.how .you .spot .whoswho .online .sucks

Deceiving Users with Generic Top-Level Domains

Richard Roberts Rachel Walter Daniela Lulli Dave Levin
University of Maryland

***Abstract*—In 2012, ICANN started allowing public applications for new generic top-level domains (gTLDs). Since then, the number of gTLDs has expanded from around two dozen to around two thousand. ICANN anticipated that this would significantly alter the web and justified the decision, stating "Unless there is a good reason to restrain it, innovation should be allowed to run free."**

In this paper, we provide what we believe to be "good reason to restrain" the creation of new gTLDs. To make safe, responsible decisions online, consumers must be able to determine with whom they are communicating, and one of the only reliable sources of information they have for this is a website's domain name. Unfortunately, we find that the creation of new gTLDs significantly complicates users' ability to evaluate who they are communicating with.. We perform an online survey showing that users become more susceptible to domain impersonation attacks when malicious domains use gTLDs, compared to country-code TLDs and "common" TLDs (.com, .net, or .org). We measure the current state of domain impersonation attacks that use gTLDs, finding a sharp increase in attackers' use of them to launch what are likely phishing websites. Finally, we discuss potential solutions and guidelines for the creation of new gTLDs.

I. Introduction

The early internet had only seven top-level domains: .com, .org, .net, .int, .edu, .gov, and .mil. The Internet Corporation for Assigned Names and Numbers, or ICANN, was formed in 1998 to manage the namespaces and numberspaces of the internet's infrastructure. ICANN became responsible for approving and managing the release of new TLDs, which trickled in through the 2000s with various rounds of proposals and opportunities for public comment. In 2011, ICANN set in motion a plan to allow for the registration of arbitrary TLDs, ushering in an unprecedented explosion of applications for new TLDs. At the time, only 22 generic top-level domains (gTLDs) were available for registration. Today, there are thousands of TLDs that can be used to host websites, with TLDs as varied as those in the title of this paper.

Chairman of ICANN's board of directors Peter Thrush said this of the decision to ease restrictions on gTLD proposals: "Today's decision will usher in a new internet age. We have provided a platform for the next generation of creativity and inspiration. Unless there is a good reason to restrain it, innovation should be allowed to run free." ICANN believed that companies would embrace the new changes, improving brand recognition by registering their trademarks as new TLDs. Instead, 79 companies signed a petition in protest of the new rules, arguing that their brands would be negatively affected under the new rules. Much public discourse centered on the net impact that ICANN's changes have had on companies, but what about the impact that gTLDs have on consumers?

In this paper, we investigate the negative impact that gTLDs have on the security posture of internet users. Through an online survey, we demonstrate that gTLDs bolster the success of *domain impersonation attacks*, where an attacker tries to convince a user that they own a domain by registering a visually similar domain. We also show that users are unable to distinguish gTLDs that companies own from those they do not own, calling into question the supposed positive benefits of gTLDs improving brand recognition. Using a longitudinal dataset, we show that the number of impersonation attacks using gTLDs is growing, and discuss what factors attackers consider when deciding wither to use a gTLD, country-code TLD (ccTLD), or "common TLD" (.com, .net, or .org) for their attack. Finally, we close with recommendations for the major players of the TLD ecosystem: ICANN, security researchers/practitioners, and companies, toward the unified goal of protecting consumers online.

II. Background and Related Work

Proposing a New gTLD ICANN provides an overview of all procedures pertaining to new gTLD proposals in their "New gTLD Applicant Guidebook" [1]. A proposing organization must demonstrate the technical and financial ability to run a compliant registry for their gTLD. The applicant must also pay a minimum $185,000 application fee. Applications progress through the following stages: Administrative Check, Initial Evaluation, Extended Evaluation, String Contention, Dispute Resolution, and Pre-delegation. An objection period solicits public comment on proposed gTLDs, allowing members of the public to file objections on several grounds: "string confusion," "legal rights," "community," or "limited public interest." ICANN advises that the entire process can take anywhere from 9 to 20 months, depending on what issues are raised during the objection period. If the proposal is successful, the applicant has full control over deciding who may register their new gTLD. Registration for some gTLDs is restricted to entities who can prove they manage an organization meeting a regulatory standard for that TLD. For example, all domains requesting a .pharmacy TLD must be a licensed pharmacy as vetted by the National Association of Boards of Pharmacy [2]. Most gTLDs however allow for anyone to register a domain, regardless of its relevance to the TLD.

