# OCRを利用したホモグラフIDNの検知法

澤部 祐太[1,a)]　千葉 大紀[2,b)]　秋山 満昭[2]　後藤 滋樹[1]

**概要**：正規サイトのドメイン名に類似したドメイン名を利用するホモグラフ攻撃が問題となっている．ホモグラフ攻撃は正規サイトのドメイン名に含まれる文字を視覚的に類似する文字列で置換したドメイン名（ホモグラフドメイン名）を生成して，ユーザを本来とは異なるサイトへ誘導する攻撃である．特に非ASCII 文字を用いることが可能な国際化ドメイン名（IDN）では類似ドメイン名を多数生成することが可能であり，攻撃者がホモグラフ攻撃で生成したIDN（ホモグラフIDN）がフィッシング攻撃で利用された事例が存在する．既存のホモグラフIDN の検知手法には固定的な変換表を用いてドメイン名を検知する手法がある．この方法では変換表に登録されていない文字を検知することができない，変換表を人手で更新する必要があるという課題がある．本稿ではホモグラフIDN が視覚的に正規サイトのドメイン名に類似しているという特徴に着目して，Optical Character Recognition (OCR) を利用してホモグラフIDN を検知する手法を提案する．評価に当たっては実際に利用されていた319 万件のIDN，および1 万件の悪性なIDN を用いて従来手法と提案手法を比較して，従来手法では検知できなかったホモグラフIDN が提案手法を用いることで検知可能であることを確認した．

**キーワード**：ドメイン名，国際化ドメイン名 (IDN)，ホモグラフ攻撃，OCR.

## 1. 序章

ドメイン名はWeb サイトやメールアドレスの一部として世界中で広く使われている．元々ドメイン名とはIP アドレスの代わりに人間が覚えやすい文字列に変換するために導入されたものであり，一般にサービス名が含まれることが多い．サイバー攻撃を行う攻撃者はこのようなドメイン名の性質を悪用して，正規のサービスで利用されているドメイン名に類似したドメイン名を用いて攻撃を行う．このような正規サービスを狙う悪性なドメイン名には大きく分けて2 種類存在する．一つは人間のタイプミスを狙ったタイポスクワッティング [2] と呼ばれる攻撃である．これは，正規サイトのドメイン名に対してキーボードの配列上で距離的に近い文字を置換・挿入することで類似するドメイン名を生成する攻撃である．もう一つは人間の視覚における判断ミスを狙ったホモグラフ攻撃 [3] と呼ばれる攻撃である．これは，正規サイトのドメイン名の一部を視覚的に類似する文字に置換することで類似するドメイン名を生成する攻撃である．本稿ではホモグラフ攻撃において生成されたドメイン名をホモグラフドメイン名と呼ぶ．国際化ドメイン名 (Internationalized Domain Name (IDN)) の導入以後，ドメイン名にUnicode に含まれる文字を用いることが可能となったため，タイポスクワッティングと比べてホモグラフ攻撃の方がはるかに多くの数の正規ドメイン名に類似するドメイン名を作り出すことができる．また，ホモグラフ攻撃で作成されたIDN（ホモグラフIDN）が実際にフィッシング攻撃で利用された事例が存在する [4].

ホモグラフIDN の検知手法には事前に作成しておいた「視覚的に近い」文字の変換表を利用する手法が存在する．この変換表には非ASCII 文字とそれに類似するASCII 文字の組が登録されており，変換表の情報を基にドメイン名中の非ASCII 文字をASCII 文字に変換することで正規サイトに類似したホモグラフIDN を検知することが可能となる．ただし，この手法では変換表に登録されていない文字は変換することができない．また変換表の更新を人手で行う必要があるという課題がある．本研究では事前に変換表を作成することなく，動的に変換表を作成してホモグラフIDN を検知する手法を提案する．我々の提案手法は任意のIDN と正規サイトのリストを入力として，ドメイン名を画像に変換してOptical Character Recognition (OCR) で読み取ることにより正規サイトと視覚的に類似するホモ

[1] 早稲田大学基幹理工学研究科
[2] NTT セキュアプラットフォーム研究所
[a)] sawabe.yuta@ruri.waseda.jp
[b)] daiki.chiba@ieee.org
本稿はAPAN Research Workshop 2018 において発表された論文 [1] の拡張版である

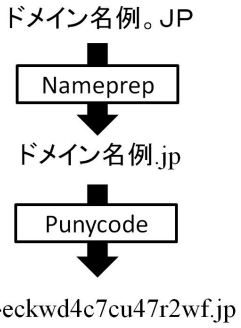


**図 1** IDNA による変換例

グラフ IDN を抽出する．本手法の主要なアイデアは，ホモグラフ IDN が正規サイトのドメイン名に類似しているという根本的な特性を逆手に取り，OCR の文字認識を利用することでホモグラフ攻撃の標的とされる正規ドメイン名を自動的に抽出することにある．本論文の貢献は次のとおりである．

- OCR を利用してホモグラフ IDN を検知する手法を初めて提案した．
- 319 万件の実在する IDN，および 1 万件の悪性な IDN を用いて提案手法の有効性を評価した．
- 提案手法は，従来手法では登録されていなかった非 ASCII 文字を含む変換表を作成することができた．

