# フィッシング検出方式の提案と JPCERT/CC フィッシングデータセットを用いた評価

YANG LIYI†

明治大学総合数理学部 先端メディアサイエンス学科 菊池研究室 †

## 1　はじめに

近年, フィッシング攻撃による被害が増加の一途をたどっている [1]. 特に, 大手 EC サイトや金融機関を模倣した偽造サイトが攻撃の主流となり, 標的に対してリンクを送り, 個人情報や認証情報を盗む手法が蔓延している. フィッシング攻撃の仕組みを図 1 に示す.

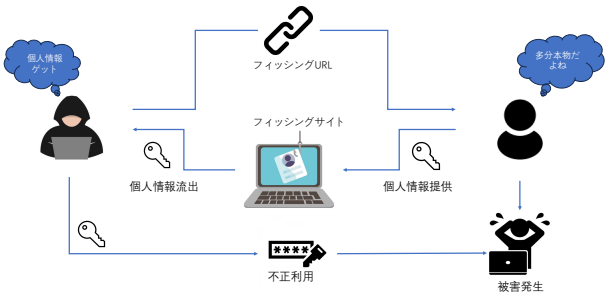


図 1　フィッシング攻撃の仕組み

そういった被害を防ぐために, 送付される URL に対して解析を行い, Netcraft[2] や VirusTotal[3] などのような多くのフィッシング URL を検出するサービスがある.

フィッシング攻撃に関する研究は非常に盛んであり, これまで多くのフィッシング URL の検出手法が提案されてきた. 中村ら [4] は Netcraft を利用した HTTP リクエスト解析によるフィッシングサイト検出システムを提案した. 彼らは既存サービスを応用し, HTTP リクエストの内容を解析し, 検出手法を設計した.

しかし, 今までの多くの研究やサービスは外国の企業が提供しており, 日本特有のフィッシングサイトを正確に検出できるかどうかはまだ不確かであった.

そこで, 本研究ではまず既存サービス Netcraft を調査し, 海外のフィッシング URL と日本のフィッシング URL の違いを明らかにする. 観察した固有の件数に基づき, URL のキーワードリスクと Splited length などの新しい特徴量をいくつか提案し, 日本のフィッシング URL に特化したフィッシング検出システムの開発を目的とする.

†Kikuchi Laboratory, Department of Frontier Media Science, School of Interdisciplinary Mathematical Science, Meiji University.

## 2　調査

本章では, 既存サービス Netcraft を使用し, 日本のフィッシング URL と海外のフィッシング URL について調査した結果を報告する. また, URL の特徴について分析する.

### 2.1　調査対象

#### 2.1.1　Netcraft

Netcraft はイギリスのロンドンに拠点を置くインターネットサービス会社であり, 主にウェブサイト調査やフィッシングサイト検出などのようなサービスを提供している.

#### 2.1.2　JPCERT/CC

JRCERT/CC[5](以下 JPCERT と略す) はインターネットによる侵入や不正アクセスなどのサイバーセキュリティインシデントに対応する機関である. 日本国内サイトからのインシデント報告を受け付けており, 発生状況の把握, 対応の支援, 手口の分析, 再発防止対策の検討や技術的な助言を行なっている. 本研究は JPCERT に掲載された 2023 年 1 月から 5 月までのフィッシング URL のデータセットを使用する.

#### 2.1.3　PhishTank

PhishTank[6](以下 Phishtank と略す) はインターネット上でのフィッシングサイトの情報を共有するウェブコミュニティ, およびサービスである. Phishtank はオープンなプロジェクトで, ユーザーからのフィッシング詐欺の詳細な報告を分析し, コミュニティの協力によって, フィッシングサイトのデータベースを構築している. 本研究は Phishtank に掲載された 2023 年 10 月 11 日のフィッシング URL のデータセットを使用する.

#### 2.1.4 Kaggle

Kaggle[7] はデータサイエンティストや機械学習エンジニア向けのプラットフォームであり, データセットの提供, 機械学習モデルの構築と評価, コンペティションの参加など, データ関連の様々な活動をサポートしている. 本研究は Kaggle が提供する機械学習用の Malicious URLs dataset を利用する. このデータセットは Manu Siddhartha によって提供され, 2021 年にアップデートされたものである.

