ドメイン名関連情報を使用した ドメイン名悪用兆候の数値化と指標化の提案

米谷嘉朗[†1]

概要：
新型染症の世界的拡大の影響で，日常生活のオンライン化が急速に普及している．その状況を逆手に取ったオンライン詐欺も増加しており，フィッシングメールや悪質な広告に誘導され，悪性サイトへの意図せぬアクセスにより個人情報や資産の詐取被害が発生している．フィッシングなどの悪性サイトに関して，ドメイン名の文字列による判定，Web サイトのコンテンツによる判定，サーバ証明書による判定，DNS トラフィック分析による判定など，既にさまざまな悪性サイトの判定手法が提案されているが，攻撃者はさまざまな社会事象に機敏に反応し，手法および対象者を変化させながら判定を回避している．
本論文は，既存研究から得られた悪性サイトの特徴を複数組み合わせ，それら特徴の分布を人気サイトと対比して差異を数値（スコア）化し，与えられたドメイン名に悪用の兆候があるかをスコアリングする手法を提案する．提案手法では，スコアリングのベースとなるテーブルを事前に作成しておくことにより，ユーザ（クライアント）側が少数の HTTP(S)クエリで取得した情報を元にドメイン名の悪用兆候スコアを軽量に計算可能である．最終的には，ユーザが SMS やメッセンジャーなどで送られてきた URL にアクセスする際に，スコアをユーザに提示し注意を促すことを目的とする．
本論文では，その手法およびと予備的評価の結果を示す．

キーワード：ドメイン名，フィッシング，悪性サイト判定

Proposal for scoring and indicating malicious use of domain names by using domain name related information

YOSHIRO YONEYA[†1]

Abstract:
Due to pandemic of COVID-19 on the world, daily life on the Internet is rapidly prevailed. On-line fraud is also increasing by using this crisis as an opportunity, to theft personal credentials or assets by navigating victims to malicious sites via phishing mail or advertisements. Various studies have been taken to detect malicious sites like phishing sites by using domain name string, web contents, server certificates and DNS traffic analysis, and countermeasures are proposed, meanwhile attackers are swift to respond social phenomenon and change methods and targets to evade from detection.
This paper proposes a method for scoring and indicating malicious use of domain names by quantifying the difference between multiple characteristics of malicious site exposed by previous studies and those of popular sites. Proposed method prepares scoring table in advance and provide to user side, the Internet users (or client applications) can calculate malicious use indicator in light weight with obtaining some of domain name related information via several HTTP(S) queries. Final objective of the proposal is to induce attention of users with indicating score when they are going to access the URL embedded in messages of SMS or Messengers.
This paper shows the proposed method and result of preliminary evaluation.

Keywords: domain name, phishing, malicious sites

1. はじめに

1.1 背景

新型感染症の世界的拡大の影響による新しい生活様式として，オンラインを中心とした日常生活が急速に普及している．その状況を逆手に取り，経理担当者や社会的弱者を狙ったフィッシング詐欺[1]，宅配便の不在通知を騙る SMS やオンラインショッピングサイトを騙るメール[2]の増加が報告されている．さまざまな対策が研究・適用されているが，攻撃者はインターネットサービスの変化や社会情勢の変化に機敏に反応して手法や対象者を変化させ，フィッシング対策を回避している状況である．

本論文では，フィッシング対策の一助となるべく，Web サイトやドメイン名の運用され方の特徴から，ユーザがアクセスしようとしているサイトの悪用兆候をユーザ（クライアント）側で軽量に判定し，ユーザに注意を促す手法を提案する．

1.2 着目した項目

提案手法では，既存研究の知見に基づいて外部から観測可能な以下の 4 項目に着目し，ドメイン名関連情報を取得してサイトの悪用兆候を判定している．

1 株式会社日本レジストリサービス
Japan Registry Services Co., Ltd.

(1) ドメイン名の文字列

典型的なフィッシングの手法として，人気サイトと視覚的に類似したドメイン名が使用されることが多い．典型的な例として, 英数字以外の文字が使用可能な IDN [3]の悪用や，攻撃者の登録ドメイン名配下にサブドメインとして人気サイトドメイン名が使用されることがある．

(2) DNSSEC の利用

DNS 情報の一部書き換えによるドメイン名ハイジャックによってフィッシングを行う攻撃がある．この場合フィッシングに使われるドメイン名は正規のドメイン名であるためドメイン名からは判定できない．正規のドメイン名が DNSSEC で署名されていると, DNS 情報の一部書き換えによるドメイン名ハイジャックは検知可能となる．

(3) サーバ証明書

2020 年第 1 四半期時点で, フィッシングサイトの 74%がサーバ証明書を使用していると報告されている[1]. そのため, サーバ証明書の有無確認だけでは判定が困難であるが, 攻撃者は自己署名証明書の使用や無償の自動発行証明書を使用する傾向がある．

(4) ドメイン名登録情報

攻撃者は新規登録ドメイン名を短期間のみ使用しその後は廃棄する傾向がある．また，ネームサーバ情報などの登録情報変更にコストが発生するレジストラロックなどのオプションは避ける傾向がある．

1.3 本論文の研究目標

既存研究では悪性サイトが使用する攻撃手法やその検知手法の分析が行われ高い正解率を得ているが，インターネットユーザが未知の URL に遭遇した際に即時にその URL が悪性サイトと関連しているかを判定し提示する手法は示されていない．本論文の提案手法は最終的にユーザが SMS やメッセンジャーなどで送られてきた URL にアクセスする際に，悪用兆候の指標をユーザに提示し注意を促すことを目的とする．そのため，以下 3 つの研究課題を設定した．

