Characterizing the Trust Dilemma: Comparing Web Security of Malicious and Benign Domains

Rebecka Lindkvist*, Linn Petersson*, Carl Magnus Bruhner*, David Hasselquist*†, Martin Arlitt‡, Niklas Carlsson*
*Linköping University, Sweden †Sectra Communications, Sweden ‡Corelight, Canada

***Abstract*—While much measurement research has highlighted the increased adoption of HTTPS and better security practices, the trustworthiness of a domain is not solely determined by the security properties of its end-to-end connections. To separate and highlight this distinction, we present a comprehensive analysis of secure communication practices across different categories of web domains, placing particular focus on the direct comparison of four classes of "malicious" domains (*fake news*, *spam*, *phishing*, and *malware*) with domains selected along three dimensions: domain popularity, service type, and geographical registration. For each domain class, we present a high-level characterization, focusing on fundamental aspects that influence connection security. Through a comparative analysis of default protocols, cipher suites, certificate authorities (CAs), certificate classifications, and vulnerability susceptibility, we illuminate disparities in the implementation of different domains' secure communications. Our findings showcase several shortcomings in safety standards among many "benign" domains and show that *fake news* and *malware* domains often resemble benign domains in their security practices. In contrast, *phishing* domains exhibit a higher level of security awareness, while *spam* domains tend to employ weaker security measures. Overall, our findings underscore the critical need to disentangle secure connections from domain trustworthiness, advocating for clearer distinctions to help safeguard Internet users and enhance Internet security.**

I. Introduction

The World Wide Web consists of a myriad of domains, serving different purposes and being operated by a wide range of both legitimate and malicious actors. With modern browsers pushing all websites to implement HTTPS and the relative ease today of obtaining X.509 certificates, it is perhaps not surprising that most domains today use HTTPS, regardless of being considered trustworthy or not. This has resulted in a trust dilemma, in which the least trustworthy websites may implement more "secure" communication than the more reputable websites.

In this paper, we present a comprehensive measurement-based comparison of four classes of malicious domains (*fake news*, *malware*, *phishing*, and *spam*) against various classes of popular domains, each selected to provide clear reference points and allow insights to be drawn with regard to their relative differences and similarities in their selection of various security related features (e.g., cipher suites, private key properties, and vulnerability to various known attacks). First, our selection of malicious domains is based on data from reputable sources (e.g., FakeNewsNet [1], URLhaus [2], PhishTank [3], and Wein [4]) and our selection of representative benign domains is defined to capture comparisons along three dimensions: continent-based splits, domain popularity, and service category. Second, our data collection methodology involved employing three distinct tools: a Selenium-based collection tool, a site parser, and the *testssl.sh* [5] command line tool. These tools enabled us to extract a multitude of connection-related statistics, certificate information, and security-related aspects from the visited domains.

Our analysis reveals intriguing patterns across different domain categories. While *fake news* and *malware* domains exhibit similarities to benign domains, *phishing* and *spam* domains emerge as outliers, showcasing distinct security profiles. *Phishing* domains, for instance, demonstrate a higher level of security awareness, whereas *spam* domains exhibit laxer security measures, despite certain outliers in both categories.

To further highlight the similarities and differences between domain categories, we perform a principal component analysis (PCA). This analysis highlights the unique positioning of *phishing* and *spam* domains as outliers, reinforcing their distinct characteristics compared to other domains. Additionally, we observe trends indicating similarities between certain domain categories and less popular domains, underscoring the nuanced relationships between domain types.

Overall, our measurement-based comparison offers valuable insights into the diverse landscape of malicious and benign domains. By analyzing security profiles, performance metrics, and behavioral patterns, we provide a comprehensive understanding of the distinct characteristics exhibited by each domain category. These findings act as a reminder of the importance of separating the use of secure communication practices from the trustworthiness of a domain. Furthermore, by showcasing differences between the different categories, including that untrustworthy domains often use up-to-date security, we provide concrete examples highlighting that security measures alone should not be used as indicators of trustworthiness. Building on these findings and examples, we advocate for clearer distinctions to enhance user protection and strengthen Internet security.

Outline: We next present our dataset and data collection methodology (Section II), followed by a high-level, comparative analysis (Section III) and a PCA-based analysis (Section IV) to further visualize and compare observed category differences and similarities. Finally, we present related works (Section V) and our conclusions (Section VI).

II. Datasets and Collection Methodology

For our analysis, we first collected data for four classes of malicious domains (Section II-A) and many classes of popular domains (Section II-B) using three different data collection methodologies (Section II-C).

A. Malicious Domain Categories

We compare four classes of domain categories that we in this paper consider malicious: (1) *fake news*, (2) *malware*, (3) *phishing*, and (4) *spam*. While this is not an exhaustive list of domain categories that can be considered "malicious" or harmful (other examples would be illegal content, scams, and pharming), we limit the scope to these four classes to better highlight examples of such domains. To account for differences in the availability of domains in each category and allow a relatively balanced analysis, we selected 500 domains for each class (slightly less than the maximum found for the least prevalent class); resulting in a combined set of 2,000 malicious domains. For *fake news* domains, we selected the first 500 domains listed on FakeNewsNet [1]. For *malware* domains, we selected the first 500 active *malware* domains from URLhaus [2]. For *phishing* domains, we selected the top 500 domains using HTTPS reported on PhishTank [3] that were active at the time of the study. Finally, for *spam* domains, we selected the 500 most common email domains listed by Wein [4] that were active at the time of the study. We acknowledge that there may be overlaps between categories, false positives [6], day-to-day variations [7], and compromised (sub)domains [8], and that many other sources can be considered. However, since the domain lists are from reputable sources, we use them as is for this study.

B. Popular Domain Categories

To provide several dimensions of comparison, we collected domains and split categories along three dimensions: (1) the most popular domains on each continent, (2) the global popularity of each domain, and (3) the category of each domain. In total, the different sets of "benign" domains combine to 6,080 domains.

Continent-based split: For each continent, we first select the four largest countries by gross domestic product and then up to 100 most visited website domains for each of these countries, as measured and reported by Netcraft [9] (on March 8, 2022). For countries with fewer than 100 domains, we include all available domains. Our continent dataset contains 1,530 domains with the following splits: Africa (AF) includes Nigeria (4), Egypt (16), South Africa (59), and Algeria (5). Europe (EU) includes Germany (100), United Kingdom (70), France (100), and Italy (99). Asia (AS) includes China (100), India (100), Japan (99), and Russia (100). North America (NA) includes the United States (100), Canada (100), Mexico (41), and Puerto Rico (3). South America (SA) includes Brazil (98), Argentina (98), Colombia (97), and Venezuela (17). Finally, Oceania (OC) includes Australia (99) and New Zealand (25).

