Vol.2020-DPS-182 No.41
Vol.2020-CSEC-88 No.41
2020/3/13

偽ショッピングサイトを起点とする攻撃の実態調査

小寺 博和[1,a)]　小出 駿[1,2]　千葉 大紀[1]　青木 一史[1]　秋山 満昭[1]

概要：ユーザから個人情報や金銭を窃取することを目的としたフィッシングによる被害が拡大している．フィッシングの 1 つとして，攻撃者が用意した正規のショッピングサイトになりすました偽のショッピングサイトにユーザを誘導し，ユーザから商品の購入代金を騙し取ることを目的とした攻撃が確認されている．偽ショッピングサイトは，存在しない企業名を名乗ることや不自然な日本語表記がある等の特徴があるが，脆弱性の悪用やマルウェアのダウンロードのような機能がないため，アンチウイルスエンジンによる検知が困難である．ユーザの被害を抑えるための対策検討のためには，偽ショッピングサイトを起点とした攻撃の実態を明らかにすることが必要である．本研究では，偽ショッピングサイトが持つ解析回避機能や攻撃の実態に関する調査結果を報告する．URL ブラックリストに掲載されたドメイン名を検索エンジンで検索することで，偽ショッピングサイトの疑いのある URL を収集した．偽ショッピングサイトに対して条件を変更した複数の HTTP リクエストを送信し，その応答を分析することで，解析回避機能や攻撃者が用いる構築ツールの実態を明らかにした．

1. はじめに

フィッシングはユーザから個人情報や金銭を窃取することを目的とした攻撃として知られている．Anti-Phishing Working Group（APWG）によると 2019 年上半期に約 36 万件のフィッシングサイトが検出され，2018 年下半期に検出された約 29 万件よりさらに増えており，フィッシングによる被害は拡大傾向にある [1].

フィッシングによる攻撃の 1 つとして，正規のショッピングサイトになりすました偽のショッピングサイトにユーザを誘導し，ユーザから購入代金を騙し取ることやユーザに偽のブランド商品を購入させることを目的とした攻撃が確認されている．日本国内においては日本人をターゲットとした正規のショッピングサイトを装った偽のショッピングサイトが多く存在している．偽ショッピングサイトはフィッシングサイトと同様にユーザから金銭を窃取することを目的としているが，ユーザがアクセスする契機が従来のフィッシングサイトとは異なる．従来のフィッシングサイトはメールに記載された URL をクリックすることでフィッシングサイトに直接アクセスすることを攻撃者は期待する．一方で，今回着目する偽ショッピングサイトはユーザが商品名や商品型番等を検索エンジンで検索し，検索結果に掲載された URL をクリックすることで偽ショッピングサイトにアクセスすることを期待する．Japan Cybercrime Control Center（JC3）によると，2017 年に 19,834 件の偽ショッピングサイト URL が確認されている [2]．JC3 は偽ショッピングサイトを “Fake Store” と定義し，その攻撃手法には 3 つの特徴があることを明らかにした．3 つの特徴を用いて犯罪者グループを分類した結果，2017 年に観測された 19,834 件の “Fake Store” による攻撃には 6 つの犯罪者グループが存在したことを明らかにした．

Drive-by Download 攻撃に利用されるウェブサイトやフィッシングサイトには，HTTP リクエストヘッダやアクセス元 IP アドレスでアクセスしたユーザの環境を判別する手法であるクローキングを用いることで解析を回避する機能があることが知られている [3], [4], [5]．偽ショッピングサイトも同様にクローキングによる解析回避機能がある可能性が考えられる．しかしながら，既存研究 [2] では改ざんされた正規のウェブサイト（以降，踏み台サイトとする）を経由して偽ショッピングサイトへ到達することが明らかにされているが，偽ショッピングサイトが有する解析回避機能については明らかにされていない．偽ショッピングサイトが有する解析回避機能を明らかにすることで，解析者はより効率的に調査することが可能になる．

本研究では，偽ショッピングサイトの URL を収集し，偽ショッピングサイトが有する解析回避機能の調査と攻撃

[1] NTT セキュアプラットフォーム研究所
NTT Secure Platform Laboratories, Musashino, Tokyo 180–8585, Japan
[2] 横浜国立大学
Yokohama National University, Yokohama, Kanagawa 240–8501, Japan
[a)] hirokazu.kodera.dh@hco.ntt.co.jp

規模の推定を実施した．具体的には，まず URL ブラックリストに掲載された URL のドメイン名を対象に，検索エンジンで改ざんされる可能性のある正規のウェブサイトのドメイン名配下の URL を収集する．次に，収集した URL に対して異なる HTTP リクエストヘッダを設定した複数の HTTP リクエストを送信した際の応答を観測することで解析回避機能を分析した．最後に，偽ショッピングサイトのドメイン名を対象に Passive DNS のデータを用いて攻撃規模を推定した．

本研究で明らかになったことは以下の通りである．

- URL ブラックリストに掲載された URL のドメイン名（14,052 件）を検索エンジンで検索した結果から偽ショッピングサイトの踏み台サイトのドメイン名を 251 件収集した．また，ドメイン名配下の URL にアクセスした際に誘導される偽ショッピングサイトのドメイン名を 875 件収集した．
- 踏み台サイトのうち 93.2%が検索エンジンサイト経由でアクセスした場合のみ偽ショッピングサイトへ誘導する解析回避のための機能を有することを明らかにした．また，87.6%が日本で利用されることが多い Google，Yahoo! JAPAN 等の検索エンジンサイトを経由した場合のみ偽ショッピングサイトへ誘導されることを明らかにした．
- Passive DNS のデータを用いて偽ショッピングサイトによる攻撃キャンペーンの規模を調査したところ，1 ドメイン名あたり平均 43.1 回，最大 302 回の名前解決があり，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．

