「フィッシング・ハンター」を１年間やってみて学んだこと

高田 哲司[1]

概要：2025 年においてもフィッシング詐欺は情報セキュリティにおける重要課題の１つである．フィッシング詐欺対策を行う上でまず必要なのはフィッシングページの情報である．これなくして対策ははじまらないが，この情報の収集活動について語られることは多くない．今回著者は，１年以上にわたり継続してフィッシングページの情報収集を手作業で行ってきた．その経験を通じて確立した独自のシンプルな情報収集法と，その活動を通じて収集した情報を５つの方法で分析し，得られた知見について報告する．

キーワード：フィッシング詐欺, データ収集, ウェブセキュリティ，データ分析

What I learned from conducting “Phishing Hunter” for a year

Tetsuji Takada[1]

Abstract: Phishing scams remain one of the critical issues in information security even in 2025. The first step in countering phishing scams is to collect data about phishing pages. Without this data, the process of countering phishing scams can not proceed; however, there is little discussion about the activities involved in the data collection. The author has been manually collecting data on phishing pages for over a year. Through this experience, the author has established a simple data collection method and analyzed the collected data using five methods. I hope to share the insights what I learned through this activity.

Keywords: phishing, scam, data collection, threat hunting, web security

1. はじめに

フィッシング詐欺の対策は 2025 年においても依然として重要な課題であると言える．フィッシング対策協議会の報告 [1] によると，2024 年のフィッシング報告件数は過去最多の 170 万件超えとなっており，2023 年と比較しても約 1.4 倍という状況にある．また 2025 年 4 月にはフィッシング詐欺によって日本国内の証券会社を対象とした不正取引の被害が発生し，金融庁から注意喚起が発出された．この資料は，以降７月にいたるまで更新され続けている [2]．この資料によると，2025 年 1 月〜6 月の間における不正アクセス件数は 13,099 件，不正な取引件数は 7,293 件と報告されており，この問題に対する対策強化が要請されている．またこの事件をきっかけに，多要素認証の設定必須化を決定した証券会社は 79 社におよぶといった変化も起きている [3].

フィッシング詐欺の研究を行うにはフィッシング詐欺を知る必要があり，そのためにはフィッシング詐欺に関するデータが必要となる．よって我々は，フィッシングページに関する情報をみずから収集することを試みた．フィッシング詐欺に対する研究の多くは，公開データベースから情報を得るか，情報収集を行なっているグループから情報提供を受けるかのどちらかでデータを入手している．著名な公開データベースとしては PhishTank, OpenPhish などがあり，日本では JPCERT/CC やフィッシング対策協議会が情報提供を行なっている ([10], [11], [12], [14]).

[1] 電気通信大学
The University of Electro-Communications

これに対して著者らは，あえてみずからフィッシングページの情報収集を試みた．理由は２つある．１つは「“生きた” フィッシングページ」の情報を必要としたためである．ここでいう「生きているページ」とは，閲覧・利用可能な状態の Web ページであることを意味する．著者が想定している研究 [6] では，フィッシングページの Document Object Model(DOM) 情報を必要とした．しかし，公開情報から URL を入手してもその時点でフィッシングページは閲覧できない状態になっていることが多く，この要件充足が困難であった．

もう一つの理由は，日本人を対象としたフィッシングページを収集対象と決めたためである．この選択は，以下の３つの理由からである．

(i) Web ページの真贋判定が可能
(ii) 情報収集が容易
(iii) Small start とするため

以降，本論文ではこのような動機で始めたフィッシングページの情報収集活動とそれによって得られた知見・経験について述べる．本報告が，今後同様の活動を行う可能性のある人への参考になることを期待するとともに，我々の活動に対するアドバイスやご指摘を頂ければ幸いである．なお論文題目にある「フィッシュハンター」とはフィッシングページの情報収集を行う人のことを指している[*1]また，以降本論文では「フィッシングページ」のことを「偽ページ」と呼ぶこととする．

2. 先行研究

フィッシングページの情報収集を目的とした研究は，著者の知る限りみあたらない．その理由は，情報源があれば収集行為自体は難しいことではないからと推測する．情報源となるのは主として電子メール，SMS または SNS などであり，偽ページへの情報はむしろユーザ側にある，とすら言えるからである．

よって以降で紹介する先行研究は，フィッシングページの情報収集が目的ではなく，その検出が目的の研究であり，その成果を使えば新たなフィッシングページも見つけられることから情報収集も促進可能というものになる．紙面の都合で紹介ははぶくが情報源という観点で見ると

- サーバ証明書の発行ログ [15]
- DNS グラフ [7]
- なりすまされる Web ページの視覚情報 [18], [19]
- Whois 情報の応用 [9]
- Web 検索エンジンの応用 [8], [17]
- Web コンテンツ [6], [16]

