DomainScouter: Understanding the Risks of Deceptive IDNs

Daiki Chiba[1], Ayako Akiyama Hasegawa[1], Takashi Koide[1],
Yuta Sawabe[2], Shigeki Goto[2], and Mitsuaki Akiyama[1]

[1]NTT Secure Platform Laboratories, Tokyo, Japan
[2]Waseda University, Tokyo, Japan

Abstract

Cyber attackers create domain names that are visually similar to those of legitimate/popular brands by abusing valid internationalized domain names (IDNs). In this work, we systematize such domain names, which we call deceptive IDNs, and understand the risks associated with them. In particular, we propose a new system called DomainScouter to detect various deceptive IDNs and calculate a deceptive IDN score, a new metric indicating the number of users that are likely to be misled by a deceptive IDN. We perform a comprehensive measurement study on the identified deceptive IDNs using over 4.4 million registered IDNs under 570 top level domains (TLDs). The measurement results demonstrate that there are many previously unexplored deceptive IDNs targeting non-English brands or combining other domain squatting methods. Furthermore, we conduct online surveys to examine and highlight vulnerabilities in user perceptions when encountering such IDNs. Finally, we discuss the practical countermeasures that stakeholders can take against deceptive IDNs.

1 Introduction

Domain names are indispensable resources or assets of online service providers on the Internet. Although the Internet was not designed to distinguish borders and languages, domain names were originally written in English only (i.e., using ASCII codes, digits, and hyphens). After some time, internationalized domain names (IDNs) were proposed to enable Internet users to create domain names in their local languages and scripts [25]. Since IDNs were successfully standardized and implemented in 2003, characters in the Unicode Standard can now be used in domain names while maintaining backward compatibility with previously implemented English-based domain names and the domain name system (DNS). The backward compatibility was implemented using the Punycode representation of the Unicode characters with a special prefix (`xn--`). For example, 例え`[.]test` in the IDN format is transformed into `xn-r8jz45g[.]test` in the ASCII-compatible format. IDNs are essential for enabling the multilingual Internet to serve culturally and linguistically diverse populations.

At the same time, cyber attackers abuse the IDN mechanism to register their domain names for cyber attacks. In fact, cyber attackers create domain names that are visually similar to those of legitimate and popular brands by abusing IDNs [36, 48, 58]. The attackers aim to trick innocent users into falsely recognizing a purposely created misleading domain name as a legitimate brand's domain name by its visual appearance. This type of attack, called an IDN homograph attack, poses a real threat to Internet users. For example, a security researcher used an IDN similar to `apple[.]com` with a valid SSL certificate to demonstrate a proof-of-concept of an almost complete phishing attack; many users could not distinguish the fake IDN from the genuine one by its appearance in April 2017 [68]. Similarly, another security researcher discovered an IDN homograph attack that used an IDN visually similar to `adobe[.]com` to distribute a fake flash player with malware [40]. Recently, a researcher reported a new vulnerability in Apple's Safari browser that renders a specific Unicode letter as a normal Latin small "d" in the browser's address bar, which can lead to IDN homograph attacks [56].

In this paper, first, we systematize such visually distorted IDNs, which we call ***deceptive IDNs***, to understand the risks associated with them. Unlike the previously reported similar studies [36, 48], the deceptive IDNs in this paper include not only homograph IDNs, wherein some of the characters in English brand domain names are replaced with visually similar characters, but also other types of lookalike IDNs targeting both English and non-English brands comprehensively. On the basis of the systematization, we propose a new system called DOMAINSCOUTER for detecting deceptive IDNs and calculating a ***deceptive IDN score*** for each IDN. This score is a new metric indicating the number of users that are likely to be misled by a deceptive IDN. The purpose of DOMAINSCOUTER is to score the suspiciousness of an attempt to deceive users on the basis of IDN characteristics. In particular, it is designed to capture distinctive visual characteristics of deceptive IDNs, consider characteristics of targeted legitimate

brand domain names, and use the domain knowledge of both IDNs and targeted domain names.

The contributions of this paper are summarized as follows.

- Propose a new system called DomainScouter to detect more various types of deceptive IDNs than previously proposed systems and calculate a deceptive IDN score, a new metric indicating the number of users likely to be misled by a deceptive IDN (Sections 3 and 4).
- Perform by far the most comprehensive measurement study on the deceptive IDNs detected by the proposed DomainScouter using over 4.4 million registered real-world IDNs under 570 top level domains (TLDs) (Section 5).
- Conduct online surveys (N=838) to examine vulnerabilities in user perceptions when encountering deceptive IDNs and evaluate that the deceptive IDN score we proposed reflects the tendency of users to be deceived by the attacks. To the best of our knowledge, this is the first user study on deceptive IDNs (Section 6).
- Discuss the practical countermeasures that stakeholders can take against deceptive IDNs (Section 7).

2 Systematization of Deceptive IDNs

We systematize all possible deceptive IDNs targeting users' visual perception. We focus on IDNs that look similar to those of legitimate brands to deceive users to take actions such as clicking links in spam emails and inputting personal information on phishing sites. To the best of our knowledge, this study is the first attempt in security research to systematize deceptive IDNs.

First, we divide deceptive IDNs into those targeting English brands and those targeting non-English brands since these two categories have quite different characteristics. Since English is the world's standard language and the Internet was originally available only in ASCII and English character sets, most globally popular brands have their websites and domain names in English. At the same time, many local brands in non-English-speaking communities have started to use their native languages and characters to create domain names. Thus, English and non-English brand names should be treated differently, especially when researching the Internet-related topics such as domain names. Whereas previous studies have focused only on deceptive IDNs targeting English brands [36, 48], IDNs targeting non-English brands have not been studied well so far.

Second, we reveal that there are three types of deceptive IDNs in theory: combosquatting (combining brand name with keywords) (***combo***), homograph (***homo***), and homograph+combosquatting (***homocombo***) IDNs. We define a combo IDN as an IDN that combines a brand domain name with some additional English or non-English phrases. Kintis et al. [30] conducted the first study to reveal English-based combosquatting domains; our paper extends this concept to IDNs. The homo IDN is an IDN wherein some of the characters of a brand domain name are replaced with characters that are visually similar. Some previous studies analyzed the characteristics of homo IDNs in 2018 [36, 48]. The homocombo IDN is defined as an IDN that does not match the above combo or homo definitions exactly but has characteristics of both the combo and homo IDNs; e.g., an IDN containing words similar to a legitimate brand name and some additional phrases. Our paper is the first to define, measure, and analyze the homocombo IDNs. Note that we do not include any non-IDN squatting domains such as typosquatting (typographical errors) [1, 29, 55, 62] or bitsquatting (accidental bit flips) [41] since our paper focuses on user misbehavior caused by deceptive IDNs.

On the basis of the above conditions, we consider six types of IDN-based attacks in this paper. In particular, when considering English brands (e.g., `example[.]test`) as targets, the brand could be targeted by combo IDNs (***eng-combo***; e.g., `example`ログイン`[.]test`), homo IDNs (***eng-homo***; e.g., `êxämplē[.]test`), and homocombo IDNs (***eng-homocombo***; e.g., `êxämplē`ログイン`[.]test`). When considering non-English brands (e.g., 例え`[.]test`), the brand could be targeted by combo IDNs (***noneng-combo***; e.g., 例えログイン`[.]test`), homo IDNs (***noneng-homo***; e.g., 亻列え`[.]test`), and homocombo IDNs (***noneng-homocombo***; e.g., 亻列えログイン`[.]test`).

In terms of creating/registering deceptive IDNs (especially combo and homocombo), attackers are free to use one or more arbitrary words as prefixes or postfixes of brands. That is, similar to non-IDN combosquatting [30], a deceptive IDN lacks a generative model. Therefore, we cannot rely on the generative model but need to design a system to grasp the nature of deceptive IDNs.

3 DomainScouter System

We propose a new system called DomainScouter to detect the six types of deceptive IDNs (eng-combo, eng-homo, eng-homocombo, noneng-combo, noneng-homo, and noneng-homocombo) defined in Section 2. Figure 1 shows an overview of DomainScouter. The inputs to DomainScouter are registered IDNs and selected brand domains. DomainScouter automatically detects deceptive IDNs on the basis of various features focusing on visual similarities, brand information, and TLD characteristics. The outputs of DomainScouter are detected deceptive IDNs, targeted brands, and deceptive IDN scores for each IDN. The deceptive IDN score is a new metric indicating the number of users likely to be deceived when encountering a deceptive IDN. DomainScouter consists of five steps: IDN extraction, brand selection, image generation, feature extraction, and score calculation. The following sections explain these steps in turn.