## 2. 背景

### 2.1 国際化ドメイン名（IDN）

インターネットにおいて当初は英語の使用を想定していたため，ドメイン名には ASCII 文字，数字，ダッシュ文字のみしか使用することはできなかった．しかし，インターネットが非英語圏でも利用されるようになったことで，Web や電子メールと同様にドメイン名も多言語で利用できるようにしたいという要望があがった．そこで，ドメイン名に非 ASCII 文字を利用できるようにするために国際化ドメイン名が考案された [5]．近年ではトップレベルドメイン名（TLD）に非 ASCII 文字を用いた国際化トップレベルドメイン名 (IDN TLD) も使用されている．実装にあたっては非 ASCII 文字に対応していない既存の仕組みと互換性を持たせるために，RFC3490 [6] で定義された IDN を ASCII 文字列に変換する IDNA (Internationalizing Domain Names in Applications) と呼ばれる仕組みが用意されている．**図 1** は IDNA を利用して日本語ドメイン名を ASCII 文字列に変換する例である．IDNA には Nameprep [7] と Punycode [8] の 2 つの処理が含まれる．変換後のドメイン名は通常のドメイン名と区別するために，頭に “xn–” をつけて IDN であることを明示する．

### 2.2 ホモグラフ攻撃

ホモグラフ攻撃とは正規サイトのドメイン名中の一部の

example.com (正規ドメイン名)
examp1e.com (数字 ‘1’ を利用したホモグラフIDN)
ëxample.com (キリル文字 ‘ë’ を利用したホモグラフIDN)
example.com (キリル文字 ‘а’ を利用したホモグラフIDN)

**図 2** ホモグラフドメイン名の例

文字を別の文字列で置換したドメイン名を生成する攻撃である．**図 2** に example.com に対するホモグラフドメイン名の例を示す．2 番目は本来のドメイン名中の文字 “l” を数字の “1” で置換したホモグラフドメイン名である．ユーザはこのドメイン名を見た際に視覚的特徴から正規のドメイン名であると判断する可能性が高いが，実際には異なるドメイン名であるため別のサイトへアクセスしてしまう恐れがある．3, 4 番目のドメイン名はドメイン名中の文字をキリル文字で置換したホモグラフ IDN である．非 ASCII 文字の中には字形が似ている文字が数多く存在するため，こうした文字を活用することで人の目では区別することが困難なドメイン名を大量に生成することが可能になる．このように，ホモグラフ攻撃では正規サイトに類似したドメイン名を用いてユーザの判断ミスを引き起こして，ユーザが意図するサイトとは異なるサイトへ誘導する．実際にフィッシング攻撃においてホモグラフ IDN が利用されていた事例が報告されている [4]．そこで，各種 Web ブラウザではドメイン名中に異なる言語の文字が混在している場合にはホモグラフ IDN であると判断して，Punycode の形式で表示する対策を取った [9]．しかし，Xudong や Wordfence らはそうしたブラウザの対策を回避するようなホモグラフ IDN を発見し，その危険性を訴えている [10], [11]．これは正規サイトのドメイン名に含まれるすべての文字を別の言語の文字で置換すれば，ブラウザではホモグラフドメイン名であると判断されなくなるためである．また正規ドメイン名の保有者の中には，類似するドメイン名を第三者に取得されないように防衛目的で取得するブランドプロテクションを行っている場合がある．このような，ドメイン名の取得・維持には費用が掛かるため一部の保有者しか対策を行っていない [12]．以上のように，ホモグラフ攻撃はその危険性が認識されながらも対策が十分とは言えない．

## 3. 提案手法

本節ではホモグラフ IDN (2.2 節) を検知する提案手法について説明する．3.1 節で提案手法の概要について説明し，3.2 節から 3.5 節にかけて各ステップの詳細を説明する．