### 2.2 調査手法

#### 2.2.1 ランダム文字列検出

本研究は文字遷移確率を算出することで, ランダム文字列を検出する. 遷移確率は学習データの文字遷移パターンに従って計算される. 与えられた文字列のパターンと学習データが近ければ近いほど遷移確率が大きくなる. 逆に遷移確率が小さいとランダム文字列の可能性が高いと意味する. 本研究は Python ライブラリ texttrans[8] を用いて, 文字遷移確率を算出した.

#### 2.2.2 類似度算出

レーベンシュタイン距離は二つの文字列がどの程度異なっているかを示す距離の一種で, 文字列の類似度計算に使われている. 本研究は Python ライブラリ rapidfuzz[9] を用いて, 類似度を算出した.

## 3 フィッシングサイトの特徴量の提案と検出システム

### 3.1 Netcraft についての調査方法

この調査では Phishtank と JPCERT に掲載されたフィッシング URL を 1,200 個ずつ Netcraft で分析し, 警告が出たかどうか, 及びリスクという項目の結果を集計した.

#### 3.1.1 警告確認

与えられた URL にマルウェアとフィッシングの危険性があると判断される場合は, 警告が提示される.

#### 3.1.2 リスク

与えられた URL に対して Netcraft が独自な評価基準でそのリスクを評価する. 評価値は 0 から 10 であり, リスクの高さを意味する.

### 3.2 URL ベースの特徴量についての調査方法

この調査では PhishTank と JPCERT と Kaggle から提供されたフィッシング URL を 19,000 個ずつ用いて, いくつかの特徴量を抽出した.

#### 3.2.1 SSL risk

指定した URL に対して, 先頭の部分が http か https かを識別する. https の場合は, 通信が暗号化されているので, SSL risk を 0, http の場合は SSL risk を 1 と定義する.

例を表 1 に示す.

表 1 SSL risk の例

| URL | 先頭部分 | SSL risk |
|---|---|---|
| http://lphvnlopuh.duckdns.org/ | http | 1 |
| https://lphvnlopuh.duckdns.org/ | https | 0 |

#### 3.2.2 Random risk

指定した URL に対して, 数字とアルファベット以外の任意の文字で URL を分割する. 4 以上の文字列長の部分文字列に対して texttrans でランダム評価を行う. 評価の値が低ければ低いほどランダム文字列である可能性が高い. 本研究では, 閾値を 0.025 と 0.015 を用いて, URL の Random risk を式 1 と算出する. $P_{URL}$ は分割後の部分文字列のランダム性評価の集合である. さらに, 0.025 より低い要素の数が 3 つ以上の場合は Random risk を 1 点増加する.

具体例を表 2 に示す.

$$Randomrisk = \begin{cases} 2 & \text{if } \min(P_{URL}) < 0.015, \\ 1 & \text{if } 0.015 \leq \min(P_{URL}) < 0.025, \\ 0 & \text{if } \min(P_{URL}) \geq 0.025. \end{cases} \quad (1)$$

表 2 Random risk の例

| URL | 評価最小値 | Random risk |
|---|---|---|
| https://wzjdayup.xyz | 0.0049 | 2 |
| https://wwwepns.tanhehe.com/jp.php | 0.0235 | 1 |
| https://www.twitter.com | 0.0260 | 0 |

#### 3.2.3 Keyword risk

JPCERT データセット 2023 のフィッシングサイトの標的となる正規サイトのドメインから抽出した銘柄を用いて, リスクを評価する. 取り出した銘柄に加

え”mypage” や ”password” などの個人情報に関するワードを用いて, キーワードリストを作成した. 与えられた URL に対して, 数字とアルファベット以外の文字でドメインを分割する. 分割後の各部分文字列に対して rapidfuzz でキーワードリストにある各キーワードとの類似度を評価する. 評価の値は 0 から 100 であり, 類似度の高さを意味する. 本研究では, 閾値を 70 と 85 の 2 つに設定し, URL の Keyword risk を式 2 と算出する. ここで, $Q_{URL}$ は分割後の各部分文字列のキーワードリストとの類似度評価の集合である.