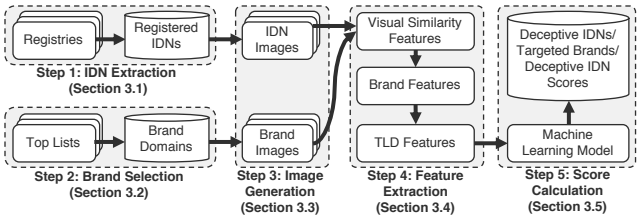

Figure 1: System Overview

3.1 Step 1: IDN Extraction

The first step involves extracting already existing IDNs from the domain registry databases. Unfortunately, since each domain registry corresponding to a TLD has been operated separately, there is no single (unified) database with all registered domains freely available for researchers. Thus, we need to collect registered domain names from more than 1,400 TLD registries to study all IDNs that exist in the world.

In general, TLDs can be divided into two categories: generic TLDs (gTLDs) and country-code TLDs (ccTLDs) [21]. In this paper, we further separate gTLDs and ccTLDs to understand the relationship between deceptive IDNs and TLDs' characteristics. We separate gTLDs into three types: ***legacy gTLD***, ***new gTLD***, and ***new IDN gTLD***. The legacy gTLD consists of 22 TLDs (`.aero`, `.asia`, `.biz`, `.cat`, `.com`, `.coop`, `.edu`, `.gov`, `.info`, `.int`, `.jobs`, `.mil`, `.mobi`, `.museum`, `.name`, `.net`, `.org`, `.post`, `.pro`, `.tel`, `.travel`, and `.xxx`) introduced before the new gTLD program started by ICANN in 2013 [24, 31]. The new gTLD is composed of 1,042 non-IDN TLDs (e.g., `.top`, `.xyz`, and `.loan`) introduced by the ICANN's program. The new IDN gTLD is made up of 84 IDN TLDs (e.g., .网址 (`.xn-ses554g`), .在线 (`.xn-3ds443g`), and .рус (`.xn-p1acf`)) also used by the program, especially for allowing the entire domain names to be represented in a local language and characters. Furthermore, we separate ccTLDs into two types: ***legacy ccTLD*** and ***new IDN ccTLD***. The legacy ccTLD is composed of 245 TLDs (e.g., `.cn`, `.jp`, and `.uk`) that were two-letter codes representing countries listed by the ISO 3166-1 standard [23]. The new IDN ccTLD consists of 42 IDN TLDs (e.g., .新加坡 (`.xn-yfro4i67o`), .한국 (`.xn-3e0b707e`), and .ڀك (`.xn-p1ai`)) registered after 2009 [22].

To collect and extract all registered IDNs under the above-mentioned TLD types, we leveraged the commercial WHOIS database [64] containing information about nearly all domains as of May 2018. Table 1 shows the breakdown of our collected dataset. In total, we processed over 294 million domains (including IDNs and non-IDNs) under 1,435 TLDs. From all domains, we extracted over 4.4 million IDNs under 570 TLDs. Note that the remaining 865 TLDs have no registered IDNs.

3.2 Step 2: Brand Selection

The second step of DomainScouter is selecting brand domains targeted by deceptive IDNs. We need to select both English and non-English brands since our paper focuses on deceptive IDNs targeting both types of brands as stated in Section 2.

Table 1: Domain Dataset

TLD type	# TLDs (IDNs)	# TLDs (Total)	# Domains (IDNs)	# Domains (Total)
Legacy gTLD	13	22	1,482,709	171,016,371
New gTLD	328	1,042	424,024	21,523,232
New IDN gTLD	84	84	599,559	599,559
Legacy ccTLD	103	245	988,963	100,398,597
New IDN ccTLD	42	42	931,062	931,062
Total	570	1,435	4,426,317	294,468,821

For English brands, we leveraged three major top domain lists (Alexa [2], Umbrella [11], and Majestic [38] top 1 million lists) that record representative Internet domains. As discussed in recent studies [35, 50], each list has its own ranking mechanism; thus, we used the three major lists in the Internet measurement community to collect English brands in an unbiased way. We extracted the top 1,000 domains from each list, removed redundant domains, and finally collected 2,310 domains in total.

For non-English brands, we used the same three top domain lists as for English brands. Since there are far fewer non-English brand domains than English ones, we extracted non-English IDNs from the top 1 million domains in each list, removed redundant domains, and finally collected 4,774 domains in total. Note that we excluded some low-ranked malicious domains accidentally listed in the top lists by referring to multiple domain blacklists such as VirusTotal [61], hpHosts [20], Google Safe Browsing [16], and Symantec DeepSight [54].

3.3 Step 3: Image Generation

The third step of DomainScouter is generating images from both registered IDNs (step 1) and brand domains (step 2) for the following calculation of visual similarities in step 4. In particular, we generate three types of images for each domain in both registered IDNs and brand domains. We select the default font used in the address bar of Google Chrome in Windows 10 since the browser/OS has the biggest market share [52].

RAW images. The first type is a *raw* image, simply generated from each domain's string without any modifications. RAW is used for specifying a very similar combination of a deceptive IDN (e.g., eng-homo and noneng-homo) and a brand domain as a whole.

PSR images. The second type is a public suffix-removed (*PSR*) image generated from substrings excluding a public suffix [39] from a domain name string. A public suffix consists of strings in domain names that cannot be controlled by individual Internet users [9]. For example, in the case of PSR images, `example` is extracted from both `example[.]com` and `example[.]co[.]jp` since `.com` and `.co.jp` are in public suffixes. PSR images can help distinguish deceptive IDNs that have different public suffixes from targeted brand domains

Table 2: List of Features

Type	No.	Feature	Importance
Visual Similarity	1	Max of SSIM indexes between RAW images	0.123
	2	Max of SSIM indexes between PSR images	0.158
	3	Max of SSIM indexes between WS images	0.391
Brand (RAW)	4	Alexa rank of identified RAW brand domain	0.019
	5	Umbrella rank of identified RAW brand domain	0.017
	6	Majestic rank of identified RAW brand domain	0.012
Brand (PSR)	7	Alexa rank of identified PSR brand domain	0.012
	8	Umbrella rank of identified PSR brand domain	0.004
	9	Majestic rank of identified PSR brand domain	0.009
Brand (WS)	10	Alexa rank of identified WS brand domain	0.041
	11	Umbrella rank of identified WS brand domain	0.046
	12	Majestic rank of identified WS brand domain	0.040
TLD	13	TLD type of Input IDN	0.085
	14	TLD type of RAW brand domain	0.024
	15	TLD type of PSR brand domain	0.006
	16	TLD type of WS brand domain	0.015

since attackers do not necessarily use the same public suffixes of the brand domains [48].

WS images. The third type is a word segmented (*WS*) image. A WS image is generated by applying word segmentation algorithms to a domain name string. For example, `example` and テスト are segmented from `example`テスト`[.]com`. We use the polyglot [44] implementation for multilingual word segmentation. The intuition behind generating WS images is to help detect combosquatting-based deceptive IDNs such as eng-combo, eng-homocombo, noneng-combo, and noneng-homocombo.

3.4 Step 4: Feature Extraction

The fourth step of DomainScouter is extracting features from registered IDNs (step 1), brand domains (step 2), and their corresponding images (step 3). This step is intended to design features that can detect the six types of deceptive IDNs defined in Section 2. In particular, we use three types of features: visual similarity, brand, and TLD features.

Visual similarity features. The visual similarity features, Nos. 1–3 listed in Table 2, are designed to grasp the most distinguishing characteristics of a deceptive IDN, the IDN's appearance. In other words, these three features are used to measure the extent to which an IDN can deceive users. We utilize image similarity between registered IDNs and brand domains as the visual similarity features. To measure similarity between two images, we use the Structural SIMilarity (SSIM) index [63] since it is reported to achieve the best performance when detecting one type of the deceptive IDNs (eng-homo) [36]. For our prototype implementation, we used pyssim [46], a python module for computing the SSIM index. The SSIM index ranges between 0.0 (non-identical) and 1.0 (perfectly identical). As explained in Section 3.3, we prepare images of three different types (RAW, PSR, and WS) to detect various deceptive IDNs; accordingly, we calculate the SSIM index for pairs of images of the same type. We use the maximums of the SSIM indexes between RAW, PSR, and WS images as features No. 1, 2, and 3, respectively. We identify the brand domain with the highest SSIM indexes as the targeted brand domain corresponding to the input IDN.