### 3.1 提案手法の概要

本稿では OCR を利用して視覚的に類似するホモグラフ IDN を検知する手法を提案する．提案手法は IDN のリス

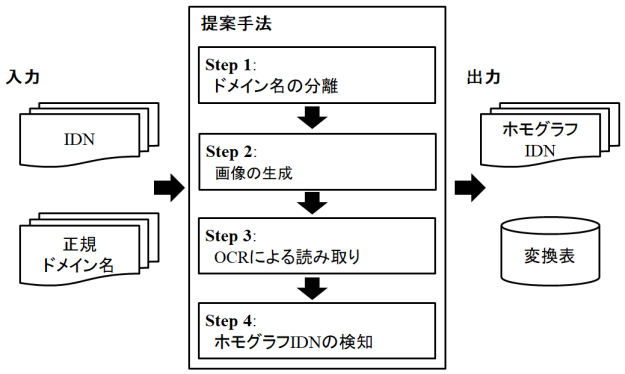


**図 3** 提案手法の概要

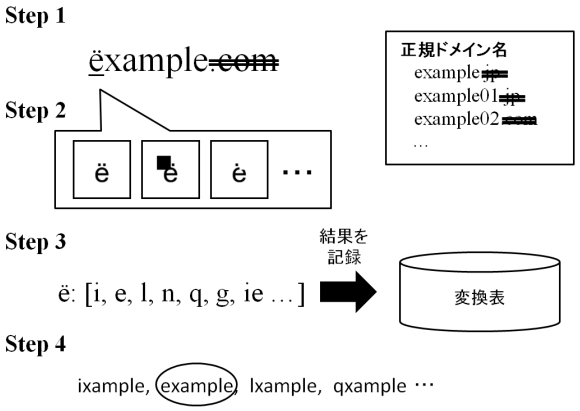


**図 4** 提案手法の適用例

トと正規サイトのドメイン名のリストを入力として，最終的に検知したホモグラフ IDN と対応する正規サイトのドメイン名の一覧，および OCR の変換結果を登録した変換表を出力する．提案手法の概要を**図 3** に，適用例を**図 4** に示す．提案手法はドメイン名の分離，画像の生成，OCR による読み取り，ホモグラフ IDN の検知の 4 つのステップで構成される．以下に各ステップを順番に説明する．

### 3.2 Step 1: ドメイン名の分離

入力した IDN，および正規サイトのドメイン名から Public Suffix を取り除く．Public Suffix とは，ドメイン名のうち個人ユーザがコントロールできない部分の文字列を指す [13]．例えば，Public Suffix は，`.com` や，`.net` のような generic top level domain （gTLD） や，`.co.jp` や `.co.uk` のような country code top level domain （ccTLD） を含む文字列で構成される．この Step を導入する理由は，攻撃者がホモグラフ IDN を生成する際に，必ずしも標的の正規サイトドメイン名と同じ Public Suffix を利用するとは限らないからである．例えば，攻撃者が `ëxample.com` を標的にしたホモグラフ IDN を生成する際に，`ëxample.com` だけでなく，`ëxample.co.jp` を用意することは十分にありうる．特に 2013 年 10 月以降は `.xyz` や `.top` のような New gTLD が導入され，Public Suffix として利用可能なドメイン名の数が急増し，それらが実際に攻撃に多用されている

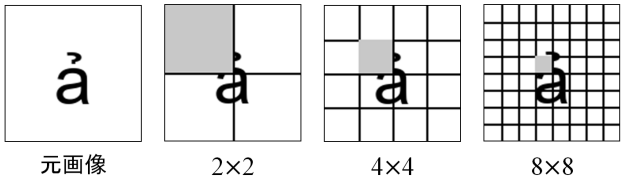


**図 5** マスク画像の例 （マスク色:白）

ことが知られている [14]．したがって，この Step を導入することは，検知可能なホモグラフ IDN の範囲を拡大させるのに役立つ．残されたドメイン名は “.” によって複数の文字列に分離しておく．

### 3.3 Step 2: 画像の生成

まず，Step 1 で作成した IDN の文字列を OCR で読みとるための画像に変換する．OCR には通常文字の切り出し機能が搭載されているが，“.” のような小さな文字や “i”, “l” のような線状の文字に対しては認識精度が落ちる．そこで，精度を落とさずに正規ドメイン名の候補を生成するために，文字列から 1 文字ずつ画像を生成する．

次に，多様なホモグラフ IDN を OCR で特定可能にするために事前の画像処理を行う．本提案手法で OCR を利用する目的は，ホモグラフ IDN で利用される Unicode 文字を攻撃者が似せようとした ASCII 文字として OCR に自動認識させることである．例えば，図 4 では文字 “ë” を OCR で読み取った際に “e” に変換されることを期待する．しかし，OCR の精度が高い場合には “ë” と “e” では字形が異なるため期待通りの読み取り結果にならない場合がある．そこで，画像の一部を塗りつぶして字形を変えた画像をあえて用意して，読み取り結果に多様性を持たせる．このような画像をマスク画像と呼ぶ．マスク画像で塗りつぶす際に用いる色には黒と白の 2 パターンを用意する．白の場合には文字の一部を消去することで，黒の場合にはノイズを加えることで，それぞれマスクをかけない画像と比べて読み取り結果に影響を与えることができる．また，マスクのサイズを 3 パターン用意する．具体的には，作成した画像を $2 \times 2$，$4 \times 4$，$8 \times 8$ マスに分割し，そのうち 1 箇所を黒または白で塗りつぶす．最終的に，1 文字に対し 2 種類の色（黒，白），84 種類のマスク箇所で 168 種類のマスク画像，およびマスクを適用しない画像の計 169 種類の文字画像が生成される．**図 5** は生成した白色のマスク画像の一例である．なお，分かりやすさのためにマスクの色をグレーで表記しており，分割したマスの罫線を書き加えている．

### 3.4 Step 3: OCR による読み取り

Step 2 で生成した画像を OCR で読み取り，ドメイン名中の文字を視覚的に類似する文字へ変換する．同じ文字の画像であってもマスクの色，大きさによって読み取り結果が異なる場合があるため，複数の読み取り結果をまとめて

変換表に記録する．この際，一つの文字に対して複数の読み取り結果を登録することで，本提案手法は多様なホモグラフ IDN の検知を実現している．また，この変換表の作成には事前の学習を必要とせず，入力した IDN を基に動的に生成することができる．