Negative Impact of gTLDs Some research has been conducted on the ethical, legal, and financial impact that individual gTLDs have had on the internet ecosystem. Proposals for the .health gTLD were met with debates about how the TLD could lend credibility to unsubstantiated health claims, and whether registration of domains with this TLD should be restricted for the public good [3], [4], [5], [6], [7]. In 2014, Halvorson et al. estimated that at most, 3.8% of domains registered with a .xxx TLD were hosting related content, with the rest serving as speculative registrations or defensive registrations intended to protect one's brand or personal reputation [8]. Similar measurements by Halvorson et al. in 2015 estimated that 15% of domains registered with new gTLDs showed patterns consistent with primary, and not speculative or defensive, registration [9].

Less research has been conducted on the security impact of new gTLDs. Chen et al. demonstrated that new gTLDs opened up users to man-in-the-middle attacks due to the leakage of Web Proxy Auto-Discovery (WPAD) queries [10]. Our paper differs from this and other prior work by taking a broader view of the security impact that gTLDs have on online consumers, providing an empirical foundation towards understanding the impact these TLDs have on users. Through an online survey, we measure the security impact that new gTLDs have when directly attacking the end user's perception and technical comprehension of domain names.

Domain Impersonation Attackers use many techniques to impersonate valuable websites, fooling users into divulging private information and taking risky actions. Many techniques fall under the umbrella of *domain impersonation*, where an attacker registers a domain name that is somehow similar to the target website's domain. Examples of domain impersonation include: *typosquatting* [11], [12], [13], [14], where an attacker obtains a domain with slight alterations to the target domain (amzon.com); *combosquatting* [15], where an attacker registers a domain that has additional tokens in its effective second-level domain (e2LD), the token to the immediate left of the TLD (amazon-sale.com); *TLD spoofing* [16], where an attacker registers their target's e2LD with a TLD that the target company does not own (amazon.pharmacy); and *target-embedding* [17], where an attacker uses subdomains to have the target domain (including its regular TLD) appear unaltered in a domain they control (amazon.com-deals.vip). Roberts et al. [17] performed a user study that showed that, among these, target-embedding is the most effective at tricking users.

While gTLDs can be used with any form of domain impersonation, they are particularly synergistic with TLD spoofing and target-embedding. More available gTLDs strain company budgets and resources, and allow more opportunities for attackers to find a TLD that a company was unwilling, unable, or forgot to register with their brand. gTLDs can also make it harder for users to comprehend the structure of domains in target-embedding attacks. Target-embedding domains are designed to make users misidentify what token is a domain's true TLD; users may be more likely to believe a familiar token such as .com is a domain's true TLD even if it is located in the wrong spot.

Target+e2LD	gTLD		Country TLD		Common TLD	
microsoft.com-security	.center	38%	.gq	20%	.org	29%
nytimes.com-daily	.news	39%	.cf	22%	.com	32%
facebook.com-login	.page	41%	.in	26%	.net	24%
amazon.com-flashsale	.store	25%	.tk	16%	.net	16%
apple.com-findmyiphone	.support	39%	.ru	16%	.com	35%
ebay.com-item	.bid	31%	.ml	22%	.com	29%
airbnb.com-request-booking	.online	34%	.us	43%	.org	27%
google.com-message	.info	26%	.me	20%	.net	19%
bankofamerica.com-account-verification	.link	33%	.ga	25%	.org	24%
Overall Percentage		34%		23%		26%

TABLE I: Impersonating domains used in our user study. Each comprised a target as a subdomain of an e2LD, followed by a varying TLD, resulting in 27 total questions for the user in this half of the study. Next to each TLD is the percentage of survey respondents who said that they believed the domain ending in that TLD belonged to the target organization (none of which actually did).

In this paper, we explore how gTLD use affects users' abilities to detect TLD spoofing and target-embedding attacks.