Q1 1.2 節の(1)〜(4)で着目した項目から悪性サイトと通常サイトの区別は付けられるか

Q2 その区別と結果確認はユーザ（クライアント）側で軽量に実施できるか

Q3 確認した結果をユーザへ効果的に伝えられるか

本論文では，研究課題を解決するアプローチとして，既知の悪性サイトの特徴を複数組み合わせ，それら特徴の分布を人気サイトと対比して差異を数値（スコア）化・テーブル化し，与えられた個々のドメイン名についてスコアリングを可能として，ドメイン名の悪用の兆候を確認可能とする手法を提案する．その応用として, ユーザが SMS 等で受信した URL をクリックした際に起動されるデフォルトブラウザへスコアを計算する拡張機能を追加し，サイトコンテンツを表示する前にスコアを提示することを想定する．

1.4 本論文の構成

以降，2 章に関連研究，3 章に提案手法，4 章に提案手法の評価，5 章に考察，6 章にまとめを示す．

2. 関連研究

フィッシングサイトなど悪性サイトに関する研究では既にさまざまな分析や検知手法の提案が行われている．

フィッシング攻撃を分析・検知する手法に関しては, Web コンテンツに含まれるドメイン名の分析 [4]，DNS 設定情報を使用した分析[5]，DNS トラフィック分析による検知[6][7]，DNS と登録情報の複合分析[8][9]，機械学習による悪性 URL 検知[10][11]などが行われている．

個別の攻撃に着目したものでは，IDN ホモグラフ攻撃に関しては使用される文字の類似性分析[12][13], ブラウザでの対策サーベイ[14]，WHOIS・パッシブ DNS・ブラックリストを併用した分析[15]，悪性 IDN の振る舞いや使われ方のサーベイ[16][17], ブランドデータベースと類似文字の突合による保護的登録のサーベイ[18]などが行われている．また，サブドメインハイジャック対策では Public Suffix List [19]（以降，PSL）と DNSSEC を使用した提案[20]がある．

サーバ証明書の分析に関しては，サブドメインに攻撃対象ドメイン名を埋め込んだサーバ証明書のサーベイ[21], Lets' Encrypt の普及度調査[22], サーバ証明書普及度調査方法のサーベイと CT ログのカバー率調査[23]などが行われている．ドメイン名登録そのものに関する分析では，ドロップキャッチのサーベイ[24][25], 悪性ドメイン名の取り消しおよび再登録状況サーベイ[26]がある．

他に，セキュリティ警告をユーザに提示する手法に関しては，各ブラウザのセキュリティインジケータのサーベイ[27]，馴化抑制方式の評価[28]などが行われている．

3. 提案手法

3.1 データセット

本論文では, 悪性ドメイン名のリストとして 2020 年 4 月 27 日の PhishTank [29] Developer 用データベース（以降，悪性リスト）を，対比のためのドメイン名リストとして 2020 年 4 月 9 日の Tranco [30]（以降, 人気リスト）を使用した．

悪性リスト，人気リスト共に 2020 年 4 月 9 日の PSL を適用してユニークなドメイン名（以降，対象ドメイン名）を抽出し, 悪性リストの 6407 対象ドメイン名, 人気リストの上位 20000 対象ドメイン名について，ドメイン名関連情報を収集した．なお，PSL の適用に当たっては，リスト中のドメイン名が PSL と完全一致したものはそのものを，PSL が後方一致したものは PSL が完全一致した部分の左側のラベルまでの部分を対象ドメイン名とした．例えばドメイン名が www.example.com の場合, com が PSL のためその左側のラベルまでを含めた example.com が対象ドメイン名となる．PSL を適用した理由は，ドメイン名登録者がドメ

イン名登録業者（レジストラ・レジストリ）に申請したドメイン名文字列（ラベル）を特定でき，ドメイン名登録情報や DNSSEC 運用情報などの周辺情報が容易に得られるためである．

3.2 収集したデータ

悪性リスト，人気リストの個々の対象ドメイン名について，1.2 節で述べた 4 項目に関するデータを収集した．

(1) ラベル情報（以降，LABEL）

対象ドメイン名の最左ラベルから表 1 の情報を抽出し特徴量とした．なお，IDN については A-Label 形式[a]ではなく U-Label 形式に変換してから情報を抽出した．また，3.3 節で後述する機械学習で使用する際は，ラベル長以外の情報はラベル長で除算し値の範囲を正規化した．

表 1 LABEL 特徴量

名称	説明
LEN	ラベル長（ラベル中の文字数）
LETTERS	アルファベット（Letter）数
DIGITS	数字（Digit）数
HYPHENS	ハイフン（Hyphen）数
SCRIPTS	出現するスクリプト [b] 数（Letters, Digits, Hyphens; LDH は除く）
SCRCHGS	スクリプトの変化回数（LDH の変化も含む）
SIMCHARS	SimChar [31]を持つ文字数
SIMCHARCHGS	SimChar を持つ文字のスクリプト変化数（LDH の変化も含む）
MINSCRCHARS	LDH を除き最も文字数の少ないスクリプトの文字数
MAXSCRCHARS	LDH を除き最も文字数の多いスクリプトの文字数
PVALIDCHARS	PVALID [32] 文字数

(2) DNSSEC 運用情報（以降，DNSSEC）

DNS フルリゾルバに対象ドメイン名のネームサーバ(NS)情報を問い合わせて取得した表 2 の情報を特徴量とした．DNS フルリゾルバには CD ビット[c]付きと DO ビット付きの 2 回問い合わせを行い，名前解決成否は CD ビット付き側の応答で，DNSSEC 署名有無は DO ビット付き側の応答で，DNSSEC 検証成否は CD ビット付きおよび DO ビット付き両側の応答比較で判定した．