本論文の構成は以下の通りである．2 節で関連研究を示す．3 節で偽ショッピングサイトによる攻撃手法の全体概要を述べる．4 節で URL ブラックリストを用いた偽ショッピングサイト URL の収集方法と，偽ショッピングサイトが持つ解析回避機能の調査方法に関して述べる．5 節で偽ショッピングサイトが持つ解析回避機能の調査結果と Passive DNS による攻撃規模の推定結果について述べる．6 節で本研究における制約，今後の課題について述べ，7 節で本論文のまとめを行う．

2. 関連研究

偽ショッピングサイトに関する研究として，偽ブランド商品の EC サイトを検出する手法に関する研究が報告されている．Wadleigh ら [6] は，WHOIS 情報，価格設定，Web サイトコンテンツをもとにした偽ブランド商品を販売するウェブサイトの検出手法を提案した．2014 年 1 月から 8 月に収集された検索エンジンでの検索結果のうち 32%が偽ブランド商品を販売するウェブサイトであることを明らかにした．Carpineto ら [7] は，検索エンジンからアクセス可能なウェブサイトのコンテンツから正規か偽のショッピングサイトであるかの判定する手法と，偽造のリスクを分析するための “Counterfeiting Charts” を生成する手法を提案した．39 種の靴のブランドを対象に検索エンジンで検索した結果，3,601 件のウェブサイトから 209 件の偽ショッピングサイトが提案手法により検出された．

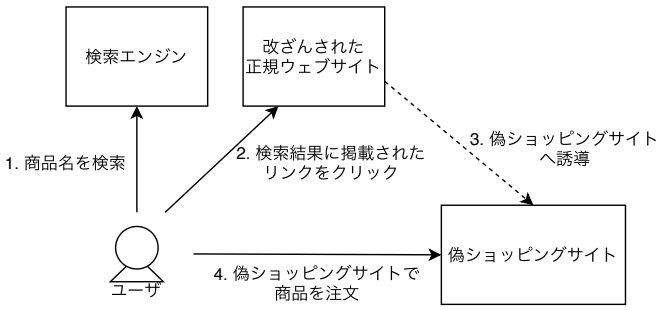

図 1 偽ショッピングサイトの全体概要

フィッシングサイトに関する研究として，フィッシングを行う攻撃者やサーバ環境等の攻撃用インフラに関する研究が報告されている．Han ら [8] は，著者らが設置したハニーポットを攻撃者がフィッシングサイトに改ざんする様子を観測することで，攻撃者の行動分析や被害者数の推定を行った．被害者数を調査した結果，2,438 人のユーザがフィッシングサイトにアクセスし，そのうち 215 人（9%）が認証情報を送信していたことを明らかにした．Oest ら [4] は，フィッシングサイトの構築ツールであるフィッシングキットを解析し，アクセスブロックの対象とされやすい IP アドレスの国別の傾向や，組織別の傾向を調査した．検索エンジンクローラ，セキュリティベンダ，フィッシングの対象となるブランドの企業（Paypal 社，Apple 社等）からのアクセスがブロックされる傾向にあることを示した．小寺ら [5] は，インターネット上に実在するフィッシングサイトを対象にアクセス妨害機能を持つフィッシングサイトを調査した．HTTP リクエストのヘッダ（User-Agent，Referer）を設定してアクセスしてその応答を観測した結果，アクセス妨害機能を持つフィッシングサイトが 10.4%存在することを明らかにした．

偽ショッピングサイトを検出する手法や，一般的なフィッシングサイトに関する攻撃者や攻撃用インフラの分析はされているが，偽ショッピングサイトに関しては解析回避機能のような攻撃インフラに関する調査は報告されていない．

3. 偽ショッピングサイトによる攻撃の概要

本節では偽ショッピングサイトの攻撃手法の概要について述べる．ユーザが偽ショッピングサイトに誘導されるまでの全体概要を図 1 に示す．ユーザが商品名や商品型番を検索エンジンで検索すると，図 2 のように検索結果に正規