Popularity-based comparisons: To compare with domains of different popularity ranges, we used three of the most popular top-1M lists (from March 22, 2022): (1) Tranco [10], (2) Alexa [11], and (3) Majestic [11]. Specifically, for each of the three top lists, we select the first 250 domains (ranks 1–250) and the last 250 domains from each magnitude sample ($[*, x]$, i.e., ranks 751–1,000, 9,751–10,000, 99,751–100,000, and 999,751–1,000,000), resulting in five popularity ranges per list, or 3,750 domains in total. We note that the lists (and hence domain selections) are assembled with different and complementing methodologies, enabling complementing comparisons of popular vs. non-popular domains (along the different methods), for example.

Figure 1 provides an overview of the rank spans (using a boxplot) for each domain category considered as per the three rankings. While the rankings for the subsets using each of the rankings themselves show a clear trend (as expected per definition), there are bigger variations when comparing across even the popularity-based selections. We also note that almost none of the *phishing* and *malware* domains are on the top-1M lists, while a noticeable subset (red areas in pie-charts) of the *fake news* and *spam* domains are, where the subsets are the largest for the Tranco and Majestic lists.

Category-based comparisons: Finally, we compare with sets of domains that represent different service categories. For this, we selected the top-50 domains from each of the following Alexa top categories (obtained on September 21, 2020, and available until near the end of 2020): Adult, Arts, Business, Computers, Games, Health, Home, Kids & Teens, News, Recreation, Reference, Regional, Science, Shopping, Society, and Sports. This results in an additional 800 domains.

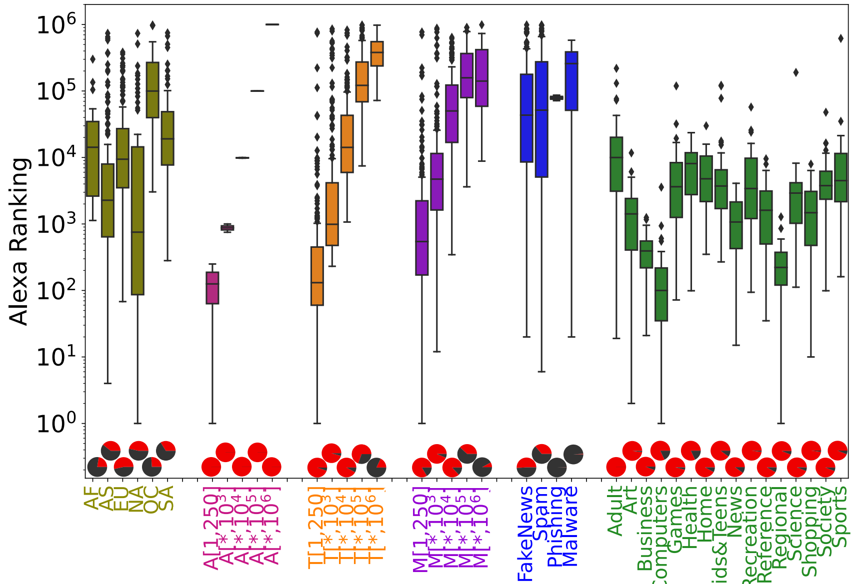
C. Collection Methodology

For the data collection, we used three different tools: a self-developed Selenium-based collection tool, a site parser tool, and a TLS/SSL test tool.

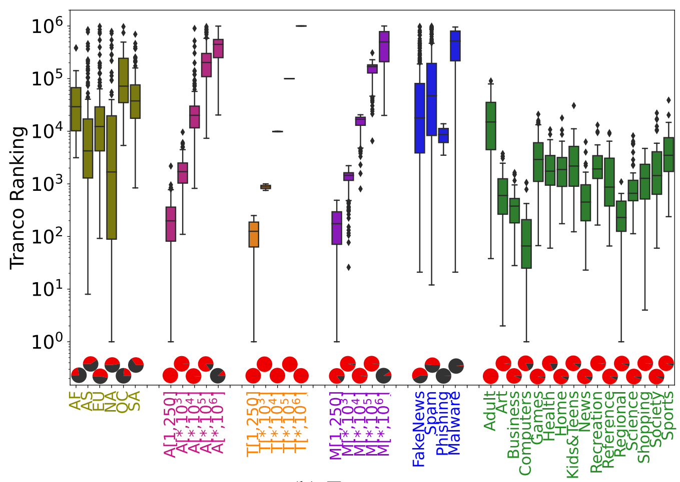
Selenium-based tool: We used our Selenium-based tool to extract various connection-related statistics, loading times, and information about the negotiated ciphers and protocols used. In total, the tool collects 38 pieces of information per domain and visit. For our experiments, we use the tool together with the Chrome browser and each domain was visited five times.

Site parser: The site parser tool, developed for revocation analysis [12], collects certificate information, including the issuing Certificate Authority (CA) and the certificate type (e.g., DV, OV, or EV). The tool gathers certificates in .pem format from the visited domains, converts them to a .txt file per certificate, and then merges the certificates into one file containing 23 columns of information per domain (and certificate). To allow parallel processing, the site parser uses GNU parallel [13] and OpenSSL [14]. We again visit each domain five times.