Brand features. The brand features, Nos. 4–12 listed in Table 2, are designed to consider characteristics of targeted brand domains. We hypothesize that more popular domains are targeted to create deceptive IDNs. Thus, we use the rank information in the three top lists (Alexa [2], Umbrella [11], and Majestic [38]) as our brand features. The reason for using multiple top lists is to measure popularity from several ranking mechanisms in an unbiased way. We refer to the Alexa, Umbrella, and Majestic ranks of the targeted brand domain identified on the basis of the visual similarity features as mentioned above in RAW, PSR, and WS images as features Nos. 4–6, 7–9, and 10–12, respectively.

TLD features. The TLD features, Nos. 13–16 listed in Table 2, are designed to use domain names' own characteristics of both input IDNs and targeted brand domains. We introduce these features since our analysis reveals that the usage of TLDs has changed dramatically in recent years, and deceptive IDNs do not always use the same TLD as the targeted brand domains. We use the TLD types defined in Section 3.1 (e.g., legacy gTLD, new gTLD, new IDN gTLD, legacy ccTLD, and new IDN ccTLD) as the TLD features for the input IDN (No. 13) and the targeted brand domain based on RAW (No. 14), PSR (No. 15), and WS (No. 16) images.

3.5 Step 5: Score Calculation

The fifth step of DomainScouter is calculating the deceptive IDN score, which is the estimated probability of the user being deceived by the corresponding input IDN. We use a supervised machine learning approach to calculate the score. The input of this step consists of the input IDN with the features listed in Table 2. We use one-hot encoding for categorical features (Nos. 13–16). Supervised machine learning is generally composed of two phases: training and testing. The training phase generates a machine learning model from training data that includes extracted features and labels. For labeling, we hypothesize that some deceptive IDNs have already been used for phishing or social engineering attacks. Thus, we rely on multiple blacklists that have phishing or social engineering categories and carefully label the input IDN *deceptive* or *non-deceptive*. Note that our aim is not labeling many known deceptive IDNs but labeling reliable deceptive IDNs for estimating the scores for unlabeled IDNs. In the testing phase, the model generated in the training phase is used to calculate the probabilities of input IDNs being deceptive IDNs. We define these probabilities as the deceptive IDN scores. The higher the score, the more likely the user is to be deceived by the IDN. Consequently, this step outputs detected deceptive IDNs, their targeting brand domains, and the deceptive IDN scores.

Among many traditional and deep learning algorithms, we select Random Forest [8] for three reasons. First, Random Forest has good interpretability, i.e., it makes clear how features contribute to the result and how they are treated. Second, the parameters of Random Forest include the number of decision

trees to employ and the features considered in each decision tree, which makes the model easy to tune. Finally, in our preliminary experiments, Random Forest outperformed other popular algorithms such as Logistic Regression, Naïve Bayes, Decision Tree, and Support Vector Machine. In Random Forest, the probability or deceptive IDN score is calculated by averaging results of each decision tree. The higher the number of decision trees predicted to be deceptive, the higher the deceptive IDN score.

3.6 Limitation

DOMAINSCOUTER has two limitations. First, it does not aim to detect various kinds of malicious domain names but only deceptive IDNs that may lead to user misbehavior. Thus, a deceptive IDN is not always used for specific malicious attacks (e.g., phising, social engineering, and malware). However, identifying deceptive IDNs itself provides incentives for various stakeholders as discussed later in Section 7. There are many previous systems aiming at detecting malicious domain names in terms of the lexical characteristics [37, 55, 67], the relationship between domains and IP addresses [3, 10, 32], and the behavior of DNS queries [4, 5, 7]. Our system complements these systems. In particular, we can combine the systems to achieve better detection coverage.

The second limitation is in the coverage of non-English brands in step 2. In particular, we selected non-English brands on the basis of the top lists; however, there could be more non-English brands for each country, region, and language. We will explore other sources such as registered trademarks or search engine results for each country in our future work.

4 Evaluation

In this section, we show the results of comparing our system DOMAINSCOUTER with those proposed in previous works in terms of system properties and detection performance.

4.1 Comparison of Properties

We compared the properties of DOMAINSCOUTER and those of two previous systems [36, 48] from four perspectives. Table 3 summarizes the results.

Dataset. When comparing datasets used in each study, there are clear gaps between our system and the other two. In particular, our system contains 570 studied TLDs, whereas that of Liu et al. [36] contains only 56. No description regarding TLDs is provided by Sawabe et al. [48]. Furthermore, our system contains many more IDNs than the two previous systems: 3 times more than that of Liu et al. [36] and 2.3 times more than that of Sawabe et al. [48]. To the best of our knowledge, our domain dataset that includes both gTLDs and ccTLDs is the most comprehensive dataset ever used in security research.

Targeted Brand. In terms of the targeted brands used in each study, DOMAINSCOUTER uses both English and non-English brand domains, and the number of these domains is much bigger than that of the Liu et al. [36] and Sawabe et al. [48] systems.

Table 3: Results of Comparing Properties

		Our System	Liu et al. [36]	Sawabe et al. [48]
Dataset	# TLDs (IDNs)	570	56	-
	# Domains (IDNs)	4,426,317	1,472,836	1,928,711
Targeted Brand	# Domains (English)	2,310	1,000	1,000
	# Domains (Non-English)	4,774	0	0
Deceptive IDN	Combo	●	◐	○
	Homo	●	●	●
	Homocombo	●	○	○
Method	Visual Similarities	●	●	●
	Brand Features	●	◐	◐
	TLD Features	●	○	○

●: Fully Covered, ◐: Partially Covered, ○: Not Covered

Deceptive IDN. DOMAINSCOUTER focuses on various deceptive IDNs (eng-combo, eng-homo, eng-homocombo, noneng-combo, noneng-homo, and noneng-homocombo), whereas Liu et al. [36] studied only eng-homo and a part of eng-combo IDNs, and Sawabe et al. [48] detected only eng-homo IDNs.

Method. Liu et al. [36] used visual similarities (the SSIM index) between IDNs and brand domains to detect eng-homo IDNs targeting the Alexa Top 1,000 brands. Sawabe et al. [48] calculated visual similarities between non-ASCII and ASCII characters using optical character recognition (OCR) to detect eng-homo IDNs. However, both methods need to tune the thresholds of either the SSIM index or OCR manually, which tends to cause false positives and false negatives, and do not consider how popular the targeted brand domain is. In addition, Liu et al. did not focus on eng-homo IDNs between different TLDs (e.g., `example[.]com` and `êxämplē[.]test`).

To solve the above problems, as stated in Section 3.4, DOMAINSCOUTER utilizes not only multiple visual similarity features but also targeted brand ranking and TLD features and applies a machine learning approach to eliminate tuning thresholds for visual similarity features.

4.2 Comparison of Detection Performance

We compared the deceptive IDN detection performance of DOMAINSCOUTER with that of the previously proposed systems [36, 48]. First, we describe the experimental setups in the other two systems and our system. Then, we illustrate the comparison results using real registered IDNs.

Setups of the Previous Systems. We replicated the previously proposed systems on the basis of their descriptions provided in the corresponding papers [36, 48] since the systems are not open-source. For the Liu et al. [36] system, we needed to set a threshold for the SSIM index to detect eng-homo IDNs. The original paper set the threshold to 0.95. However, in our re-implemented system, the 0.95 threshold caused non-negligible false positives, which may be due to the differences in the font and image settings between the original system and our re-implemented one. We manually verified the SSIM index results to determine the threshold of 0.99, which caused only few false positives. For the Sawabe et al. [48] system, we used the mappings between non-ASCII and correspond-

ing similar ASCII characters kindly provided by Sawabe et al. themselves [48] to re-implement the detection method of eng-homo IDNs. To match the brand domains employed in the previous works, we used all English brand domains shown in Section 3.2 for fair evaluation, even though the original papers used only the top 1,000 brand domains.

Setup of DomainScouter. Section 3 describes the implementation of our system, DomainScouter. As stated in Section 3.5, we need to set up a labeled training dataset. In our evaluation, we used 10,000 labeled IDNs consisted of 242 deceptive (positive) and 9,758 non-deceptive (negative) IDNs for building our machine learning model. The positive IDNs were labeled by referring to the latest three blacklists (hpHosts [20], Google Safe Browsing [16], and Symantec DeepSight [54]) as of November 2018 and manually verified by the authors. The 242 positive IDNs are composed of only eng-homo deceptive IDNs since even the latest blacklists do not cover other types of deceptive IDNs. However, we design our proposed features to grasp the nature of various deceptive IDNs, thus DomainScouter can identify other types of deceptive IDNs other than eng-homo. The negative IDNs were randomly sampled from the IDNs shown in Table 1 and manually verified by the authors.