表 2 DNSSEC 特徴量

名称	説明
NSRES	対象ドメイン名の名前解決成否
SIGNED	DNSSEC 署名有無
VALIDATE	DNSSEC 検証成否

(3) サーバ証明書情報（以降，SRVCERT）

対象ドメイン名の TCP ポート 443 に HTTP リクエストを送付してサーバ証明書を取得し，CT ログ[d]に記録されているシリアル番号と突合した．また，DNS フルリゾルバに対象ドメイン名の CAA レコード[e]を問い合わせ，取得した issue ドメイン名を CCADB [33]に登録されている認証局ドメイン名[34]と突合した．結果として得られた表 3 の情報を特徴量とした．CT ログ上の記録有無確認は CRT.SH [35]でサーバ証明書のシリアル番号を検索して ID を取得し，その ID を再度 CRT.SH で検索してシリアル番号を取得し，サーバ証明書のシリアル番号と突合している．

表 3 SRVCERT 特徴量

名称	説明
CERT	対象ドメイン名のサーバ証明書取得成否
CERTTYPE	サーバ証明書タイプ（DV，OV）
CTLOG	CT ログ上のサーバ証明書発行記録有無
CAA	対象ドメイン名の CAA レコード取得成否
CCADB	CCADB の認証局ドメイン名一覧に CAA レコードのドメイン名有無

(4) ドメイン名登録情報（以降，REG）

対象ドメイン名を RDAP [36]サーバに問い合わせ登録情報を取得し表 4 の特徴量とした．クライアント禁止ステータスおよびサーバ禁止ステータスはそれぞれ 4 つずつ定義されている[37]が，特徴量としてはそれらが設定されている数とした．

表 4 REG 特徴量

名称	説明
RDAP	RDAP 情報取得成否
DAYSPASTREG	ドメイン名新規登録後の経過日数
DAYSPASTUPD	登録情報最終更新後の経過日数
CLIENTHOLD	クライアントホールドステータス
CLIENTPRHBS	クライアント禁止ステータス数
SERVERHOLD	サーバホールドステータス
SERVERPRHBS	サーバ禁止ステータス数

3.3 特長量と分類閾値の決定

悪性リストを分析し，寄与する特徴量およびクラスタ分類の閾値を決定した．

(1) 寄与する特徴量の抽出

悪性リストについて収集した 4 項目の各特徴量を，項目毎に教師なし機械学習（ランダムフォレスト）でジニ係数を計算し，表 5 に示す特徴量を寄与するものとして抽出した．なお，人気リストは必ずしも良性であるとは言えないため，正解データとした教師あり機械学習はしていない．

(2) クラスタ数と特徴量毎のクラスタリング閾値の決定

抽出した特徴量をまとめ，再度教師なし機械学習（Partition Around Medoids ; PAM）でクラスタリングし，クラスタ数および分類の閾値を得た．具体的には，クラスタ数は作成したシルエット図を比較しその結果から 8，特徴量毎のクラスタリング閾値は PAM で得られた当該特徴量

a) A-Label は xn--で始まる ASCII 互換形式の IDN ラベル表現，U-Label は Unicode 文字列形式の IDN ラベル表現
b) 言語を表記する書き文字（用字）のことで，アルファベット，漢字，平仮名，ハングルなど
c) フルリゾルバに DNSSEC 検証の無効化（Checking Disabled; CD）および有効化（DNSSEC OK; DO）を指示する拡張フラグ
d) サーバ証明書の発行記録で，証明書の透明性（Certificate Transparency; CT）を監視可能とする
e) サーバ証明書の発行を許可する認証局を記述する DNS レコード（Certification Authority Authorization; CAA）で，ドメイン名登録者が設定する

表 5 各項目において寄与する特徴量

項目	特徴量
LABEL	LETTERS，DIGITS，HYPHENS，SCRCHGS，SIMCHARCHGS
DNSSEC	NSRES，SIGNED，VALIDATE
SRVCERT	CERT，CERTTYPE，CTLOG
REG	RDAP，DAYSPASTREG，CLIENTHOLD，CLIENTPRHBS，SERVERHOLD，SERVERPRHBS

の最小値および最大値とした．閾値を用いたクラスタリングでは，主にクラスタ F および H に他のクラスタとの重複が見られたが，その割合は非常に小さく，閾値の選択は妥当であると考えられる．各クラスタの傾向と悪性リストのクラスタリング結果を表 6 に示す．傾向は，クラスタ毎に作成した箱ひげ図を比較し，顕著に差異が見られる特徴量およびその中位値から得た．

3.4 数値化と指標化

悪性リストの分析から得られた特徴量と閾値を使用し，人気リストをクラスタリングした．人気リストも悪性リストと同様にクラスタ間の重複割合は非常に小さく，また，悪性リストの閾値では分類外となる対象ドメイン名の割合は 2.1%であった．

図 1 は横軸にクラスタ（N は分類外），縦軸にクラスタ毎の割合を示しており，悪性リストと人気リストでは各クラスタに分類される対象ドメイン名の割合に差が見られる．提案手法ではこの差に着目し，数値（スコア）化した．

最初に，クラスタの傾向を顕著に表す条件（表 7）を定義した．なお，名前解決失敗と DNSSEC 検証失敗はドメイン名ハイジャックなどクリティカルな攻撃下にある可能性があるため，条件定義に追加している．

次に，悪性リストおよび人気リストのクラスタ毎にそれら条件を複数含む対象ドメイン名の割合を算出し，スコアテーブル（表 8）を作成した．スコアテーブルの最左カラムは条件の組み合わせを示しており，白抜き数字は条件の否定を表している．例えば❶④は RDAP 取得失敗かつスクリプト変化数 0 回を示す．

最後に，悪性リスト側のスコアを悪用の兆候が「よりありそう」（以降，more likely），人気リスト側のスコアを悪用の兆候が「よりなさそう」（以降，less likely）と定義し，対象ドメイン名を個別にクラスタリングして more likely と less likely のスコアをスコアテーブルに基づきそれぞれ合算し，結果として得た 2 つの値を悪用兆候の指標とした．例えば，悪性リストおよび人気リストどちらにも現れた対象ドメイン名 google.com の場合は，収集したドメイン名関連情報からクラスタ B に分類され，また，条件①から⑦までを満たしていたため more likely は 29，less likely は 58 となり，悪性リストに現れた iapple-support.net はクラスタ D に分類され，条件①⑤⑥を満たしていたため more likely は 42，less likely は 5 となる（図 2 上にピンで注釈した）．