PhishZoo は，Web ページの視覚的特徴を用いることでフィッシングページを自動的に検出する方法を提案している [19]．正規の Web ページを事前にプロファイリングし，ページ内テキストと画像，ロゴを組み合わせることで検出制度を評価した．その際，画像を用いた判定では単純一致ではなく類似度による検出も行なった．その結果，検出率 90.2%，偽陽性率 0.5%を達成した．したがって，視覚的特徴に類似性が見られるフィッシングページの検出・収集が可能になる．

3. 情報収集方法

本章では著者が行なっているフィッシングページの情報収集法について述べる．概要は以下の５ステップである．

(1) 情報源から URL を入手
(2) Web ブラウザで上記 URL にアクセス
(3) 必要に応じて対話操作実施
(4) フィッシングページが表示されたら URL を入手．その後，いくつかの追加情報を取得
(5) スクリーンショット取得

以降では，これらの各ステップについて述べる．

3.1 情報源から URL 入手

本活動では「なりすましメール」を情報源とした．なりすましメールとは，既存のブランドや組織を語りフィッシングページに誘導しようとする電子メールのことを指す．今回，このメールの提供について３名の協力者に協力いただいた．協力者のもとに届いたなりすましメールと思われるメールを著者に転送していただき，そのメール本文内に存在する URL を抽出することで URL を入手した．なお本活動では，メール本文が日本語で書かれたもののみを対象とした．この対応により抽出された URL を「入口 URL」と呼ぶ．

3.2 Web ブラウザでアクセス

入手した入口 URL を Web ブラウザに入力し，実際にアクセスを行なった．本作業は VMWare による仮想マシンの上に Windows 11 の環境を用意し，その環境に Google chrome と Firefox をインストールしてアクセスを行なった．

3.3 必要に応じて対話操作実施

前節による対応の結果，偽ページが表示された場合は次のステップに進む．しかし，そうでない場合は以下のような対応を追加で行い，偽ページに到達できるかを試みた．

Op11：エラーや空白ページが表示されたら，再読み込みを何度か実施
Op12：Desktop 用の User Agent(UA) でエラーページが表示された場合，Mobile 用の UA に変更してアクセス
Op13：Google Safe Browsing(GSB)[13] や，Cloudflare, 短縮 URL サービス等による警告が表示された場合，「危険を理解した上で安全でない Web ページへのアクセ

[*1] この言葉は，JPAAWG General Meeting のセッション名で用いられていたものであり，それを流用させていただいている．

表 1 収集情報と収集方法
Table 1 Collected data and method

収集情報	収集方法
入口 URL	手動
PPURL	半自動
タイトル文字列	半自動
ブランド名	手動
警告状況	手動
User agent	半自動
Serialized DOM	半自動

収集情報	収集方法
データ収集日時	自動
IP アドレス	半自動
ドメイン登録日	半自動

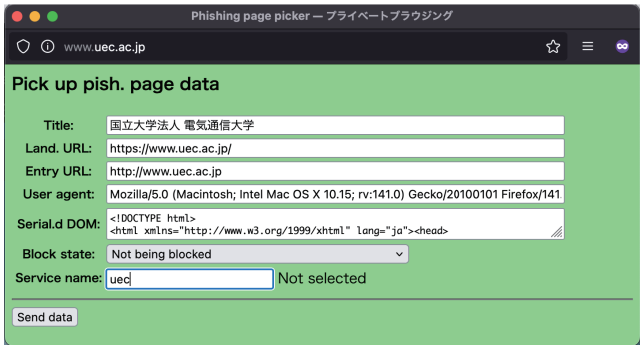


図 1 情報抽出用 Bookmarklet の実行画面
Fig. 1 A screenshot of data collection program

ス」を実施

Op14：エラー内容によっては，異なる network（回線）からのアクセスを実施

これらの対応でも Web ページが表示されない場合はここで収集作業を終了とした．また Op13 の状況になった場合，どの仕組み・サービスによって警告が発生したかを記録した．また偽ページとは言えない Web ページが表示された場合，必要に応じて以下の対応を実施した．

Op21：要求されている対話処理を実行（時間待機，ボタン押下，文字入力，計算実行，パズル実行，契約者情報入力など）

なお ステップ (2，3) の実施回数は，1つの入口 URL に対して1回のみと限定してはいない．ステップ (3) でアクセス検証に失敗した場合，時間をおいてあらためて検証を行なっている場合もある．これらの対応を通じて，最終的に偽ページに到達するかどうかを検証した．

3.4 各種情報抽出

このステップは，偽ページに到達した時に行う処理である．なお本活動では「個人認証，クレジットカード，個人情報を入力する入力フォームが存在する Web ページ」を偽ページ (フィッシングページ) と定義した．偽ページに到達したら，**表 1** 左に示す7情報を抽出し，それらを情報収集サーバに送信した．なお到達した偽ページの URL を「PPURL」（Phishing page URL）と呼ぶ．