We performed 10-fold cross-validation (CV) on the training dataset and achieved a true positive rate of 0.981, a true negative rate of 0.998, a false positive rate of 0.002, a false negative rate of 0.019, and an F-measure of 0.972 on average. Regarding the two parameters in Random Forest, we set the number of decision trees as 100 and the number of sampled features in each individual decision tree as 6 on the basis of the best results in our preliminary experiments. Note that, as explained in Section 3.6, DomainScouter does not aim to detect malicious IDNs but only deceptive IDNs. Thus, the *positive* does not mean *malicious* but *deceptive*. Similarly, *negative* does not mean *legitimate/benign* but *non-deceptive*. This has been a typical evaluation setting regarding detecting deceptive IDNs (e.g., eng-homo). Similar to ours, previous studies [36, 48] did not provide true positive/negative rates in terms of detecting *malicious* IDNs since they focused on detecting eng-homo IDNs.

The last column of Table 2 shows relative feature importance of all features. The higher the importance score, the more the feature contributed to the correct detection. The results demonstrated that the three visual similarity features (Nos. 1–3) are more effective than the other features. In particular, the visual similarity based on word segmented images (feature No.3) appeared to contribute to the correct detection the most. The remaining proposed features (Nos. 4–16) were confirmed to contribute to detecting deceptive IDNs as well.

Detection Performance. Here, we compare the detection performance of the three systems. The input IDNs for each system were the same 4,426,317 IDNs described in Table 1.

Unfortunately, there is no ground truth to label all IDNs. Thus, we used the re-implemented previous systems and the trained DomainScoter, which proved to be accurate in the CV evaluation, to explore unknown deceptive IDNs in the dataset. Of course, there could be unavoidable false negatives or missed deceptive IDNs. We manually excluded false positives or falsely detected non-deceptive IDNs from the results of the three systems.

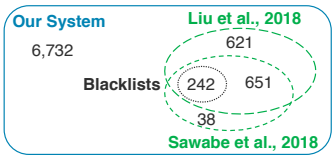

Figure 2: Venn Diagram of Detected Deceptive IDNs

We did not exclude the 242 positive IDNs used for the training dataset of DomainScouter from the input IDN in this evaluation for two reasons. One is that the goal of our paper is not just to compare the detection performance but also to conduct a comprehensive measurement study of deceptive IDNs (shown later in Section 5). The other is all the 242 positive IDNs were confirmed to be easily detected by the three systems since they were easily identifiable eng-homo deceptive IDNs.

Figure 2 is a Venn diagram showing intersections of deceptive IDNs detected by the three systems and the 242 positive IDNs labeled using blacklists. The Liu et al. system detected 1,514 deceptive IDNs (=621+651+242) and the Sawabe et al. system detected 931 (=38+651+242). Our analysis revealed that the difference between the coverage achieved by the Liu et al. and Sawabe et al. systems originated from the difference in handling the input IDN by each system: the Liu et al. system handled an IDN string as one image, whereas the Sawabe et al. system handled each non-ASCII character contained in an IDN.

Surprisingly, DomainScouter fully covered the 1,552 (=621+38+651+242) deceptive IDNs detected by the two previous systems. Moreover, DomainScouter detected 6,732 further deceptive IDNs that were not detected by the two systems. The extra detected deceptive IDNs mainly consisted of our new targets such as eng-combo, eng-homocombo, noneng-combo, and noneng-homocombo. The results of the 8,284 IDNs detected in total are explained in the next section.

5 Measurement Study

So far, we have evaluated the detection performance of DomainScouter compared with those of the two previously proposed systems. This section focuses on the 8,284 deceptive IDNs detected by DomainScouter. To the best of our knowledge, this is the most comprehensive study in terms of the numbers of both the input IDNs (more than 4.4 million registered IDNs under 570 TLDs as shown in Table 1) and the detected deceptive IDNs. In the following sections, we describe our measurement results in terms of the characteristics of deceptive IDNs, the impacts caused by deceptive IDNs, and the brand protection of deceptive IDNs.

Table 4: Breakdown of Detected Deceptive IDNs

Type	# IDNs
eng-combo	368
eng-homo	1,547
eng-homocombo	3,697
noneng-combo	144
noneng-homocombo	2,528
Total	8,284

5.1 Characteristics of Deceptive IDNs

Deceptive Types. We begin by investigating the types of deceptive IDNs found in the registered IDNs as of May 2018. The identified deceptive IDNs were grouped into the defined types on the basis of the information we obtained when extracting our proposed features, i.e., identified targeted brands (eng/noneng) and which SSIM index of images (RAW/WS/PSR) is the highest. Table 4 provides a breakdown of the detected deceptive IDNs. Our system found 368 eng-combo, 1,547 eng-homo, 3,697 eng-homocombo, 144 noneng-combo, and 2,528 noneng-homocombo IDNs. As explained in Section 2, some eng-homo IDNs were already analyzed in the previous studies [36, 48]. We successfully revealed that there were many deceptive IDNs other than eng-homo IDNs, which were found in the research literature for the first time. We defined a noneng-homo IDNs; however, our system did not detect any noneng-homo IDNs that targeted our selected non-English brand domains from the input IDNs.

Targeted Brands. Next, we focused on the targeted brands among the detected deceptive IDNs. Table 5 lists the 10 most targeted English brands, along with their Alexa ranks, among the detected deceptive IDNs. The results highlight three major outcomes. First, more popular brand domains (i.e., those with higher Alexa ranks) are targeted for creating deceptive IDNs as hypothesized in Section 3.4. Second, all websites of the top 10 targeted brands offer user accounts and user login functions. A possible explanation for this is attackers targeted these websites to obtain sensitive information such as user IDs and passwords via phishing or social engineering attacks. Finally, DomainScouter successfully detected many eng-combo and eng-homocombo IDNs that were defined in this paper for the first time. For example, we found that Amazon was targeted the most (56 eng-combo IDNs, 64 eng-homo IDNs, and 843 eng-homocombo IDNs).

Table 6 lists the 10 most targeted non-English brands, along with their Alexa rankings and English meanings. The result proves the existence of many noneng-combo and noneng-homocombo IDNs that are defined and studied in this paper for the first time. Noneng-combo IDNs were found for only one target brand in the top 10 brands. We found many noneng-homocombo IDNs that targeted place names (e.g., Austria, Pattaya, and Antalya) and common words (e.g., sport, flights, and weather) in non-English languages.

Creation Dates. We examined when the detected deceptive IDNs were registered and started to be used. To this end, we leveraged the WHOIS database [64] explained in Section 3.1. From the WHOIS database, we extracted the dates

Table 5: Top 10 Targeted English Brands

Target	Alexa	eng-combo	eng-homo	eng-homocombo	Total
amazon[.]com	10	56	64	843	963
hotels[.]com	622	2	13	457	472
google[.]com	1	14	122	100	236
apple[.]com	71	20	59	129	208
facebook[.]com	3	18	78	58	154
target[.]com	410	0	6	135	141
youtube[.]com	2	23	37	61	121
bet365[.]com	274	79	0	22	101
office[.]com	38	5	6	84	95
yahoo[.]com	7	7	18	64	89

Table 6: Top 10 Targeted Non-English Brands

Target	Alexa	Meaning	noneng-combo	noneng-homocombo	Total
xn-sterreich-z7a[.]at	487,222	Austria	0	1,032	1,032
xn-nlabebi[.]kz	479,087	sport	0	307	307
xn-o3cnn2dh[.]ws	977,559	Pattaya	0	159	159
xn-flge-1ra[.]de	199,379	flights	0	155	155
xn-80ahnhrfk[.]shop	419,929	presents	0	42	42
xn-mto-bmab[.]fr	58,899	weather	42	0	42
xn-72c0ao2e4bzd[.]com	475,666	cash	0	28	28
xn-80abmi5aecftcl4j[.]su	459,704	security	0	26	26
xn-90acjmnnc1hybf[.]su	900,952	ad	0	23	23
xn-hgbkak5kj5a[.]net	234,297	Antalya	0	23	23

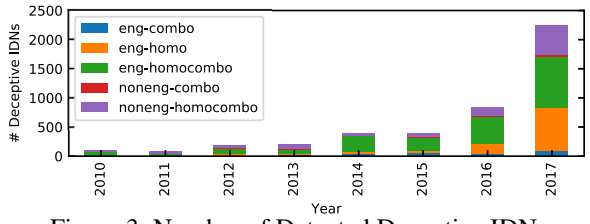

Figure 3: Number of Detected Deceptive IDNs

of registration corresponding to deceptive IDNs. Due to some limitations of the WHOIS dataset (e.g., dates of registration were not provided in some registries), we were able to extract the dates for only 62% (=5,176 / 8,284) of the detected deceptive IDNs.