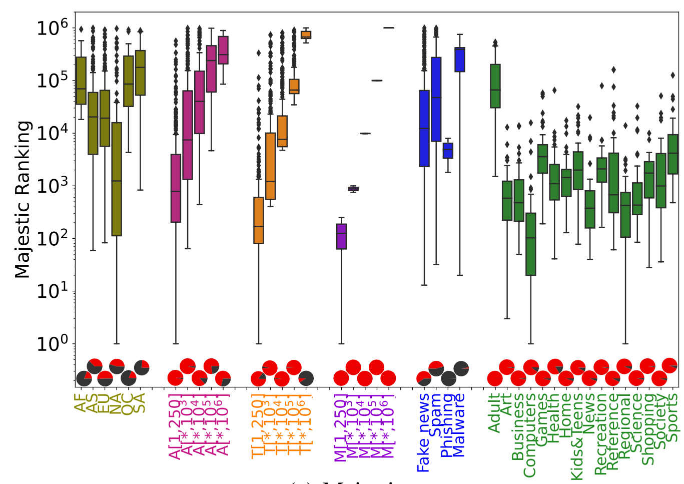
Testssl.sh script, known vulnerabilities, and security classifications: Finally, to gather more information on security-related aspects of each domain, including negotiated ciphers and protocols, and susceptibility to known vulnerabilities, we used the testssl.sh command line tool [5]. This is a



(a) Alexa



(b) Tranco



(c) Majestic

Fig. 1: Domain ranking for all considered domain categories, where the rankings are determined using each of the top-1M lists. Here, for domains where we find a ranking, we show boxplots and pie charts contrasting ranked and total domains. The interval $L[*, x]$ denotes the 250 positions ending with x on the top-1M list $L = \{A, T, M\}$. To improve readability of labels, one of the plots is made bigger.

free, open-source tool that allowed us to gather information for about 900-1,300 different headers.

Using Ciphersuite [15], we classify the ciphers collected by testssl.sh as *Recommended*, *Secure*, *Weak*, or *Insecure*, where the *Recommended* category is the strongest and supports perfect forward secrecy [16]. In addition to the cipher suites, testssl.sh also analyzes the domains and determines whether they are potentially susceptible to the attacks through known vulnerabilities, of which we include some of the most recognized ones: *LUCKY13* [17], [18], *BEAST* [19], [20], *POODLE (SSL)* [21], [22], *BREACH* [23], [24], *DROWN* [25], [26], *SWEET32* [27], [28], *FREAK* [29], and *LOGJAM* [30].

Using testssl.sh, we extract the key types employed by the different domains. For domains in our dataset, we observe six different key types: *RSA-1024*, *RSA-2048*, *RSA-3072*, *RSA-4096*, *ECDSA-256*, and *ECDSA-384*. We use the security strengths estimated by NIST [31] to classify their security level as *weak*, *okay*, *good*, *better*, or *best* (limited to the observed key types). RSA-1024 is classified as the least secure, i.e., *weak*, and RSA-2048 as *okay*. Despite their different bit sizes, RSA-3072 and ECDSA-256 are considered to offer similar security, and both are classified as *good*. RSA-4096 is classified as *better*, and ECDSA-384, considered the most secure key type, is classified as *best*.

III. High-level Comparison

As an initial high-level comparison of the different domain categories, Figure 2 shows a heatmap of some examples of security-related properties (rows) for the different domain

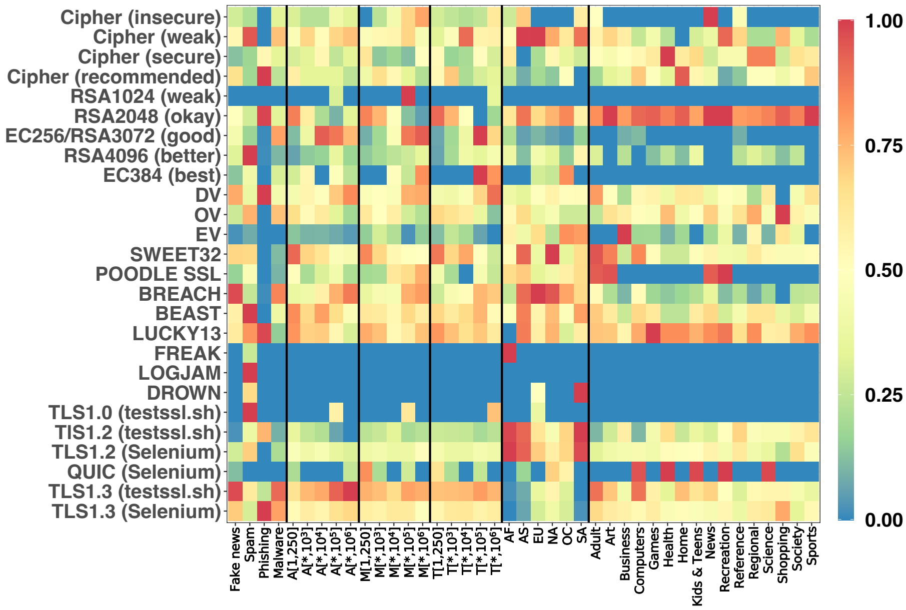


Fig. 2: A high-level comparison between domain categories using a heatmap of security-related properties (rows) across different domain categories (columns).

categories (columns). Here, all values are normalized to fall between 0 and 1, where 0 represents the smallest observed value and 1 the biggest (for a given property; i.e., row). With this choice, the blue cells represent the cases with the lowest values in that specific category (i.e., normalized value close to 0) and the red cells represent the highest value (i.e., normalized values close to 1).

Considering the different rows, we report the cipher suite negotiated (as classified), the public key (including their classification), the certificate type, the domains' vulnerability to a number of different known attacks, and the TLS version observed both with testssl.sh and our Selenium-based tool.

We note that both the *fake news* domain category and the *malware* domain category look relatively similar to most of the other domain categories (e.g., comparing these two columns to those of the different Alexa rank categories reveals only small differences). For *malware* this can in part be due to us focusing on the subset of such domains that implement HTTPS (as several such domains did not), whereas in the case of *fake news* domains, we note that these domains may aim to appear as much like benign domains as possible.

Outlier categories: We observe substantially bigger differences for *phishing* and *spam* domains, but in opposite directions. Perhaps the most obvious outlier is *spam*, which, for example, has the highest value for TLS 1.0 (most other domains have no value for TLS 1.0) and is the only category with domains that were still vulnerable to *LOGJAM*. Other vulnerabilities presented in *spam* but not considerably shown in the other domain categories were *FREAK* and *DROWN*. *FREAK* is present only in the Africa and *spam* categories, while *DROWN* is present in the Europe, South Africa, and *spam* categories. These observations all suggest that *spam* domains put very little emphasis on security. However, *spam* also stands out in the opposite direction, as it is the category with the highest use of RSA-4096 (classified as *better* strength keys). We note that the key strength most commonly used at this time was RSA-2048 (classified as *okay*).