### 3.5 Step 4: ホモグラフ IDN の検知

Step 3 で作成した変換表の情報を基に，入力 IDN をホモグラフ IDN と考えた場合に想定される正規ドメイン名の候補を生成する．生成された正規サイトのドメイン名リストに含まれるドメイン名と一致するものが存在した場合には，入力 IDN がホモグラフ IDN であると検知し，対応する正規サイトのドメイン名と共に出力する．

Step 3 で OCR の読み取り結果の全部を登録すると，Step 4 において候補ドメイン名の数が膨大になる可能性がある．また登録数が多い場合には，Step 4 において視覚的に類似していない正規ドメイン名まで候補に含まれて誤検知となる場合がある．一方で登録数が少ない場合には，多様なホモグラフ IDN を検知することができない．Step 3 で適切な登録数となるように調整する必要がある．本評価における具体的な調整法を 4.4.2 節で説明する．

## 4. 評価

本節では 3 章で説明した提案手法の有効性について評価を行う．具体的には，従来手法と提案手法を用いて検知されるホモグラフ IDN の数について比較を行う．

### 4.1 従来手法

従来手法では固定的な変換表を利用する．変換表には非 ASCII 文字とそれに視覚的に類似する ASCII 文字列をあらかじめ登録しておく．変換表は二つの変換表を用いる．一つは dnstwist [15] で用いられている対応表から作成した変換表である．dnstwist はドメイン名の類似性を利用した攻撃で用いられる悪性ドメイン名の探索に用いられるツールである．dnstwist の内部ではすべての英字とそれに対して視覚的に類似する文字列の対応表が用意されており，ホモグラフドメイン名の探索には対応表を利用する．この対応表を逆に利用することで非 ASCII 文字と視覚的に類似する ASCII 文字を登録した変換表を作成することができる．もう一つは Unicode Consortium が提供している類似文字の対応表である．この対応表には Unicode の文字とそれに類似する文字列が登録されており，今回は 2018 年に作成された Version 11.0.0 の変換表 [16] の中から変換後の文字が英字のもののみを抽出して利用する．二つの変換表を利用することで，IDN 中の非 ASCII 文字を視覚的に類似する ASCII 文字に変換することができる．ここで，提案手法との性能比較を行うために，入力を IDN のリストと正規サイトのドメイン名のリストとし，変換表の情報を基に正規サイトのドメイン名の候補を生成し，検知されたホモグラフ IDN を出力する仕組みを用意した．

**表 1** データセットの概要

| データセット | データ取得日 | #IDNs |
|---|---|---|
| Project Sonar Dataset | 2018-08-03–2018-08-04 | 3,198,144 |
| Malicious Dataset | 2017-10-29–2017-11-10 | 16,341 |

### 4.2 データセット

評価には 2 つのデータセットを利用する．一つは Project Sonar で公開されている Forward DNS のデータセット [17] である．このデータセットは DNS リクエストに対して A, AAAA, ANY レコードを返したドメイン名のみ登録されており，データセット中のドメイン名はすべて実際に IP アドレスと紐づいていたドメイン名である．このデータセットの中から IDN を抽出して評価に用いる．もう一つは悪性ドメインのデータセットである．これは，公開ブラックリスト hpHosts [18], MalwareDomains [19], Phishtank [20] と商用ブラックリスト Spamhaus [21], URIBL [22] を組み合わせたものである．このデータセットの中から IDN を抽出して評価に用いる．データセットの概要を**表 1** に示す．

### 4.3 評価観点

本評価では 4.2 節で説明したデータセット中の IDN を入力として，従来手法と提案手法で検知されるホモグラフ IDN の数や性質を比較する．評価においては以下の条件を満たすホモグラフ IDN を検知対象とする．

(1) 視覚的に英字に類似する非 ASCII 文字を用いている
(2) アクセス数の多い正規サイトのドメイン名を標的にしている

条件 1 より，OCR の読み取り結果は英字のみに限定し，ドメイン中の ASCII 文字は OCR による変換を行わずにそのまま利用する．条件 2 より，入力する正規サイトのドメイン名には Alexa Topsites [23] の上位 1,000 件に含まれるドメイン名を用いる．

### 4.4 事前準備

#### 4.4.1 正規ドメイン名

評価で用いる正規サイトのドメイン名を準備する．正規サイトのドメイン名で用いる Alexa Topsites のデータ中には，Public Suffix を除いた部分が一致するドメイン名が複数含まれている．例えば，1 位は `google.com` であるが，Public Suffix のみ異なるドメイン名（例，`google.co.jp`）が上位 100 件の内 24 件，上位 1,000 件の内 75 件存在する．そこで Public Suffix の部分を除いたドメイン名が重複するドメイン名は最上位のドメイン名だけを残す．

また，Public Suffix 以外のドメイン名について，極端に

ドメイン長が短い場合はホモグラフドメイン名であるかどうか判断することが難しくなる．そこで，今回は適切な評価を行うために，ドメイン長が 4 文字以下の正規ドメイン名は評価の対象外とする．以上の手順で作成した 723 件のドメイン名を正規サイトのドメイン名として利用する．

#### 4.4.2 OCR の設定

OCR にはオープンソースの Tesseract OCR [24] を利用する．Tesseract OCR は精度向上のために事前に読み取る対象の言語を指定したり，ホワイトリストと呼ばれる読み取り結果で利用できる文字を指定したりすることが可能である．今回の評価ではホワイトリストに英字を指定して，読み取り結果が英字のみになるように設定した．