Figure 3 illustrates the number of deceptive IDNs in each year by the deceptive type. The results revealed two major facts about deceptive IDNs. First, the number of deceptive IDNs increases year by year. Second, many deceptive IDNs that are newly defined in this paper (e.g., eng-combo, eng-homocombo, noneng-combo, and noneng-homocombo) were registered after 2014.

5.2 Impacts of Deceptive IDNs

Accesses. To understand the impacts of the detected deceptive IDNs, we investigate how many accesses or queries to the detected deceptive IDNs were observed over time. To this end, we leveraged the passive DNS database (DNSDB) [13] covering the period from 6-24-2010 to 9-19-2018. The DNSDB enabled us to investigate statistics of DNS queries to the deceptive IDNs such as the dates of first- and last-seen queries and the number of queries. Note that, because the provider could not identify specific users from aggregated DNS queries, the DNSDB data inevitably counted queries from victims, attackers, and security researchers. Table 7 lists the total number of DNS queries to each type of deceptive IDNs. For the deceptive IDNs targeting English brands, 1,547 eng-homo IDNs were queried over 1 million times in total. For those targeting non-English brands, 2,528 noneng-homocombo IDNs were

Table 7: Accesses to Deceptive IDNs

	# Deceptive IDNs	Sum of Queries
eng-combo	368	226,546
eng-homo	1,547	1,019,613
eng-homocombo	3,697	737,696
noneng-combo	144	317,043
noneng-homocombo	2,528	1,440,388
Total	8,284	3,741,286

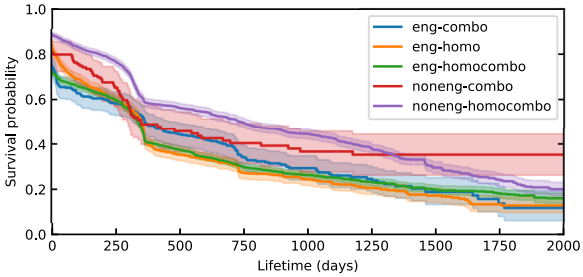

Figure 4: Lifetime of Deceptive IDNs

queried over 1.4 million times in total. These results show that all types of deceptive IDNs accumulated a non-negligible number of accesses over time.

Lifetime. Next, we focus on the lifetime of the detected deceptive IDNs. In this paper, we define the lifetime of deceptive IDNs as the period from the first-seen to the last-seen DNS queries on the basis of the DNSDB data. To analyze the lifetime, we used a survival analysis based on the Kaplan-Meier estimator [28], commonly used to estimate lifespan in cyber security [15, 33, 42]. Figure 4 shows the results of the survival analysis. The x-axis is the lifetime (in days), while the y-axis shows the survival probability, meaning the probability that the deceptive IDN remains after the elapsed number of days. From the figure, we can see that the characteristics of the deceptive IDNs targeting English brands and those targeting non-English brands are different. In particular, eng-combo, eng-homo, and eng-homocombo IDNs have shorter survival probability than noneng-combo and noneng-homocombo IDNs.

5.3 Brand Protection

So far, we have shed light on the characteristics of the detected deceptive IDNs overall. This section focuses on the deceptive IDNs that are now protected by their legitimate domain owners or rightsholders. In particular, we investigated each detected deceptive IDN to identify whether it is protected by its targeted brand's owner. To this end, we used the WHOIS dataset to extract registrant emails of both the deceptive IDN and the targeted brand domain. In this work, a deceptive IDN is considered to be protected if both emails are the same and the domain part of the email (e.g., `@example[.]com`) is identical to the targeted brand domain (e.g., `example[.]com`). This identification process has two limitations. One is the process does not work when an email address is not properly extracted from the WHOIS dataset. The other is the process cannot properly identify a protected deceptive IDN if the legitimate domain owner uses different email addresses for the brand domain and its deceptive IDN, or if the owner uses a WHOIS privacy protection service to hide their email addresses.

Table 8: Top 10 Protected Brands

Brand Domain	Alexa	# Protected	# Detected	Protective Ratio
`amazon[.]com`	10	42	963	4.4%
`google[.]com`	1	35	236	14.8%
`gmail[.]com`	536	18	81	22.2%
`skype[.]com`	456	17	56	30.4%
`android[.]com`	990	16	45	35.6%
`blogger[.]com`	299	16	26	61.5%
`bet365[.]com`	274	15	101	14.9%
`cloudflare[.]com`	256	14	16	87.5%
`youtube[.]com`	2	14	121	11.6%
`symantec[.]com`	310	14	16	87.5%

Using the above identification processes, we revealed that only 3.8% (=316 / 8,284) of the detected deceptive IDNs were protected by their targeted brand owners. Table 8 lists the top 10 protected brands; it contains the Alexa rank, the number of protected deceptive IDNs, the number of all detected deceptive IDNs, and the protective ratio. From the table, one can derive two noteworthy facts regarding brand protection. One is no brand domain in the top 10 or the world's most popular Internet companies protected themselves from all of its corresponding detected deceptive IDNs. This strongly indicated that the deceptive IDN problem is difficult for one company to solve by itself. The other fact is only the few companies offering Internet security services (e.g., Cloudflare and Symantec) protected themselves from the deceptive IDNs more than other companies did.

6 User Study

The attacks that use deceptive IDNs target the perceptions of the users accessing websites. In this section, we examine whether the deceptive IDN score we proposed reflects the tendency of users to be deceived by the attacks. Understanding the impact of the attacks on users helps stakeholders to discuss more practical countermeasures. We conducted two separate online surveys on Amazon Mechanical Turk (MTurk): *User Study 1* to investigate users' knowledge of IDNs, and *User Study 2* to examine the extent to which users are deceived by deceptive IDNs. Our Institutional Review Board (IRB) approved both surveys. Participants were limited to U.S. residents with an approval rating over 97% and more than 50 tasks approved. We conducted these surveys in November 2018.

6.1 User Study 1

The first survey was designed to ask participants about their demographics and knowledge of IDNs.

Method. The survey consisted of 12 closed-ended questions. We asked the participants about the characters used in domain names. This was a multiple-choice question with the following options: English (Upper case), English (Lower case), digit, hyphen, punctuation, Cyrillic, Greek, Chinese, Japanese, and Korean. IDNs can contain all these characters except for punctuation.

Table 9: Participants' Misunderstanding of the Characters That Can be Used in Domain Names

	# Participants (all)	# Participants (Computer Eng. / IT Pro.)
Correct Answer	20 (5.5%)	7 (13.5%)
Incorrect Answer	344 (94.5%)	45 (86.5%)
Total	364 (100.0%)	52 (100.0%)

The median time to complete the survey was 3.2 minutes, and we compensated the participants $0.50 each. After removing 15 participants who gave incomplete or careless answers, we analyzed the remaining 364 participants. 61.0% of the participants were male, and their ages ranged from 19 to 71, with a median of 33 (mean 36.3). Our sample had a wide range of education levels (from high school to graduate degree) and various occupations.

Results. Each language other than English was selected by only one-fourth of the participants at most, whereas English, hyphen, and digit were selected by over a half of the participants. As shown in Table 9, a small number, only 5.5% (=20 / 364), of the participants knew enough about IDNs (i.e., they selected all choices except for punctuation). Only 11.3% (=41 / 364) of the participants seemed to have *some* knowledge about IDNs, i.e., they selected some languages other than English and did not select punctuation. Surprisingly, only 13.5% of the computer engineers or IT professionals answered the question correctly.

In summary, the majority of the participants did not know enough about IDNs, even those engaged in IT-related occupations.

6.2 User Study 2

In the second survey, we aimed to examine the extent to which users are deceived by attacks employing deceptive IDNs.

Method. The survey consisted of 18 closed-ended questions regarding users' demographics and visual perception of deceptive IDNs.

To measure how many users are not aware of the deceptive IDNs that disguise domains of popular online services, we prepared 70 actual deceptive IDNs for seven popular brands (online services): Google, YouTube, Facebook, Amazon, Twitter, Instagram, and PayPal. We prepared five high-scoring deceptive IDNs (with the score of 1.0) and five low-scoring deceptive IDNs (with the scores ranging from 0.06 to 0.56) for each target brand.

After demographic questions, the participants were first asked which services they used more than once a month. The list of the seven popular brands mentioned above was used to formulate this question. After a few dummy questions, we then gave the participant a deceptive question, asking "Have you ever visited `[SERVICE].com`?" as a closed-ended question, which could be answered with "yes" or "no." Note that `[SERVICE].com` was actually replaced by a deceptive IDN in this question. For example, `êxample[.]test` would be used instead of `example[.]test`. The displayed deceptive IDNs and their order were randomized for each participant. We defined *potential* victims of the attack as those who answered "yes" in the deceptive questions about a certain brand's service among those who used the service more than once a month in the previous question. We assumed that the participants who answered "no" recognized the deceptive IDNs.