Phishing stands out in a different way, as it sometimes is the most security-aware and up-to-date class. For example, it has the highest use of *recommended* ciphers, lowest use of *insecure* and *weak* ciphers, is the largest user of TLS 1.3, and is not vulnerable to any of the tested vulnerabilities (except LUCKY13, for which it had the largest share of vulnerable domains). The fact that *phishing* has none or the lowest values for the vulnerabilities *BEAST*, *SWEET32*, and *BREACH* are noteworthy, as almost all other categories have at least some domains that are vulnerable to these attacks. These results may capture that *phishing* domains due to their ephemeral nature often use the latest recommendations. For example, to quickly create new websites for their phishing campaign, the attackers may use services/applications that use the latest security best practices and likely do not consider backward compatibility. Not surprisingly, we also note that this category has the largest use of DV certificates; i.e., the validation type with the least validation checks (and that can be obtained for free).

Correlation between properties: When interpreting the property correlations in Figure 3, calculated using Pearson correlation, we observe that some values are naturally correlated. For example, since each connection negotiates only one cipher suite with a domain, one would expect a category with higher usage of insecure or weak ciphers (top two rows) to be negatively correlated with the use of secure and recommended ciphers (columns 3 and 4), as indicated by the red patterns for these cells in the figure.

Similarly, domain categories with higher use of weak and insecure ciphers (first two rows) are positively correlated

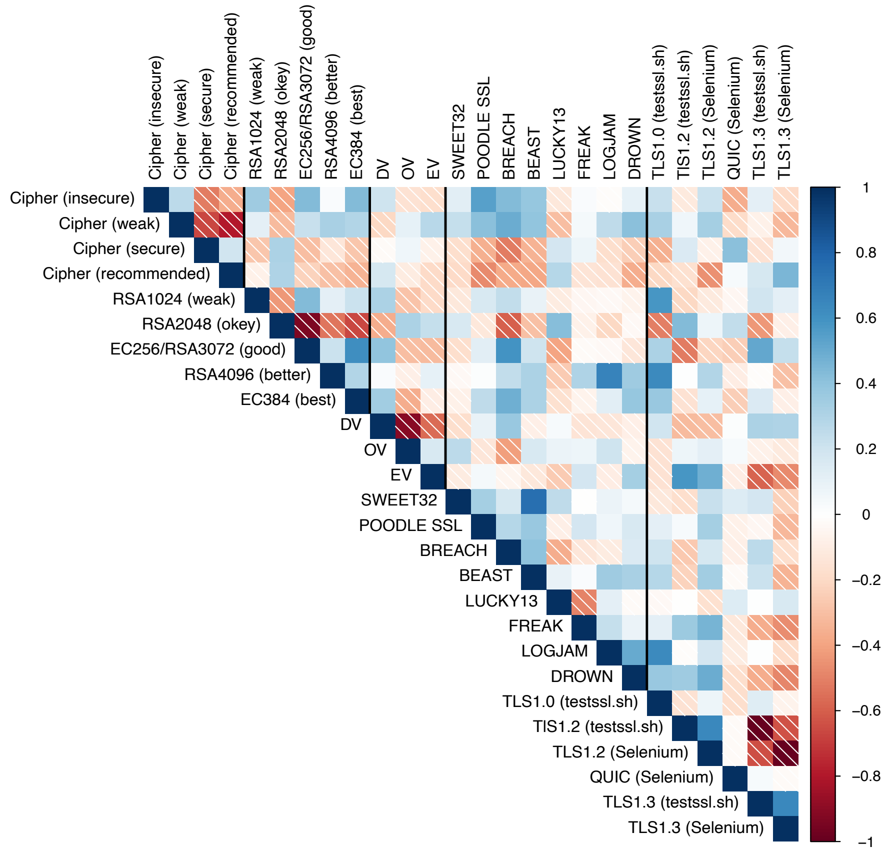


Fig. 3: Correlation plot showing how the investigated aspects relate to each other.

TABLE I: Total number of domains per CA found.

Certificate Authority (CA)	Number of Domains
DigiCert Inc	1,481
Let's Encrypt	1,428
Cloudflare, Inc.	782
Google Trust Services LLC	544
Sectigo Limited	533
GlobalSign nv-sa	486
Amazon	473
GEANT Vereniging	343
GoDaddy.com, Inc	220
Entrust, Inc.	176
Other	729

with being sensitive to most of the known vulnerabilities tested (except LUCKY13), and the reverse holds for domain categories with higher usage of secure and recommended ciphers. As expected, we also observed positive correlations between insecure/weak ciphers and TLS 1.0 usage but a negative correlation with QUIC/TLS 1.3 (Selenium).

Certificate Authority (CA): Next, we look at the issuing CAs. These results are summarized in Figure 4, where we include the top-10 CAs observed in our dataset as well as an "Other" category (aggregate of all other CAs). Table I shows the total number of domains per CA (top 10) based on the collection from the site parser. A few things stood out here. First, the most commonly used CA was DigiCert, followed by Let's Encrypt, and Cloudflare. While at first it may be surprising that Let's Encrypt does not see a bigger share, we note that this aligns with other studies showing a lower share among the most popular domains (top 100/1K/10K) compared to the top 1M overall [32], [33]. Other studies might also consider logged certificates, where numbers for Let's Encrypt are further inflated by their shorter 90-day validity times than the typically longer-validity (often up to 398-day [34]) certificates issued by DigiCert, for example.

Second, we note that the most used CA, DigiCert, is much less used for *phishing* and *malware* than for any of the other categories. Instead, Google Trust Services (GTS) and GÉANT see much higher usage among these domain classes (but less for the other domain categories). The low usage of DigiCert does to some extent match the observation that DigiCert typically is more frequently used for more popular domains according to Alexa, Majestic, and Tranco. Therefore, it is surprising that *spam* shows a relatively high usage of DigiCert-issued certificates.

Third, *phishing* stands out, as it sees much less diversity in CA usage (three dominant CAs: GTS, Let's Encrypt, and GÉANT) than the other categories, with an exceptionally high usage of GTS (75%). This may be because many *phishing* domains are operated by the same organizations, resulting in greater similarities between domains within this category. We note that similar reuse has been observed with style sheets of *phishing* domains [35].

