名と異なるドメイン名の URL にリダイレクトが発生し，偽ショッピングサイトへ到達するケース，踏み台サイトのドメイン名と異なるドメイン名の URL にリダイレクトが発生し，クライアントエラーやサーバエラーを返すケースがあったことが分かる．さらに，踏み台サイトにアクセスしようとするとアクセス拒否が発生するケースがあったことが分かる．

表 9 より，Referer を設定しない場合（表 8 の通番 1），踏み台サイトとは異なるドメイン名の URL へのリダイレクトが行われないケースが多いことが分かった．また，Referer として，Google と Yahoo!を指定した場合（表 8 の通番 2，および通番 4），偽ショッピングサイトへ遷移するケースが多いことが分かった．さらに，モバイル端末の User-Agent について，Android と iOS（表 7 の通番 2，および通番 3）ではアクセス結果が異なる場合があり，機種の違いなどの粒度で解析回避機能が起動していると推察する．

自動化ツール (Selenium) を用いない場合は偽ショッピングサイトに到達するが，自動化ツールを用いる場合は踏み台サイトから別のドメイン名の URL へのリダイレクトが行われず，商品説明のページを表示するケースが 5 件あった．このことから，User-Agent や Referer 以外の情報が解析回避機能に利用されている可能性があり，今後さらに調査が必要と考えられる．

表 7 ユーザに模倣するための User-Agent の文字列

通番	User-Agent
1	`Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/126.0.0.0 Safari/537.36Edg/126.0.0.0`
2	`Mozilla/5.0 (Linux; Android 13; Pixel 7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/107.0.0.0 Mobile Safari/537.36`
3	`https://www.bing.com/Mozilla/5.0 (iPhone; CPU iPhone OS 16_2 like Mac OS X) AppleWebKit/605.1.15 (KHTML, like Gecko)GSA/247.0.501893814 Mobile/15E148 Safari/604.1`

表 8 ユーザに模倣するための Referer の文字列

通番	Referer
1	なし
2	`https://www.google.com/`
3	`https://www.bing.com/`
4	`https://www.yahoo.co.jp/`
5	`https://duckduckgo.com/`

5. 研究倫理

論文投稿による悪影響の有無を確認するためにサイバーセキュリティ研究倫理に関するチェックリスト [16] に基づくセルフチェックを行った．

踏み台サイトは正規の Web サイトが改ざんされた可能性がある．このため，踏み台サイトの URL の記載は脆弱な Web サイトに対する攻撃の標的につながる．また，誤判定による正常なショッピングサイトの URL の公表は名誉の毀損につながる．以上のことから，踏み台サイトと判定した URL は記載していない．

解析回避機能の分析を行うために踏み台サイトへの複数回の自動アクセスを行った．この際，短期間で集中したアクセスは DoS 攻撃となりうる．このため，本調査ではアクセスの実行間隔を意図的に開けるように設定した．また，調査では，継続したアクセスは行わず 1 回の実験限りの調査にとどめた．

6. おわりに

偽ショッピングサイトへ誘導する踏み台サイトの検出を実現するために，踏み台サイトの実態調査を行った．実態調査として，Web 検索に基づいて踏み台サイトを効率的に収集する手法を調査し，検索サイトとして Google を用いることや TLD と商品名に関するキーワードをクエリとすることが効率的であることを明らかにした．また，収集し

表 9 ユーザを模倣してアクセスした結果（件数）

User-Agent（表 7 の通番に対応）	表 7 の通番 1					表 7 の通番 2					表 7 の通番 3				
Referer（表 8 の通番に対応）	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
踏み台サイトで商品説明を表示	10	7	9	7	9	10	6	9	5	9	10	6	9	5	9
踏み台サイトでエラーを表示	121	21	56	29	116	120	22	55	27	119	118	20	56	26	116
偽ショッピングサイトを表示	2	95	62	88	9	3	95	63	93	7	5	97	63	94	9
偽ショッピングサイトでエラーを表示	0	12	8	11	1	0	12	8	10	0	0	12	7	10	1
アクセス拒否	2	0	0	0	0	2	0	0	0	0	2	0	0	0	0