Table 10: The Ratio of the Participants Who Were Aware of Deceptive IDNs. We Prepared Five High-scoring Deceptive IDNs and Five Low-scoring Deceptive IDNs for Each Brand

Brand (Score)	Score Min	Score Max	Score Mean	# Potential Victims Min	# Potential Victims Max	# Potential Victims Mean	# Participants* Min	# Participants* Max	# Participants* Mean	Insensible Rate Min	Insensible Rate Max	Insensible Rate Mean
Google (H)	1.0	1.0	1.0	40	45	42.8	41	47	43.8	0.96	1.0	0.98
Google (L)	0.07	0.40	0.26	23	41	35.0	46	50	47.2	0.50	0.85	0.74
YouTube (H)	1.0	1.0	1.0	34	42	38.6	40	48	44.8	0.79	0.93	0.86
YouTube (L)	0.28	0.40	0.35	29	42	35.8	40	46	43.6	0.71	0.91	0.82
Facebook (H)	1.0	1.0	1.0	30	37	33.2	33	39	36.4	0.82	0.95	0.91
Facebook (L)	0.11	0.40	0.26	29	36	31.4	36	41	38.6	0.76	0.90	0.81
Amazon (H)	1.0	1.0	1.0	40	46	44.0	41	47	44.8	0.96	1.0	0.98
Amazon (L)	0.25	0.40	0.36	28	43	35.6	39	47	43.2	0.72	0.91	0.82
Twitter (H)	1.0	1.0	1.0	22	28	24.2	23	31	27.2	0.77	0.96	0.89
Twitter (L)	0.38	0.53	0.41	15	25	20.6	25	31	27.4	0.60	0.84	0.75
Instagram (H)	1.0	1.0	1.0	20	26	22.2	22	28	25.0	0.75	1.0	0.89
Instagram (L)	0.39	0.56	0.42	15	23	19.0	16	26	23.6	0.71	0.94	0.81
PayPal (H)	1.0	1.0	1.0	23	28	25.6	25	28	26.4	0.92	1.0	0.97
PayPal (L)	0.06	0.38	0.26	18	26	21.4	24	30	27.0	0.68	0.88	0.79

	# Potential Victims	# Participants*	Insensible Rate
All (H)	1,153	1,242	0.92
All (L)	994	1,253	0.79
Total	2,147	2,495	0.86

*Participants who answered that they use the brand's service more than once a month
H: High Score, L: Low Score

The median time to complete the survey was 4.3 minutes, and we compensated the participants $0.75 each. After removing 17 participants who gave incomplete or careless answers, we analyzed the remaining 474 participants. The participants' ages ranged from 18 to 72, with a median of 34 (mean 35.7). 59.7% of the participants were male. Similar to the first survey, the sample of the second survey had a wide range of education levels and occupations.

A limitation of this user study is that we did not measure the actual *success rate* of the attacks. As an ethical consideration, we did not provide the hyperlinks of the actual deceptive IDNs in the questionnaires to avoid harming the participants. Another limitation is that the study was limited to 70 deceptive IDNs. However, we believe this study can provide unique and adequate results to show the risks of deceptive IDNs.

Results. We defined the *insensible rate* = v/p, where p is the number of participants who answered that they used a certain brand's service once a month, and v is the number of potential victims who answered that they visited the deceptive IDN disguising the brand's service. The results are shown in Table 10. Most participants did not really notice the deceptive IDNs with high scores; the insensible rate for the IDNs with high scores was 0.92 (=1,153 / 1,242). Surprisingly, many participants did not notice deceptive IDNs even if their scores were not high; the insensible rate for the IDNs with low scores was 0.79 (=994 / 1,253) in total. The insensible rate of all IDNs was 0.86 (= (1,153+994) / (1,242+1,253)). Some participants who noticed deceptive IDNs commented: "*[...] I marked these as no because they contained these special characters*" and "*[...] questions are supposed to be phishing or intentionally fake sites but I marked no on the ones that aren't plainly the*

Table 11: Correlation between Deceptive IDN Score and Insensible Rate

Brand	Correlation Coefficient (γ)	p-value
Google	0.83	0.0027*
YouTube	0.35	0.31
Facebook	0.74	0.014*
Amazon	0.84	0.0021*
Twitter	0.73	0.016*
Instagram	0.46	0.18
PayPal	0.87	0.0011*
All	0.68	<0.0001*

We note statistically significant differences with asterisks.

real domain." Unfortunately, the participants who were IT professionals and computer engineers were also likely not to notice deceptive IDNs, similar to other participants.

Overall, as shown in Table 11, we found a positive correlation between the deceptive IDN score and the insensible rate of the attacks (γ=0.68, p-value<0.0001), although the correlation was not significant for YouTube and Instagram. This result indicates that the proposed system can successfully measure the reasonable scores that reflect the tendency of users to be deceived by deceptive IDN attacks.

In summary, our user study newly revealed deceptive IDNs are difficult for end users to recognize even if they are IT-professionals or computer engineers. Through correlation analysis, we confirmed that the deceptive IDN score successfully reflects the tendency of users to be deceived by the considered type of cyber attacks.

7 Discussion

In the previous section, the user studies revealed that most end users do not notice deceptive IDNs. To mitigate the risks of deceptive IDNs and enhance cultural and linguistic diversity on the Internet with IDNs, various stakeholders should take countermeasures against deceptive IDNs. We believe that our findings based on the measurements and user studies can help improve countermeasures for stakeholders. Now, we briefly provide discussions and suggestions for client applications, domain registrars/registries, domain owners, and certificate authorities (CA) on how to reduce the spread of deceptive IDNs.

7.1 Client Application

Client applications such as web browsers and other applications displaying URLs or domain names can prevent users accessing deceptive IDNs by detecting them. For example, to mitigate eng-homo deceptive IDNs, many web browsers have original policies/rules about whether to display IDNs in Unicode or Punycode format in their address bars [18, 45]. Moreover, very recently, the Google Chrome browser has implemented a new experimental feature for warning against lookalike URLs including eng-homo deceptive IDNs [51]. DomainScouter found many newly defined deceptive IDNs other than simple eng-homo, thus, DomainScouter can help improve the rules/functions for providing better detection coverage of deceptive IDNs.

Unfortunately, the mitigation in client applications can only prevent users from accessing deceptive IDNs and does not address the root cause that such deceptive IDNs exist. The existence of a deceptive IDN similar to a legitimate brand is a risk of brand defamation, especially for companies. Therefore, not only client applications but also the other stakeholders should take other countermeasures against them.

7.2 Registrar and Registry

The guidelines for implementing IDNs [17] for mainly TLD registries describe that visually confusing characters from different scripts must not be allowed to co-exist in a single IDN label unless a corresponding IDN policy and IDN Table [12, 47] are defined to minimize confusion between domain names. The majority of eng-combo and eng-homocombo exhibit the prohibited pattern, mixing cross-script code points in a single label. According to Table 4, eng-combo and eng-homocombo account for 49% (=(368+3,697) / 8,284) of all 8,284 detected deceptive IDNs. If registries strictly followed the guidelines prohibiting the mixture of cross-script code points, approximately half of the discovered deceptive IDNs could have been avoided.

Registrars and registries make an effort to enable rightsholders to protect their rights when registering domain names; however, they do not investigate IDNs comprehensively. Although the trademark clearinghouse (TMCH) [57] contributes to protecting domains, deceptive IDNs are beyond its technical scope. The TMCH serves as a database for verified trademark rights information. Trademarks are submitted to the TMCH by rightsholders. Verified marks are provided with a priority-registration period and the Trademark Claims service for all new gTLDs. The Trademark Claims service identifies potentially abusive registrations by comparing TMCH-recorded trademark strings to domain names and sends a notice to rightsholders. The technical problem is a domain name is considered as an exact match to a TMCH-recorded string. This method results in false negatives when detecting deceptive IDNs. Our system discovered various deceptive IDNs unexplored by other methodologies. This means that registrars and the TMCH should broaden the scope of the detection to include IDNs and adopt the method proposed in this paper to prioritize defending high-scoring deceptive IDNs. Furthermore, the TMCH should serve not only new gTLDs but also legacy ccTLDs and new IDN ccTLDs.