Fourth, similar to the above analysis, the CA usage for *fake*

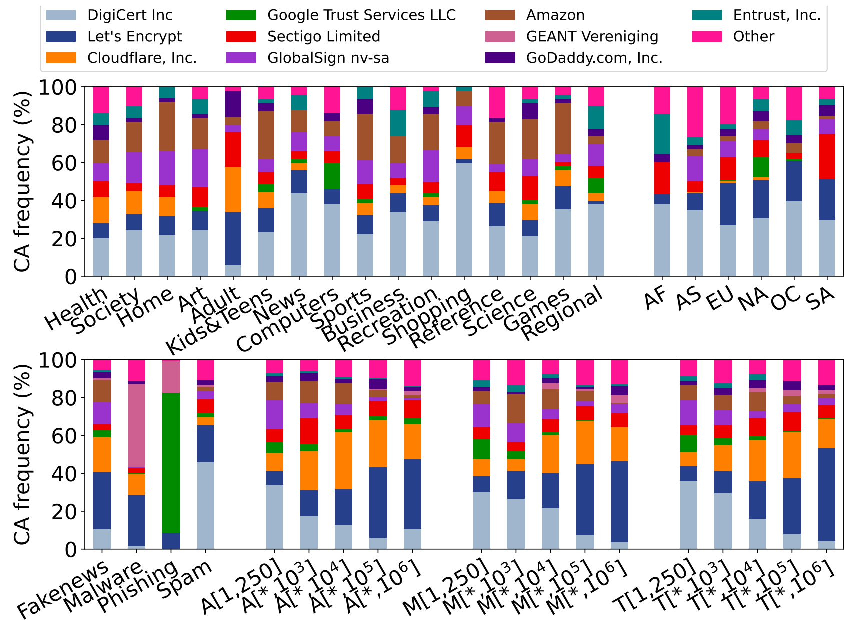


Fig. 4: CA frequency for every domain category considered (split into two rows).

news and *spam* looks relatively similar to the various "benign" domain categories, with the first resembling more that of less popular domains and the second resembling more that of more popular domains (e.g., based on DigiCert and Let's Encrypt usage).

Finally, we also note some patterns for the other categories. For example, as perhaps expected, there is a relatively higher usage of "Other" CAs for most of the non-American categories AF, AS, EU, OC). This can be explained by a higher usage of regional CAs. There are also some clear trends visible across the three popularity rankings that extend from the shift from DigiCert (most used among top-ranked domains) to Let's Encrypt (most used among the less popular domains), as also Cloudflare and the "Other" category tend to see relatively higher usage towards less popular domains, whereas GTS primarily shows up among the top-ranked domains (making the observation that GTS is highly used among the *phishing* domains particularly interesting).

IV. Principal Component Analysis

To further visualize and compare the similarities and differences between the four malicious domain categories and the other "benign" domain categories, we performed a principal component analysis (PCA). A scree plot (omitted) revealed a sharp knee after the second component: the first two principal components were responsible for 44.1% of the variations (23.6% and 20.5%, respectively), while the rest of the components were each responsible for less than 10% (e.g., third and fourth were responsible for 9.8% and 8.2%, respectively).

Motivated by this, Figure 5 shows a scatter plot of the domain categories using the first two principal components. Interestingly, *phishing* (blue diamond) and *spam* (blue square) again stand out as two clear outliers. This is consistent with what we have seen in the previous section, where these two categories showed substantially different behaviors than many of the other categories. The observation that they are at different ends of the spectrum for principal component one also matches what we observed regarding their adoption of security-related properties (*phishing* scoring high and *spam* scoring low). This suggests that this component may capture some of these aspects. A closer look at the individual terms of the principal component confirms this (e.g., the top-contributing terms of this component relate to the usage of strong keys or negotiated cipher suites, with BREACH also being a noticeable contributor).

In comparison, *fake news* and *malware* are relatively similar to the less popular domains (e.g., closest to the last 250 of the top-1M domains on the Alexa list). The placement of these domains is also consistent with the observation that the rank-based domains for the most part appear to have a "roughly" downward trend moving from left to right in the plot (regardless of the list) as we go from top-ranked (e.g., [1,250]) to less popular (e.g., [*,10^6]).

Another interesting observation here is that Europe is much further away from the top domains of Alexa, Tranco, and Majestic, than where you would expect to find it given that Europe contains many popular domains. Some of the biggest differences between Europe and the top domains are that the European domains have a larger number of ciphers classified as *weak*, BREACH is also much higher for Europe than the top domains of Alexa, Tranco, and Majestic. Here, it should be noted that several of the other regional categories (for example, Africa, South America, and Asia) can also be considered outliers. They are all relatively closer to each other and further away from the main clusters of domains.

Finally, we note that *Adult* is the furthermost to the right

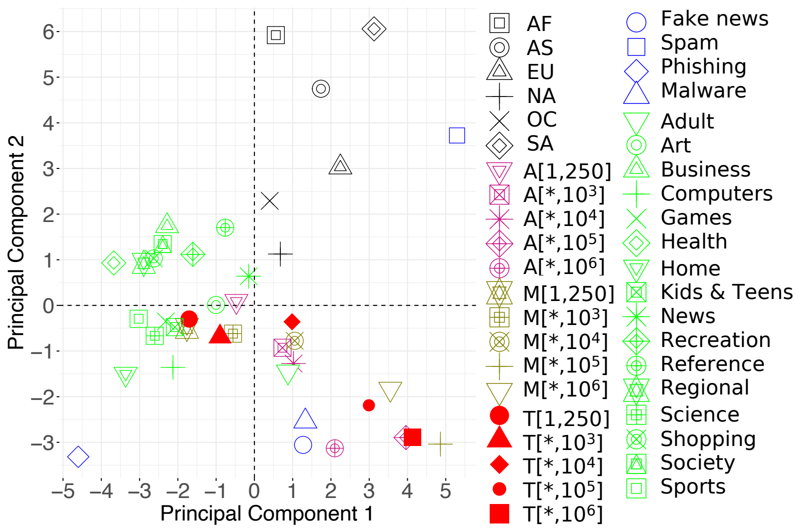


Fig. 5: PCA plot with all categories using the first two principal components.

(and only positive) among the 13 top-50 categories. Although this category is not explicitly studied as a malicious domain here, we note that this class still represents a special category of domains, which makes it noteworthy that it emerged as an outlier.

V. Related Work

Previous research has examined domains [36], protocols [37]–[41], vulnerabilities [42]–[45], and certificate authorities [46]–[48], to identify risks and vulnerabilities. The cipher suites and key types align with the protocols used. Springall et al. [39] found that the key types DHE and ECDHE were reused for the Alexa top-1M domains, leading to 38% of the connections being vulnerable to attacks 24 hours after implementation and 10% vulnerable 30 days after, regardless of the cipher suite used. Alabduljabbar et al. [49] found in their comparative study of SSL certificates on free content websites that 36% had major weaknesses, with 17% being invalid. ECDSA-256 was predominately used by 38% of these websites, compared to 20% of sites requiring payment, even though ECDSA offers greater security than the most commonly used public key type RSA-2048. Only 15 months after the TLS 1.3 standardization in 2018 [50], Holz et al. [37] observed it to be used in 20% of the cases. For popular domains, TLS 1.3 was used by 30%, while only 10% among top-level domains.