また，本評価では OCR の読み取り結果の多い順から上位 10 件のみを変換表に登録した．ただし誤検知を防ぐために，Tesseract OCR が出力する confidence とよばれる読み取り結果の精度を表すスコアを利用する [25]．検知した IDN に含まれる各文字の読み取り結果に対する confidence を記録し，confidence の平均値が閾値以下になる候補ドメイン名を除外した．ただし，非 ASCII 文字は今回の評価では変換を行わないため 100.0 として計算し，閾値は事前に複数の非 ASCII 文字に関して調査を行い 80.0 と設定した．

### 4.5 評価結果

**表 2** は Project Sonar のデータセットを用いて従来手法と提案手法でホモグラフ IDN の検知数を比較した結果である．正規サイトのドメイン名ごとに従来手法のみで検知されたホモグラフ IDN，従来手法と提案手法の両方で検知されたホモグラフ IDN，提案手法のみで検知されたホモグラフ IDN の個数を数え，検知されたホモグラフ IDN の総数の多い順に上位 20 件の正規サイトのドメイン名のみ表に記載した．また，提案手法のみで検知されたホモグラフ IDN の中には一部ホモグラフでない IDN が含まれていたため，正しく検知されたことが確認された IDN のみカウントした．これを見ると，提案手法を用いると従来手法では検知することのできないホモグラフ IDN が多数検知できていることが分かる．一方で，従来手法で検知できるものの提案手法では検知できていないドメイン名が一定量存在することも分かる．また，**表 3** は悪性のデータセットを用いて同様の評価を行った結果を表す．悪性のデータセットの場合にはホモグラフ IDN の検知数が少なかったため，検知されたすべてのホモグラフ IDN について記載している．この表より，ホモグラフ IDN が悪性の目的に利用されており，提案手法を用いることで従来手法では検知できない悪性なホモグラフ IDN が検知可能であることが分かる．検知されたホモグラフ IDN の詳細は 4.6 節で説明する．

**表 4** は検知したホモグラフ IDN の中に含まれる非 ASCII 文字の個数についてまとめたものである．これを見ると，正規のドメイン名から 1 文字置換して生成された IDN が

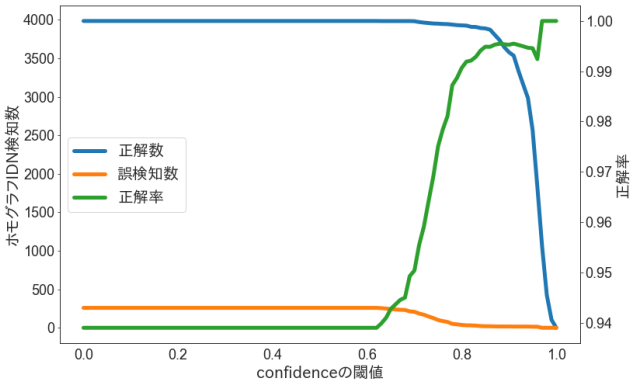


**図 6** confidence の閾値と検知した IDN 数の関係

多くの割合を占めていることが分かる．一方で，5 文字以上置換した IDN も一定数あり，特にドメイン名中のすべての文字が非 ASCII 文字で置換されたホモグラフ IDN は 53 件存在した．

表 2 には，検知されたホモグラフ IDN に対してブランドプロテクションに関する調査を行った結果も載せている．この調査には 2018 年 8 月 16 日に取得した WHOIS のデータを利用した．これを見ると，多くの正規ドメイン名は防衛目的でホモグラフ IDN を登録していないことが分かる．また，対策を取っている一部のドメイン名に関しても，今回の実験で検知したホモグラフ IDN の数に比較すると登録数が少ない．したがって，現状では正規ドメイン名の保有者がブランドプロテクションを目的にホモグラフ IDN を取得する対策が十分ではない．このような状況において，本提案手法はユーザ側でホモグラフ IDN を検知することが可能であり，ユーザが意図しないサイトへアクセスすることを防ぐために有効である．

次に，4.4.2 節で説明した confidence の閾値とホモグラフ IDN の検知数の関係を調べた．**図 4.5** は Project Sonar Dataset を用いて，提案手法において confidence の閾値と正しく検知された IDN，および誤検知した IDN の数の関係性を表している．これを見ると，confidence の閾値を高く設定すると類似度の低い文字列が除外されるため，誤検知数を減らし正解率を上げることが可能になることが分かる．一方で，confidence の閾値を上げることで正しく検知されるホモグラフ IDN の数も減少するため，実運用においては目的に応じて適切な閾値を設定することが必要である．

最後に，表 2, 表 3 に含まれる正規サイトのドメイン名に着目すると，二つ特徴がある．一つは Alexa Topsites で上位にあるドメイン名が複数含まれていることである．特に表 2 では Alexa Topsites の上位 100 件に含まれるドメイン名が 13 件存在する．これは，より多くのユーザを本来とは異なるサイトへ誘導するために，攻撃者がアクセス数が多いサイトに類似するホモグラフ IDN を多数作成していると考えられる．もう一つは，すべての正規サイトがログイン機能が存在するサイトであるということである．

