InternetWeek2020 DNSDAY ドメイン名の事故例(悪用例)その2

2020年11月26日
長崎県立大学 教授 岡田 雅之
JPNIC IPv6教育専門家チーム メンバー
(2020年3月31日までJPNIC 技術部)

自己紹介

研究情報詳細

氏名

岡田　雅之（オカダ　マサユキ）
MASAYUKI OKADA

所属

情報システム学部 情報セキュリティ学科

職名

教授

学外略歴

1. 東邦大学 理学部 非常勤講師 暗号と情報セキュリティ 非常勤 2018年9月 ～ （継続中）
2. 東邦大学 理学部 訪問研究員(金岡研究室) 非常勤 2016年10月 ～ （継続中）
3. 一般社団法人日本ネットワークインフォメーションセンター(JPNIC) 技術部 課長 常勤 2014年6月 ～ 2020年3月
4. 一般社団法人日本ネットワークインフォメーションセンター(JPNIC) 技術部 常勤 2004年2月 ～ 2016年5月

昨年、不正アクセスの認知件数は急増

- 2019年(令和元年)の不正アクセスの認知件数は、2,960件と前年の約2倍に増加
- 内訳を見ると、インターネットバンキングでの不正送金等が1,808件と半数以上を占める。
- 不正送金は、金銭を目的として、組織的かつ計画的に行われている。

想定を超える事態を常に意識しないといけないかもしれません。私もありました。(後述)

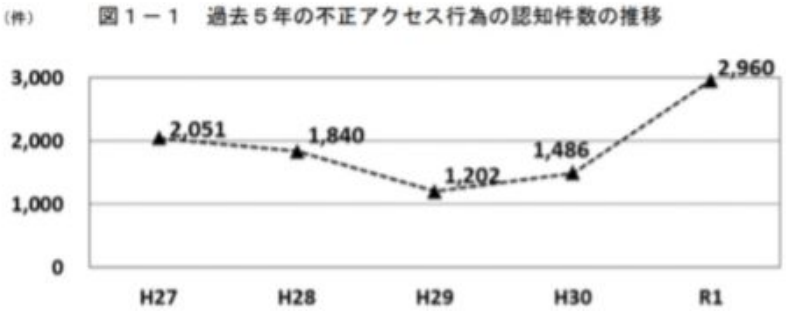


図１－１　過去５年の不正アクセス行為の認知件数の推移

不正アクセス行為の発生状況　出典：警察庁

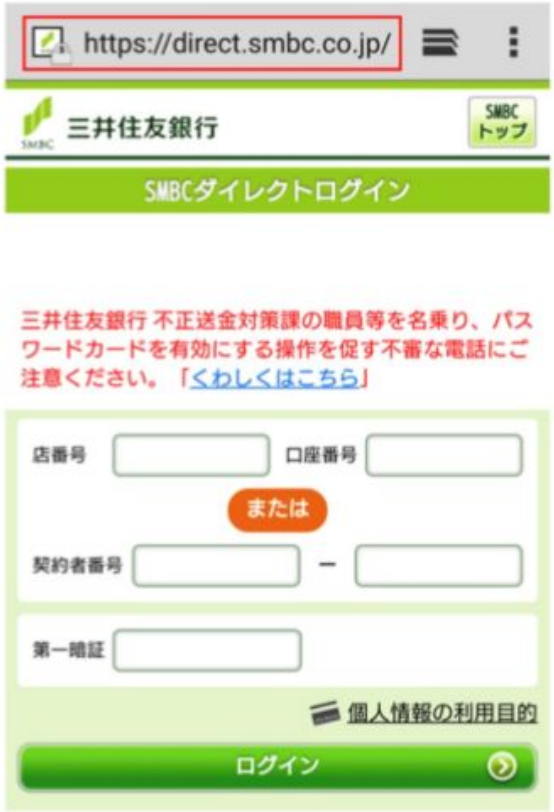
表１－３　過去５年の不正アクセス後の行為別認知件数

区分＼年次	平成27年	平成28年	平成29年	平成30年	令和元年
インターネットバンキングでの不正送金等	1,531	1,305	442	330	1,808
インターネットショッピングでの不正購入	167	172	133	149	376
メールの盗み見等の情報の不正入手	92	91	146	385	329
オンラインゲーム・コミュニティサイトの不正操作	96	124	83	199	60
インターネット・オークションの不正操作	20	34	28	29	47
知人になりすましての情報発信	83	25	110	24	30
仮想通貨交換業者等での不正送信			149	169	22
ウェブサイトの改ざん・消去	34	6	14	13	19
その他	28	83	97	188	269
計（件）	2,051	1,840	1,202	1,486	2,960