Simos et al. [51] analyzed cipher suite recommendations, finding that organizations recommend varying numbers of cipher suites based on their security standards. NSA Suite B recommended two cipher suites, the European Union Agency for Network and Information Security (ENISA) recommended 24, the Federal Office for Information Security recommended 16, and Mozilla 22.

Passive and active measurements have been used to study HTTPS connections and the web certificate ecosystem [52]–[54], revealing poor certificate quality among top-ranked domains [52], weak key types and certificate practices [53], and more than 25% of monitored HTTPS sessions using weak security [54].

Kim et al. [48] studied phishing attacks in combination with CAs. They analyzed large numbers of phishing domains that used HTTPS from the database eCrimeX and found that around the same time Let's Encrypt and began issuing free certificates, the number of phishing URLs rose. Phishing domains began using more CAs that were automated.

Several studies have examined phishing websites, comparing compromised domains to malicious registration [8], [55], [56], suggesting that more than 70% of phishing websites are hosted on compromised domains [8], [55]. Bayer et al. [6] proposed an enhancement to phishing blocklist accuracy, showing that phishing blocklists contain a small fraction of benign domains classified as malicious.

Other works have compared the security practices seen across domain categories, including wildcard certificates [57], ad-blocker performance [58], and many other aspects [11], [59], [60]. However, none of these works compare the security aspects of domain categories, especially "malicious" vs. "benign" domains.

VI. Conclusion

In this paper, we have conducted a comprehensive measurement-based comparison of various classes of malicious domains alongside popular domains. First, after assembling a diverse dataset that encompasses four classes of malicious domains and numerous classes of popular domains, we collected statistics using three complementary data collection methodologies. Second, facilitated by this unique dataset, we conducted a detailed analysis of domain attributes that provides notable observations and highlights the importance of understanding domain types for bolstering cybersecurity measures.

For example, our analysis revealed intriguing patterns across domain categories, with *fake news* and *malware* domains exhibiting similarities to popular domains, while *phishing* and *spam* domains emerged as outliers. Here, *phishing* domains showcased a higher level of security awareness, while *spam* domains exhibited laxer security measures. Principal component analysis (PCA) further elucidated the unique positioning of these two domain classes, emphasizing their distinct characteristics compared to other domains. Additionally, we observed trends indicating similarities between certain domain categories and less popular domains, underscoring the nuanced relationships between domain types.

Our measurement-based comparison captures the diverse characteristics of malicious and benign domains. By analyzing security profiles, performance metrics, and behavioral patterns, we provide an improved understanding of the distinct traits exhibited by each domain category. Our findings reinforce the importance of separating secure communication practices from domain trustworthiness, emphasizing that untrustworthy domains often employ up-to-date security measures. By highlighting these differences, our research underscores the need for clearer distinctions and informed decision-making to enhance user protection and strengthen Internet security against evolving cyber threats.

REFERENCES

[1] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "Fakenewsnet: A data repository with news content, social context and dynamic information for studying fake news on social media," *arXiv preprint*, 2018.

[2] URLHaus, https://urlhaus.abuse.ch/downloads/text_online/, 2024.

[3] PhishTank, https://phishtank.com/, 2024.

[4] J. Wein, https://joewein.de/sw/bl-text.htm, 2024.

[5] testssl.sh, https://github.com/drwetter/testssl.sh, 2024.

[6] J. Bayer, S. Maroofi, O. Hureau, A. Duda, and M. Korczynski, "Building a resilient domain whitelist to enhance phishing blacklist accuracy," in *Proc. APWG Symposium on Electronic Crime Research (eCrime)*, 2023.

[7] S. Bell and P. Komisarczuk, "An analysis of phishing blacklists: Google safe browsing, openphish, and phishtank," in *Proc. Australasian Computer Science Week Multiconference (ACSW)*, 2020.

[8] R. D. Silva, M. Nabeel, C. Elvitigala, I. Khalil, T. Yu, and C. Keppitiyagama, "Compromised or Attacker-Owned: A large scale classification and study of hosting domains of malicious URLs," in *Proc. USENIX Security Symposium (USENIX Security)*, 2021.

[9] Netcraft, "Most visited websites," https://trends.netcraft.com/topsites, 2023.

[10] V. L. Pochat, T. Van Goethem, S. Tajalizadehkhoob, M. Korczyński, and W. Joosen, "Tranco: A research-oriented top sites ranking hardened against manipulation," in *Proc. Network and Distributed System Security (NDSS)*, 2018.

[11] Q. Scheitle, O. Hohlfeld, J. Gamba, J. Jelten, T. Zimmermann, S. D. Strowes, and N. Vallina-Rodriguez, "A long way to the top: Significance, structure, and stability of internet top lists," in *Proc. Internet Measurement Conference (IMC)*, 2018.

[12] A. Halim, M. Danielsson, M. Arlitt, and N. Carlsson, "Temporal analysis of X.509 revocations and their statuses," in *Proc. IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 2022.

[13] G. Parallel, https://www.gnu.org/software/parallel/, 2024.

[14] OpenSSL, https://www.openssl.org, 2024.

[15] H. C. Rudolph and N. Grundmann, "Ciphersuite info," https://ciphersuite.info/, 2024.

[16] D. Adrian, K. Bhargavan, Z. Durumeric, P. Gaudry, M. Green, J. A. Halderman, N. Heninger, D. Springall, E. Thomé, L. Valenta *et al.*, "Imperfect forward secrecy: How diffie-hellman fails in practice," in *Proc. ACM Computer and Communications Security (CCS)*, 2015.

[17] N. I. of Standards and T. (NIST), "CVE-2013-0169," https://nvd.nist.gov/vuln/detail/CVE-2013-0169.

[18] N. J. Al Fardan and K. G. Paterson, "Lucky thirteen: Breaking the TLS and DTLS record protocols," in *Proc. IEEE Symposium on Security and Privacy (S&P)*, 2013.

[19] N. I. of Standards and T. (NIST), "CVE-2011-3389," https://nvd.nist.gov/vuln/detail/CVE-2011-3389.