例を表 4 に示す. 作成したキーワードリストの一部を表 3 に示す.

表 3　キーワードリストの一部

| Description | Keyword |
|---|---|
| エポスカード | eposcard |
| 楽天 | rakuten |
| ログイン | login |

$$Keywordrisk = \begin{cases} 2 & \text{if } \max(Q_{URL}) > 85, \\ 1 & \text{if } 70 \leq \max(Q_{URL}) \leq 85, \\ 0 & \text{if } \max(Q_{URL}) < 70. \end{cases} \quad (2)$$

表 4　Keyword risk の例

| URL | 評価最大値 | Keyword risk |
|---|---|---|
| https://www-cr-mufg-jp.kia8k.com/mufgcard/newsplus/ | 100.0 | 2 |
| https://www.tmall.com | 72.0 | 1 |
| https://qwepo.xyz/ | 67.5 | 0 |

### 3.2.4　TLD count

Tranco[10] が提供する信頼できる Top1,000,000 のドメインのデータセットから, 出現回数が 100 回以上の Top Level Domain(TLD) を TLD リストとした. 与えられた URL に対して, 数字とアルファベット以外の文字でドメインを分割する. 分割後の各部分文字列を TLD リストと対照し, 一致する数をカウントする. その結果を TLD count と定義する.

例えば, ”https://newplenty.com.cn/jp” の場合はカウントされる部分文字列は com, cn, jp であり, TLD count は 3 である.

### 3.2.5　Splited length

指定した URL に対して, 数字とアルファベット以外の文字で URL を分割する. 分割された部分文字列数を Splited length と定義する.

例えば, ”https://info-e-orico.nftsgiant.com/” のドメインは info, e, orico, nftsgiant, com に分割され, Splited length は 5 である.

### 3.2.6　Whitelist risk

Tranco の Top1,000,000 のドメインのデータセットをホワイトリストとして利用する. 与えられた URL に対して, ドメインがホワイトリストにあるかどうかを判断する. ホワイトリストに載っている場合は Whitelist risk を 0, 載っていない場合は 1 と定義する.

具体例を表 5 に示す.

表 5　Whitelist risk の例

| URL | ドメイン | Whitelist risk |
|---|---|---|
| http://lphvnlopuh.duckdns.org/ | lphvnlopuh.duckdns.org | 1 |
| https://www.twitter.com/ | twitter.com | 0 |

## 3.3　検出システム

本研究では既存サービスが日本のフィッシング URL に弱い問題を解決するため, 日本のフィッシング URL に特化した検出システムを開発した.

システム構成を図 2 に示す.

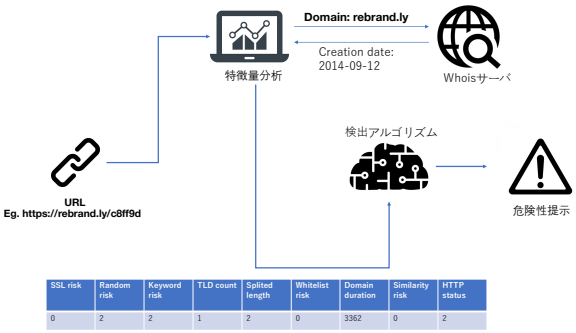


図 2　システム構成図

本システムは入力された URL に対して, 特徴量分析を行う. SSL risk, Random risk, Keyword risk, TLD count, Splited length, Whitelist risk のような SVM 用の特徴量以外, HTTP status と Domain duration と Similarity risk も検出用の特徴量として定義される.

## 3.4　検出アルゴリズム

本システムは点数加算方式でフィッシング URL を検出する. HTTP status と SVM 分類器と Domain duration と Similarity risk の 4 つの点数付け項目を設ける. 閾値

を 55 に設定し, 閾値を超えた場合はフィッシング URL をみなす.