Vol.2020-DPS-182 No.41
Vol.2020-CSEC-88 No.41
2020/3/13

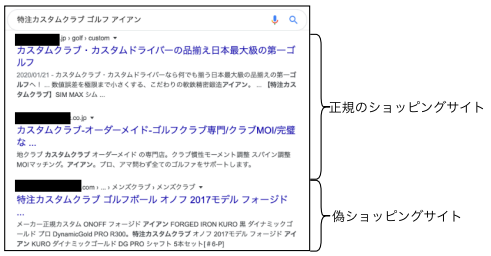

図 2　検索エンジンの検索結果

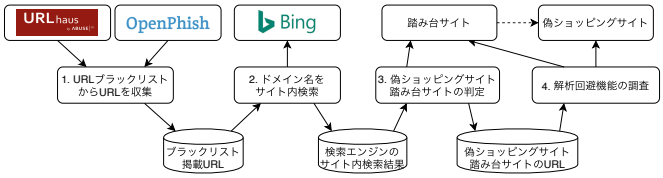

図 3　偽ショッピングサイト収集方法

のショッピングサイト以外に偽ショッピングサイトへ誘導するためのリンクが含まれる場合がある．検索結果に掲載された URL は偽ショッピングサイトのものではなく，攻撃者がリダイレクタとして利用するために改ざんした正規のウェブサイト（踏み台サイト）であることが知られている [2]．ユーザが踏み台サイトへのリンクをクリックすると，踏み台サイトから偽ショッピングサイトへ誘導される．偽ショッピングサイトは正規のショッピングサイトと同様に商品の注文機能があり，利用ガイド等のユーザ向けのマニュアルも掲載されていることから見た目の区別が付きにくく，ユーザは偽ショッピングサイトと気付かずに注文してしまう．偽ショッピングサイトでユーザが商品を注文し，購入代金を攻撃者に支払うことで攻撃が成立し，攻撃者に購入代金を窃取される．

4. 調査方法

偽ショッピングサイトが有する解析回避機能を調査するために，URL ブラックリストに掲載された URL を用いて偽ショッピングサイトを収集した．本節では，調査方法の詳細を述べる．

4.1 全体の概要

調査の全体概要を図 3 に示す．まず，検索エンジンで検索するドメイン名を収集するために，URL ブラックリストに掲載された URL を収集した．次に，入手した URL のドメイン名を検索エンジンでサイト内検索することで当該ドメイン名配下の URL を収集した．次に，サイト内検索で収集した URL に対してアクセスし，偽ショッピングサイトへ誘導されるかを確認することで偽ショッピングサイトの踏み台サイトであるかを判定した．最後に，HTTP リクエストに特定のヘッダを付加した複数の HTTP リクエストを送信し，その応答を確認することで偽ショッピングサイトが有する解析回避機能を調査した．

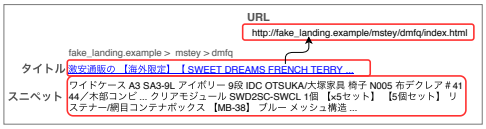

図 4　検索エンジンの検索結果の各部説明

4.2 URL ブラックリストに掲載された URL の収集

ユーザが検索エンジンで商品名等を検索すると偽ショッピングサイトに誘導する検索結果が掲載される場合があるが，3 節で述べたようにそれらの URL は偽ショッピングサイトの踏み台サイトとなる．踏み台サイトは改ざんされた正規のウェブサイトであるため，改ざんされる可能性が高いウェブサイトのドメイン名を検索エンジンでサイト内検索することで，踏み台サイトの可能性が高い URL を効率的に収集できる．

マルウェアの配布サイトや C&C サーバは正規のウェブサイトが攻撃者によって改ざんされたケースが存在することが知られている [9]．また，フィッシングサイトも同様に正規のウェブサイトの改ざんによるケースが存在することが知られている [10]．そこで，改ざんされる可能性が高いウェブサイトのドメイン名を収集するために，マルウェアを対象としたブラックリストである URLhaus [11] と，フィッシングサイトを対象としたブラックリストである OpenPhish [12] に掲載された URL を収集した．

4.3 検索エンジンを用いた踏み台サイト URL の収集

4.2 節で得た URL を用いて，それぞれのドメイン名配下の踏み台サイトの候補となる URL をサイト内検索を用いて収集した．本研究では，Bing Web Search API を用いてサイト内検索を実施した．サイト内検索により得られた検索結果は，図 4 のようにタイトル，URL，スニペットから主に構成される．検索結果のうち，踏み台サイトへのリンクはユーザを偽ショッピングサイトに誘導するために，タイトルやスニペットには日本語で記述された商品説明が含まれる．そこで，本研究ではスニペットに日本語が含まれている URL を踏み台サイトの疑いのある URL として収集した．

4.4 偽ショッピングサイト判定

4.3 節で得た踏み台サイトの疑いのある URL にアクセスすることで踏み台サイトであるかを判定し，踏み台サイトから誘導される偽ショッピングサイトの URL を収集した．踏み台サイトから偽ショッピングサイトへのリダイレクトは HTTP レスポンスの Location ヘッダによるものだけではなく，JavaScript や HTML を用いたものも想定される．そこで，Web ブラウザ自動化ツールである Selenium [13]

表 1 Selenium に付加した Referer ヘッダ

No.	Referer
1	https://www.google.co.jp/
2	https://www.bing.com/
3	https://www.yahoo.co.jp/
4	https://www.yahoo.com/

を用いて踏み台サイトにアクセスすることで，JavaScript や HTML によるリダイレクトを観測した．

ユーザが検索エンジンサイト経由で踏み台サイトへアクセスした場合のみ，偽ショッピングサイトにユーザを誘導する挙動を有することが知られている [2]．そこで，検索エンジンサイトの URL を Referer ヘッダに付加した状態で踏み台サイトにアクセスすることでリダイレクタを動作させる方式を取った．特定の検索エンジン経由の場合のみ誘導される場合も想定されるため，本研究では踏み台サイトであるかの判定をするために表 1 の 4 種類の Referer ヘッダをそれぞれ付加して踏み台サイトに計 4 回アクセスした．