※　平成28年以前は、「仮想通貨交換業者等での不正送信」を分類して集計していない。

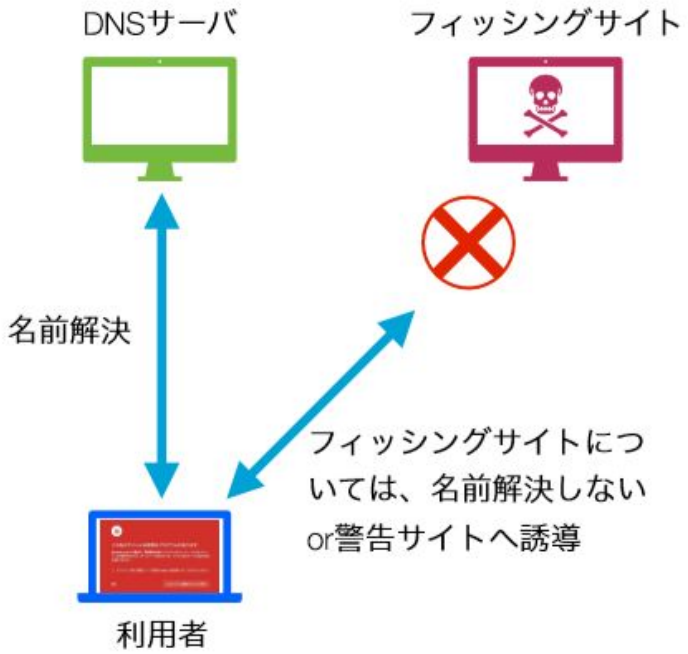
不正アクセスの入り口はフィッシングサイト

- 不正アクセスに利用される識別符号は、フィッシングサイトを介して被疑者に渡っているものが多い。
- フィッシングサイトは、正規のサイトと外見は同一である。
- ドメイン名からフィッシングサイトを見破ることはできるが、見破られないように利用権者を急かして焦らせる場合が多い。
- 多くの方は騙されないが、それでも少数の騙される方がいる限り、サイバー犯罪者はフィッシングサイトの構築をやめることはない。

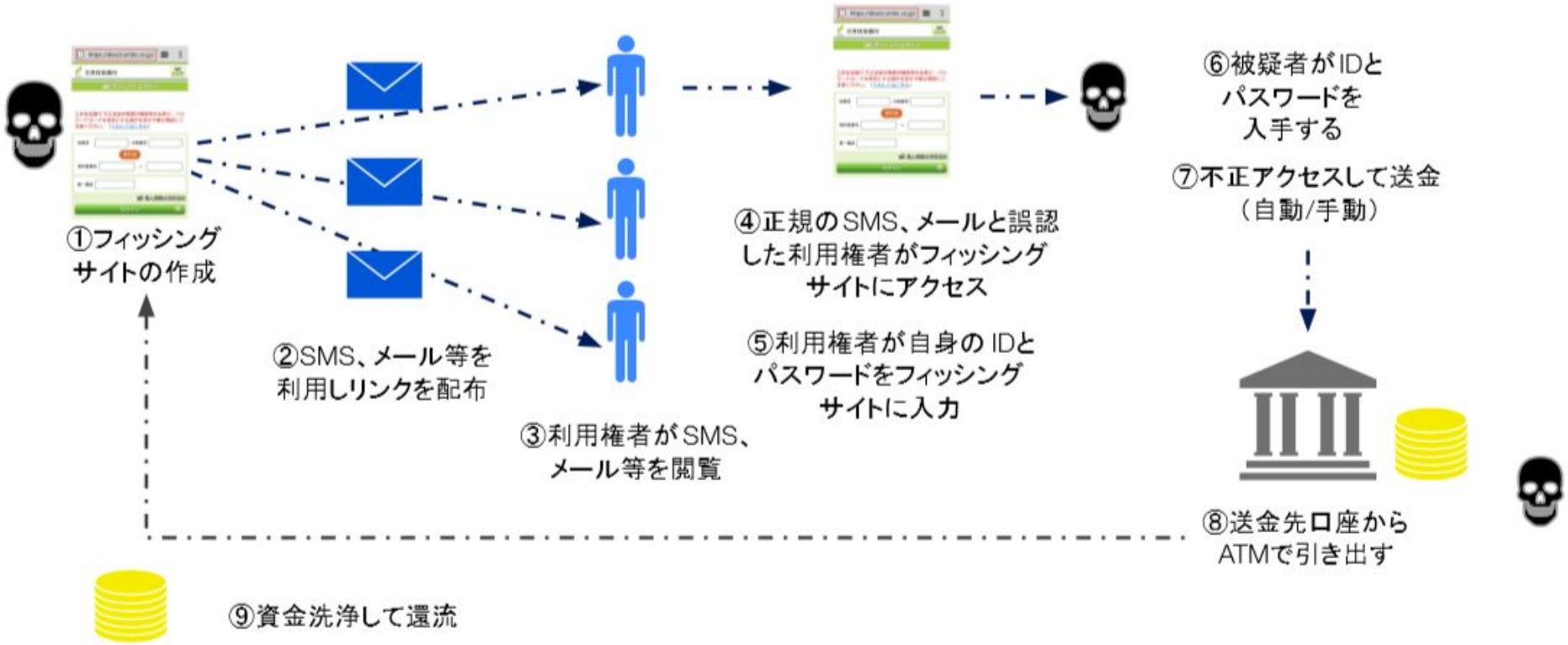


ネットワーク経路上でのフィッシングサイトの遮断

- ネットワーク上で遮断する方法としては、いくつか考えられる。
- 例えばパケットを監視してフィッシングサイトとの通信を発見したら、その通信を遮断する、といったやり方では明確に通信の秘密を害してしまう。
- IDNSフィルタリングが有効ではないか。
- 正規サイトの管理者に鑑定を依頼して、フィッシングサイトであることを確認できたサイトについてフィルタリングルールを設定する。
- フィルタリングルールを用いて、フィッシングサイトとの通信を遮断する。
- DNSを理解しているユーザーは、自ら設定を変更することでフィルタリングを回避することもできる。



