A Peek into Top-Level Domains and Cybercrime

By Janos Szurdi
November 11, 2021 at 6:00 AM
Category: Unit 42
Tags: Cybercrime, DNS, Phishing

This post is also available in: 日本語 (Japanese)

Executive Summary

Top-level domains (TLDs), such as `.com`, `.net`, `.xxx` and `.hu`, sit at the highest level of the domain name system (DNS) naming hierarchy. When users want to acquire domain names (e.g., `paloaltonetworks.com`), typically, they need to register them under a TLD directly or one level lower (e.g., `google.co.uk`). Properties and policies of TLDs such as pricing, registration restrictions, security practices and the lexical similarity to other TLDs (`.cm` vs `.com`) influence how attractive criminals will find these TLDs for their endeavors.

Out of more than 1,000 TLDs, the top 25 TLDs (by number of malicious domains) account for more than 90% of all malicious domain names. While these 25 TLDs are not malicious, they are well-positioned to help mitigate malicious domain registrations. We find that TLDs offering free domain registration are among the top preferred TLDs for phishing domains. We hypothesize that the above-mentioned properties and policies play a germane role in making a TLD favorable for criminal enterprises.

When we explore TLDs with the highest rate of malicious domains, we find that six out of the top 10 are TLDs of developing countries. One especially disreputable TLD, `.zw`, is seven standard deviations above the average rate of malicious domains for a TLD. Surprisingly, we found that `.zw` and a few other TLDs have a bad reputation, at least partially, because domains in these TLDs are frequently compromised rather than registered with malicious intent. Another example is `.pw`, with over seven standard deviations above the average rate of phishing registrations. TLD reputation based on our research can be used as one of the features to decide whether a domain name is malicious or not.

Studying domains in TLDs specifically concocted for sensitive topics such as adult and gambling sites, including `.xxx`, `.casino`, `.poker` and `.porn`, we confirm that they host content expected given their TLDs. Using our External Dynamic List feature, Palo Alto Networks customers have the opportunity to block individual TLDs entirely when they deem them inappropriate, for example, by setting custom wildcard rules (e.g., *.xxx) to block adult and gambling TLDs identified in this blog.

Palo Alto Networks offers multiple security subscriptions, including Advanced URL Filtering and DNS Security, that can be used to block malicious and sensitive category domains, including those discussed in this blog.

Top-Level Domains and the DNS Ecosystem

We start with a brief introduction to the hierarchical structure of the domain name system and why we study top-level domains (TLDs).

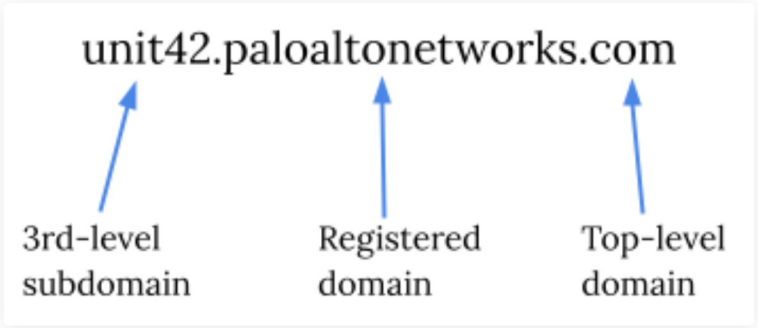

Figure 1. The structure of a domain name.

A domain name like `unit42.paloaltonetworks.com` consists of three parts. The `.com` part is the top-level domain (TLD), which is

at the highest level of the DNS naming hierarchy. Usually, users looking to buy domain names can register under these TLDs. Domain names acquired by users are called registered domains. When Palo Alto Networks Inc. bought the second-level name `paloaltonetworks` in the `.com` namespace, it gained ownership of all names under `paloaltonetworks.com`. Therefore when Palo Alto Networks Inc. decided to form a sub-organization tasked with providing "the answer to the ultimate question of life, the universe and everything," it could freely create the third-level domain name `unit42`. In this blog, the numbers presented are based on unique registered domain counts. For example, in the case of `www.google.co.uk`, we consider the third-level registered domain `google.co.uk` and not the second-level `co.uk` or fourth-level `www.google.co.uk`.