#### 3.4.1 HTTP status

本システムでは, requests モジュールを使用し, 対象 URL の HTTP ヘッダーの情報を取得する. リクエスト成功した場合は, URL リダイレクトを確認する. 確認できた場合は危険点数を 10 点増加する. 確認できなかった場合は危険点数を増加しない. 一方, リクエスト失敗した場合は, SVM 分類器だけで URL を判定する.

#### 3.4.2 SVM 分類器

本システムでは 2.3 節で定めた 6 つの URL ベースの特徴量を利用し, Kaggle の安全な URL 20,000 個と JPCERT のフィッシング URL 19,714 個を学習し, 分類器を作る. 7 次元の機械学習用データセットを用いて, SVM 分類モデルに学習させる. SVC のパラメーターは, C=100, kernel='rbf' と設定する. 対象 URL を訓練した SVM モデルで判定し, 陽性が出た場合は危険点数を 40 点増加する. 陰性が出た場合は危険点数を増加しない.

#### 3.4.3 Domain duration

本システムでは Whois サービスを利用し, 対象サイトの作成日時を調べる. 対象サイトが作成されてから, 経過した日数を Domain duration と定義する. 先行研究 [11] によると, 正規サイトの最頻値が 3,650 日であるのに対し, フィッシングサイトは 7 日と非常に期間が短い. Domain duration が 30 日以下の場合は危険点数を 25 点増加する. Domain duration が 30 日以上で 365 日以下の場合は危険点数を 15 点増加する. Domain duration が 365 日以上の場合は危険点数を増加しない.

#### 3.4.4 Similarity risk

本システムでは, rapidfuzz モジュールを使用し, 対象 URL のドメイン部を 2.3.6 で説明したホワイトリストに載っている 1000000 個のドメインと比較し, 類似度を算出する. 類似度が 90 以上の項が存在すれば, 危険点数を 10 点増加する. 類似度が 90 以上の項が存在しなければ, 危険点数を増加しない.

## 4 評価実験

### 4.1 Netcraft についての調査結果

Netcraft の警告確認の結果を表 6 に示す.

表 6　警告確認結果

| | positive | % |
|---|---|---|
| JPCERT | 141 | 11.8 |
| Phishtank | 676 | 56.3 |

Netcraft のリスク値の分布を図 3 に示す.

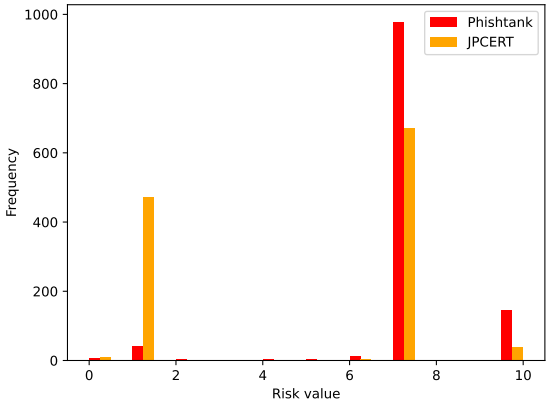


図 3　Netcraft のリスク値の分布

### 4.2 URL ベースの特徴量の調査結果

全 19,000URL の SSL リスクについての評価を表 7 に示す.

表 7　SSL リスク

| | positive | % |
|---|---|---|
| JPCERT | 2442 | 12.9 |
| Phishtank | 1551 | 8.2 |
| Kaggle | 0 | 0 |

ランダムリスクの分布を図 4 に示す.

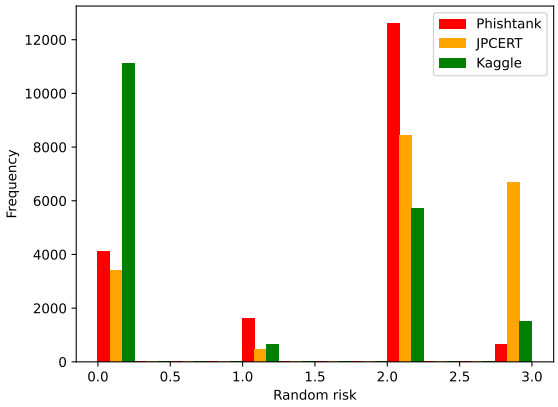


図 4　ランダムリスクの分布

キーワードリスクについての分布を図 5 に示す.