**表 2** Project Sonar Dataset を用いた評価結果

| 正規ドメイン名 | Alexa 順位 | #IDNs | | | | ブランドプロテクション |
|---|---|---|---|---|---|---|
| | | 従来手法でのみ検知された IDN | 従来手法，提案手法両方で検知された IDN | 提案手法でのみ検知された IDN | 合計 | |
| google | 1 | 0 | 149 | 74 | 223 | 30 |
| facebook | 3 | 1 | 101 | 66 | 168 | 0 |
| apple | 58 | 1 | 67 | 86 | 155 | 0 |
| icloud | 368 | 1 | 28 | 69 | 98 | 0 |
| amazon | 10 | 2 | 69 | 27 | 98 | 32 |
| bittrex | 386 | 2 | 18 | 76 | 96 | 1 |
| blockchain | 884 | 2 | 40 | 32 | 74 | 0 |
| instagram | 17 | 7 | 25 | 28 | 60 | 0 |
| yahoo | 6 | 0 | 43 | 13 | 56 | 31 |
| twitter | 13 | 4 | 32 | 19 | 55 | 5 |
| paypal | 61 | 0 | 23 | 28 | 51 | 3 |
| youtube | 2 | 0 | 22 | 23 | 45 | 0 |
| hotels | 748 | 0 | 9 | 35 | 44 | 0 |
| whatsapp | 69 | 0 | 33 | 3 | 36 | 0 |
| coinbase | 357 | 2 | 8 | 23 | 33 | 0 |
| skype | 345 | 0 | 18 | 14 | 32 | 30 |
| microsoft | 47 | 1 | 17 | 10 | 28 | 0 |
| wikipedia | 5 | 1 | 24 | 3 | 28 | 0 |
| steamcommunity | 169 | 1 | 8 | 18 | 27 | 0 |
| linkedin | 30 | 2 | 11 | 9 | 22 | 0 |
| 合計 | | 118 | 1,108 | 955 | 2,181 | 166 |

**表 3** Malicious Dataset を用いた評価結果

| 正規ドメイン名 | Alexa 順位 | #IDNs | | | |
|---|---|---|---|---|---|
| | | 従来手法でのみ検知された IDN | 従来手法，提案手法両方で検知された IDN | 提案手法でのみ検知された IDN | 合計 |
| paypal | 61 | 0 | 4 | 2 | 6 |
| apple | 50 | 0 | 4 | 0 | 4 |
| facebook | 3 | 0 | 2 | 1 | 3 |
| icloud | 368 | 0 | 2 | 1 | 3 |
| google | 1 | 0 | 2 | 0 | 2 |
| steamcommunity | 169 | 0 | 0 | 2 | 2 |
| elpais | 405 | 0 | 1 | 0 | 1 |
| 合計 | | 0 | 15 | 6 | 21 |

ホモグラフ攻撃では攻撃者はユーザを正規のサイトとは異なるサイトへ誘導することが可能である．そこで，攻撃者がログイン機能が存在する正規のサイトに対してホモグラフドメイン名を作成し，誘導先にコンテンツを正規のサイトに偽装した悪性サイトを用意しておくことで，ユーザの ID やパスワードを含む個人情報を盗み取ることが可能となる．

### 4.6 ケーススタディ

本節では 4.5 節で確認した評価結果を基に，実際に検知されたホモグラフ IDN について説明する．

#### 4.6.1 従来手法でのみ検知されるホモグラフ IDN

**図 7** は，従来手法でのみ検知されたホモグラフ IDN 中に含まれていた非 ASCII 文字の一例である．この中には，OCR の認識精度が下がる傾向にある線状の文字に類似した文字が含まれており，提案手法ではこれらの文字を OCR で読み取った結果に期待する文字列が含まれていなかったため検知できなかったものと考えられる．今後はマスク画像で用いるマスクの種類を変更したり，読み取り結果から変換表に登録する文字の選定アルゴリズムを変更したりすることで，提案手法でより多くのホモグラフ IDN を検知できるようになると期待される．また，従来手法と提案手法は実用上は併用可能であるため，実際に運用する際には従来手法のみで検知されるホモグラフ IDN が存在しても問題はない．

#### 4.6.2 提案手法でのみ検知されるホモグラフ IDN

**図 8** は提案手法によって新しく検知されたホモグラフ IDN の一例である．1 番目のドメイン名では “e” に類似す

表 4　ホモグラフ IDN に含まれる非 ASCII 文字の個数

| 非 ASCII 文字の個数 | #IDNs |
|---|---|
| 1 | 1,528 |
| 2 | 484 |
| 3 | 109 |
| 4 | 3 |
| 5 | 24 |
| 6 | 20 |
| 7 | 5 |
| 8 | 6 |
| 9 | 0 |
| 10 | 2 |
| 合計 | 2,181 |