There are two main types of TLDs. Generic TLDs (gTLDs) are owned and operated by private companies or organizations and are regulated by the Internet Corporation for Assigned Names and Numbers (ICANN). Examples of gTLDs include `.com`, `.xxx`, and `.google`. Differently, country code TLDs (ccTLDs) are owned and regulated by countries (though often still operated by private companies). ccTLDs include domains such as `.us`, `.cn`, `.de` and `.hu`.

Organizations operating TLDs and maintaining records of registered domains directly under TLDs are called registries. These registries do not just manage domain names under TLDs, but in the case of gTLDs, they also determine the policies regarding pricing, necessary identity verification and restrictions on purchasing domain names. These policies impact how much criminals will favor a given TLD for abusive domain registrations. Free or cheap registration and lax policies make TLDs favorable for malicious behavior.

Studying TLDs can provide a better understanding of criminal preferences and where malicious domains reside. TLD reputation can aid us in deciding if a domain is malicious, and can be used to nudge the operators of ill-famed TLDs toward curbing some of the abuse. Additionally, we can identify TLDs created for certain sensitive topics such as adult entertainment and gambling that are best entirely blocked in certain use cases. For example, educational institutions likely want to block both adult and gambling TLDs.

Malicious Domains and TLDs

Methodology

To understand both malicious and benign TLD usage, we use the fine-grained categories provided by our Advanced URL Filtering service. First, we only study domains categorized by the Advanced URL Filtering service, and we only consider registered domains (also called root domains). Additionally, we validate whether domains existed the past one year by checking zone files and passive DNS, and by issuing active DNS queries. We do not consider domains that we categorize as parked, insufficient content or unknown for our calculations. Further, when calculating reputation scores, we don't consider domains sinkholed for preemptive measures as malicious. Finally, we only consider TLDs with at least a hundred domains, as smaller TLDs likely have policies in place restricting entities allowed to register domain names. This blog post is based on data collected on Oct. 7, 2021.

We study four malicious categories defined by Palo Alto Networks: malware, phishing, command and control (C2), and grayware. When we discuss malicious domains in general, we consider the union of these four malicious categories.

Next to malicious content, we examine domains registered to host sensitive content that might be illegal or inappropriate in a corporate setting, at educational institutions or for governments.

In our recommendations, we propose the blocking of these categories when deemed appropriate by our customers. The list of sensitive categories for the purpose of this blog includes Dynamic DNS, Abused Drugs, Adult, Gambling, Peer-to-Peer, Hacking, Questionable, Cryptocurrency, Proxy Avoidance and Anonymizers, and Copyright Infringement.

TLDs With the Highest Number of Malicious Domains

To study cybercriminals' TLD preference, we first look at the number of unique malicious domains registered at each TLD. As expected, the direct count of malicious domains will be highest at TLDs such as `.com`, which is by far the most popular TLD – however, it only has an average ratio of malicious domains. Hence, such big TLDs are not considered malicious, though they still have a huge responsibility since some of them provide a home for a large fraction of all malicious domains. For example, nearly half of malicious domains are `.com` domains.