Selenium で踏み台サイトへアクセスし，異なるドメイン名の URL へリダイレクトされた場合はリダイレクト先のコンテンツを手動で確認した．リダイレクト先が偽ショッピングサイトであった場合は，それぞれ踏み台サイトと偽ショッピングサイトの URL であるとそれぞれ判定した．

4.5 解析回避機能の調査

4.4 節で特定した踏み台サイトに対して条件を変更した HTTP リクエストを複数回送信することで，踏み台サイトが有する解析回避機能を調査した．踏み台サイトに対して，表 2 に示す User-Agent と Referer のパターンを用いて踏み台サイトにアクセスすることで，踏み台サイトが有するアクセス回避機能を調査した．

No.1–6 でそれぞれアクセスすることで，検索エンジンボットによるアクセスの場合にのみ検索エンジンにインデックスするためのコンテンツを応答するかの判定ができる．No.6–13 でそれぞれアクセスすることで，ユーザがどの検索エンジン経由でのアクセスの場合に偽ショッピングサイトへ誘導されるかの判定ができる．調査対象とする検索エンジンサイトとして，国内で利用されることが多い Google，Yahoo! JAPAN，Bing に加え，海外で利用されることが多い Yandex，Yahoo!，Baidu をそれぞれ用いた．Google については海外で利用される検索エンジンサイトとして，TLD が.uk である google.co.uk も調査対象として用いた．これ以降，google.co.jp，google.co.uk をそれぞれ Referer としてアクセスした場合の表記を Google（JP），Google（UK）とする

5. 調査結果

本節では，4 節の手法による偽ショッピングサイトの収集結果と解析回避機能の分析結果について述べる．次に，偽ショッピングサイトの構築ツールの分析結果について述べる．最後に，Passive DNS のデータをもとにした偽ショッピングサイトの攻撃キャンペーン規模の推定結果を述べる．

5.1 偽ショッピングサイトの収集結果

URLhaus，OpenPhish に 2019/12/20–2020/2/7 に掲載された URL をもとに収集した偽ショッピングサイトの収集結果を表 3 に示す．URL ブラックリストから取得した URL は OpenPhish と URLhaus から合計 75,610 件で，ユニークなドメイン名 14,052 件を Bing Web Search API によるサイト内検索を実施した．サイト内検索の検索結果のうち，スニペットに日本語が含まれていたものは 1,788 件あった．これらの URL に対して 4.4 節の手法でアクセスし，偽ショッピングサイトの踏み台サイトとして動作していると確認されたドメイン名の数は 251 件で，スニペットに日本語が含まれるサイト内検索結果のドメイン名のうち 14.0%に留まった．この理由の 1 つとして，検索結果のスニペットに日本語を含む場合であっても，すでに改ざんが修正された場合やウェブサイトが閉鎖された場合があったことが考えられる．

251 件の踏み台サイトから 875 件の偽ショッピングサイトに到達した．同一ドメイン名の踏み台サイトから複数の偽ショッピングサイトへ誘導されるということが確認された．また，異なる踏み台サイトから共通の偽ショッピングサイトに誘導される場合も確認され，最も多いもので 10 件の踏み台サイトから誘導される偽ショッピングサイトが確認された．

踏み台サイトとされたウェブサイトのドメイン名の TLD の割合を表 4 に示す．.com の TLD を持つドメイン名が最も多く 117 件確認され，その他 49 種の TLD を持つドメイン名が確認された．検索エンジンの検索結果のスニペットに日本語が含まれるものを対象としたが，踏み台サイトには.jp の TLD を持つドメイン名は確認されなかった．偽ショッピングサイトのドメイン名の TLD の割合を表 5 に示す．.xyz，.asia，.online の順に多く，TLD の種類も 12 種と踏み台サイトと比較して少ない．

5.2 偽ショッピングサイトの解析回避機能

偽ショッピングサイトと踏み台サイトが有する解析回避機能の分析結果について述べる．

（1）踏み台サイトの解析回避機能

表 2 の HTTP リクエストヘッダを用いて，踏み台サイトにアクセスした結果，解析回避機能を持つ踏み台サイトがどの程度存在したかを述べる．表 6 に検索エンジンサイトの URL を Referer に設定してアクセスした場合の調査結果を示す．検索エンジンサイトの URL を Referer に設定してアクセスした場合にのみ偽ショッピングサイトへ誘導する踏み台サイトは 93.2%存在した．87.6%は日本国内で利用されることが多い Google（JP），Bing，Yahoo ！ JAPAN

表 2　解析回避機能調査のための HTTP リクエストヘッダ

No.	User-Agent	Referer
1	Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)	無し
2	Mozilla/5.0 (compatible; bingbot/2.0; +http://www.bing.com/bingbot.htm)	無し
3	Mozilla/5.0 (compatible; YandexBot/3.0; +http://yandex.com/bots)	無し
4	Mozilla/5.0 (compatible; Yahoo! Slurp; http://help.yahoo.com/help/us/ysearch/slurp)	無し
5	Mozilla/5.0 (compatible; Baiduspider/2.0; +http://www.baidu.com/search/spider.html)	無し
6	Mozilla/5.0 (Windows NT 10.0; WOW64; Trident/7.0; rv:11.0) like Gecko	無し
7	Mozilla/5.0 (Windows NT 10.0; WOW64; Trident/7.0; rv:11.0) like Gecko	https://www.google.co.jp/
8	Mozilla/5.0 (Windows NT 10.0; WOW64; Trident/7.0; rv:11.0) like Gecko	https://www.google.co.uk/
9	Mozilla/5.0 (Windows NT 10.0; WOW64; Trident/7.0; rv:11.0) like Gecko	https://www.bing.com/
10	Mozilla/5.0 (Windows NT 10.0; WOW64; Trident/7.0; rv:11.0) like Gecko	https://yandex.com/
11	Mozilla/5.0 (Windows NT 10.0; WOW64; Trident/7.0; rv:11.0) like Gecko	https://www.yahoo.co.jp/
12	Mozilla/5.0 (Windows NT 10.0; WOW64; Trident/7.0; rv:11.0) like Gecko	https://www.yahoo.com/
13	Mozilla/5.0 (Windows NT 10.0; WOW64; Trident/7.0; rv:11.0) like Gecko	https://www.baidu.com/