7.3 Domain Owner

Brand protection is an essential way for rightsholders to fight against the violation of their rights. The mindset of those owning famous domain names (or trademarks) should be to make an effort to protect their brands and not to allow visually confusing domain names to be operated by other parties. The owners of famous domains (or trademarks) can take preventive actions to protect their brands. They can proactively

register additional domain names that are similar to their own brands to prevent abusive registrations by other parties. They can also use brand protection services (e.g., the TMCH) or take measures by themselves. According to our measurement results, only 3.8% of the visually confusing domain names that we discovered as deceptive IDNs were legitimately registered for brand protection. We assume that most domain owners (and also brand protection services) are not aware of such IDNs because they were unexplored by other existing methodologies; thus, domain owners should broaden the scope of brand protection to include IDNs.

When domain owners find squatted domain names (e.g., deceptive IDNs) targeting their brands, they can use the Uniform Domain-Name Dispute-Resolution Policy (UDRP) [59] to confiscate or cancel such domain names. The UDRP, a policy for resolving disputes regarding the registration of domain names, has been adopted by all ICANN-accredited registrars of gTLDs [26]. Many registrars of ccTLDs also adopt the UDRP or regionally localized policies based on it (e.g., JP-DRP [27]). Dispute resolution services based on the UDRP are widely used by rightsholders. The World Intellectual Property Organization (WIPO), one such service provider, handled over 73,000 cases from 1999 to 2017 and successfully transferred the rights to rightsholders [65, 66]. A case filed with the WIPO is normally concluded within two months. The Uniform Rapid Suspension System (URS) [60], which complements the UDRP, provides rightsholders with a quick and a low-cost process to take down squatted domain names. The fees of the URS start from almost $1,000 less than those of the UDRP ($1,500 [49]). The identified invalid domain names are suspended by the registry within two or three weeks; however, they are not deleted or transferred to the rightsholders. To counter deceptive IDNs, domain owners can select one of the two services (the UDRS or the URS) by taking both the urgency and the monetary costs into consideration.

7.4 Certificate Authority

Outreach efforts to spread HTTPS by security engineers, researchers, and browser vendors made many large websites serve HTTPS by default. The major browsers also require HTTPS; e.g., Google Chrome started to mark all HTTP sites as "not secure" in July 2018.

Certificate authorities (CAs) should not issue certificates to suspicious domain names (websites) to protect end users from deceptive IDNs. However, in reality, many CAs have issued certificates to squatted domain names, including deceptive IDNs [58]. The baseline requirement for the issuance and management of publicly trusted certificates published by the CA Browser Forum [6] mentions that CAs should do additional verification activities for high-risk certificate requests. We recommend that CAs accommodate the brand-protection policies and procedures that are followed by domain registrars. If all responsible CAs proactively shared trademark information similar to the TMCH, they would NOT issue certificates to squatted domain names. In addition, CAs would be able to revoke certificates for the domain names that violate trademarks if they received such claims from rightsholders. Domain owners are able to explore certificates of squatted domain names in the log server of certificate transparency [34] because all CAs are now encouraged to submit new certificates to it. Many responsible CAs receive claims from rightsholders.

8 Related Work

We summarize related research literature in terms of deceptive IDNs and non-IDN squattings.

Deceptive IDNs. Gabrilovich and Gontmakher first mentioned an IDN homograph attack using non-ASCII characters in 2002 [14]. In 2006, Holgers et al. investigated a campus network traffic to find eng-homo IDNs targeting the Alexa top 500 [19]. As mentioned in Section 4, in 2018, Liu et al. proposed an eng-homo IDN detection method using the SSIM index between IDNs and brand domains [36]. Sawabe et al. proposed using OCR-based similarities between non-ASCII and ASCII characters [48]. In 2019, Le Pochat et al. explored candidate IDNs that brand owners may want to register [43]. Suzuki et al. developed a framework to identify IDN homographs in an automated manner [53]. Whereas the above studies focused mainly on eng-homo IDNs using a smaller number of IDNs under a limited number of TLDs, our work has advanced these studies by focusing on various deceptive IDNs (e.g., eng-combo, eng-homocombo, noneng-combo, and noneng-homocombo), analyzing more IDNs under almost all TLDs, and studying the extent to which users are deceived by deceptive IDNs.

Non-IDN Squattings. In addition to deceptive IDNs, many previous studies analyzed a wide range of domain squatting methods in non-IDN (ASCII) domains such as combosquatting (combining brand name with keywords) [30], bit squatting (accidental bit flips) [41], and typosquatting (typographical errors) [1, 29, 55, 62].

9 Conclusion

This paper proposed a system called DomainScouter to detect deceptive internationalized domain names (IDNs) and calculate the deceptive IDN score. We performed the most comprehensive measurement study to show that (1) there are many previously unexplored deceptive IDNs, (2) their number has kept increasing since 2014, and (3) only 3.8% of them are protected by their targeted brand owners. Moreover, we conducted online surveys to reveal that the majority of users cannot recognize deceptive IDNs and confirm that the deceptive IDN score successfully reflects the tendency of users to be deceived. To reduce the risk of deceptive IDNs, we provided suggestions for client applications, domain registrars/registries, domain owners, and certificate authorities. We hope that our results can be used to enable a secure and multilingual Internet for all users.

References

[1] Pieter Agten, Wouter Joosen, Frank Piessens, and Nick Nikiforakis. Seven months' worth of mistakes: A longitudinal study of typosquatting abuse. In *Proc. 22nd Annual Network and Distributed System Security Symposium (NDSS)*, 2015.

[2] Alexa Top Sites. http://www.alexa.com/topsites/.

[3] Manos Antonakakis, Roberto Perdisci, David Dagon, Wenke Lee, and Nick Feamster. Building a dynamic reputation system for DNS. In *Proc. 19th USENIX Security Symposium*, pages 273–290, 2010.

[4] Manos Antonakakis, Roberto Perdisci, Wenke Lee, Nikolaos Vasiloglou II, and David Dagon. Detecting malware domains at the upper DNS hierarchy. In *Proc. 20th USENIX Security Symposium*, 2011.

[5] Manos Antonakakis, Roberto Perdisci, Yacin Nadji, Nikolaos Vasiloglou II, Saeed Abu-Nimeh, Wenke Lee, and David Dagon. From throw-away traffic to bots: Detecting the rise of DGA-based malware. In *Proc. 21st USENIX Security Symposium*, pages 491–506, 2012.

[6] Baseline Requirements for the Issuance and Management of Publicly-Trusted Certificates, v.1.1.7. https://cabforum.org/wp-content/uploads/BRv1.1.7.pdf.

[7] Leyla Bilge, Engin Kirda, Christopher Kruegel, and Marco Balduzzi. EXPOSURE: finding malicious domains using passive DNS analysis. In *Proc. 18th Network and Distributed System Security Symposium (NDSS)*, 2011.

[8] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[9] Daiki Chiba, Mitsuaki Akiyama, Takeshi Yagi, Kunio Hato, Tatsuya Mori, and Shigeki Goto. DomainChroma: Building actionable threat intelligence from malicious domain names. *Computers & Security*, 77:138–161, 2018.

[10] Daiki Chiba, Takeshi Yagi, Mitsuaki Akiyama, Toshiki Shibahara, Takeshi Yada, Tatsuya Mori, and Shigeki Goto. DomainProfiler: Discovering domain names abused in future. In *Proc. 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, pages 491–502, 2016.

[11] Cisco Umbrella 1 Million. https://umbrella.cisco.com/blog/2016/12/14/cisco-umbrella-1-million/.

[12] K. Davies and A. Freytag. Representing Label Generation Rulesets Using XML. RFC 7940 (Proposed Standard), August 2016.

[13] Farsight Security, Inc. DNSDB. https://www.dnsdb.info/.

[14] Evgeniy Gabrilovich and Alex Gontmakher. The homograph attack. *Commun. ACM*, 45(2):128, 2002.

[15] Carlos Gañán, Orcun Cetin, and Michel van Eeten. An empirical analysis of ZeuS C&C lifetime. In *Proc. 10th ACM Symposium on Information, Computer and Communications Security (ASIACCS)*, pages 97–108, 2015.

[16] Google Safe Browsing. https://developers.google.com/safe-browsing/.

[17] Guidelines for the Implementation of Internationalized Domain Names Version 4.0. https://www.icann.org/en/system/files/files/idn-guidelines-10may18-en.pdf.

[18] Guidelines for URL Display. https://chromium.googlesource.com/chromium/src/+/master/docs/security/url_display_guidelines/url_display_guidelines.md.

[19] Tobias Holgers, David E. Watson, and Steven D. Gribble. Cutting through the confusion: A measurement study of homograph attacks. In *Proc. USENIX Annual Technical Conference (ATC)*, pages 261–266, 2006.

[20] hpHosts. http://www.hosts-file.net/.

[21] IANA. Root zone database. https://www.iana.org/domains/root/db.