フィッシングの流れ（不正送金）



フィッシングサイトの検出数

Googleのセーフブラウジングの透明性レポートによれば、不正なソフトウェアサイトには減少傾向が見られるものの、フィッシングサイトの検出数は増加傾向

具体的な機械学習の方法

content baseの指標
CANTINA(2007)

age of domain	フィッシングサイトの生存期間は短い
known images	正規サイトのロゴをパクる
suspicious url	疑わしいURL　@や-の有無
suspicious links	疑わしいリンク
ip adress	そもそもドメイン使ってないかも
dots in url	ドットが多いと怪しい
forms	入力フォームがある
tf-idf	単語の重要度

Table 2. The results of Experiment 2 showing what weights should be used for the various heuristics. Note that link check generates a negative effect, which means it is nearly useless, so we assign 0 to its effect.

Heuristic	True Positive	False Positive	Effect	Weight
Age of Domain	87%	30%	57.0	0.18
Known Images	37%	0%	37.0	0.12
Suspicious URL	6%	3%	3.0	0.01
Suspicious Links	8%	25%	0.0	0.00
IP Address	22%	0%	22.0	0.07
Dots in URL	45%	3%	42.0	0.13
Forms	94%	27%	67.0	0.21
TF-IDF-Final	99%	10%	89.0	0.28

機械学習に使用する識別子

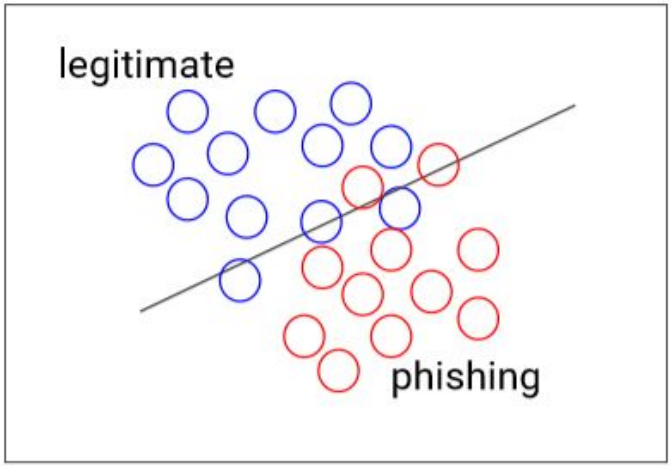
url length 長いことが多い
ip address ドメインがないかも
shortening service 使われることがある
suspicious characters 「@」や「-」など
double slash redirecting リダイレクト
prefix or suffix -でドメイン区切って騙す
subdomain サブドメインもフィッシングかも
website traffic 偽サイトは少ない
favicon
port
https token
request url 本物なら全部同じ
url of anchor 上に同じ
link of tags 同じドメインのはず
submitting to email ちゃんとmailtoしてる?
abnormal or nonstandard urls
iframe 見えないフォーム
age of domain 偽サイトは短い

dns record
on mouseover 偽サイトは隠す
right click マウスオーバーと同様
pop up window 正規サイトはポップアップはあまり使わない
page rank 偽サイトはない
google index 偽サイトはない
links pointing to page