た踏み台サイトの分析の結果，日本国内での使用率が高い検索エンジンのクローラでは，踏み台サイト内にて，インデックス登録のための商品説明を記載した Web ページを多く表示することを明らかにした．さらに，ユーザによるアクセスでは，機種などの違いによるアクセス制御が行われていることを明らかにした．

今後は，Referer と User-Agent 以外の解析回避機能の起動条件を調査するとともに，明らかになった特徴から偽ショッピングサイトへ誘導する踏み台サイトを検出するための有効な手法を検討する．

参考文献

[1] 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（2023 年）, (Online), available from ⟨https://www.jc3.or.jp/threats/topics/article-555.html⟩ (accessed 2024-8-22).

[2] 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).

[3] 一般社団法人セーファーインターネット協会: 悪質 EC サイトホットライン 通報フォーム, (Online), available from ⟨https://www.saferinternet.or.jp/akushitsu_ec_form/⟩ (accessed 2024-8-22).

[4] 才納明英, 高田一樹, 藤田彬, 小出駿, 金井文宏, 秋山満昭, 田辺瑠偉, 吉岡克成, 松本勉：Web 検索から偽ショッピングサイトへの誘導の実態調査, コンピュータセキュリティシンポジウム 2023(CSS2023) 論文集, Vol. 2023, pp.620–627 (2023).

[5] Sakai, K. Takeshige, K. Kato, K. Kurihara, N. Ono, K. Hashimoto, M.：An Automatic Detection System for Fake Japanese Shopping Sites Using fastText and LightGBM, *IEEE Access*, Vol. 11, pp. 111389–111401 (2023).

[6] Bitaab, M. Cho, H. Oest, A. Lyu, Z. Wang, W. Abraham, J. Wang, R. Bao, T. Shoshitaishvili, Y. Doupé, A.：Beyond Phish: Toward Detecting Fraudulent e-Commerce Websites at Scale, *2023 IEEE Symposium on Security and Privacy (SP)*, pp.2566–2583 (2023).

[7] Carpineto, C. Romano, G.:Learning to detect and measure fake ecommerce websites in search-engine results, *WI' 17: International Conference on Web Intelligence 2017*, pp.403–410 (2017).

[8] Mostard, W. Zijlema, B. Wiering, M.:Combining Visual and Contextual Information for Fraudulent Online Store Classification, *WI'19: IEEE/WIC/ACM International Conference on Web Intelligence*, pp.84–90 (2019).

[9] Manek, A.S. Shenoy, P.D. Mohan, M.C. Venugopal, K. R.:Detection of fraudulent and malicious websites by analysing user reviews for online shopping websites, *International Journal of Knowledge and Web Intelligence*, Vol.5, pp.171–189 (2016).

[10] Beltzung, L. Lindley, A. Dinica, O. Hermann, N. Lindner, R.:Real-Time Detection of Fake-Shops through Machine Learning, *2020 IEEE International Conference on Big Data*, pp.2254–2263 (2020).

[11] StatCounter:Statcounter Global Stats - Browser, OS, Search Engine including Mobile Usage Share, (Online), available from ⟨https://gs.statcounter.com/⟩ (accessed 2024-8-22).

[12] 警察庁: 偽ショッピングサイト・詐欺サイト対策, (Online), available from ⟨https://www.npa.go.jp/bureau/cyber/countermeasures/fake-shop.html⟩ (accessed 2024-8-22).

[13] JPDirect:CO.JP ドメインの取得, (Online), available from ⟨https://jpdirect.jp/domain/register/cojp/⟩ (accessed 2024-8-22).

[14] Selenium:Selenium, (Online), available from ⟨https://www.selenium.dev/⟩ (accessed 2024-8-22).

[15] 株式会社 NTT ドコモ モバイル社会研究所: モバイル社会白書 2023 年版, (Online), available from ⟨https://www.moba-ken.jp/whitepaper/wp23/pdf/wp23_all.pdf⟩ (accessed 2024-8-22).

[16] IWSEC: サイバーセキュリティ研究倫理に関するチェックリスト, (Online), available from ⟨https://www.iwsec.org/csec/ethics/checklist.html⟩ (accessed 2024-8-22).