III. Survey Design

In this section, we describe the design of our two-part user survey to assess whether gTLD usage improves the effectiveness of (1) target-embedding and (2) TLD spoofing attacks. At a high level, we presented users with various domain names (e.g., microsoft.com-security.center) and asked them if they truly belonged to the organization they appeared to be (Microsoft, in this case). Note that we are measuring whether an attentive, motivated user is able to detect domain impersonation. We are not measuring the end-to-end efficacy of such attacks; rather, we want to know whether gTLDs negatively affect users' ability to comprehend domain names. Since the users have been motivated to look for discrepancies, we are measuring a lower bound on user susceptibility.

Participants (N=249) were recruited from Amazon's Mechanical Turk platform, and limited to subjects from the United States (and territories) who were at at least 18 years of age and had an MTurk HIT rate over 95%. They were compensated $2 for an expected 15 minutes task. Our study was approved by an Institutional Review Board (IRB). No impersonating domains were hosting content at the time of the survey's publication, though as a precaution participants were told not to visit any of the domains they were shown.

A. Target-Embedding

The first part of our survey was designed to measure the impact that gTLDs have on target-embedding attacks, when compared to common TLDs (.com, .net, or .org) and country code TLDs (.us, .ml, .tk, etc.). We began with 9 e2LDs from target-embedding attacks seen in the wild, and selected a target website that each was suited to impersonate. For each of the 9 target+e2LD pairs, we selected three TLDs: one

gTLD with semantic relevance to the target, one ccTLD, and one common TLD. Country-code and common TLDs were assigned randomly. The 27 resulting fully-qualified domain names (FQDNs) can be found in Table I.

Participants were shown each of the 27 domains alongside the name of the company or organization that domain was targeting. They were instructed to answer "Yes" or "No" to the question: "Do you believe this is the organization's URL?" The 27 questions were shown in a random order to mitigate ordering effects.

B. TLD Spoofing

For each of the 9 organizations in part one, we showed three domains in part two: the organization's primary TLD (all were .com), the organization's e2LD with a gTLD that the organization controlled, and the e2LD with a gTLD that the organization did not control. These 27 FQDNs are shown in Table II. Participants were shown a triplet of domains at a time, all three with the same e2LD. For each domain, they were again asked to answer Yes/No to "Do you believe this is the organization's URL?" The order that the triplets were shown, and the order that the 3 domains within each triplet were shown, were both randomized. Participants were told that for each triplet, there may be more than one correct answer.

We used whois data, registrar information, and DNS information to determine whether or not a domain was controlled by the organization. In some cases, the controlled gTLDs redirected to the domain with the primary TLD. In others, registrars allowed organizations to block the registration of their e2LD on different TLDs. No content is hosted at those domains, but the organization is responsible for preventing the domain's use and so they still control the domain.

IV. Survey Results

A. Targe-Embedding

With the first part of the survey, we wanted to know whether users were more likely to fall for target-embedding attacks if they use a semantically relevant gTLD, compared to a country-code TLD or a common TLD. Table I shows the percentage of respondents who incorrectly said they believed a domain was owned by the organization that the domain was actually impersonating.

We perform pairwise chi-squared tests (with $\alpha = 0.0167$ after applying the Bonferroni Correction) to determine if significantly more errors were made using one type of TLD compared to another. Our results show that more errors are made with gTLDs than both common TLDs ($p < .0001$) and ccTLDs ($p < .0001$). Surprisingly, more errors are also made when using common TLDs than ccTLDs ($p < .002$). We speculate that users' lack of familiarity with some country-code TLDs may raise red flags that "something in this domain does not belong," a concern that the presence of .com, .org, or .net would not raise.

From our survey results, we can see the emergence of three different types of users. 94 users correctly identified every question they responded to in part one of the survey.

e2LD	Primary TLD		Controlled gTLD		Not controlled gTLD	
microsoft	.com	98%	.vodka	6%	.secure	27%
nytimes	.com	98%	.biz	16%	.fm	8%
facebook	.com	98%	.sucks	6%	.bank	8%
amazon	.com	98%	.party	10%	.pharmacy	13%
apple	.com	99%	.help	25%	.safe	8%
ebay	.com	98%	.bid	20%	.reviews	20%
airbnb	.com	98%	.business	14%	.review	19%
google	.com	99%	.observer	9%	.winners	6%
bankofamerica	.com	98%	.zone	8%	.rugby	5%
Overall Percentage		98%		13%		13%

TABLE II: TLD-testing URLs in our user study. Each question included a domain name, which comprised an e2LD of a popular organization followed by either its primary TLD (.com in all cases), a gTLD the organization controls, or a gTLD it does not control. This resulted in 27 total questions for the user in this half of the study. Next to each TLD is the percentage of survey respondents who said that they believed the domain ending in that TLD belonged to the target organization. Those in the "Not Controlled" column did *not* belong to the given organizations.