表 6 各クラスタの傾向と悪性リストの分類結果

クラスタ	項目	傾向	割合・重複
A	LABEL	英字	割合 22.3% 重複 D 0.3% H 0.1%
	DNSSEC	署名無	
	SRVCERT	DV，CT ログ有	
	REG	登録後 1000 日以上，クライアント禁止ステータス 1 以下	
B	LABEL	英字	割合 19.5% 重複 F 0.6% H 0.9%
	DNSSEC	署名無	
	SRVCERT	DV，CT ログ有	
	REG	登録後 1000 日以上，クライアント禁止ステータス 3 以上	
C	LABEL	英字	割合 18.6% 重複 G 0.0%
	DNSSEC	署名無	
	SRVCERT	DV，CT ログ有	
	REG	RDAP 無	
D	LABEL	英字以外が複数混在	割合 14.9% 重複 F 1.9% H 0.2%
	DNSSEC	署名無	
	SRVCERT	DV，CT ログ有	
	REG	登録後 1000 日未満	
E	LABEL	英字以外が複数混在	割合 9.4% 重複 なし
	DNSSEC	署名無	
	SRVCERT	OV，CT ログ有	
	REG	RDAP 無	
F	LABEL	英字以外が複数混在	割合 5.9% 重複 H 1.3%
	DNSSEC	ほぼ署名無	
	SRVCERT	CERT 無	
	REG	登録後 1000 日以上，クライアント禁止ステータス 1 以下	
G	LABEL	英字以外が複数混在	割合 6.1% 重複 なし
	DNSSEC	署名無	
	SRVCERT	CERT 無	
	REG	RDAP 無	
H	LABEL	英字以外が複数混在	割合 3.2% 重複 なし
	DNSSEC	名前解決失敗	
	SRVCERT	CERT 無	
	REG	登録後 100 日未満，CLP1 以下	

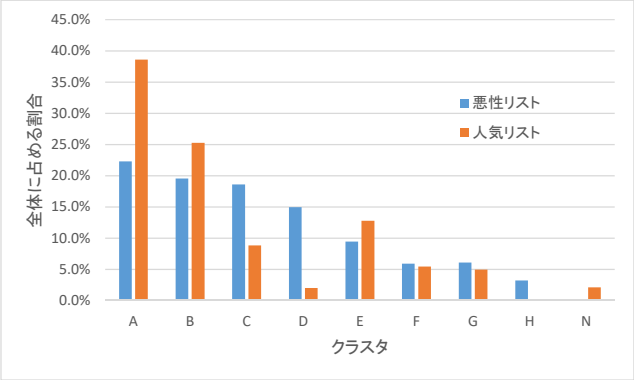

図 1 悪性リストと人気リストのクラスタ毎の割合

表 7 クラスタの傾向を顕著に表す条件

条件	内容
①	ドメイン名登録情報（RDAP）取得成功
②	新規登録後の経過日数が 1000 日以上
③	クライアント禁止ステータスが 3 以上
④	スクリプトの変化回数が 0 回
⑤	サーバ証明書取得成功
⑥	CT ログ有
⑦	証明書タイプが OV
⑧	名前解決失敗
⑨	DNSSEC 検証失敗

表 8 スコアテーブル

スコア / 条件組み合せ	A		B		C		D		E		F		G		H		N	
	more likely	less likely	more likely	less likely	more likely	less likely	more likely	less likely	more likely	less likely	more likely	less likely	more likely	less likely	more likely	less likely	more likely	less likely
❶❹	0	0	0	0	4	1	0	0	5	1	0	0	2	0	0	0	0	0
❶④	0	0	0	0	14	8	0	0	4	12	0	0	4	5	0	0	0	0
❶❺	0	0	0	0	0	0	0	0	0	0	0	0	6	5	0	0	0	0
❶⑤❻	0	0	0	0	5	2	0	0	4	2	0	0	0	0	0	0	0	0
❶⑤⑥❼	0	0	0	0	13	7	0	0	0	0	0	0	0	0	0	0	0	0
❶⑤⑥⑦	0	0	0	0	0	0	0	0	5	11	0	0	0	0	0	0	0	0
❶④⑤⑥⑦	0	0	0	0	0	0	0	0	3	10	0	0	0	0	0	0	0	0
①❷	5	2	9	1	0	0	15	2	0	0	4	1	0	0	6	0	0	0
①②	18	37	11	25	0	0	0	0	0	0	5	7	0	0	0	0	0	2
①❽	22	39	20	25	0	0	15	2	0	0	8	8	0	0	6	0	0	2
①③	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
①❹	3	3	3	2	0	0	5	0	0	0	3	1	0	0	3	0	0	0
①④	20	35	16	23	0	0	10	2	0	0	6	6	0	0	3	0	0	2
①❺	0	0	4	3	0	0	3	0	0	0	8	7	0	0	6	0	0	1
①⑤❻	6	6	6	4	0	0	3	0	0	0	0	0	0	0	0	0	0	0
①⑤⑥❼	14	16	8	8	0	0	7	1	0	0	0	0	0	0	0	0	0	0
①⑤⑥⑦	3	17	2	10	0	0	3	1	0	0	0	0	0	0	0	0	0	1
①②③④⑤⑥⑦	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
⑧	10	0	10	0	10	0	10	0	10	0	10	0	10	0	10	0	10	0
⑨	100	0	100	0	100	0	100	0	100	0	100	0	100	0	100	0	100	0

3.5 実装方式

提案手法の実装は以下 2 つの要素から構成される．

(1) スコアテーブル生成部

3.1～3.4 節で述べた手法で機械学習データの収集，クラスタ分類の閾値およびスコアの算出を行い，静的なスコアテーブルを生成する．次に述べるユーザアプリケーション部の開発者側が実行し結果をアプリケーション内に保持すればよく，ユーザ側での生成は不要である．