表 1 左内の各情報について説明する．タイトル文字列は偽ページ内の title tag のコンテンツ文字列である．ブランド名は，フィッシングページがなりすましているブランド・会社名であり，ユーザが Web ページを見て判断し入力する．警告状況とは，Web ブロックリストや他の仕組みで警告されたかどうかを示すもので，ステップ (3) を通じて以下の3種の仕組みによる警告があったかどうかを記録する．なお複数の仕組みから警告があった場合，数値の少ない方の仕組みを記録した．

BL1：Google Safe Browsing[13]
BL2；短縮 URL サービスによる警告
BL3：その他の枠組みによる警告
(Cloudflare, Web browser extension 等)

User agent は，使用した Web ブラウザの User agent 文字列を抽出しており，Serialized DOM は，フィッシングページの Document Object Model(DOM) をシリアライズ化した文字列を記録している．

これら情報抽出は，Bookmarklet と呼ぶ JavaScript プログラムで行っている．表 1 の「収集方法」列にて「半自動」と記載のデータは，本プログラム実行により自動で抽出される．一方，入口 URL，ブランド名，警告状況の3情報は手動入力が必要である．**図 1** は Bookmarklet 実行時の画面例で，上記3情報を手動で入力完了した状況を示したものである．

3.5 スクリーンショット取得

Web ブラウザのスクリーンショット機能を利用してフィッシングページを画像データとして取得した．この際，可能な限り全画面領域を保存するようにしたが，なんらかの理由で全画面取得がうまくいかない場合は，Web ブラウザで表示されている領域だけを Operating System のスクリーンキャプチャ機能を使用して取得した．

3.6 サーバ側での追加処理

Web ブラウザから送信されたデータはサーバ側でデータベースに保存した．その際，以下に述べる2つの追加処理を行った．

(1) 重複確認
(2) 属性情報の追加収集

「重複確認」とは，同一データを繰り返し収集しないようにするための処理である．重複条件は「過去 48 時間以内に収集されたデータ内に，PPURL が完全一致するデータが存在する」である．この判定結果が真の場合「重複データあり」と判断し，受信データは保存せずに破棄した．一方，判定条件が偽の場合，データ受信時刻を「データ収集日時」として追加し，データベースに保存した．

「属性情報の追加収集」とは，データベースに保存された PPURL の値を入力とし，IP アドレスとドメイン名の登録年月日を収集する処理である．処理内容を**図 2** に示す．

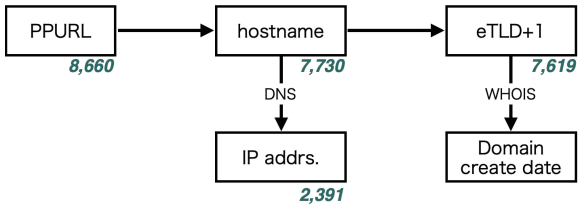


図 2 属性情報の追加収集
Fig. 2 Process on appending related info. from URL

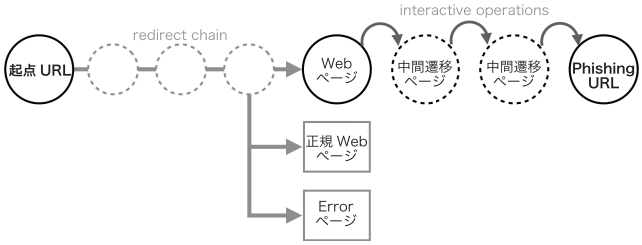


図 3 フィッシングページへの画面遷移
Fig. 3 A transition of web pages to Phishing page

PPURL からホスト名を抽出し，DNS の仕組みを用いてホスト名から IP アドレスを取得した．またホスト名から eTLD+1 を取得し，それをドメイン名とみなして WHOIS の仕組みを利用しドメイン登録日を取得した．本処理は Python を用いて実装し，手動で日に 1 回以上不定期に実行した．したがって最終的には受信した 7 情報に表 1 右の３情報が追加され，計１０種の情報がデータベースに保存されることとなる．

3.7 情報収集のポイント

今回の情報収集のポイントは 2 点ある．

- フィッシングページの実在確認
- (入口 URL, PPURL) の組み情報として記録

１つはフィッシングページの実在確認をしてから情報収集を行っている点にある．これをルールとした理由は，フィッシングページが表示されなかった場合，それをフィッシング詐欺と断定することはできないという考えからである．なお情報収集の最後にスクリーンショットを取得している理由は，フィッシングページが稼働していた証拠を残すためでもある．またフィッシングページがなりすましていたブランド名も，到達した偽ページの内容から判断して手動入力しており，なりすましメールの内容は利用していない．

もう１つは，２つの URL を収集している点にある．１つはなりすましメールに記載されていた URL（入口 URL）であり，もう１つは到達したフィッシングページの URL（PPURL）である．フィッシング詐欺では，入口 URL から PPURL に到達するまでに複数の URL を経由することがある．よって始点と終点を明確に記録しておく方が良いと考えた．本活動では，これら 2URLs を組情報として記録している．

表 2 分析対象となったなりすましメール数と収集期間
Table 2 The number of provided E-mails

	# of Emails	(%)	Period	
Provider A	13833	96.9	2024-06-01	2025-07-31
Provider B	36	0.2	2025-05-26	2025-07-11
Provider C	405	2.8	2025-05-31	2025-07-23

表 3 フィッシングデータに対する警告状況
Table 3 Warning status to Phishing URL.