These subjects are sufficiently skeptical enough to avoid target-embedding attacks as long as they vigilantly inspect their browser's URL bar while browsing the internet. Next, 14 respondents answered "Yes" to all 27 questions, trusting every target-embedding domain in our survey. We speculate that these users' mental models of the structure and usage of domains may be fundamentally flawed, causing them to trust any domain that includes their expected target website somewhere within the domain. The majority of subjects (141) were fooled some, but not all, of the time. The median user in this group was fooled by 5 domains with semantically relevant gTLDs, compared to 4 domains with common TLDs and 3 with ccTLDs. We speculate that these users make judgements based on instinct more than concrete technical knowledge, and have partially accurate but incomplete mental models surrounding domain impersonation.

B. TLD Spoofing

With part two, we want to see if users behave differently when shown a company's e2LD and a gTLD that the company owns versus a gTLD that they do not own. Ideally, users should express a belief of ownership more often for TLDs that the company owns than those it does not. The percentage of users who said they believe each domain was owned by the associated company can be found in Table II.

Comparing the controlled and non-controlled gTLD responses, we see that users are not able to discern whether a company controls a given gTLD for their brand. In both categories, users on average responded that 13% of the gTLDs were owned by the given organization (which is equivalent to 87% success rate over the controlled-gTLD category and 13% success rate over the non-controlled category). User *behavior* is the same regardless of whether the company actually owns a gTLD.

Looking at the gTLDs which users most commonly believed belonged to the given organization, we note that they tend to be highly relevant to the organization. The top five are: microsoft.secure (27%), apple.help (25%), ebay.bid (20%), ebay.reviews (20%), and airbnb.review (19%). Conversely, the five domains the users least believed to belong to the given organization have no obvious connection: nytimes.fm (8%), facebook.sucks (6%), google.winners (6%), microsoft.vodka (6%), and bankofamerica.rugby (5%). These results indicate that relevance to the organization appears to be a strong predictor for whether users will believe a gTLD belongs to that organization. Unfortunately, relevance is not necessarily a predictor for whether an organization registers under a given gTLD, nor is it a prerequisite for obtaining most gTLDs.

A false negative occurs when a company owns a domain with a gTLD, but the user does not believe the company owns it. A suspicious user presumably would leave the benign website in search of a domain with another TLD that they actually trust. A false positive occurs when a user believes a company owns a domain when they actually do not. This opens the user up to more severe attacks, like phishing. Both false positives and false negatives affect users' ability to know who they are communicating with online. These results may feel "obvious" but that makes the situation even more alarming; if it is obvious to us that users have a hard time comprehending gTLDs, surely ICANN must have thought of these concerns at some point as well, yet pushed forward with easing restrictions on gTLD applications. In summary, gTLDs confuse users. Next, we measure whether attackers have been exploiting this user confusion.

V. Longitudinal Analysis of TLD Usage in Target Embedding

We now look at how TLD use in target-embedding attacks has changed over time. We use the target-embedding dataset provided by Roberts et al. [17] to collect a list of certificates with target-embedding domains collected from public Certificate Transparency (CT) logs. We split domains into three categories based on what real TLD the domain has: a two-letter country-code TLD, a common TLD (.com, .org, or .net), or a gTLD that is an English word (excluding "net" and all two-letter country-codes). We determine the first time that each domain appeared on a TLS certificate by finding the certificate with the earliest issuance date. Figure 1 plots the cumulative number of unique domains seen up to each day, ending on the date of data collection (May 18, 2019).

The number of target-embedding domains is growing steadily for each category of TLD. Adoption of gTLDs that are English words appears to be growing at a slower rate than both common and country TLDs. This begs the question: if gTLDs with semantic relevance to their targets are more effective at fooling users, why aren't attackers utilizing them more? Why are country-code TLDs the most popular for attacks, when they are also the least effective?