(2) ユーザアプリケーション部

対象ドメイン名に関する 4 項目の情報取得と各特長量の値の計算，内部に持つスコアテーブルに基づくスコアリングとユーザへの悪用兆候指標注釈を行う．

このように構成を分けることで，ユーザアプリケーションは動作が軽量で即時に結果を提示可能な実装とすることができる．

4. 提案手法の評価

4.1 悪性リストおよび人気リストの評価

提案手法を用い，悪性リストおよび人気リストを評価した．それぞれのリストに含まれる個々のドメイン名について more likely および less likely の値をスコアリングし，クラスタ毎に分布をプロットした．図 2 に悪性リストの分布，図 3 に人気リストの分布を示す．図の横軸は more likely，縦軸は less likely のスコアであり，円は各クラスタ，円の大きさは当該クラスタが全体に占める割合，斜線は more likely と less likely の値が同じ点を結んだ対角線を表現している．クラスタ A に関しては悪性・人気の判別は困難である．その他のクラスタに関しては，悪性リストでは more likely のスコアが less likely より高い傾向が見られ（図 2 で斜線の下側に分布しているクラスタが多い），人気リストでは more likely のスコアが less likely より低い傾向が見られる（図 3 で斜線の上側に分布しているクラスタが多い）．ユーザがメッセージ中のリンクをクリックした際に，当該リンクのドメイン名（対象ドメイン名）をスコアリングし，リファレンスとなるリスト（ここでは悪性リストを仮定する）の分布図上に対象ドメイン名のスコアをマップアプリケーションのピンのようにプロットして注釈を与えることで，ユーザに悪用兆候指標を示すことができると考えられる（図 2 上に google.com および iapple-support.net のスコアをピンで注釈した）．

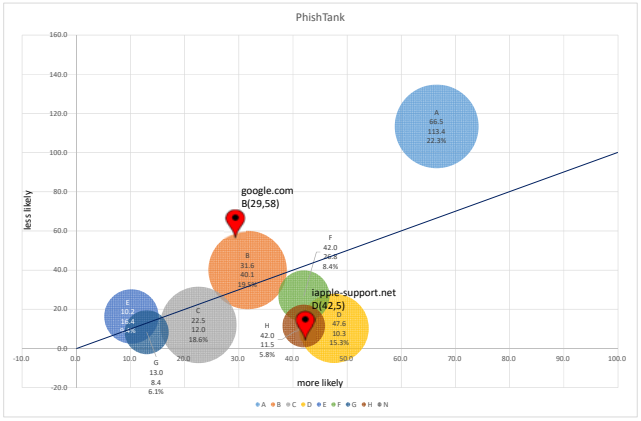

図 2 悪性リストのスコア分布

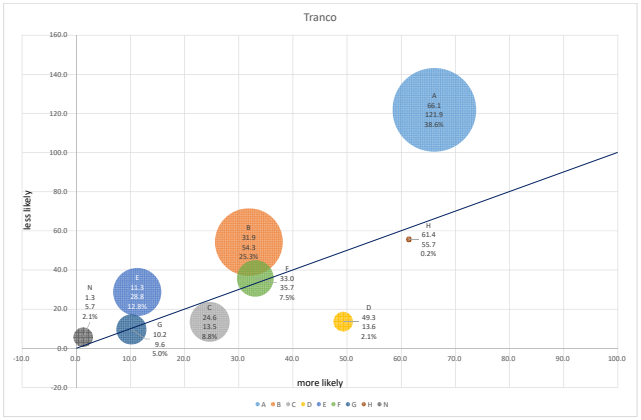

図 3 人気リストのスコア分布

4.2 他のドメイン名リストでの評価

他の悪性ドメイン名リストおよび JPRS が管理しているドメイン名リスト（ランダムサンプリング）で提案手法を評価した．

(1) 他の悪性ドメイン名リストでの評価

他の悪性ドメイン名リストとして，2020 年 7 月 15 日の Spam404 [38]（以降，Spam404）および 2020 年 7 月 29 日の Malware Domains Blocklist [39]（以降，MalwareDomains）を使用した．いずれのリストも悪性リストと同様に PSL を事前適用してユニークな対象ドメイン名を抽出し，4 項目の情報を収集してスコアリングした．Spam404，MalwareDomains ともにリスト全体としては悪性リストと