表 3　偽ショッピングサイトの収集結果概要

	件数
URL ブラックリスト掲載 URL 数	75,610
URLhaus	33,548
OpenPhish	42,062
サイト内検索実施ドメイン名数	14,052
スニペットに日本語含むドメイン名数	1,788
偽ショッピングサイトの踏み台サイトドメイン名数	251
偽ショッピングサイトドメイン名数	875

表 4　踏み台サイトドメイン名の TLD

No.	TLD	件数
1	.com	117
2	.vn	16
3	.ru	12
4	.org	11
5	.br	7
6	.net	6
7	.in	6
8	.ar	4
9	.uk	4
10	.id	4
-	その他 (40 種)	68

表 5　偽ショッピングサイトドメイン名の TLD

No.	TLD	件数
1	.xyz	365
2	.asia	131
3	.online	101
4	.club	80
5	.icu	66
6	.site	50
7	.fun	37
8	.wang	29
9	.space	8
10	.website	6
11	.life	1
12	.me	1

表 6　解析回避機能（検索エンジンサイト経由アクセス）の調査結果

	到達可能数
検索エンジン経由（Google（JP），Yahoo! JAPAN，Bing）	220（87.6%）
検索エンジン経由（上記以外含む）	14（5.6%）
対象 URL への直接接続	17（6.8%）

のうちいずれかからのアクセスでないと偽ショッピングサイトへ誘導されないことが分かった．また，Yahoo ！ JAPAN 経由のみ偽ショッピングサイトへ誘導する場合，Google（JP），Yahoo ！ JAPAN 経由のみ誘導する場合等の踏み台サイトによって異なるパターンが確認された．一方で，6.8%は直接アクセスすることでも偽ショッピングサイトへ誘導され，解析回避機能を持たない踏み台サイトがあることが確認された．

表 7 に検索エンジンボットの User-Agent を設定してアクセスした場合の調査結果を示す．検索エンジンボットの User-Agent を設定してアクセスした場合にのみ検索エンジンにインデックスさせるための商品紹介を記載したページを応答する踏み台サイトは 94.1%存在した．検索エンジンボットの中でも，日本で利用されることが多い Google，Bing の検索エンジンボットにのみ商品紹介を記載したページを応答する踏み台サイトが 8.6%存在した．表 6，表 7 の

表 7　解析回避機能（検索エンジンボット）の調査結果

	到達可能数
検索エンジンボット（Google，Bing）	19（8.6%）
検索エンジンボット（上記以外含む）	188（85.5%）
対象 URL への直接接続	13（5.9%）
データ取得失敗（サーバエラー等）	31

表 8　フィッシングサイトとの解析回避機能の比較

	偽ショッピングサイト	フィッシングサイト
検索エンジン経由（国内）	○	●
検索エンジン経由（海外）	▲	●
対象 URL への直接接続	▲	○
検索エンジンボット（国内）	○	●
検索エンジンボット（海外）	▲	●

（○：アクセス許可　●：アクセス不許可　▲：正規コンテンツを応答）

結果から，既存研究 [2] で明らかにされた検索エンジン経由のアクセスであるかの確認だけではなく，日本国内のユーザが良く利用する検索エンジンからのアクセスを攻撃対象としている可能性が推測される．

表 8 にフィッシングサイトと偽ショッピングサイトの踏み台サイトが有する解析回避機能の比較結果を示す．一般的にフィッシングサイトはメールに記載した URL をユーザがクリックすることでアクセスすることを想定している．そのため，フィッシングサイトは直接のアクセスの場合のみアクセスを許可し，検索エンジン経由や検索エンジンボットのアクセスは不許可としている場合が多く存在する [5]．一方で，偽ショッピングサイトは検索エンジン経由でユーザがアクセスすることを想定しており，検索エンジン経由のアクセスを許可し，直接のアクセスの場合は改ざん対象のウェブサイト自体が持つコンテンツを応答する．また，検索エンジンボットによるアクセスの場合は検索エンジンにインデックスさせるための商品紹介のページを応答する．このようにそれぞれユーザの流入経路が異なるため，解析回避機能も異なっている．

（2）偽ショッピングサイトの解析回避機能

偽ショッピングサイトの中には，Proxy を経由する際に HTTP リクエストヘッダに付加されることがある `X-Forwarded-For` ヘッダの有無に応じて偽ショッピングサイトへアクセスさせるかを判定する機能を有するものが確認された．`X-Forwarded-For` ヘッダを持つ HTTP リクエストの送信時点で HTTP ステータスコード 302 が応答され，`https://www.yahoo.co.jp/` へリダイレクトされて