	データ数	
警告なし	7,455 件	(86.09%)
Google safe browsing	1,101 件	(12.71%)
短縮 URL サービス	0 件	(0.0%)
その他	104 件	(1.20%)

4. 収集データとその分析

本章では，活動を通じて収集したデータとその分析によって分かったことについて述べる．

まずはじめに収集データについて述べる．３名の協力者から，**表 2** に示す期間に受信した「なりすましメール」を提供いただいた．期間は最大で 14 ヶ月である．その結果，合計で 14,274 通のメールを受領した[*2]．このメール群を情報源とし，前章で述べた方法で情報収集を行なった結果，計 8,660 件のフィッシングページ情報を収集した．なお図 2 内の 4 つの数字は，本活動によって得られた情報の重複なし情報数を示している．

4.1 警告状況について

収集データに対する警告状況の分析結果を**表 3** に示す．３カテゴリーの仕組みによって偽ページのアクセス時に警告が発出されたデータは総計で 1,205 件であり，割合としては 13.91%であった．

4.2 ドメイン登録日と攻撃実施日の関係

PPURL に含まれるホスト名は，攻撃者がドメイン名を取得して用意していると考えられる．よってドメイン名登録から攻撃実行までは短期間であると推測される．これについて検証を試みた．

「データ収集日時」と「ドメイン登録日」を用いてドメイン名登録日から攻撃実施までの経過日数を算出し，その値をもとに 30 日を単位区間としてデータを集計した．集計対象となったデータはドメイン登録日が取得できたデータで計 8,577 件である．集計結果を**図 4** に示す．横軸のラベルは年であり，縦軸の値は集計されたデータ数の常用対数値を示している．

この分析結果から，攻撃の多くは１年以内に行われているといえる．しかしその一方で，少数ではあるがドメイン

[*2] データ数から，実質的には 1 名の協力者によるデータである

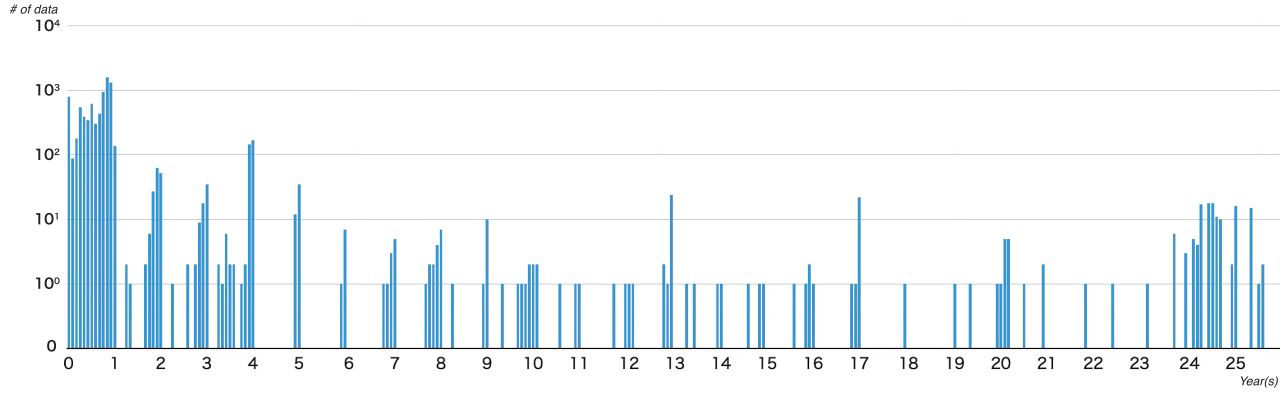
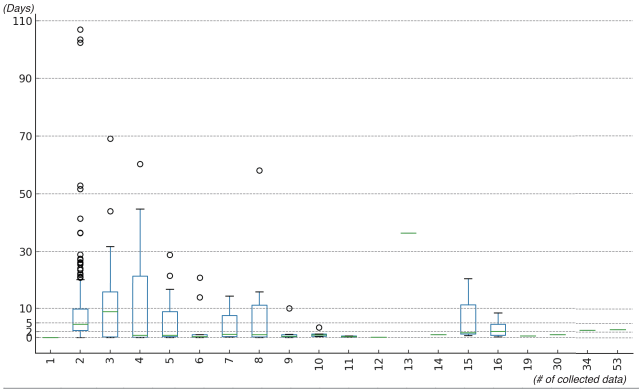