[20] T. Duong and J. Rizzo, "Here come the ⊕ ninjas," *Unpublished manuscript*, 2011.

[21] N. I. of Standards and T. (NIST), "CVE-2014-3566," https://nvd.nist.gov/vuln/detail/CVE-2014-3566.

[22] B. Möller, T. Duong, and K. Kotowicz, "This POODLE bites: Exploiting the SSL 3.0 fallback," *Google Security Advisory*, 2014.

[23] N. I. of Standards and T. (NIST), "CVE-2013-3587," https://nvd.nist.gov/vuln/detail/CVE-2013-3587.

[24] Y. Gluck, N. Harris, and A. Prado, "BREACH: Reviving the crime attack," https://www.breachattack.com/resources/BREACH%20-%20SSL,%20gone%20in%2030%20seconds.pdf, 2013.

[25] N. I. of Standards and T. (NIST), "CVE-2016-0800," https://nvd.nist.gov/vuln/detail/CVE-2016-0800.

[26] N. Aviram, S. Schinzel, J. Somorovsky, N. Heninger, M. Dankel, J. Steube, L. Valenta, D. Adrian, J. A. Halderman, V. Dukhovni, E. Käsper, S. Cohney, S. Engels, C. Paar, and Y. Shavitt, "DROWN: Breaking TLS using SSLv2," in *Proc. USENIX Security Symposium (USENIX Security)*, 2016.

[27] N. I. of Standards and T. (NIST), "CVE-2016-6329," https://nvd.nist.gov/vuln/detail/CVE-2016-6329.

[28] K. Bhargavan and G. Leurent, "On the practical (in-)security of 64-bit block ciphers: Collision attacks on http over tls and openvpn," in *Proc. ACM Computer and Communications Security (CCS)*, 2016.

[29] N. I. of Standards and T. (NIST), "CVE-2015-2319," https://nvd.nist.gov/vuln/detail/CVE-2015-2319.

[30] ——, "CVE-2015-4000," https://nvd.nist.gov/vuln/detail/CVE-2015-4000.

[31] K. McKay and D. Cooper, "Recommendation for key management: Part 1 – general (NIST SP 800-57 part 1 rev. 5)," National Institute of Standards and Technology (NIST), Tech. Rep., 2017.

[32] J. Aas, R. Barnes, B. Case, Z. Durumeric, P. Eckersley, A. Flores-López, J. A. Halderman, J. Hoffman-Andrews, J. Kasten, E. Rescorla *et al.*, "Let's encrypt: an automated certificate authority to encrypt the entire web," in *Proc. ACM Computer and Communications Security (CCS)*, 2019.

[33] M. Döberl, Y. F. von Wangenheim, C. M. Bruhner, D. Hasselquist, M. Arlitt, and N. Carlsson, "Chain-sawing: A longitudinal analysis of certificate chains," in *Proc. IFIP Networking Conference*, 2024.

[34] C. M. Bruhner, O. Linnarsson, M. Nemec, M. Arlitt, and N. Carlsson, "Changing of the guards: Certificate and public key management on the internet," in *Proc. Passive and Active Measurement (PAM)*, 2022.

[35] D. Hasselquist, E. K. Gawell, A. Karlström, and N. Carlsson, "Phishing in style: Characterizing phishing websites in the wild," in *Proc. Network Traffic Measurement and Analysis Conference (TMA)*, 2023.

[36] T. Libert, "Exposing the hidden web: An analysis of third-party HTTP requests on 1 million websites," *International Journal of Communication*, vol. 9, 2015.

[37] R. Holz, J. Hiller, J. Amann, A. Razaghpanah, T. Jost, N. Vallina-Rodriguez, and O. Hohlfeld, "Tracking the deployment of TLS 1.3 on the web: a story of experimentation and centralization," *SIGCOMM Comput. Commun. Rev.*, vol. 50, no. 3, p. 3–15, July 2020.

[38] K. McKay and D. Cooper, "Guidelines for the selection, configuration, and use of transport layer security (TLS) implementations (NIST SP 800-52 rev. 2)," National Institute of Standards and Technology, Tech. Rep., 2017.

[39] D. Springall, Z. Durumeric, and J. A. Halderman, "Measuring the security harm of TLS crypto shortcuts," in *Proc. ACM Internet Measurement Conference (IMC)*, 2016.

[40] O. Alrawi and A. Mohaisen, "Chains of distrust: Towards understanding certificates used for signing malicious applications," in *Proc. International Conference Companion on World Wide Web (WWW Companion)*, 2016.

[41] Y. Chen and Z. Su, "Guided differential testing of certificate validation in SSL/TLS implementations," in *Proc. Joint Meeting on Foundations of Software Engineering (ESEC/FSE)*, 2015.

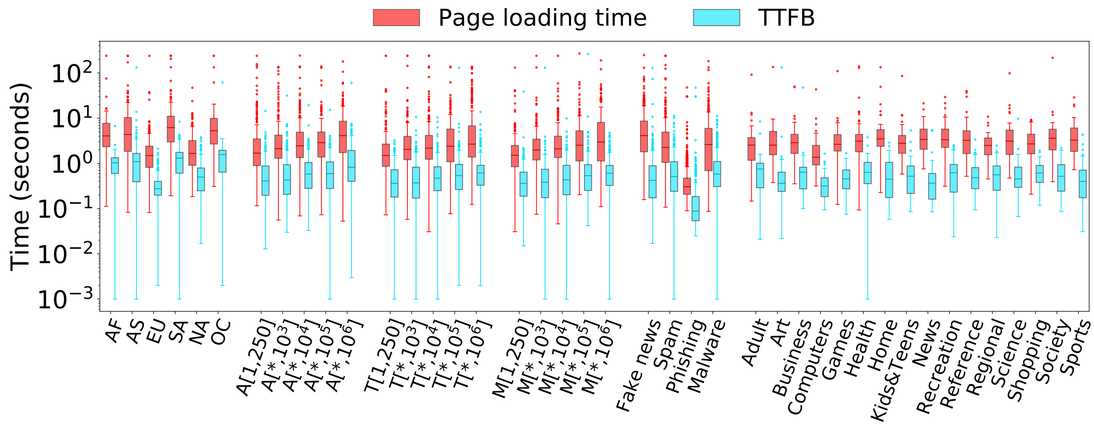


Fig. 6: Page load times and time to first byte (TTFB).