おり，サーバサイドでのクローキングがされていた．収集した偽ショッピングサイトの 62.2%に本機能があることが確認された．この機能を有している理由の 1 つとして，解析者等が Proxy を経由したアクセスで調査を試行する場合に解析を回避するための機能であると推測される．

5.3 偽ショッピングサイト構築ツール分析

収集した偽ショッピングサイトと踏み台サイトのコンテンツの分析結果をそれぞれ述べる．

(1) プロキシを利用した画像の読み込み

偽ショッピングサイトには多くの商品の画像が掲載されている．画像を HTML の `img` タグで読み込み元を指定する際に，正規のショッピングサイトから読み込むようにしている場合がある．調査の結果，偽ショッピングサイトでは図 5 に示す手法で画像を正規のショッピングサイトから読み込むことが確認された．

図 5（a）に `src` 属性に正規のショッピングサイトの画像の URL を直接指定した場合を示す．この場合，外部の URL が Referer ヘッダに設定されたユーザの HTTP リクエストが正規のショッピングサイトの Web サーバのアクセスログに残る．そのため，ショッピングサイト事業者は画像の URL のアクセスログの Referer を確認することで，画像の参照元となる偽ショッピングサイトの URL を知ることができると考えられる．

図 5（b）に `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 を指定した場合を示す．画像プロキシサーバの URL のクエリストリングに URL を設定することで画像を代理で取得することを実現している．この場合，画像プロキシサーバの IP アドレスからの HTTP リクエストが正規のショッピングサイトの Web サーバのアクセスログに残る．画像プロキシサーバは偽ショッピングサイトと同サーバに配置される場合と，別のサーバ（ドメイン名：fwma-umbrella[.]bid）に配置される場合が確認された．そのため，参照元の偽ショッピングサイトの URL を特定することが（a）と比較して困難になる．

(2) 踏み台サイトの検索エンジン掲載手法

改ざんされる可能性の高い正規のウェブサイトのドメイン名と “販売” 等のショッピングサイトで用いられるキーワード合わせてを検索エンジンで検索すると，数万件の検索結果が得られる場合がある．通常，検索エンジンに URL を掲載するためには検索エンジンのクローラに全ての URL を巡回させる必要がある．しかしながら，攻撃者が正規のウェブサイトを改ざん後に数万件の新規の URL を個別に検索エンジン事業者に巡回要求することは現実的ではない．そこで，攻撃者は検索エンジンクローラに効率的に巡回させるために sitemap.xml と，踏み台サイトの内部リンクを利用しているものと考えられる．

sitemap.xml とは検索エンジンのクローラにどの URL を巡回させるか指示するために用いられる．踏み台サイトの sitemap.xml を調査したところ，図 6 のように 46.3% の踏み台サイトが sitemap.xml を持ち，そのうち 61.0%が 5,000 件以上の URL を巡回させるように記述している．7 件の踏み台サイトにおいて複数ファイルの sitemap.xml を用意することで，1 万件以上の URL を検索エンジンクローラに巡回するように要求するものが確認され，最大で約 1,000 万件の URL を sitemap.xml に定義していた．

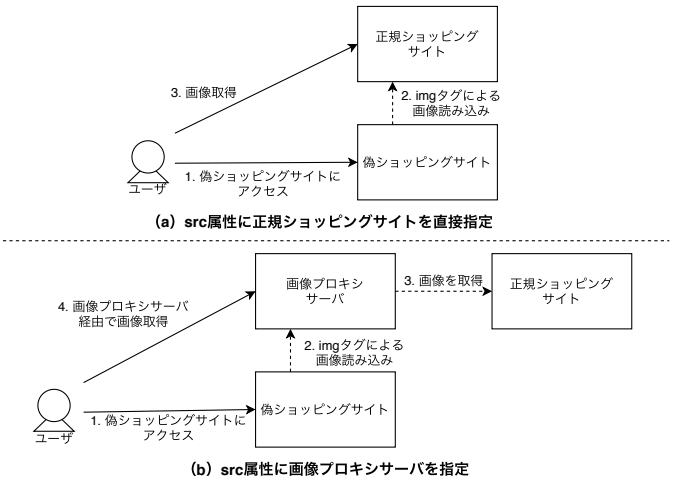

図 5 偽ショッピングサイトの画像の読み込み手法

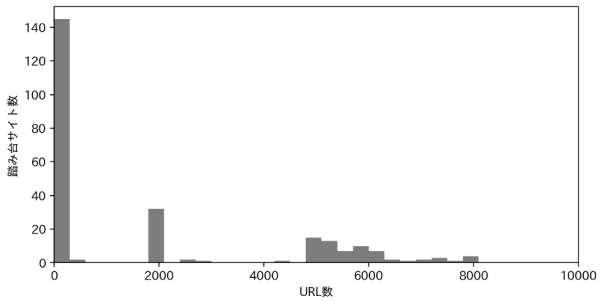

図 6 sitemap.xml に含まれる URL 数（10,000URL 以下）

踏み台サイトには図 7 のように検索エンジンのスニペットに掲載するための商品の説明と，同一ドメイン名内に存在する他の商品説明のページの URL への内部リンクが多数含まれている．検索エンジンクローラは内部リンクを辿るため，多数の内部リンクを相互に掲載することで効率的にインデックスに掲載させているものと考えられる．