図 4　ドメイン登録日からデータ収集日時までの経過日数によるデータの集計結果



攻撃回数	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	19	30	34	53
eTLD+1数	7235	236	56	23	11	9	9	10	7	5	4	1	1	1	3	4	1	1	1	1
(%)	94.96	3.1	0.74	0.3	0.14	0.12	0.12	0.13	0.09	0.07	0.05	0.01	0.01	0.01	0.04	0.05	0.01	0.01	0.01	0.01

図 5　eTLD+1 を基準とした攻撃回数と攻撃期間の分布

登録から 20 年を経過したドメインも攻撃に使用されていることが明らかになった．

4.3　同一ドメインによる攻撃回数と攻撃期間の分布

PPURL に含まれるホスト名から eTLD+1 を取得し，重複なしリストを生成した．その結果，7,619 件の eTLD+1 を得た．このリストを用い，各 eTLD+1 ごとに収集データをグループ化した．その結果，以下のようなデータが集計できる．

(eTLD+1，該当データ数 (攻撃回数)，最小攻撃日時，最大攻撃日時)

ここから，攻撃期間 = 最大攻撃日時 - 最小攻撃日時を求め，(eTLD+1, 攻撃回数，攻撃期間) によるデータ群を求めた．このデータ群から同一ドメインによる攻撃回数とその攻撃期間を集計・視覚化したものが**図 5** である．横軸は同一 eTLD+1 を利用した攻撃回数，縦軸は攻撃期間の分布である．

今回の分析から，2 回以上攻撃に悪用された eTLD+1 の数は合計で 384 ドメイン (5.04%) であり，同一 eTLD+1 による攻撃回数の最大値は 53 回であった．また少数ではあるが，攻撃期間が 30 日を超えるものも確認された．なお攻撃期間の最大値は 106.94 日 (攻撃回数 2 回) であった．

表 4　入口および PPURL の eTLD によるデータ集計:Top 10

	Entry URL	#	PPURL	#	(Entry, PP)	#
1	cn	5055	cn	4945	('cn', 'cn')	4922
2	com	1659	com	2191	('com', 'com')	1552
3	cc	424	top	671	('cc', 'com')	343
4	translate.goog	359	shop	178	('top', 'top')	243
5	top	245	net	131	('translate.goog', 'top')	239
6	workers.dev	187	icu	109	('shop', 'shop')	171
7	shop	173	sbs	60	('workers.dev', 'top')	143
8	net	127	com.cn	42	('net', 'net')	127
9	sbs	58	cc	39	('cn', 'com')	125
10	com.cn	42	duckdns.org	35	('translate.goog', 'com')	85

4.4　攻撃に使われた 2 つの URLs の関係

入口 URL と PPURL に含まれるホスト名から eTLD を抽出し，その値をもとにデータ集計を試みた．分析結果を**表 4** に示す．表内のデータは，入口 URL，PPURL および (入口 URL, PPURL) 毎に集計したデータのうち，データ数の多い順に Top 10 を示したものである．

入口 URL, PPURL ともに No.1 eTLD は “cn” ドメインであり，全データに対する割合も双方ともに 50%を超えている．一方，フィッシング詐欺に悪用されているとおぼしきサービスのドメイン名も Ranking に散見されている．

eTLD の組情報による集計では (“cn”, “cn”) が Top であり，全データに対する割合は 50%を超えている．一方，組情報で eTLD が異なるデータでは (“cc”, “com”) が Top であった．また (“translate.goog”, “*”) や (“*”, “top”) といったドメインペアも eTLD が異なる組み合わせとして上位にくることが明らかになった．

4.5　PPURL のホスト名と IP アドレスの関係

本活動では，DNS 利用により PPURL に含まれるホスト名から IP アドレスを取得した．この際，IP アドレスが複数個取得できることもあった．そこで 1 つのホスト名から得られる「IP アドレスの数」に注目し，データ集計を試みた．またその値毎に得られた IP アドレスの集合を作成し，その要素数を調査した．その分析結果を**表 5** に示す．表の読み方について説明する．一番上の行は 1 ホスト名から得られた IP アドレスの数を示す．左から 4 列目の場合，

表 5 DNS 正引きに得られた IP addr 数と攻撃データの分布
Table 5 # of IP addrs. from 1 hostname and data distribution.

DNS正引きによる IP addrs.数	0	1	2	3	4	5	6	7
# of data	530	6171	1421	5	1	0	0	532
# of unique IPs	-	558	1816	6	4	-	-	7

表 6 IP addr に割り当てられているホスト名数の分布
Table 6 # of hostnames from 1 IP addr and data distribution.

# of hostnames	1	2 - 50	51-100	101-150	151-200	201-250	251-300
# of IP addrs	1987	363	19	11	4	0	7

表 7 IP addr の 2 octets 別データ集計 Top 10
Table 7 Top 10 2 octets of IP addrs.