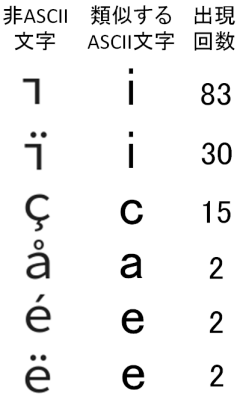


図 7　従来手法で検知された文字の例


図 8　提案手法で検知されたホモグラフドメイン名の例

る文字として “θ” が利用されている．従来手法では “o” に類似する文字として変換表に登録されていたが，提案手法では OCR を用いて，“o”, “e” 双方に類似する文字として登録したことによってホモグラフ IDN を検知することができた．2 番目のホモグラフ IDN で利用されている非 ASCII 文字はは O と E の合字であるが，文字の意味とは関係なく “ce” に見えるという視覚的特徴に着目して作成されている．3〜5 番目のドメイン名は類似度の低い文字を利用したホモグラフ IDN である．これは，有名サイトの場合は既に多数のホモグラフ IDN が取得されているため，攻撃者が新たにホモグラフ IDN を取得する際にはあまり類似しない文字を使わざるを得ない状況にあるためと考えられる．提案手法では複数の変換結果を登録することでこうしたドメイン名も検知可能である．6 番目のドメイン名は発音記号を利用してホモグラフ IDN を作成している．これも先ほどと同様の理由で，これまでに使われていない文字を使用する傾向があるからであると考えられる．

以上のように，攻撃者はホモグラフ IDN を作成する際に非常に多くの非 ASCII 文字を分析し活用しているため，変換表を人手で作成・更新を行い運用を行うことは極めて困難であることが分かる．今回の提案手法ではドメイン名を入力として自動的に変換表が作成されるため，有効な手法である．

## 5. 議論

本稿の評価では英字に類似する非 ASCII 文字を利用した IDN に限定して提案手法の評価を行った．今後，英字同士の類似性を利用したホモグラフドメイン名，数字を利用したホモグラフドメイン名を検知する手法についても検討していきたい．

また，評価においては正規サイトのドメイン名に Alexa Topsites の上位 1,000 件を使用した．ホモグラフ攻撃と同様にドメイン名の類似性を利用したタイポスクワッティングでは，順位の低いドメイン名であっても悪用されることが知られている [26]．そこで，入力する正規ドメイン名の数を増加させることで，より多くのホモグラフ IDN を検知できる可能性がある．

## 6. 関連研究

ホモグラフ攻撃は Gabrilovich が 2002 年に書いた論文で初めて言及された [3]．論文内では 2000 年 4 月に発生した PairGain 社のドメイン名に対するホモグラフ攻撃の事例を取り上げて，偽のサイトに掲載された虚偽の情報を投資家が信じた結果深刻な損失を被ったことが記されている．また，ホモグラフ攻撃の対策としてブラウザが国際文字を強調して表示する他，ブラウザが類似する文字のマッピングを行い，有名サイトのドメイン名と類似するドメイン名を検知する手法についても言及している．

Holgers らは 2006 年にホモグラフドメイン名に関する大規模な調査を行っている [27]．調査では大学内のトラフィックを観測し，ユーザがアクセスした正規のドメイン名に対するホモグラフドメイン名を生成し，IP アドレスに紐づいているかどうか確認した．ホモグラフドメイン名の生成には固定的な変換表を用い，英数字に類似する非 ASCII 文字で置換した．これに対して，我々はホモグラフ IDN から正規サイトのドメイン名を特定している点，動的に変換表を作成している点において彼らの研究とは異なる．

Dhamija らの研究 [28] ではフィッシングサイトを模した複数のサイトを用意し，ユーザが正規のサイトと区別できるかどうか調査を行った．調査では正規サイトのドメイン名の “w” を “vv” に置換したホモグラフドメイン名を利用したサイトが最も正答率が低く，91%の被験者が悪性サイトであることを見抜けなかった．

Baojun らの研究 [12] では，複数の TLD のゾーンファイルの情報を基に IDN の利用状況の実態に関して調査を行った．この中で，ホモグラフ攻撃において利用される IDN を自動で検知する仕組みを考案しており，Structural Similarity（SSIM）の値を基に正規ドメイン名と IDN の類似度を測りホモグラフ IDN を検知した．これに対して，我々は OCR を利用して類似度を測った点，読み取り結果を複数用意することでより幅広くホモグラフ IDN を検知

できた点においてこの研究とは異なる．

タイポスクワッティングに関しては Agten が 2013 年に 7 カ月間にわたって調査を行っている [2]．このなかで，防御目的でタイポスクワッティングドメイン名を取得する商標権者が少ないことを指摘している．また，すでにドメイン長が短い有名サイトに対しては，想定されるタイポスクワッティングドメイン名の 75%以上が取得されており，今後はより長いドメイン名が標的となると考えられる．

Szurdi らの研究 [26] では Alexa Topsites の下位のドメイン名に対してもタイポスクワッティングが行われており，.com ドメイン名の約 20%がタイポスクワッティングに起因するものであることを発見した．また，不正なドメイン名の取得に対する介入策や，ユーザ側でタイポスクワッティングを検知する仕組みについても議論している．

本稿で提案した OCR を利用したホモグラフ IDN の検知手法は上記のいずれの研究とも異なり，OCR を利用して視覚的に類似する文字の変換表を動的に生成し，マスク画像を利用して複数の読み取り結果を生成することでより多くのホモグラフ IDN の検知に成功している．