The top five most popular ccTLDs for target-embedding (.cf, .ga, .ml, .tk, and .gq) all allow for free domain registration,

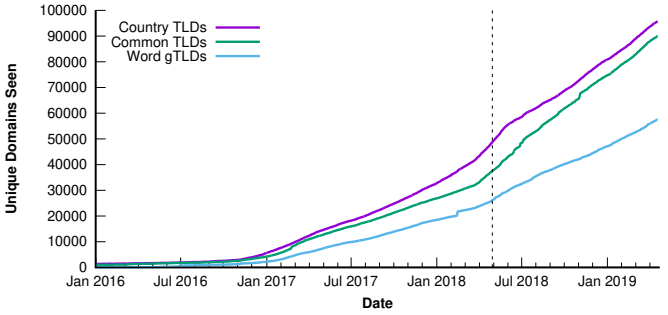

Fig. 1: Cumulative number of unique target-embedding domains seen up to each date in CT logs, separated by the type of TLD the domain has. Google Chrome requires all certificates issued after April 30, 2018 (noted by the vertical black line) to be included in CT logs.

and are flagged by Spamhaus and the Anti-Phishing Working Group as some of the most abused TLDs for spam and phishing [18], [19]. Attackers are faced with an economic trade-off: use a free TLD that will fool fewer users, or spend money for a domain with a more persuasive gTLD. These measurements show that attackers approach this trade-off differently. Recall from Section IV-A that a subpopulation of users will likely be deceived regardless of what type of TLD an attacker uses; some attackers are content to use free ccTLDs to target these users for minimal cost. Other attackers are willing to pay to cast a wider net, and register domains with gTLDs in order to target users who use their instincts to make judgements about domains on a case-by-case basis.

The fact that common TLDs are still so popular in target-embedding attacks, despite costing money and being less effective than other gTLDs, merits further investigation. Attackers may be swayed by other non-economic and non-user-based incentives which are outside the scope of this paper.

VI. Discussion

Section IV showed that many users lack the technical knowledge to make correct decisions when presented with impersonating domains, especially gTLDs. One subject commented at the conclusion of our survey that "I thought the URL's that had .com twice or .com then .net were fake URL's. Those were the only ones I felt fairly certain about." In Section V, we demonstrated that attackers are increasingly making use of gTLDs (and other TLDs) in their attempts to manipulate users. In this section, we discuss positive steps that various players in the TLD ecosystem can take to reduce the harm that ultimately falls on the shoulders of consumers.

ICANN ICANN solicits feedback on new gTLDs during a public objection period before accepting them. Unfortunately, consumer security and online safety are discussed infrequently in these proposals when compared to other reasons for rejection. Objections on safety grounds would fall under String Confusion Grounds in ICANN's objection model, meaning a string "is confusingly similar to an existing TLD or to

another applied-for gTLD string - delegating two or more similar TLDs could cause user confusion." [1] Of the 12,834 available public comments, 186 were filed as String Confusion objections [20]. Only 23 of those objections cite security concerns, on applications for .home (12), .corp (6), .mobile (4), and .zip (1). String Confusion is a broad category that encapsulates all concerns of string similarity. We feel that user safety is important enough to merit its own objection category. We recommend ICANN recognize online safety as a top-level priority, and encourage the public to scrutinize potential gTLDs' capability for abuse by impersonators and attackers.

Our survey demonstrated that gTLDs with relevance to their target website are especially effective at tricking users. Some gTLDs are relevant to a small number of websites, such as .bid with auction sites like ebay.com or .news and news sources like nytimes.com. However, there are some gTLDs that are relevant to *all online activities*. These gTLDs are especially problematic as they are related to the act of internet browsing itself, and could be used to target virtually any website on the internet. Examples of these include .click, .company, .computer, .contact, .download, .help, .info, .link, .online, .page, .protection, .safe, .secure, .security, .services, .site, .systems, .tech, .trust, and .website. ICANN should be exceptionally critical toward generic TLDs that are so generic they are relevant to nearly any company or website, *especially* words that could be used to convince consumers that their connection is more secure than it actually is.