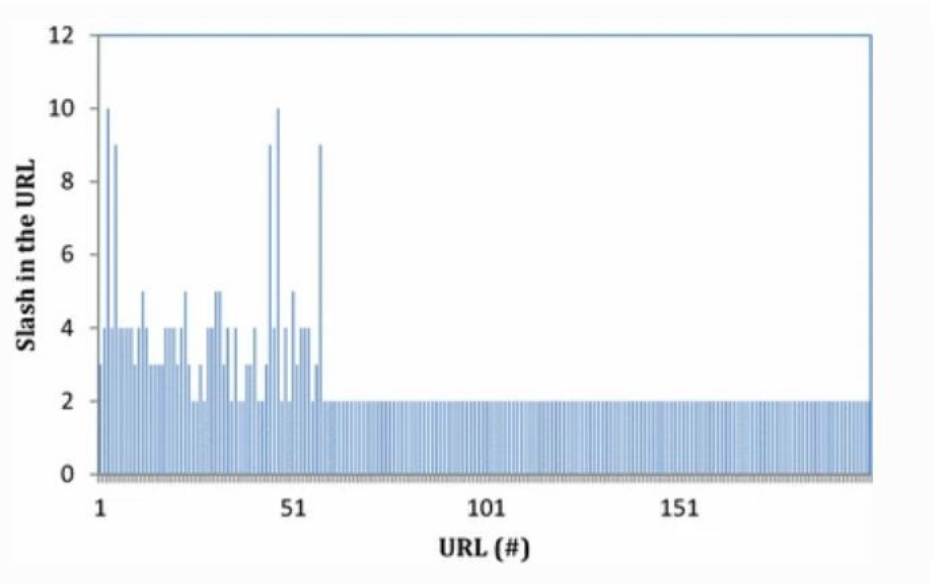
検出に関する課題

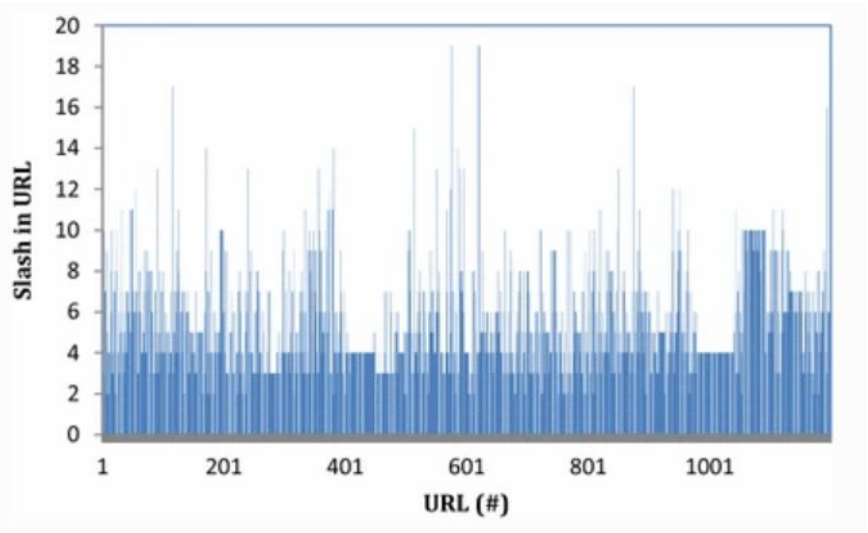
誤検出
- false positive
- false negative

検出を回避するカウンター技術
- APWGのレポートによれば、2020の第2四半期のフィッシングサイトの58%は詐称元とは無関係のURLを使用（字句ベースの検出を回避）
- 78%はHTTPS



フィッシングサイトに関する研究より(URL)

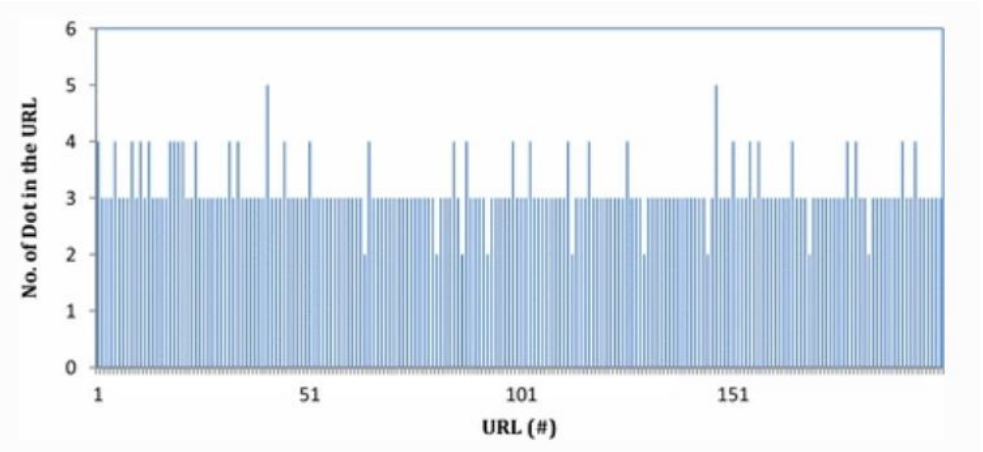


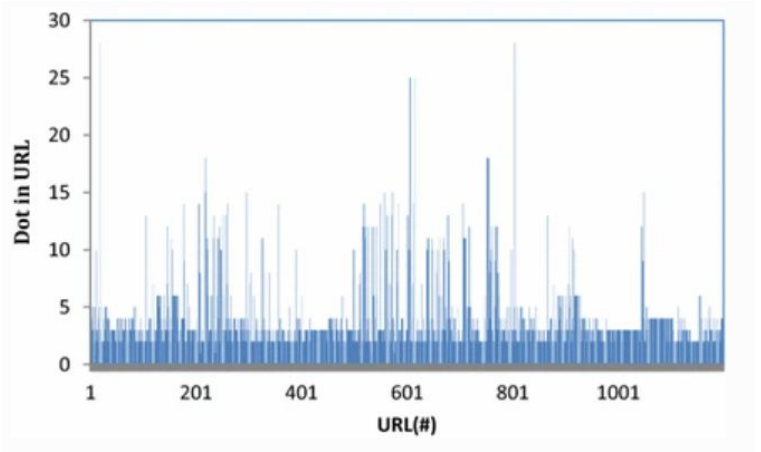


URLに含まれる"/"(スラッシュ)の数が通常のURLは2~4(左)、フィッシングサイトのURLは大量の"/"が存在している。

https://link.springer.com/article/10.1186/s13673-016-0064-3より

フィッシングサイトに関する研究より(URL)