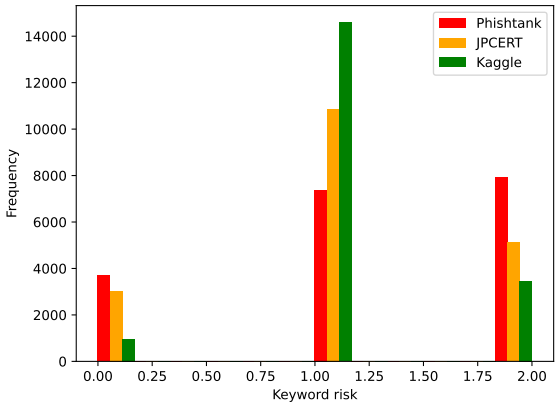


図 5　キーワードリスクの分布

TLD count の分布を図 6 に示す.

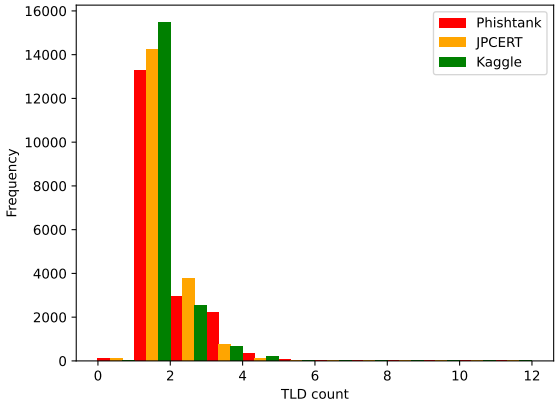


図 6　TLD count の分布

Splited length の分布を図 7 に示す.

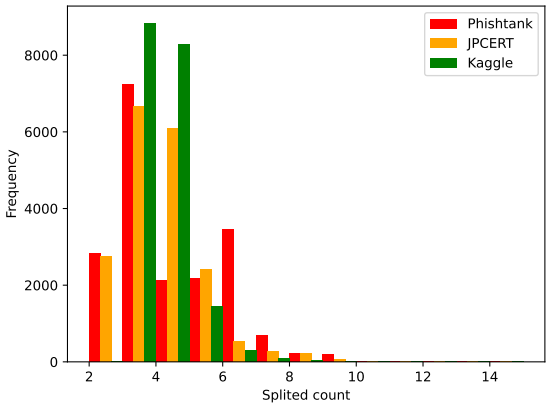


図 7　Splited length の分布

ホワイトリストリスクについての評価を表 8 に示す.

表 8　Whitelist リスク

| | positive | % |
|---|---|---|
| JPCERT | 18564 | 97.7 |
| Phishtank | 14790 | 77.8 |
| Kaggle | 18256 | 96.1 |

### 4.3　実験結果

5 分割交差検証を行った結果として, 平均精度は 0.84 であった.

本実験では JPCERT, Phishtank, Kaggle データセットから URL をランダムに 500 個ずつサンプリングして用いて, 5 のクロスバリデーションを行う.

実験結果を表 9 に示す. 実験結果についての精度評価を表 10 に示す.

表 9　検出精度

| | 平均値 | 最小値 | 最大値 | 標準偏差 |
|---|---|---|---|---|
| FN(JPCERT) | 16.8 | 13 | 22 | 2.9 |
| FP(JPCERT) | 21.8 | 16 | 25 | 3.1 |
| FN (Phishtank) | 65 | 41 | 92 | 17.3 |
| FP(Phishtank) | 21.8 | 16 | 25 | 3.1 |

表 10　精度評価

| | JPCERT | Phishtank |
|---|---|---|
| Precision | 0.79 | 0.62 |
| Recall | 0.83 | 0.35 |
| F score | 0.81 | 0.45 |
| Accuracy | 0.81 | 0.57 |

### 4.4 誤分類の例

誤分類の例を表 11 に示す.