Security Researchers and Practitioners Our results identified three categories of users; those who aren't tricked when actively engaged, those who misidentify all target-embedding domains, and those who can identify some attacks but fall prey to more sophisticated forms of impersonation. Security researchers and practitioners should take these groups into account when designing user protection mechanisms. We recommend the design of tools that make use of users' underlying mental models to help them reach the safest correct outcome. Instead of searching for one-size-fits-all solutions, we recommend a multi-pronged approach with individual complementary techniques designed to target each group's needs. Users who are fooled by many forms of impersonation would benefit from hard interventions, whereas users capable of identifying improper domains should be guided to use their knowledge by checking their URL bar. Further work is needed to understand the mental models of those who rely on instinct and make their judgements on a case-by-case basis.

Security researchers should also stay up to date on what new gTLDs are proposed, and offer objections when they see fit. We recommend monitoring proposals for strings that have semantic relevance to many websites, or to websites which are known to be frequently targeted in other impersonation attacks. We also recommend the development of a standardized method to measure the impact that a proposed gTLD may have on user perception. A standardized battery of survey questions, including questions similar to those found in our survey, would help evaluate the impact of new gTLDs compared to existing ones and determine whether proposed gTLDs would ultimately cause more public harm than good.

Companies Companies have already had to deal with TLD spoofing and domain squatters trying to capitalize on their branding. Today's imperfect solution is for companies to purchase a new domain for every new gTLD that is approved.[2] ICANN offers sunrise periods where these companies can buy domains before the general public, but this approach can be costly. Not all companies can afford to constantly spend money on defensive domain registrations that they have no intention of using. Our results show that companies on a budget should prioritize registering gTLDs that have relevance to their product or service. Some registrars like Donuts Inc. offer reservation packages that allow companies to prevent registrations of their e2LD from others, at a cheaper price than registering all domains for active use [22]. Unfortunately, there is little a company can do to prevent target-embedding. Like security researchers, companies can and should participate actively during ICANN's public objection periods on new gTLDs. If a company notices a proposed gTLD that is relevant to their business, they should file an objection on String Confusion Grounds.

VII. Conclusion

Innovation sometimes has unforeseen costs. While companies have taken advantage of new creative opportunities provided by the rapid explosion of new generic top-level domains, so have attackers. In this paper, we showed multiple ways in which gTLDs can be harmful to users. We designed and ran an online survey which showed that gTLDs allowed for more effective domain impersonation attacks and resulted in a higher number of misidentified websites than common TLDs or country-code TLDs. Our survey also showed that users are unable to differentiate between gTLDs that a company owns and those the company does not own, calling into question supposed positive benefit that gTLDs have for brand recognition. Our survey revealed three different kinds of users: those who are able to correctly identify attempts at domain impersonation, those who are unable to identify any impersonating domains, and a group in the middle who can sometimes catch domain impersonation but often fall victim to more sophisticated attacks. We analyzed a longitudinal dataset to show how different categories of TLDs were utilized in traget-embedding attacks in the wild, and discussed some of the practical trade-offs attackers make when deciding what TLD to use to register impersonating domains. Sadly, there is no one clear fix to the problems that gTLDs pose for internet security. Instead, we outlined several steps that companies, security researchers, and ICANN can take in order to better

[1] We note that ICANN has approved many TLDs that are very similar to one another, including .accountant(s), .auto(s), .career(s), .coupon(s), .deal(s), .fan(s), .game(s), .gift(s), .loan(s), .new(s), .review(s), and .work(s).

[2] Some have argued that this is what motivated ICANN to add more gTLDs in the first place [21].

protect consumers from the malicious use of generic top-level domains.

REFERENCES

[1] The Internet Corporation for Assigned Names and Numbers, "New gtld applicant guidebook," https://newgtlds.icann.org/en/applicants/agb.

[2] National Association of Boards of Pharmacy, ".pharmacy verified websites, program eligibility and policies," https://nabp.pharmacy/programs/dotpharmacy/standards.

[3] T. K. Mackey, G. Eysenbach, B. A. Liang, J. C. Kohler, A. Geissbuhler, and A. Attaran, "A call for a moratorium on the .health generic top-level domain: preventing the commercialization and exclusive control of online health information," *Globalization and Health*, vol. 10, no. 1, p. 62, 2014.