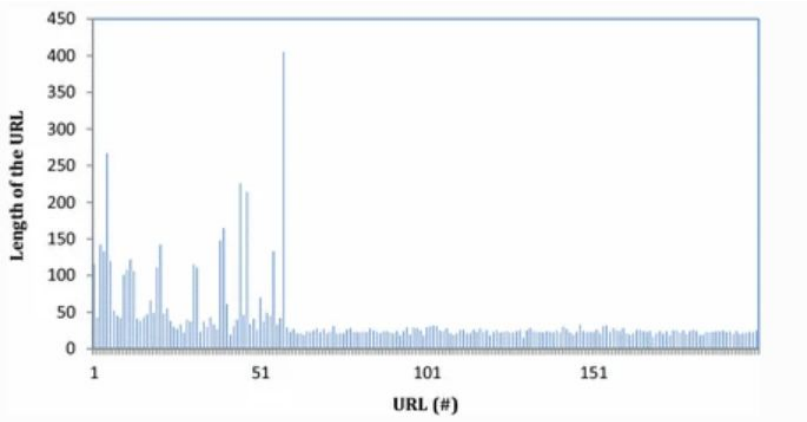


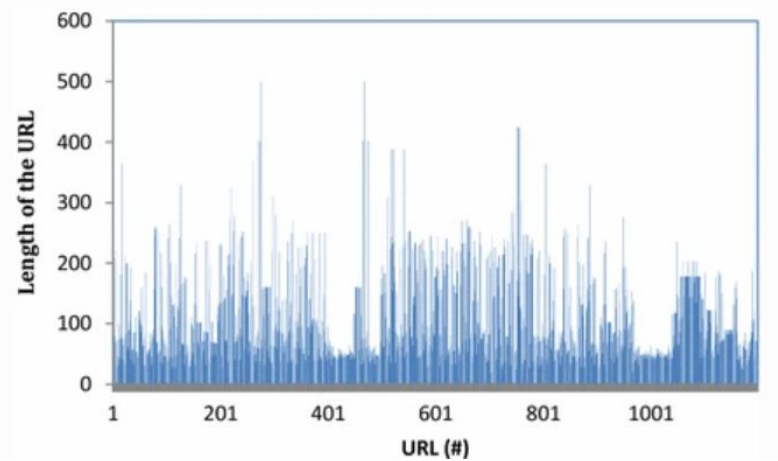


URLに含まれる"."(ドット)の数。通常のURLは多くても3、4(左)、フィッシングサイトは5以上など、最大30などもある。(右)

https://link.springer.com/article/10.1186/s13673-016-0064-3より

フィッシングサイトに関する研究より(URL)