[42] S. Lee, Y. Shin, and J. Hur, "Return of version downgrade attack in the era of TLS 1.3," in *Proc. ACM International Conference on Emerging Networking EXperiments and Technologies (CoNEXT)*, 2020.

[43] K. Bhargavan, C. Brzuska, C. Fournet, M. Green, M. Kohlweiss, and S. Zanella-Béguelin, "Downgrade resilience in key-exchange protocols," in *Proc. IEEE Symposium on Security and Privacy (S&P)*, 2016.

[44] A. E. W. Eldewahi, T. M. H. Sharfi, A. A. Mansor, N. A. F. Mohamed, and S. M. H. Alwahbani, "SSL/TLS attacks: Analysis and evaluation," in *Proc. IEEE International Conference on Computing, Control, Networking, Electronics and Embedded Systems Engineering (ICCNEEE)*, 2015.

[45] L. Zhang, D. Choffnes, D. Levin, T. Dumitraş, A. Mislove, A. Schulman, and C. Wilson, "Analysis of ssl certificate reissues and revocations in the wake of heartbleed," in *Proc. ACM Conference on Internet Measurement Conference (IMC)*, 2014.

[46] K. Hageman, E. Kidmose, R. Rydhof Hansen, and J. Myrup Pedersen, "Can a TLS certificate be phishy?" in *Proc. INSTICC International Conference on Security and Cryptography (SECRYPT)*. SciTePress, 2021.

[47] Y. Sakurai, T. Watanabe, T. Okuda, M. Akiyama, and T. Mori, "Discovering HTTPSified phishing websites using the TLS certificates footprints," in *Proc. IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 2020.

[48] D. Kim, H. Cho, Y. Kwon, A. Doupé, S. Son, G.-J. Ahn, and T. Dumitras, "Security analysis on practices of certificate authorities in the HTTPS phishing ecosystem," in *Proc. ACM Asia Conference on Computer and Communications Security (AsiaCCS)*, 2021.

[49] A. Alabduljabbar, R. Ma, S. Choi, R. Jang, S. Chen, and D. Mohaisen, "Understanding the security of free content websites by analyzing their SSL certificates: A comparative study," in *Proc. ACM Workshop on Cybersecurity and Social Sciences (CySSS)*, 2022.

[50] E. Rescorla, "The Transport Layer Security (TLS) Protocol Version 1.3," RFC 8446, 2018.

[51] D. E. Simos, K. Kleine, A. G. Voyiatzis, R. Kuhn, and R. Kacker, "TLS cipher suites recommendations: A combinatorial coverage measurement approach," in *Proc. IEEE International Conference on Software Quality, Reliability and Security (QRS)*, 2016.

[52] R. Holz, L. Braun, N. Kammenhuber, and G. Carle, "The ssl landscape: a thorough analysis of the x.509 pki using active and passive measurements," in *Proc. ACM Internet Measurement Conference (IMC)*, 2011.

[53] Z. Durumeric, J. Kasten, M. Bailey, and J. A. Halderman, "Analysis of the https certificate ecosystem," in *Proc. ACM Internet Measurement Conference (IMC)*, 2013.

[54] G. Ouvrier, M. Laterman, M. Arlitt, and N. Carlsson, "Characterizing the HTTPS trust landscape: A passive view from the edge," *IEEE Communications Magazine*, vol. 55, no. 7, pp. 36–42, July 2017.

[55] S. Le Page, G.-V. Jourdan, G. V. Bochmann, I.-V. Onut, and J. Flood, "Domain classifier: Compromised machines versus malicious registrations," in *Proc. International Conference on Web Engineering (ICWE)*, 2019.

[56] S. Maroofi, M. Korczyński, C. Hesselman, B. Ampeau, and A. Duda, "Comar: Classification of compromised versus maliciously registered domains," in *Proc. IEEE European Symposium on Security and Privacy (EuroS&P)*, 2020, pp. 607–623.

[57] D. Hasselquist, L. Bolin, E. Carlsson, A. Hylander, M. Larsson, E. Voldstad, and N. Carlsson, "Longitudinal analysis of wildcard certificates in the webpki," in *Proc. IFIP Networking Conference*, 2023.

[58] P. Gunnarsson, A. Jakobsson, and N. Carlsson, "On the impact of internal webpage selection when evaluating ad blocker performance," in *Proc. International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS)*, 2022.

[59] A. Rasaii, D. Gosain, and O. Gasser, "Thou shalt not reject: Analyzing accept-or-pay cookie banners on the web," in *Proc. ACM Internet Measurement Conference (IMC)*, 2023.

[60] K. Ruth, D. Kumar, B. Wang, L. Valenta, and Z. Durumeric, "Toppling top lists: Evaluating the accuracy of popular website lists," in *Proc. ACM Internet Measurement Conference (IMC)*, 2022.

Appendix

Ethical Considerations

This paper does not raise any ethical concerns. First, for data collection, we exclusively use public infrastructure and adhere strictly to public protocols. Second, we report only aggregate values and do not discuss vulnerabilities for specific domains. Finally, our measurements contribute only to a small portion of the overall server load.

Load Times and TTFB

Here, we report on the time to first byte (TTFB) and the page load times, as calculated from when the Selenium tool began its navigation until the DOM is complete. Figure 6 shows these results. As expected, comparing the regional domain sets, given our measurement location (Europe), the fastest load times (and TTFB times) are for the European domains, followed by the North American domains. We also see a clear trend in that more popular domains load faster and have shorter TTFB than less popular domains (i.e., going from ranks [1,250] to [*,10^6] for the three rank categories). Part of these differences are due to better infrastructure and higher content replication (e.g., through the use of effective CDNs), as well as more careful design of webpages. With the above reference points, it is therefore interesting to consider the four classes of malicious domains. Here, we note that *phishing* domains have by far the lowest load times and TTFB of the four classes, suggesting that their infrastructure may be close by. However, we have also found these sites to be relatively lighter (e.g., not as reliant on objects being delivered through a long range of third-party domains), effectively reducing their load times. With these domains—as seen in this paper—more frequently being up-to-date with the latest TLS versions, cipher suites, and good protection against known attacks, we expect them to often be hosted on up-to-date infrastructure.

Finally, we note that these performance metrics are biased by server and testing location. While multi-perspective measurements would help strengthen the significance of our observations, we do not expect such measurements to affect the high-level observations presented in the paper's main part.