[4] A. E. Solomonides, "ICANN, Health Information and the "Dot Health" Top Level Domain," in *2014 IEEE 27th International Symposium on Computer-Based Medical Systems*. IEEE, 2014, pp. 460–462.

[5] G. Eysenbach, "The new health-related top-level domains are coming: will cureforcancer .health go to the highest bidder?" *Journal of Medical Internet Research*, vol. 16, no. 3, p. e73, 2014.

[6] T. K. Mackey, B. A. Liang, J. C. Kohler, and A. Attaran, "Health domains for sale: the need for global health internet governance," *Journal of Medical Internet Research*, vol. 16, no. 3, p. e62, 2014.

[7] J. B. Walther, Z. Wang, and T. Loh, "The effect of top-level domains and advertisements on health web site credibility," *Journal of Medical Internet Research*, vol. 6, no. 3, p. e24, 2004.

[8] T. Halvorson, K. Levchenko, S. Savage, and G. M. Voelker, "XXXtortion? Inferring Registration Intent in the .XXX TLD," in *International World Wide Web Conference (WWW)*, 2014.

[9] T. Halvorson, M. F. Der, I. Foster, S. Savage, L. K. Saul, and G. M. Voelker, "From .academy to .zone: An analysis of the New TLD Land Rush," in *ACM Internet Measurement Conference (IMC)*, 2015.

[10] Q. A. Chen, E. Osterweil, M. Thomas, and Z. M. Mao, "MitM Attack by Name Collision: Cause Analysis and Vulnerability Assessment in the New gTLD Era," in *IEEE Symposium on Security and Privacy*, 2016.

[11] Y.-M. Wang, D. Beck, J. Wang, C. Verbowski, and B. Daniels, "Strider typo-patrol: Discovery and analysis of systematic typo-squatting," in *USENIX Workshop on Steps to Reducing Unwanted Traffic on the Internet (SRUTI)*, 2006.

[12] J. Szurdi, B. Kocso, G. Cseh, J. Spring, M. Felegyhazi, and C. Kanich, "The long "taile" of typosquatting domain names," in *USENIX Security Symposium*, 2014.

[13] T. Moore and B. Edelman, "Measuring the perpetrators and funders of typosquatting," in *Financial Cryptography (FC)*, 2010.

[14] P. Agten, W. Joosen, F. Piessens, and N. Nikiforakis, "Seven months' worth of mistakes: A longitudinal study of typosquatting abuse," in *Network and Distributed System Security Symposium (NDSS)*, 2015.

[15] P. Kintis, N. Miramirkhani, C. Lever, Y. Chen, R. Romero-Gómez, N. Pitropakis, N. Nikiforakis, and M. Antonakakis, "Hiding in plain sight: A longitudinal study of combosquatting abuse," in *ACM Conference on Computer and Communications Security (CCS)*, 2017.

[16] A. Banerjee, D. Barman, M. Faloutsos, and L. N. Bhuyan, "Cyber-fraud is One Typo Away," in *IEEE Conference on Computer Communications (INFOCOM)*, 2008.

[17] R. Roberts, Y. Goldschlag, R. Walter, T. Chung, A. Mislove, and D. Levin, "You Are Who You Appear to Be: A Longitudinal Study of Domain Impersonation in TLS Certificates," in *ACM Conference on Computer and Communications Security (CCS)*, 2019.

[18] "The spamhaus project: The world's most abused tlds," https://www.spamhaus.org/statistics/tlds/, 2018.

[19] Anti-Phishing Working Group, "Phishing attack trends report – 4th Quarter, 2018," https://docs.apwg.org/reports/apwg_trends_report_q4_2018.pdf.

[20] The Internet Corporation for Assigned Names and Numbers, "Application comment forum," https://gtldcomment.icann.org/applicationcomment/viewcomments.

[21] Stilgherrian, "New top-level domains a money grab and a mistake: Paul Vixie," https://www.zdnet.com/article/new-top-level-domains-a-money-grab-and-a-mistake-paul-vixie/.

[22] Donuts Inc., "Enhanced brand protection," https://donuts.domains/what-we-do/brand-protection.