表 11　誤分類の例

| No | URL | SSL risk | Random risk | Keyword risk | TLD count | Splited length | Whitelist risk | 誤分類の種類 |
|---|---|---|---|---|---|---|---|---|
| 1 | https://funfun-kids.com/contact/23/ | 0 | 0 | 2 | 1 | 3 | 1 | FN |
| 2 | https://tersyesg.yourtrap.com | 0 | 0 | 1 | 1 | 3 | 1 | FN |
| 3 | https://www.lostcanadianchildren.com/MP.html | 0 | 2 | 1 | 1 | 3 | 1 | FP |

表 11 の 2 番目と 3 番目の誤例から見ると, Random risk 以外の特徴量が全部等しく, 検出システムの判断は Random risk に大きく影響されていることが分かった. URL に識別できない文字列が含まれる場合は Random risk が高くなり, 誤分類される可能性も高くなる.

### 4.5 考察

調査結果の図 3 から, Phishtank の URL の殆どはリスク値が 7 以上で, JPCERT の URL の半数ぐらいはリスク値が 2 以下であることが分かった. 故に, Netcraft は日本のフィッシングサイトに弱いと考えられる. Netcraft 以外にも VirusTotal のような人気なサービスがあるため, 他の既存サービスに対して調査を行う必要がある.

また, 提案したシステムの実験結果の表 10 から, Phishtank のフィッシング URL には精度が低いが, JPCERT のフィッシング URL は 8 割以上の精度で検出された. 実用できるほど精度は高くないが, 日本のフィッシング URL に対して, 海外の既存サービス Netcraft より検出精度が高いため, 日本のフィッシング URL に特化したフィッシング検出システムの開発である本研究の目的に達成できたと言えるだろう.

精度を上げるには, より適切な特徴量を選ぶべきである. 例えば, 本研究では TLD count を特徴量として利用したが, 実は調査結果の図 6 によると, 安全な URL とフィッシング URL の分布が差が僅かしかないため, あまり適切ではなかったと考えられる. 文字列のランダム判定に用いた texttrans は分割なしの長い文字列に弱い. 例えば, lostcanadianchildren という文字列はランダム文字列に認識される. それも精度が上がらない原因の一つであると推測される.

## 5　おわりに

本研究では既存のサービス Netcraft を調査し, Netcraft が日本のフィッシング URL に弱いことを明らかにした. その問題を解決するために日本のフィッシング URL に特化した検出システムを開発した. 本研究を通じて, フィッシング対策は一つのサービスに頼らず, 地域に対応した複数のサービスを運用すべきだと提案したい.

今後, URL ベースだけでではなく, 多様な方式で適切な特徴量を定め, より質の高いシステムを構築していきたいと考える.

## 参考文献


[1] 読売新聞オンライン, ”「フィッシング」の不正送金が急増、2 月以降の被害 9 億 6000 万円” (https://www.yomiuri.co.jp/national/20230426-OYT1T50155/, 2023 年 10 月参照)

[2] Netcraft (https://www.netcraft.com/, 2023 年 10 月参照)

[3] VirusTotal(https://www.virustotal.com/, 2023 年 10 月参照)

[4] 中村元彦, 寺田真敏, 千葉雄司, 土井範久, ”プロキシを利用した HTTP リクエスト解析によるフィッシングサイト検出システムの提案”, 情報処理学会論文誌 Vol.48 No.10, pp3365-3374, 2007.

[5] JPCERT/CC (https://github.com/JPCERTCC/phishurl-list/, 2023 年 10 月参照)

[6] PhishTank (https://www.phishtank.com, 2023 年 10 月参照)

[7] Kaggle (https://www.kaggle.com/datasets/sid321axn/malicious-urls-dataset, 2023 年 10 月参照)

[8] texttrans (https://pypi.org/project/texttrans/, 2023 年 10 月参照)

[9] rapidfuzz (https://pypi.org/project/rapidfuzz/, 2023 年 10 月参照)

[10] Tranco (https://tranco-list.eu/, 2023 年 10 月参照) https://tranco-list.eu/

[11] 桜井啓多, “ドメイン情報と HTTP レスポンスヘッダに基づくフィッシングサイトの識別と評価”, 2018 年度菊池研究室卒業論文, 2018.