	2 octets of IP addr.	data count
1	104.21	5120
2	172.67	1396
3	43.133	849
4	43.167	536
5	43.165	525
6	43.163	467
7	43.153	455
8	43.134	385
9	43.130	324
10	165.154	242

1つのホスト名から2つの IP アドレスが取得できたデータを示しており，集計の結果 1,421 件のデータがこれに該当するとともに，該当データに含まれる重複なし IP アドレスの数は 1,816 個であった，ということを示している．

分析結果から，データ数に対して使用されている IP アドレスの数は少ないことがみてとれる．特に1つのホスト名から7つの IP アドレスが得られた攻撃データは 532 件あるのに対し，その攻撃に使用されている IP アドレスは7つのみであることは興味深い．

一方，この組データ (ホスト名 → IP アドレス) を逆に分析し，各 IP アドレスに対してホスト名がいくつ割り当てられているかを調査した．結果を**表 6** に示す．3列目のデータを例に表の読み方を説明する．1行目のデータは1つの IP アドレスに割り当てられていたホスト名の数を示している．よって左から3列目のデータの場合，1つの IP アドレスに 2〜50 個のホスト名が割り当てられている IP アドレスが 363 件あったことを示している．

分析結果から，1つのホスト名しか割り当てられていない IP アドレスが 1,987 件と大半であるのに対し，複数のホスト名が割り当てられている IP アドレスは 404 件であった．また1つの IP アドレスに割り当てられていたホスト数の最大値は 260 で，該当する IP アドレスが 7 件確認された．

最後に，攻撃に悪用された IP アドレスを基準として，攻撃データの集計を試みた．**表 7** は IP アドレスの上位 2 octets を集計単位とし，その IP アドレスを利用していた攻撃データを集計したものである．なお表内3列目の集計データ数は，1つの攻撃データが複数の IP アドレスでカウントされている可能性があることに留意する必要がある．

分析結果から，多くの攻撃が特定の IP アドレスレンジで行われていることが分かる．なお 1,2 位の IP アドレスレンジは，DNS 逆引きから Cloudflare が管理している IP アドレスと推測される．よって多くの偽ページは Cloudflare のサービスを利用して展開されていると言える．なお，3〜9 位のデータも集約して解釈すると[*3]，データ数の合計は 3,541 件となり，相当数の攻撃が特定の IP アドレスレンジで行われていることがうかがえる．

5. 考察

本章では，前章で述べた収集データの分析に関する考察と，情報収集活動を通じて得た経験について述べる．

5.1 分析結果の考察

4.1 節の分析結果から，日本のサービスを対象とした GSB の有効性は低いといえる．比較対象として，日本語のフィッシングページに対する Web ブロックリストの警告率は 38.9%という値が言及されている [6]．今回の分析結果は，この値と比較してもその半分以下という低警告率となっている．なおこれら2つの警告率の違いは，検証タイミングの違いにあると考える．文献 [6] では SNS を情報源としているのに対し，本研究ではなりすましメールを情報源としたため，本活動の方が情報収集のタイミングが早く，それゆえ警告率が低い値になっていると推測する．

4.2 節の分析結果からは，ドメイン登録日から攻撃実施までの期間はその多くが1年以内であり，ドメインを登録してから時間をおかずに攻撃に利用していることが確認された．その一方で少数ではあるが，過去に登録したドメイン名を攻撃に再利用している事例も見てとれる．特に5年以上前に登録されたドメイン名の利用については，ドメイン名の更新タイミングにあわせて攻撃に利用していることがデータからうかがわれる[*4]．ただし，ドメイン名登録から数年以上経過しているドメイン名については正規サイトの悪用も考えられるので追加調査が必要である．また Phisher によって長く保持されるドメイン名と，そうではないドメイン名の違いについても分析を検討する．

4.3 節の分析から，取得したドメイン名を攻撃で複数回利用する傾向は低いと言える．また複数回利用する場合も攻撃期間は数日と短期間で攻撃を行なっていることが明らかになった．一方，少数ではあるが，30 日を超える期間でドメイン名を攻撃に再利用する事例も確認された．なお今回の収集データは最大でも 14 ヶ月であることと，4.2 節の分析結果からドメイン名の更新タイミングで攻撃を行う傾

[*3] IP addrs: 43.[130-167].*.*
[*4] ドメイン名の更新は年単位であると仮定

向が見られることから，さらに長期間のデータがあれば，本分析は違った結果が得られるかもしれない．

4.4 節の eTLD 分析から，cn ドメインを用いた攻撃が半数以上を占めることが明らかになった．また入口 URL と PPURL の eTLD 名は同一であるものが多数であるものの，これら 2 つの eTLD が異なるデータも上位に存在することは興味深い．とくに Google translate と Cloudflare workers を入口 URL として利用し，なりすましメールでの検出を回避しつつ，フィッシングページ自体は他の eTLD で用意するという攻撃形態が確認された．ただしこれらは既報での事実でもある [4], [5].