[22] ICANN. ICANN IDN ccTLD Fast Track Process. https://www.icann.org/resources/pages/fast-track-2012-02-25-en.

[23] ICANN. ICANN IDN Glossary. https://www.icann.org/resources/pages/glossary-2014-02-04-en.

[24] ICANN. ICANN new gTLDs delegated strings. https://newgtlds.icann.org/en/program-status/delegated-strings.

[25] ICANN. Internationalized domain names. https://www.icann.org/resources/pages/idn-2012-02-25-en.

[26] ICANN-Accredited Registrars. https://www.icann.org/registrar-reports/accredited-list.html.

[27] JP Domain Name Dispute Resolution Policy (JP-DRP). https://www.nic.ad.jp/en/drp/.

[28] E. L. Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481, 1958.

[29] Mohammad Taha Khan, Xiang Huo, Zhou Li, and Chris Kanich. Every second counts: Quantifying the negative externalities of cybercrime via typosquatting. In *Proc. 36th IEEE Symposium on Security and Privacy (SP)*, pages 135–150, 2015.

[30] Panagiotis Kintis, Najmeh Miramirkhani, Charles Lever, Yizheng Chen, Rosa Romero Gómez, Nikolaos Pitropakis, Nick Nikiforakis, and Manos Antonakakis. Hiding in plain sight: A longitudinal study of combosquatting abuse. In *Proc. 24th ACM SIGSAC Conference on Computer and Communications Security (CCS)*, pages 569–586, 2017.

[31] Maciej Korczynski, Maarten Wullink, Samaneh Tajalizadehkhoob, Giovane C. M. Moura, Arman Noroozian, Drew Bagley, and Cristian Hesselman. Cybercrime after the sunrise: A statistical analysis of DNS abuse in new gTLDs. In *Proc. 13th ACM Asia Conference on Computer and Communications Security (ASIACCS)*, pages 609–623, 2018.

[32] Marc Kührer, Christian Rossow, and Thorsten Holz. Paint it black: Evaluating the effectiveness of malware blacklists. In *Proc. 17th International Symposium on Research in Attacks, Intrusions and Defenses (RAID)*, volume 8688, pages 1–21, 2014.

[33] Tobias Lauinger, Kaan Onarlioglu, Abdelberi Chaabane, William Robertson, and Engin Kirda. WHOIS lost in translation: (mis)understanding domain name expiration and re-registration. In *Proc. 16th ACM on Internet Measurement Conference (IMC)*, pages 247–253, 2016.

[34] B. Laurie, A. Langley, and E. Kasper. Certificate Transparency. RFC 6962 (Experimental), June 2013.

[35] Victor Le Pochat, Tom Van Goethem, Samaneh Tajalizadehkhoob, Maciej Korczyński, and Wouter Joosen. Tranco: A research-oriented top sites ranking hardened against manipulation. In *Proc. 26th Annual Network and Distributed System Security Symposium (NDSS)*, 2019.

[36] Baojun Liu, Chaoyi Lu, Zhou Li, Ying Liu, Hai-Xin Duan, Shuang Hao, and Zaifeng Zhang. A reexamination of internationalized domain names: The good, the bad and the ugly. In *Proc. 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, pages 654–665, 2018.

[37] Justin Ma, Lawrence K. Saul, Stefan Savage, and Geoffrey M. Voelker. Beyond blacklists: learning to detect malicious web sites from suspicious URLs. In *Proc. 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 1245–1254, 2009.

[38] Majestic Million. https://majestic.com/reports/majestic-million.

[39] Mozilla foundation. Public suffix list. https://publicsuffix.org/list/.

[40] NewSky Security. Fake Adobe website delivers BetaBot. https://blog.newskysecurity.com/fake-adobe-website-delivers-betabot-4114d1775a18.

[41] Nick Nikiforakis, Steven Van Acker, Wannes Meert, Lieven Desmet, Frank Piessens, and Wouter Joosen. Bitsquatting: exploiting bit-flips for fun, or profit? In *Proc. 22nd International World Wide Web Conference (WWW)*, pages 989–998, 2013.

[42] Arman Noroozian, Maciej Korczynski, Carlos Hernandez Gañán, Daisuke Makita, Katsunari Yoshioka, and Michel van Eeten. Who gets the boot? analyzing victimization by DDoS-as-a-Service. In *Proc. 19th International Symposium on Research in Attacks, Intrusions, and Defenses (RAID)*, volume 9854, pages 368–389, 2016.

[43] Victor Le Pochat, Tom van Goethem, and Wouter Joosen. Funny accents: Exploring genuine interest in internationalized domain names. In *Proc. 20th International Conference on Passive and Active Measurement (PAM)*, volume 11419, pages 178–194, 2019.

[44] Polyglot. http://polyglot.readthedocs.org.

[45] The Chromium Projects. IDN in Google Chrome. https://www.chromium.org/developers/design-documents/idn-in-google-chrome.

[46] Pyssim. https://github.com/jterrace/pyssim.

[47] Repository of IDN Practices. https://www.iana.org/domains/idn-tables.

[48] Yuta Sawabe, Daiki Chiba, Mitsuaki Akiyama, and Shigeki Goto. Detecting homograph IDNs using OCR. In *Proc. Asia-Pacific Advanced Network (APAN) Research Workshop*, volume 46, pages 56–64, 2018.

[49] Schedule of Fees under the UDRP (valid as of December 1, 2002). https://www.wipo.int/amc/en/domains/fees/.

[50] Quirin Scheitle, Oliver Hohlfeld, Julien Gamba, Jonas Jelten, Torsten Zimmermann, Stephen D. Strowes, and Narseo Vallina-Rodriguez. A long way to the top: Significance, structure, and stability of Internet top lists. In *Proc. 18th ACM Internet Measurement Conference (IMC)*, pages 478–493, 2018.

[51] Emily Stark. The URLephant in the room. In *USENIX Enigma 2019*. https://www.usenix.org/conference/enigma2019/presentation/stark.

[52] StatCounter. Browser Market Share Worldwide. http://gs.statcounter.com/browser-market-share.

[53] Hiroaki Suzuki, Daiki Chiba, Yoshiro Yoneya, Tatsuya Mori, and Shigeki Goto. ShamFinder: An automated framework for detecting IDN homographs. In *Proc. 19th ACM Internet Measurement Conference (IMC)*, 2019.

[54] Symantec. Deepsight intelligence. https://www.symantec.com/services/cyber-security-services/deepsight-intelligence.

[55] Janos Szurdi, Balazs Kocso, Gabor Cseh, Jonathan Spring, Márk Félegyházi, and Chris Kanich. The long "taile" of typosquatting domain names. In *Proc. 23rd USENIX Security Symposium*, pages 191–206, 2014.

[56] Tencent Security Xuanwu Lab. Spoof All Domains Containing 'd' in Apple Products [CVE-2018-4277]. https://xlab.tencent.com/en/2018/11/13/cve-2018-4277/.

[57] The Trademark Clearinghouse. http://trademark-clearinghouse.com.

[58] Touched by an IDN: Farsight Security shines a light on the Internet's oft-ignored and undetected security problem. https://www.farsightsecurity.com/2018/01/17/mschiffm-touched_by_an_idn/.

[59] Uniform Domain-Name Dispute-Resolution Policy. https://www.icann.org/resources/pages/help/dndr/udrp-en.

[60] Uniform Rapid Suspension (URS). https://www.icann.org/resources/pages/urs-2014-01-09-en.

[61] VirusTotal. https://www.virustotal.com/.

[62] Yi-Min Wang, Doug Beck, Jeffrey Wang, Chad Verbowski, and Brad Daniels. Strider typo-patrol: Discovery and analysis of systematic typo-squatting. In *Proc. 2nd USENIX Workshop on Steps to Reducing Unwanted Traffic on the Internet (SRUTI)*, 2006.

[63] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Processing*, 13(4):600–612, 2004.

[64] Whois XML API. https://www.whoisxmlapi.com/.

[65] WIPO Cybersquatting Cases Reach New Record in 2017. https://www.wipo.int/pressroom/en/articles/2018/article_0001.html.

[66] WIPO UDRP Domain Name Decisions (gTLD). https://www.wipo.int/amc/en/domains/decisionsx/index-gtld.html.

[67] Sandeep Yadav, Ashwath Kumar Krishna Reddy, A. L. Narasimha Reddy, and Supranamaya Ranjan. Detecting algorithmically generated malicious domain names. In *Proc. 10th ACM SIGCOMM Internet Measurement Conference (IMC)*, pages 48–61, 2010.

[68] Xudong Zheng. Phishing with Unicode Domains. https://www.xudongz.com/blog/2017/idn-phishing/.