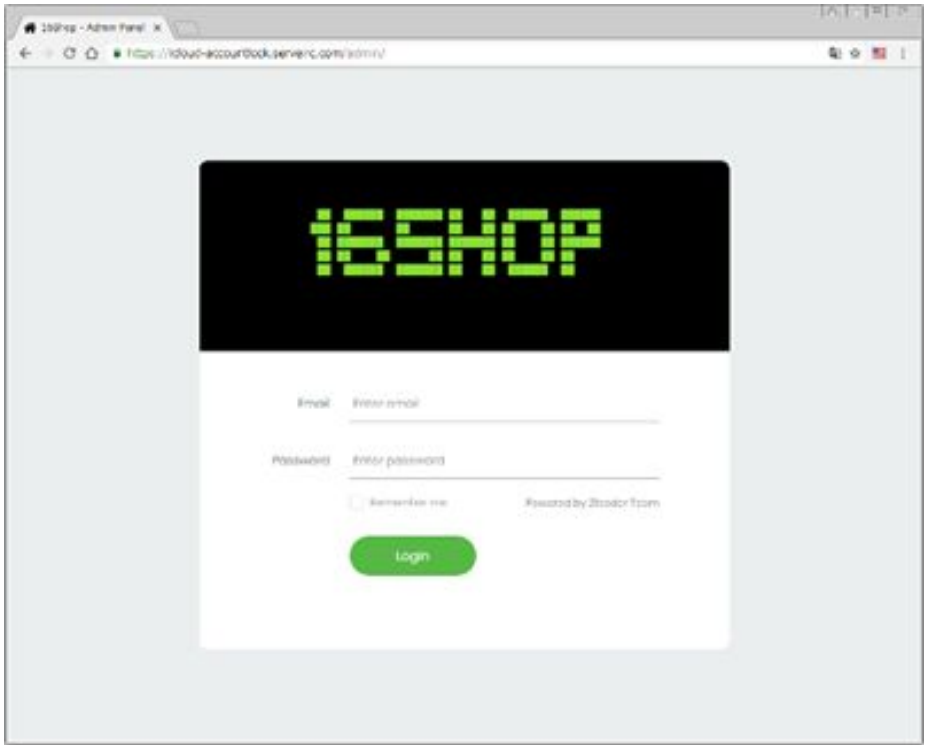


URLの文字列の長さ、左が正常サイト、右が悪性サイト

https://link.springer.com/article/10.1186/s13673-016-0064-3より

と対策を考えてもなかなかいたちごっこな状態

Phishing as a Service？参入障壁が下がっている？



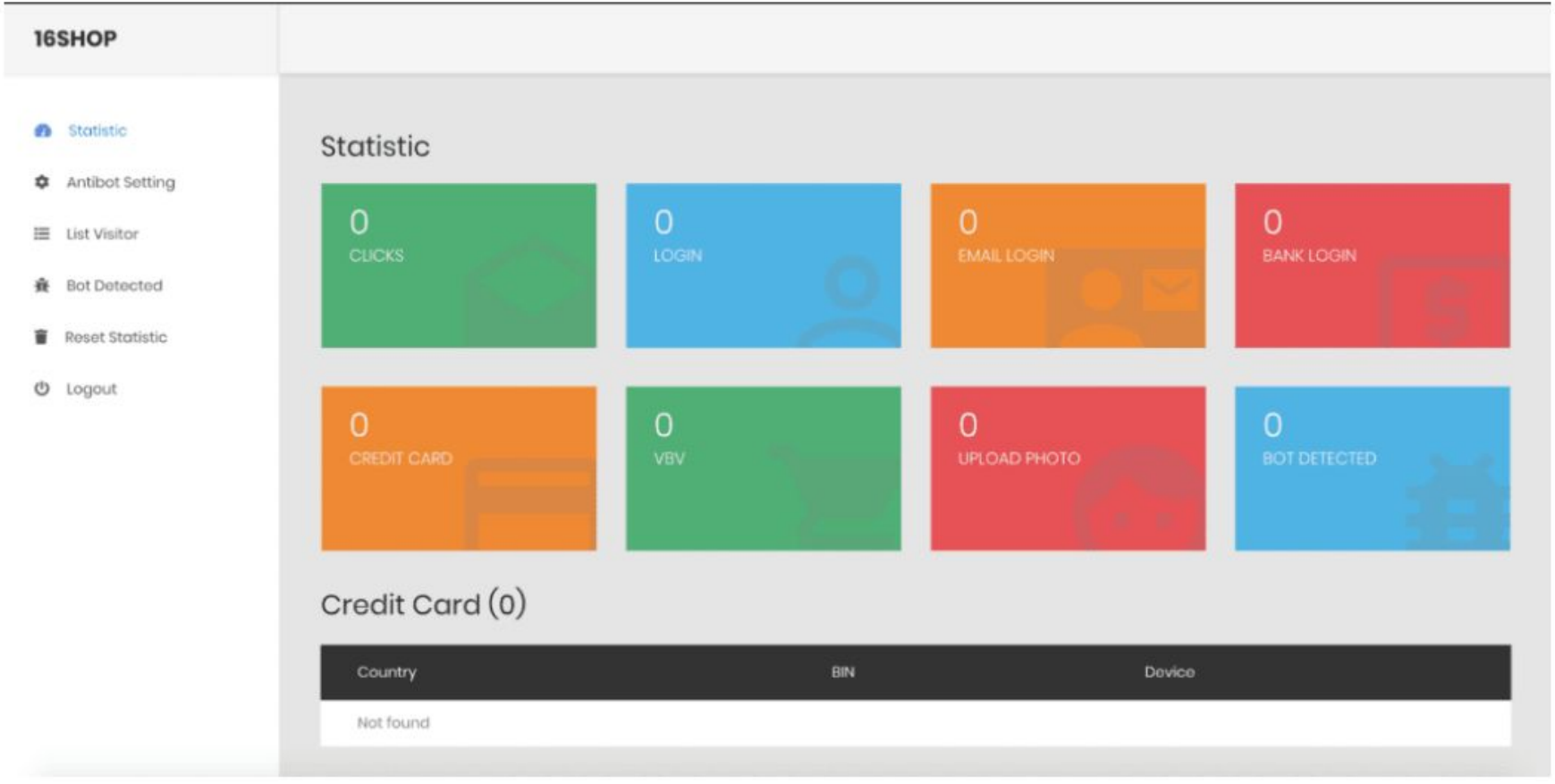
16SHOP Phishing tool kit

- 偽サイトのひな型
- フィッシングメール送信システム
- フィッシング対象管理インターフェイス
- などのTotal Phishing Solutionが出回っている
- 多くの人が利用可能で日々改悪されている

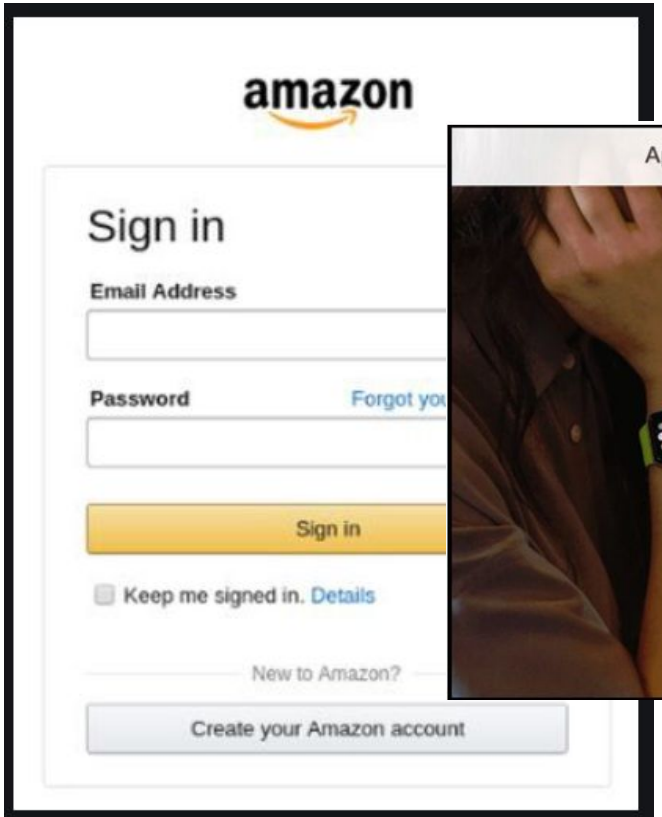
偽サイトのひな型選択→

Product Name	Version
Apple Scampage	2.1.0
Amazon Scampage	1.7.0
Amex Scampage	1.0.0
PayPal Scampage	1.2.0
Sender	2.0.0
Apple Validator CLI	1.1.0