## 7. 結論

本稿では OCR を用いてホモグラフ IDN を検知する新たな手法を提案した．従来手法では固定的な変換表を利用することでホモグラフ IDN を検知していたが，変換表を手動で更新しなくてはならないという課題があった．そこで，提案手法ではホモグラフ IDN が視覚的に正規のサイトに類似しているという特徴に着目して，OCR を利用して動的に変換表を作成しホモグラフ IDN を検知する手法を提案した．評価では 319 万件の実際に利用されていた IDN，および 1 万件の悪性 IDN を利用して，従来手法では検知できないホモグラフ IDN を提案手法では検知できることを確認した．本提案手法を活用することで，ユーザがアクセスするドメイン名がホモグラフ IDN であるかどうか判断することが可能となる．

**謝辞** 本研究の一部は JSPS 科研費 16H02832 の助成を受けたものです．

### 参考文献

[1] Sawabe, Y., Chiba, D., Akiyama, M. and Goto, S.: Detecting Homograph IDNs Using OCR, *Proc. Asia-Pacific Advanced Network (APAN)* (2018).
[2] Agten, P., Joosen, W., Piessens, F. and Nikiforakis, N.: Seven months' worth of mistakes: A longitudinal study of typosquatting abuse, *Proc. Network and Distributed System Security Symp. (NDSS)* (2015).
[3] Gabrilovich, E. and Gontmakher, A.: The homograph attack, *Communications of the ACM*, Vol. 45, No. 2, p. 128 (2002).
[4] Symatec: Bad Guys Using Internationalized Domain Names (IDNs), https://www.symantec.com/connect/blogs/bad-guys-using-internationalized-domain-names-idns.
[5] 宇井隆晴：日本語ドメイン名 インターネット標準策定の軌跡，インプレス R&D (2006).
[6] Faltstrom, P., Hoffman, P. and Costello, A.: Internationalizing Domain Names in Applications (IDNA), RFC 3490, RFC Editor (2003).
[7] Hoffman, P. and Blanchet, M.: Nameprep: A Stringprep Profile for Internationalized Domain Names (IDN), RFC 3491 (2003).
[8] Costello, A.: Punycode: A Bootstring encoding of Unicode for Internationalized Domain Names in Applications (IDNA), RFC 3492 (2003).
[9] McElroy, T., Hannay, P. and Baatard, G.: The 2017 homograph browser attack mitigation survey (2017).
[10] Zheng, X.: Phishing with Unicode Domains, `https://www.xudongz.com/blog/2017/idn-phishing/` (2017).
[11] Wordfence: Chrome and Firefox Phishing Attack Uses Domains Identical to Known Safe Sites, https://www.wordfence.com/blog/2017/04/chrome-firefox-unicode-phishing/ (2017).
[12] Liu, B., Lu, C., Li, Z. et al.: A Reexamination of Internationalized Domain Names: The Good, the Bad and the Ugly, *Proc. IEEE/IFIP Int. Conf. on Dependable Systems and Networks (DSN)*, pp. 654–665 (2018).
[13] Mozilla Foundation: Public Suffix List, `https://publicsuffix.org/list/`.
[14] Chiba, D., Yagi, T., Akiyama, M. et al.: DomainProfiler: Discovering Domain Names Abused in Future, *Proc. 46th Annu. IEEE/IFIP Int. Conf. on Dependable Systems and Networks (DSN)*, pp. 491–502 (2016).
[15] Ulikowski, M.: dnstwist, `https://github.com/elceef/dnstwist/`.
[16] The Unicode Consortium: Unicode Security Mechanisms for UTR #39, `https://www.unicode.org/Public/security/11.0.0/confusables.txt` (2018).
[17] Rapid7: Project Sonar Forward DNS, `https://opendata.rapid7.com/sonar.fdns_v2/` (2017).
[18] hpHosts: Ad and Tracking servers only, `https://hosts-file.net/ad_servers.txt`.
[19] RiskAnalytics: DNS-BH Malware Domain Blockilist, `http://www.malwaredomains/com/`.
[20] PhishTank: PhishTank, `https://www.phishtank.com`.
[21] The Spamhaus Project Ltd.: The Domain Block List., `https://www.spamhaus.org/dbl`.
[22] URIBL.COM: URIBL, `https://uribl.com/about.html`.
[23] Alexa Internet, I.: Alexa Topsites, `https://www.alexa.com/topsites`.
[24] Google: Tesseract OCR, `https://opensource.google.com/projects/tesseract/`.
[25] Smith, R.: An overview of the Tesseract OCR engine, *Proc. Int. Conf. Document Analysis and Recognition (ICDAR)*, Vol. 2, pp. 629–633 (2007).
[26] Szurdi, J., Kocso, B., Cseh, G. et al.: The Long "Taile" of Typosquatting Domain Names., *Proc. USENIX Security Symp.*, pp. 191–206 (2014).
[27] Holgers, T., Watson, D. E. and Gribble, S. D.: Cutting through the Confusion: A Measurement Study of Homograph Attacks., *Proc. USENIX Annual Technical Conf.*, pp. 261–266 (2006).
[28] Dhamija, R., Tygar, J. D. and Hearst, M.: Why phishing works, *Proc.SIGCHI Conf. Human Factors in Computing Systems*, pp. 581–590 (2006).