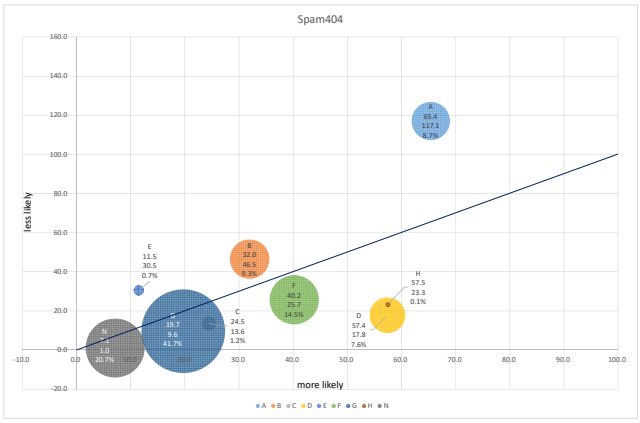

図 4 Spam404 のスコア分布

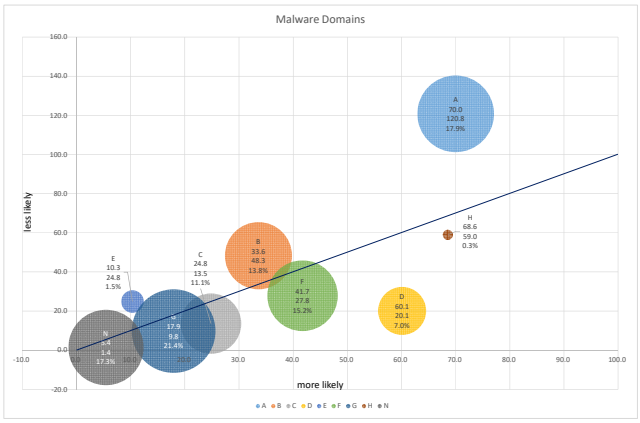

図 5 MalwareDomains のスコア分布

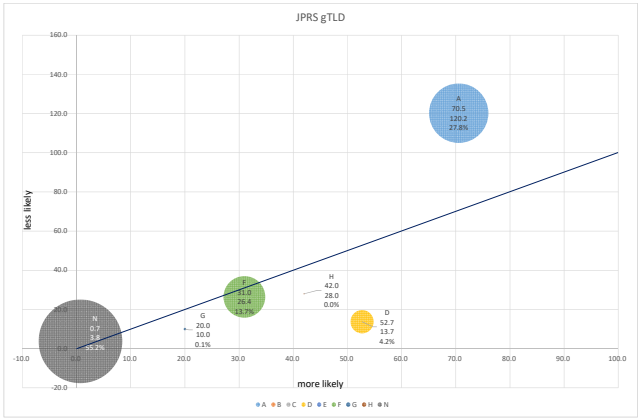

図 6 JPRS gTLD のスコア分布

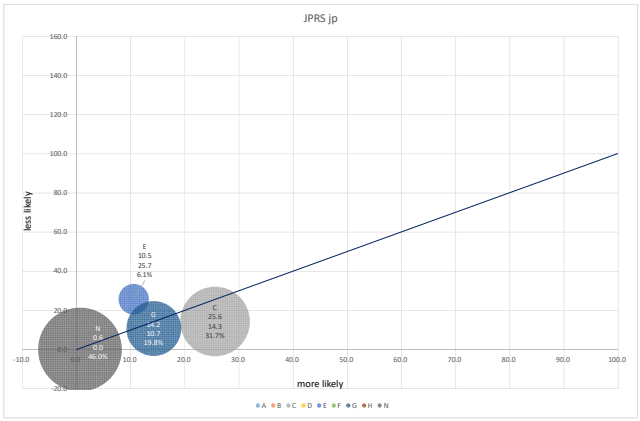

図 7 JPRS jp のスコア分布

同様の傾向を示している（図 4，図 5）.

(2) JPRS 管理ドメイン名での評価

さらに他のドメイン名リストとして，2020 年 7 月 1 日の JPRS 管理ドメイン名（gTLD 取次からランダムサンプリング，以降 JPRS gTLD，および JP ドメイン名からランダムサンプリング，以降 JPRS jp）を使用した．いずれもリスト全体としては人気リストとは異なる傾向を示しており（図 6，図 7），提案手法で決定した閾値では分類外となるものが多い．人気リストはむしろ悪性リストとの類似の傾向があるが，1.2 節（3）で述べたサーバ証明書使用率が高いことなど悪性ドメイン名は人気ドメイン名を模倣したドメイン名運用を行うためと考えられる．

また，JPRS gTLD と JPRS jp の傾向の違いは，前者は REG を RDAP で取得可能なことに対し，後者は RDAP が提供されていないため取得できていないことによる．そのため，クラスタリング結果が大きく異なっている．JPRS jp に関しては，WHOIS 参照で得た情報から新規登録後の経過日数（DAYSPASTREG）を補うと，JPRS gTLD と同様の傾向であることがわかった（図 8）．このことから，REG 有無は悪用兆候の指標化に特に大きく寄与していることがわかる．

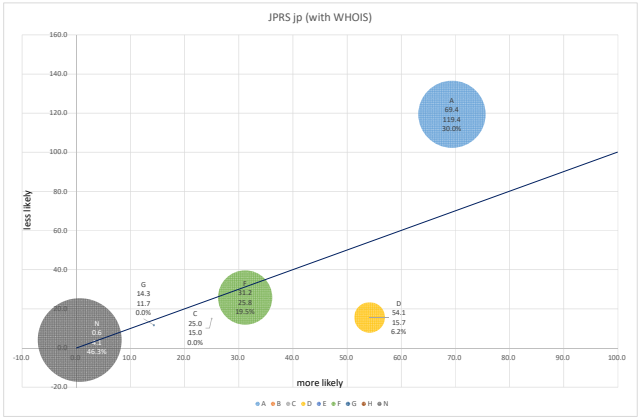

図 8 JPRS jp のスコア分布（WHOIS で補足）

5. 考察

5.1 指標とユーザへの注釈について

提案手法によるスコアリングは対象ドメイン名が悪性か良性かを判定しているわけではなく，ドメイン名運用情報の傾向から悪用の兆候が高いかどうかだけを表している．そのため，ユーザに兆候を示す際には，ユーザにとって馴化を起こしにくい内容および頻度での注釈が肝要である．内容に関しては，4.1 節で述べたように悪性のスコア分布と対象ドメイン名のスコアを示すとともに，Google Safe Browsing [40]など既存の悪性サイト判定サービスへの誘導も必要であると考えられる．また，頻度に関してはユーザが他のアプリケーションでリンクをクリックしブラウザを起動するタイミングなど，ユーザ行動が変わる境界に限定することが考えられる．

5.2 クラスタリング閾値とスコアの算出について

提案手法におけるクラスタリング閾値の算出およびスコアの算出はヒューリスティックである．悪性ドメイン名としてフィッシングドメイン名リストを，対比するドメイン名として人気ドメイン名リストを使用しているが，それらが悪性ドメイン名および良性ドメイン名を代表しているかは検証していない．ドメイン名を悪用する第三者の即応性を考慮すると，運用的視点からはスコアテーブルの更新頻度を高めた方がよい可能性がある．