4.5 節については分析を通じて分かったことについて言及する．表 5 による分析から，7 つの IP アドレスを使用し，532 件の攻撃を行なっていた事例が明らかになった．これらの IP アドレスを調べたところ，すべて [104.21.x.x] であり，DNS 逆引きから Cloudflare のサービスを利用しているものと推測される．またこれらの攻撃では 260 ものホスト名を使用し，2024-12-25 から 2025-07-31 の期間にわたり散発的に攻撃を行っている記録が確認された．さらにこれらの攻撃でなりすまされたサービスも 43 種におよんでいる．このことから，この 7 つの IP アドレス群は Bad guy によるフィッシング攻撃用のインフラであったと推測され，協力者の 1 人はこのインフラを利用した攻撃を継続的に受けていた，と考えられる．

5.2 活動を通じた経験について

ページ数の都合で詳細は書けないが，今回の活動を通じて経験した事象について列挙する．

- (ほぼ) 同じ文面のなりすましメールが短時間に複数届く．これはユーザが異常さに気づく機会になるので，攻撃側がこういったメール配布をする理由が理解困難
- 対話操作を通じて複数ページを遷移するフィッシングページ：本物を忠実に模倣した場合と，「安全確認」をうたったページをはさむ場合が見られた．これらは偽ページの検出回避に寄与するものと推測．
- アクセス元に応じたアクセス制限：検証に毎日使用していた network からは徐々にフィッシングページに到達できなくなった．また海外からのアクセスではフィッシングページに到達できない．
- Web ブラウザによるアクセス制限：Firefox では表示されないが，Google chrome では表示される．Desktop 版の Web ブラウザでは表示されないが，UA を Mobile に変更すると表示される．
- アクセス回数によるアクセス制限：一定回数までは繰り返しアクセスしても偽ページが閲覧できるが，それ以上になるとエラーページとなる．
- なりすましメールと偽ページの運用時間不整合：なりすましメールはすでに届いているが，偽ページはメール到着後数時間経過後でないとアクセスできるようにならない．
- 個人認証に失敗する偽ページの存在：これはリアルタイム（透過型）・フィッシングの事例だろうか？
- Unicode による URL：今回の活動ではごく少数しか確認できなかった

なお，本活動で収集したフィッシングページの情報については Google Safe Browsing team[13] に精力的に報告した．

5.3 今後の課題

まずはじめに情報収集処理の自動化を進める．現状では多くても 40 通/Day 前後なので手作業でなんとかなっているが，収集情報量を増やすためには作業者を増やすか自動化を実現するかのどちらかが必要となる．今後，さらなるシステム化や生成 AI・AI エージェントの活用を試みる予定である．

次に必要なのは「何を情報収集し，それをどう分析して対策に活かせば良いのか？」の再考である．現状の対策手段は事実上 GSB 1 択であると言える．よって GSB による警告効果向上を目標とするのであれば，なるべく多くの情報を収集し，かつ迅速に報告することが重要であり，収集したデータの分析は不要である．一方，今回の分析結果を対策に活用するというのであれば，別の対策方法を考える必要があると言える．IP アドレスや eTLD を用いた警告システムや，警告対象の偽ページを日本のサービス利用者や，1 個人・組織に限定した対策システムの実現を考えるのも一考かと考えている．

また今回の活動を通じて感じたのは，なりすましメールをスパム判定し，ユーザの目につかない状況にしている現状は，フィッシング詐欺対策として見た場合「よい対応」と言えるのか？ということであった．なりすましメールを見ないということは，フィッシング詐欺を学ぶ機会を損なっていないかという視点である．ユーザを「知らなかった」「気づかなかった」という状況から意識変化させるため，スパム判定されたフィッシングメールを見える状況にすること，そしてそこから対策に繋げる方法はないか？という点についても考察を進めたい．

5.4 制限事項

本論文における分析は，数名の協力者宛に届いたなりすましメールのデータに基づいている．よって，今回の知見は偽ページ分析の 1 例でしかなく，その点は留意する必要がある．また，今回の収集データに以下の 2 種に関する偽ページ情報は含まれていない：(1) QRcode を媒介とする詐欺：キャッシュレスサービスで支払いを促すものや投資情報の提供をかたり LINE にて友達登録を促すものなどがこれに該当する．これらのケースは，QRcode を表示する

ページが正規サービス上の Web ページであると判断したため，それが詐欺であるとは断定できず，情報収集は行わなかった．(2) ショッピングサイト：なりすましメールの送信者ドメイン名や到達した Web ページのドメイン名から詐欺が疑われるが，正規サービスのなりすましではないことから詐欺と判断できず，情報収集は行わなかった．

なお収集データの一つである「データ収集日時」にも留意する必要がある．このデータは著者の活動時刻に依存しており，なりすましメールや偽ページの属性等に基づいた情報ではない．この点も留意する必要がある．ただし，日に１度以上は情報収集活動するように努めてきたことは申し添えておく．