(3) 偽ショッピングサイトへのリダイレクト機能

5.2 節で示したように，踏み台サイトから偽ショッピングサイトへリダイレクトさせるかはユーザのアクセスが検索エンジン経由であるかをもとに判断している．そのため，JavaScript コードを実行しない curl コマンドや wget コマンドで踏み台サイトのコンテンツを収集し，HTML を解析することでも偽ショッピングサイトの URL を特定することができる．しかし，踏み台サイトのコンテンツを収集するだけでは偽ショッピングサイトの URL を特定すること

Vol.2020-DPS-182 No.41
Vol.2020-CSEC-88 No.41
2020/3/13

図 7　踏み台サイトの内部リンク

Listing 1　JavaScript の実行を要するリダイレクトコード

```
<script>
    function set_cookie(){
        var now = new Date(); var time = now.getTime()
            ;
        time += 19360000 * 1000; now.setTime(time);
        document.cookie='beget=begetok'+';
        expires='+now.toGMTString()+'; path=/';
    }
    set_cookie(); location.reload();;
</script>
```

が困難な踏み台サイトが確認された．

6 つの踏み台サイトにおいて Listing 1 のような JavaScript コードを含むコンテンツが応答された．Listing 1 の JavaScript コードが実行されると，名前が `beget`，値が `begetok` となる Cookie を付加した状態で reload される．Cookie が付加された状態で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の誘導するためののコンテンツが応答されるという挙動が確認された．この場合，curl コマンド等で踏み台サイトのコンテンツを取得しただけでは偽ショッピングサイトの URL を特定できないため，Selenium 等の JavaScript が実行できる環境での調査が必要となる．

5.4　偽ショッピングサイトの攻撃規模推定

DNS での名前解決要求回数がアクセスしたユーザ数であると仮定し，偽ショッピングサイトによる攻撃規模を Passive DNS のデータを用いて推定した．Passive DNS のデータは FarSight 社の DNSDB [14] を用いた．調査結果の制約として，キャッシュ DNS の影響や観測範囲の影響を考慮すると，実際のアクセス数よりも少ない結果である可能性がある．また，本研究を含む調査通信によって発生した名前解決も含まれていることも制約として考慮する必要がある．

図 8 に名前解決が観測され始めた時期別の偽ショッピングサイト名前解決数の集計結果を示す．名前解決が観測され始めた時期が過去であるほど名前解決数が多く，最も多いドメイン名で 302 回の名前解決があった．全体で見ると，1 ドメイン名あたり平均 43.1 回の名前解決がされており，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が分かった．

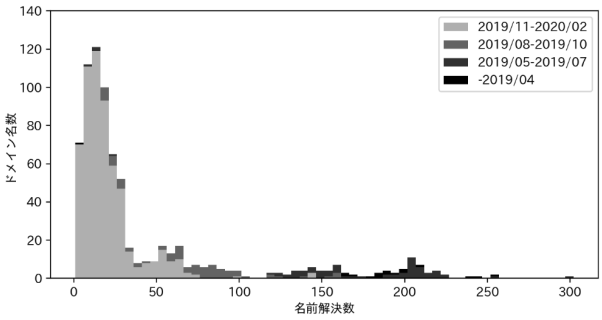

図 8　偽ショッピングサイトドメイン名の名前解決数（名前解決の観測開始時期別）

6.　議論

本節では本研究での制約，今後の課題に関して述べる．

6.1　解析回避機能調査の制約

本調査では，HTTP リクエストヘッダ（User-Agent，Referer）を変更することで解析回避機能の調査を実施した．フィッシングサイトには IP アドレスでの解析回避機能があることが知られており [4]，偽ショッピングサイトにも同様の機能がある可能性が考えられる．例えば，検索エンジンのクローラであるかの判断をアクセス元 IP アドレスで制御している場合，本調査では検出できていないものと考えられる．

6.2　海外ユーザをターゲットとした偽ショッピングサイト

本調査では，4.3 節で述べたようにスニペットに日本語が含まれている検索結果を踏み台サイトの疑いのある URL としている．既存研究 [6], [7] において海外ユーザをターゲットとした偽ブランド商品の販売サイトが存在することが明らかになっているが，海外ユーザをターゲットとした偽ショッピングサイトも存在することが考えられる．しかしながら，本調査では海外ユーザをターゲットした偽ショッピングサイトは調査対象に含まれていない．今後の課題として，日本国内だけではなく海外ユーザをターゲットした偽ショッピングサイトの実態を明らかにする必要がある．

6.3　商品名を用いた検索による調査の大規模化

本調査では，URL ブラックリストに掲載された URL をもとに踏み台サイトの URL を収集した．偽ショッピングサイトに誘導する商品は複数の偽ショッピングサイトで共通していることが見受けられ，商品名を検索エンジンで検索することで踏み台サイトを収集するという手法が考え

られる．しかしながら，商品名を検索する場合には正規のショッピングサイトも検索結果に多く含まれるため，正規のショッピングサイトのドメイン名をホワイトリスト化することや，偽ショッピングサイトのタイトルやスニペットに多く見られる単語を用いて偽ショッピングサイトであるかの判定をすることが必要であると考えられる．今後の課題として，商品名を用いた検索により調査を大規模化をする必要がある．