5.3 攻撃者によるスコア操作可能性について

提案手法では表 7で示したように4項目のうち特にREG特徴量と SRVCERT 特徴量が寄与している．ドメイン名登録後の経過日数はレジストリが管理する情報であり攻撃者の操作は不可能である．また，信頼性の高い認証局が OV 証明書を発行する際は申請組織の実在証明を求めるため，存在を秘匿したい攻撃者が使用する可能性は低い．そのため，攻撃者によるスコア操作の可能性は低いと考えられる．

5.4 実装の軽量性について

提案手法で取得する対象ドメイン名の運用データは，LABEL は外部クエリを必要とせず，SRVCERT は HTTP(S)クエリが 3 回（CAA レコード取得を除く），REG は HTTP(S)クエリが 1 回で取得できる．DNSSEC は DNS クエリが 2 回，CAA レコードは DNS クエリが 1 回必要で，従来は DNS パケットでの送受信を要し困難であったが，DoH [41]および DNS メッセージの JSON フォーマット[42]が標準化されたことにより，現在は HTTPS クエリで取得できる．そのため，すべての情報取得と処理は JavaScript などのライブラリが豊富でブラウザ拡張やスマートフォンアプリケーション作成に使用されている言語で記述でき，スコアリングもテーブルの参照で可能なため，軽量に実装できる．

5.5 静的テーブルの更新について

提案手法でユーザ側アプリケーションが参照する外部の静的テーブルは PSL，SimChar リスト，CCADB の CAA 一覧である．いずれも数 KB～数百 KB 程度のサイズであり，アプリケーションが内部に保持しても近年のユーザ側の実行環境では支障はない．最も頻繁に更新されるのは PSL であるが，クラウドサービスなどのサブドメイン追加が主な理由のため毎回追従していなくとも影響は少ない．SimChar リストの更新は Unicode の改版（年 1 回）と同程度の頻度であると考えられる．CCADB の CAA 一覧は認証局の事情に依存するが顧客サービスに直接影響する運用パラメータであるため更新頻度は低いと考えられる．提案手法で作成するスコアテーブルは 5.2 節で述べた通り一定の更新頻度が必要と考えられるため，アプリケーションが保持する外部の静的テーブル更新もそのタイミングで実施すれば十分と考えられる．

5.6 実装時の限界について

提案手法のユーザ側実装に関して，現状ではいくつか特定のサービスに依存している部分がある．

(1) JSON で応答を得られる DoH サーバは Google Public DNS [43]か Cloudflare 1.1.1.1 [44]のみである．これらパブリックリゾルバサービスは特定の国や地域ではアクセスがブロックされることがある[45][46].

(2) 一括して CT ログが確認できるサービスは CRT.SH のみである．CRT.SH への連続アクセスではリトライが必要になることもあり，タイムアウトとエラーハンドリングを考慮しなければならない.

(3) ccTLD では RDAP を提供していないレジストリが多くある．WHOIS は必ず提供されているが，出力形式が統一されておらずパーサの実装が複雑となる．また，RDAP を提供しているレジストリにおいても，サーバ側で問い合わせ数制限を設けていることがあり，頻繁な問い合わせはできない．一部のレジストリにおいては，JSON フォーマットが標準に則っていないことがあり，情報取得に失敗する可能性がある．

5.7 関連業界への提言

(1) レジストリ・レジストラ業界

ドメイン名の REG 参照を容易にする RDAP の導入・提供を早急に進めるとよい．その際，ドメインオブジェクトの登録・更新・管理主体変更などの日付記録を標準に則った形式で出力すること，正当な参照を妨げないレートリミットの採用を検討するとよい.

(2) Web ホスティング業界

サブドメインハイジャックを減らす工夫をするとよい．その際，管理ドメイン名の PSL への登録を行うこと，ワイルドカード証明書の使用を適切にすること，サブドメインの DNSSEC 署名を有効にすることを検討するとよい.

(3) ドメイン名登録者

攻撃者の模倣を困難とするため，Web サービスのサービスドメイン名は固定し深い階層は使わず，レジストリロックまたはレジストラロックをかけるとよい．また，証明書は DV ではなく OV を使うとよい.

6. まとめ

悪性ドメイン名リストおよび人気ドメイン名リストを使用し，ドメイン名およびドメイン名運用に関連する複数の情報を取得・対比することでドメイン名悪用兆候をスコア化した．そのスコアテーブルに基づきユーザ側アプリケーションが個々のドメイン名をスコアリングし，ユーザに悪用兆候の指標を示す手法を提案した.

ユーザ側アプリケーションが個々のドメイン名に関する情報を取得するために使用するプロトコルは HTTP(S)のみであり，既存のライブラリを使用することが容易である．また，テーブルの規模は小さく，計算アルゴリズムも単純であるためユーザ側アプリケーションは軽量に実装できる.