DASHボード

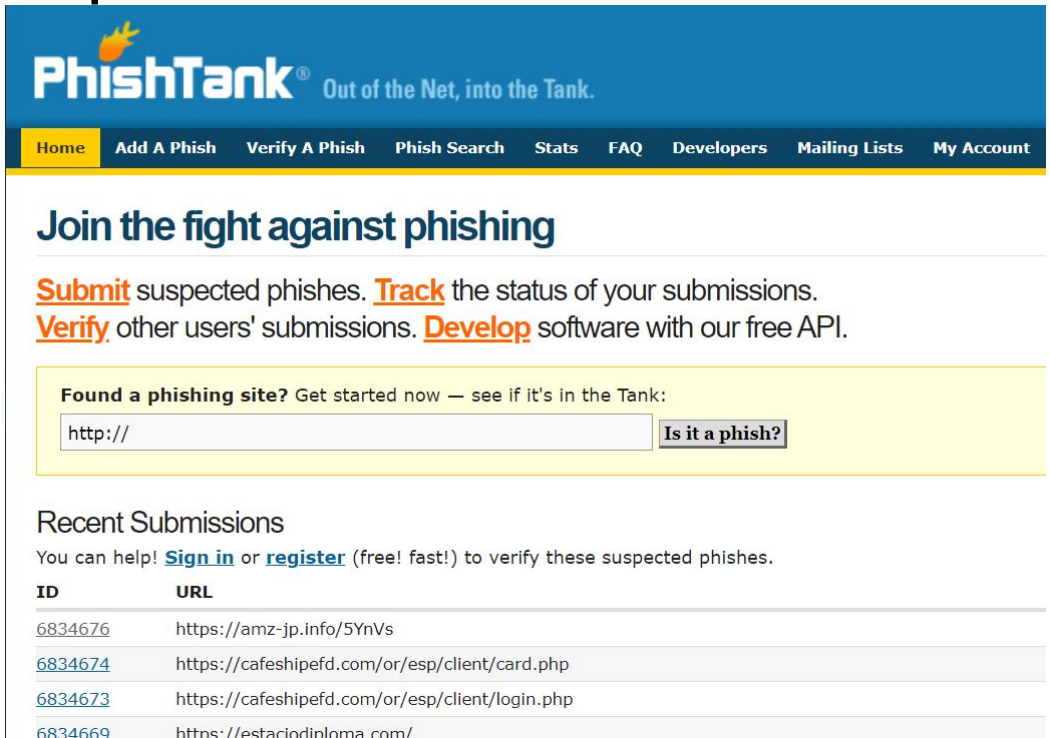

16SHOP
Statistic
Antibot Setting
List Visitor
Bot Detected
Reset Statistic
Logout
Statistic
0
CLICKS
0
LOGIN
0
EMAIL LOGIN
0
BANK LOGIN
0
CREDIT CARD
0
VBV
0
UPLOAD PHOTO
0
BOT DETECTED
Credit Card (0)
Country
BIN
Device
Not found

これはどちらも16SHOPが生成したサイト



参考：フィッシングサイトの共有サイト

https://www.phishtank.com/



フィッシングサイトで使われるgTLDの変化 2015年

Top 10 Phishing TLDs by Domain Score, 2016

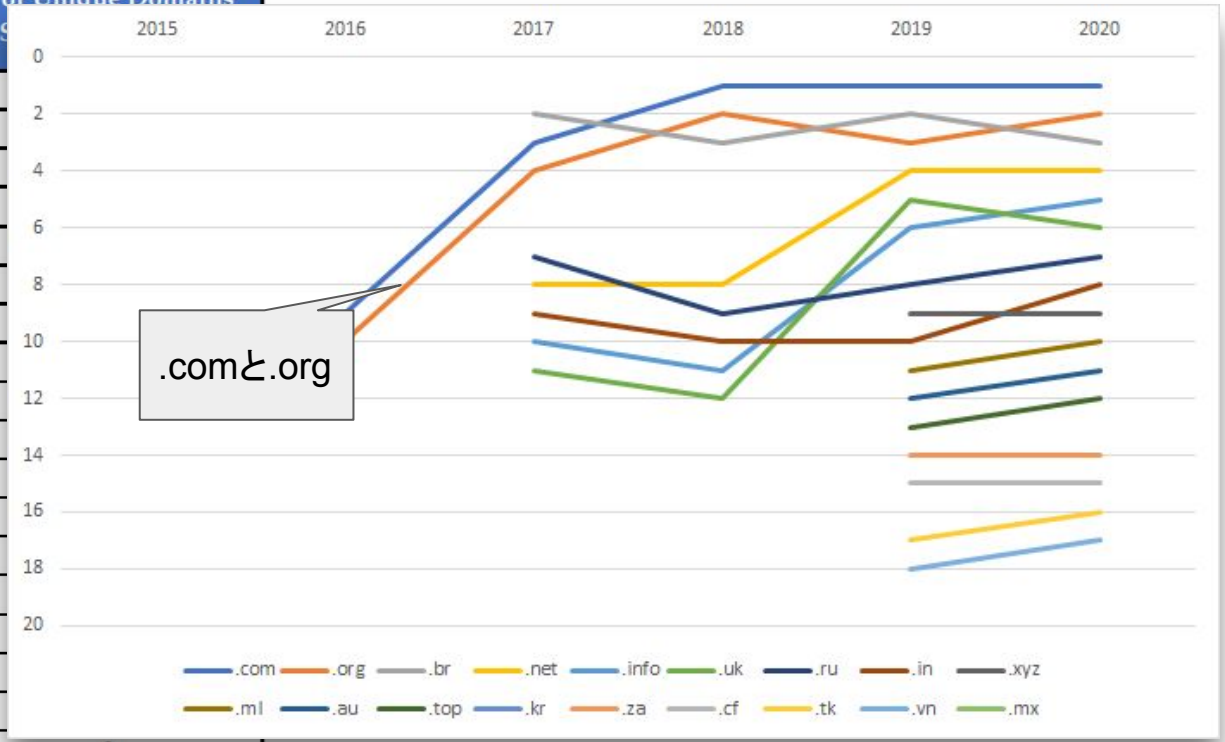
Minimum 25 phishing domains and 30,000 domain names in registry