6. おわりに

本論文では，フィッシング詐欺対策の研究に必要となるフィッシングページの情報収集活動と収集したデータの分析結果について報告した．日本語で書かれたなりすましメールを協力者から提供いただき，それを起点に実際にフィッシングページへのアクセスを実施した．その結果，フィッシングページが閲覧できたものを対象に情報収集を行い，結果として 14 ヶ月で 8,660 件のフィッシングページ情報を収集した．また収集情報を五つの方法で分析し，その結果を議論した．結果の大半は既知の内容ではあるものの少数のデータから興味深い攻撃傾向も確認された．

フィッシング詐欺対策の現状は，情報収集，情報共有，検出器開発，そしてユーザへの警告発出の四段階である．しかし，この対策も情報収集を行わないことには始まらない．本論文では情報収集手段の１例と，収集されたデータの分析結果について報告した．この知見を共有することにより，意志・能力のある有志がフィッシング詐欺の情報収集に興味を持ち，活動に関与するようになることを期待する．またそれを支援しうるシステムの実現について検討を進めていく．

謝辞 なりすましメールを提供していただいた３名の協力者に深く感謝いたします．

参考文献

[1] フィッシング対策協議会 技術・制度検討 WG, フィッシングレポート 2025, 入手先 ⟨*https://www.antiphishing.jp/report/phishing_report_2025.pdf*⟩, (Accessed 2025-07-30)

[2] 金融庁, インターネット取引サービスへの不正アクセス・不正取引による被害が急増しています, 入手先 ⟨https://www.fsa.go.jp/ordinary/chuui/chuui_phishing.html⟩, (Accessed 2025-07-30).

[3] 日本証券業協会, 多要素認証の設定必須化を決定した証券会社, 入手先 ⟨https://www.jsda.or.jp/about/hatten/inv_alerts/alearts04/list_tayouso/index.html⟩, (Accessed 2025-07-30)

[4] PR TIMES, ネットスコープ、Cloudflare Workers を悪用したフィッシング手法．透過的フィッシングと HTML スマグリングに関する調査結果を発表, 入手先 ⟨https://prtimes.jp/main/html/rd/p/000000014.000137550.html⟩, (Accessed 2025-08-20)

[5] Kaspersky daily, Google 翻訳を悪用したフィッシング詐欺, 入手先 ⟨https://blog.kaspersky.co.jp/google-translate-scheme/32912/⟩, (Accessed 2025-08-20)

[6] 渡邉祐貴, et al., Web ページ表示後のリアルタイム検出によるフィッシング対策の強化, 情報処理学会論文誌（採録決定）, To appear.

[7] 石田裕貴, et al., DNS グラフを用いたフィッシングサイト検知手法, 情報処理学会論文誌, Vol. 66, Issue 8, (2025).

[8] 山本雄心, et al., Web ブラウザにおけるフィッシングページ検出の実現可能性調査, コンピュータセキュリティシンポジウム, Oct., (2024).

[9] 小島大輝, et al., WHOIS のドメイン登録情報に基づくフィッシングサイトの検出, コンピュータセキュリティシンポジウム, Oct., (2024).

[10] PhishTank, PhishTank, available from ⟨https://phishtank.org/⟩, (Accessed 2025-07-30).

[11] OpenPhish, OpenPhish Database, available from ⟨https://www.openphish.com/ phishing_database.html⟩, (Accessed 2025-07-30).

[12] JPCERT/CC, Phishing URL dataset from JPCERT/CC, 入手先 ⟨https://github.com/JPCERTCC/phishurl-list/⟩, (Accessed 2025-07-30).

[13] Google, Google Safe Browsing, available from ⟨https://safebrowsing.google.com/, https://safebrowsing.google.com/safebrowsing/report_phish/⟩, (Accessed 2025-08-20).

[14] フィッシング対策協議会, フィッシングサイト URL 提供, 入手先 ⟨https://www.antiphishing.jp/enterprise/url.html⟩, (Accessed 2025-07-30).

[15] Y. Sakurai, et al., "Discovering HTTPSified Phishing Websites Using the TLS Certificates Footprints,", IEEE European Symp. on Security and Privacy Workshops (EuroS&PW), pp. 522-531, (2020).

[16] Y. Zhang, J. I. Hong, and L. F. Cranor, Cantina: a content-based approach to detecting phishing web sites, int'l conf. on World Wide Web (WWW '07), pp.1681–1698, (2020).

[17] J. H. Huh and H. Kim, Phishing Detection with Popular Search Engines: Simple and Effective. Foundations and Practice of Security (FPS 2011), (2011).

[18] S. Abdelnabi, K. Krombholz, and M. Fritz, VisualPhishNet: Zero-Day Phishing Website Detection by Visual Similarity, ACM SIGSAC Conference on Computer and Communications Security (CCS '20), pp.1681–1698, (2020).

[19] S. Afroz and R. Greenstadt, "PhishZoo: Detecting Phishing Websites by Looking at Them," IEEE Fifth International Conference on Semantic Computing, pp. 368-375, 2011.