今後，考察で述べた検討課題への対応と，ユーザ側アプリケーションの PoC（Proof of Concept）実装を行う．

謝辞 本研究への有益なアドバイスを頂いた国立情報学研究所福田健介准教授にこの場を借りて感謝いたします．

参考文献

[1] APWG (Anti Phishing Working Group), Phishing Activity Trends Report 1st Quarter 2020
[2] フィッシング対策協議会，フィッシングに関するニュース，https://www.antiphishing.jp/news/alert/
[3] John Klensin, Internationalized Domain Names for Applications (IDNA): Definitions and Document Framework, RFC 5890, August 2010.
[4] Hossein Shirazi et al., "Kn0w Thy Doma1n Name": Unbiased Phishing Detection Using Domain Name Based Features, 2018, 23nd ACM on Symposium on Access Control Models and Technologies
[5] Olivier van der Toorn et al., Melting the Snow: Using Active DNS Measurements to Detect Snowshoe Spam Domains, 2018, IEEE/IFIP Network Operations and Management Symposium
[6] Michael Weber et al., Unsupervised Clustering for Identification of Malicious Domain Campaigns, 2018, The 13th ACM ASIA Conference on Computer and Communications Security
[7] Giovane C. M. Moura et al., nDEWS: a New Domains Early Warning System for TLDs, IEEE/IFIP International Workshop on Analytics for Network and Service Management, 2016
[8] Giovane C. M. Moura et al., Domain names abuse and TLDs: from monetization towards mitigation, 3rd IEEE/IFIP Workshop on Security for Emerging Distributed Network Technologies, 2017
[9] Egon Kidmose et al., Heuristic methods for efficient identification of abusive domain names, 2018, International Journal On Cyber Situational Awareness, Volume 3, Number 1
[10] Malicious URLs Detection Using Decision Tree Classifiers and Majority Voting Technique, 2018, Cybernetics and Information Technologies, Volume 18: Issue 1
[11] Hong Zhao et al., Malicious Domain Names Detection Algorithm Based on N-Gram, 2019, Journal of Computer
[12] Jonathan Woodbridge et al., Detecting Homoglyph Attacks with a Siamese Neural Network, 2018, 1st Deep Learning and Security Workshop, co-located with the 39th IEEE Symposium on Security and Privacy
[13] Hiroaki Suzuki et al., ShamFinder: An Automated Framework for Detecting IDN Homographs, 2019, ACM Internet Measurement Conference 2019
[14] Tyson McElroy et al., The 2017 homograph browser attack mitigation survey, 2017, Australian Information Security Management Conference 2017
[15] Baojun Liu, A Reexamination of Internationalized Domain Names: the Good, the Bad and the Ugly, 2018, IEEE/IFIP International Conference on Dependable Systems and Networks
[16] Florian Quinkert et al., It's Not What It Looks Like: Measuring Attacks and Defensive Registrations of Homograph Domains, 2019, IEEE Conference on Communications and Network Security
[17] Victor Le Pocha et al., Funny Accents: Exploring Genuine Interest in Internationalized Domain Names, 2019, Passive and Active Measurement conference 2019
[18] Daiki Chiba et al., DomainScouter: Understanding the Risks of Deceptive IDNs, 2019, 22nd International Symposium on Research in Attacks, Intrusions and Defenses
[19] LEARN MORE ABOUT THE PUBLIC SUFFIX LIST, https://publicsuffix.org/learn/
[20] Hijacking DNS Subdomains via Subzone Registration: A Case for Signed Zones, Peter Thomassen et al., Open Journal of Web Technologies (OJWT) Volume 5, Issue 1, 2018
[21] Richard Roberts et al., You Are Who You Appear to Be A Longitudinal Study of Domain Impersonation in TLS Certificates, 2019, The 26th ACM Conference on Computer and Communications Security
[22] Maarten Aertsen at el., No domain left behind: is Let's Encrypt democratizing encryption?, 2017, Applied Networking Research Workshop 2017
[23] Benjamin VanderSloot et al., Towards a Complete View of the Certificate Ecosystem, 2016, ACM Internet Measurement Conference 2016
[24] Chaz Lever et al., Domain-Z: 28 Registrations Later Measuring the Exploitation of Residual Trust in Domains, 2016, IEEE Symposium on Security and Privacy
[25] Najmeh Miramirkhani, Panning for gold.com: Understanding the Dynamics of Domain Dropcatching, 2018, World Wide Web Conference 2018
[26] Eihal Alowaisheq et al., Cracking the Wall of Confinement: Understanding and Analyzing Malicious Domain Take-downs. 2019, The Network and Distributed System Security Symposium
[27] Adrienne Porter Felt et al., Rethinking Connection Security Indicators, 2016, 2016 Symposium on Usabile Privacy and Security
[28] 皆川諒 他, 馴化を抑制しうる新たなセキュリティ警告の探求:「かわいい」とその付加刺激の効果に関する評価, 2017, コンピュータセキュリティシンポジウム
[29] What is phishing?, https://phishtank.com/what_is_phishing.php
[30] Victor Le Pochat et al., TRANCO: A Research-Oriented Top Sites Ranking Hardened Against Manipulation, 2019, Network and Distributed System Security Symposium
[31]https://github.com/shamfinder/shamfinder/blob/master/simchar.json
[32] Patrik Faltstrom, The Unicode Code Points and Internationalized Domain Names for Applications (IDNA), RFC 5892, August 2010.
[33] Common CA Database, https://www.ccadb.org/
[34] All CAA Identifiers Report, https://ccadb-public.secure.force.com/ccadb/AllCAAIdentifiersReport
[35] Certificate Search, https://crt.sh
[36] Andrew Newton et al., HTTP Usage in the Registration Data Access Protocol (RDAP)", RFC 7480, March 2015.
[37] RDAP JSON Values, https://www.iana.org/assignments/rdap-json-values/rdap-json-values.xhtml
[38] Spam404, https://github.com/Spam404/lists/
[39] URLhaus Malicious URL Blocklist, https://gitlab.com/curben/urlhaus-filter/
[40] Google Safe Browsing, https://transparencyreport.google.com/safe-browsing/overview
[41] Paul Hoffman et al, DNS Queries over HTTPS (DoH), RFC 8484, October 2018
[42] Paul Hoffman, Representing DNS Messages in JSON, RFC 8427, July 2018
[43] Googe Public DNS, JSON API for DNS over HTTPS (DoH), https://developers.google.com/speed/public-dns/docs/doh/json
[44] Cloudflare, DNS over HTTPS Using JSON, https://developers.cloudflare.com/1.1.1.1/dns-over-https/json-format/
[45] Baojun Liu et al., Who Is Answering My Queries: Understanding and Characterizing Interception of the DNS Resolution Path, 2018, 27th USENIX SECURITY SYMPOSIUM
[46] Chaoyi Lu et al., An End-to-End, Large-Scale Measurement of DNS-over-Encryption: How Far Have We Come?, 2019, ACM Internet Measurement Conference 2019