	TLD	TLD Location	# Unique Phishing attacks 2016	Unique Domain Names used for phishing 2016	Domains in registry, Dec 2016	Score: Phishing domains per 10,000 domains 2016
1	.VE	Venezuela	1,206	1,045	77,555	134.7
2	.CC	Cocos (Keeling) Islands	13,708	13,061	1,750,000	74.6
3	.ML	Mali	1,448	1,434	220,000	65.2
4	.BD	Bangladesh	272	205	33,000	62.1
5	.KE	Kenya	260	207	50,000	41.4
6	.CENTER	new gTLD	136	135	32,809	41.1
7	.NG	Nigeria	307	267	65,000	41.1
8	.PW	Palau	2,782	2,702	675,000	40.0
9	.PK	Pakistan	301	227	62,000	36.6
10	.RU	Russia	4,340	2,433	735,000	33.1

https://docs.apwg.org/reports/APWG_Global_Phishing_Report_2015-2016.pdfより

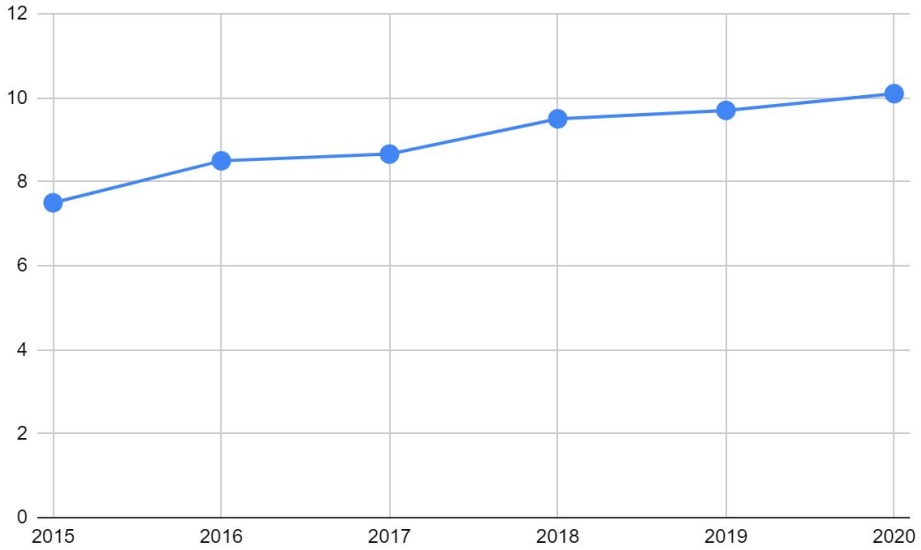
2020年1Q 徐々にgTLDが使われるケースが増加

Rank	TLD	Category	# of Unique Domains in S
1	.COM	generic	
2	.ORG	generic	
3	.BR	ccTLD	
4	.NET	generic	
5	.INFO	generic	
6	.UK	ccTLD	
7	.RU	ccTLD	
8	.IN	ccTLD	
9	.XYZ	nTLD	
10	.ML	ccTLD	
11	.AU	ccTLD	
12	.TOP	nTLD	
12	.KR	ccTLD	
13	.ZA	ccTLD	
14	.CF	ccTLD	
14	.TK	ccTLD	
14	.VN	ccTLD	
15	.MX	ccTLD	9



URLの文字列から悪性サイトを推測できないか？

- フィッシングサイトのURLから辞書ワードとのレーベンシュタイン距離を計測してみる。
- 辞書ワードとハイプロファイルなサイトのリストを作る



思いつく対策といたちごっこの例

- 悪いサイトのドメイン名をWHOISで引いてTake Down
  - GDPRの影響などでWHOISの情報はどんどん隠蔽
  - フィッシングサイトの寿命は 2時間程度と短命
- ドメイン名のレジストラで対応してもらえれば？
  - 悪い人御用達のレジストラが存在しそう
  - 完全匿名でドメイン名を運用可能か？
    - 本人確認 なし！
    - 課金 匿名決済OK!
    - 連絡先 到達しなくてもスルー
  - 結果としてコンタクト不能
- フィッシングのサイトはID/PASSがでたらめでもログインできるので試す？
  - フィッシングメールに偽サイトの ID/PASSがつけられるようになりました
-

もう少しドメイン名に注目して調査予定です。

ご清聴ありがとうございました。