6.4 偽ショッピングサイト対策への提言

本調査の結果から考えられる偽ショッピングサイトに対する対策方法を提言する．

ショッピングサイト事業者: 5.3 節において，偽ショッピングサイトの画像は正規のショッピングサイトから読み込まれることがあることを示した．ショッピングサイト事業者が Web サーバのアクセスログを確認し，表 5 に掲載された TLD を含むドメイン名が画像読み込みリクエストの Referer に含まれている場合，その URL が偽ショッピングサイトである可能性が高いと判断して発見することができ，テイクダウン等の対処を早期に実施できる．

セキュリティ事業者: 偽ショッピングサイトは攻撃コードやマルウェアを利用しないため，シグネチャによる検知や攻撃時の振る舞い検知では発見することが困難である．よって本研究の発見方法もしくは前述のショッピングサイト事業者による発見方法に基づいて，偽ショッピングサイトのブラックリストを作成し，セキュリティ製品・サービスに活用することで，エンドユーザが偽ショッピングサイトにアクセスすることを抑制できる．

検索エンジン事業者: 5.3 節において，sitemap.xml を用いて検索エンジンの検索結果に踏み台サイトを掲載させる手法があることを示した．通常，sitemap.xml によって巡回対象の URL 数が急増することは考えにくい．そこで，検索エンジン事業者はあるドメイン名の過去の巡回対象 URL 数と比較して大幅に増加している等の sitemap.xml に掲載される URL 数の傾向が異なる場合に，踏み台サイトの疑いのあるドメイン名と判断して巡回を抑制することで，踏み台サイトが検索エンジンの検索結果に掲載されることを防ぐことができる．

7. まとめ

本研究では，URL ブラックリストに掲載された URL と検索エンジンを用いて偽ショッピングサイトとその踏み台サイトを収集し，偽ショッピングサイトに関する実態調査を実施した．偽ショッピングサイトの踏み台サイトが有する解析回避機能を調査した結果，検索エンジンを経由したアクセスの場合のみ偽ショッピングサイトに誘導する踏み台サイトが 93.2%あった．また，検索エンジンボットによるアクセスの場合のみ検索エンジンにインデックスさせるためのページを応答する踏み台サイトが 94.1%あった．既存研究 [2] では，検索エンジン経由の場合に偽ショッピングサイトへ誘導されることが示されていたが，本研究ではその中でも日本国内のユーザをターゲットとしていることを明らかにした．Passive DNS のデータを用いて偽ショッピングサイトによる攻撃キャンペーンの規模を調査したところ，1 ドメイン名あたり平均 43.1 回の名前解決があったことが確認された．今後は海外ユーザをターゲットとした偽ショッピングサイトの調査や定常的な観測を実施することで，より大規模な調査を実施することにより偽ショッピングサイトの実態を明らかにすることが必要である．

参考文献

[1] Anti-Phishing Working Group: Phishing Activity Trends Report 2nd Quarter 2019, `https://docs.apwg.org/reports/apwg_trends_report_q2_2019.pdf`.
[2] Japan Cybercrime Control Center: Revealed Threat of Fake Store, `https://www.jc3.or.jp/about/pdf/JC3_APWG_Revealed_Threat_of_Fake_Store.pdf`.
[3] Akiyama, M., Yagi, T., Yada, T., Mori, T. and Kadobayashi, Y.: Analyzing the ecosystem of malicious URL redirection through longitudinal observation from honeypots, *Computers & Security*, Vol. 69, pp. 155–173 (2017).
[4] Oest, A., Safei, Y., Doupé, A., Ahn, G.-J., Wardman, B. and Warner, G.: Inside a phisher's mind: Understanding the anti-phishing ecosystem through phishing kit analysis, *APWG Symposium on Electronic Crime Research (eCrime)*, pp. 1–12 (2018).
[5] 小寺博和，芝原俊樹，千葉大紀，青木一史，波戸邦夫，秋山満昭：動的解析を利用したフィッシングサイトのアクセス妨害機能の実態解明，情報処理学会論文誌，Vol. 61, No. 3.
[6] Wadleigh, J., Drew, J. and Moore, T.: The E-Commerce Market for" Lemons" Identification and Analysis of Websites Selling Counterfeit Goods, *Proceedings of the 24th International Conference on World Wide Web*, pp. 1188–1197 (2015).
[7] Carpineto, C. and Romano, G.: Learning to detect and measure fake ecommerce websites in search-engine results, *Proceedings of the International Conference on Web Intelligence*, pp. 403–410 (2017).
[8] Han, X., Kheir, N. and Balzarotti, D.: Phisheye: Live monitoring of sandboxed phishing kits, *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS)*, pp. 1402–1413 (2016).
[9] Invernizzi, L., Comparetti, P. M., Benvenuti, S., Kruegel, C., Cova, M. and Vigna, G.: Evilseed: A guided approach to finding malicious web pages, *2012 IEEE symposium on Security and Privacy*, IEEE, pp. 428–442 (2012).
[10] Corona, I., Biggio, B., Contini, M., Piras, L., Corda, R., Mereu, M., Mureddu, G., Ariu, D. and Roli, F.: Deltaphish: Detecting phishing webpages in compromised websites, *European Symposium on Research in Computer Security*, Springer, pp. 370–388 (2017).
[11] abuse.ch: `https://urlhaus.abuse.ch/`.
[12] OpenPhish: `https://openphish.com/`.
[13] Selenium: `https://selenium.dev/`.
[14] Farsight Security, Inc.: `https://www.dnsdb.info/`.