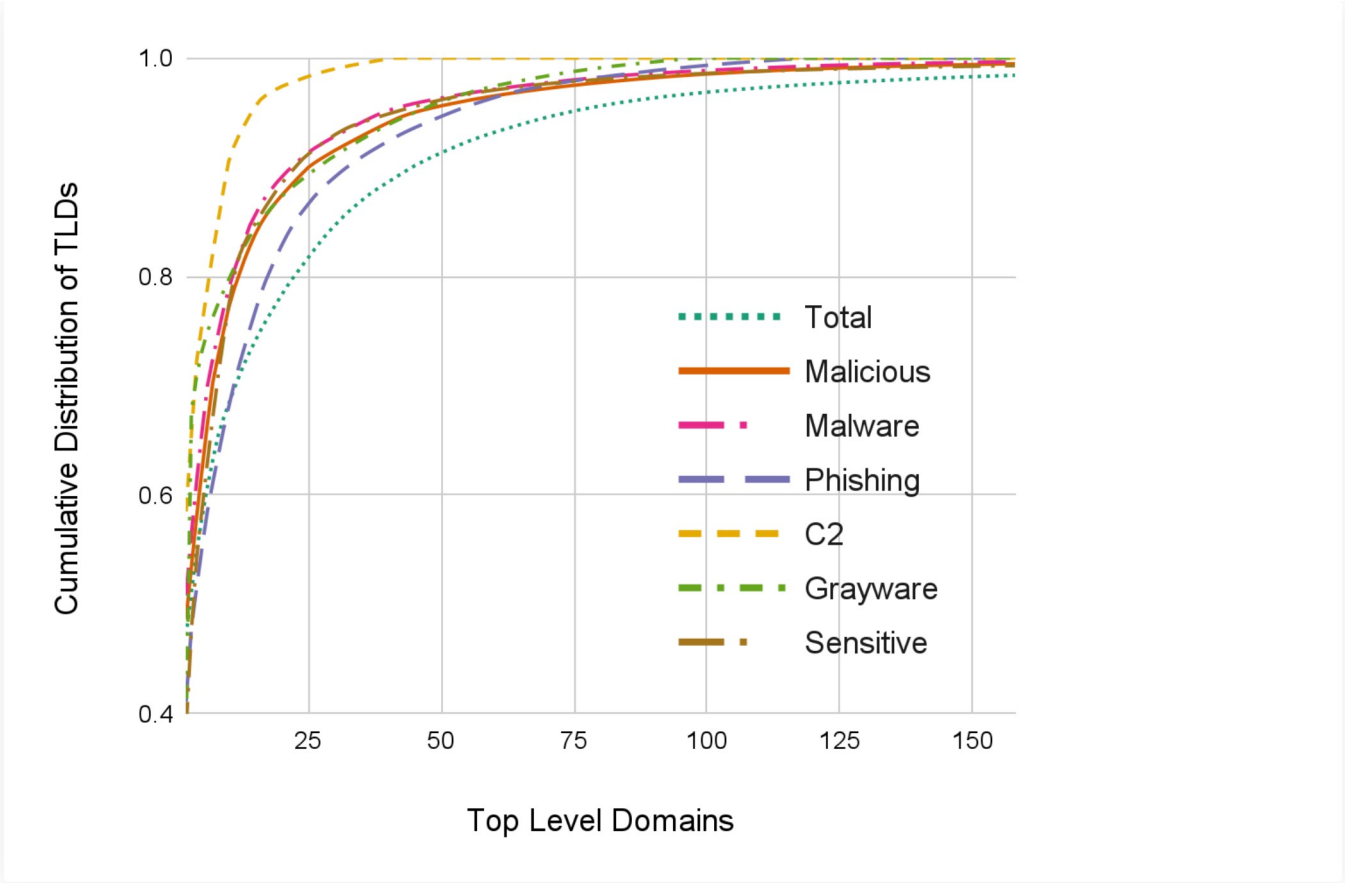

Figure 2. The cumulative distribution of the number of domains across TLDs for several categories.

In Figure 2, we look at the cumulative distribution of TLDs for the total number of domains registered, various malicious categories and sensitive domains. Having a small area under the curve (a "flat" curve) means that the domains are evenly distributed across TLDs for that category. The total domain line is the flattest curve compared to malicious and sensitive domains, implying that criminals prefer certain TLDs above others. For example, more than 99% of all C2 domains are concentrated at only 29 TLDs. At the same time, 99% of all domains are concentrated at 219 TLDs. Phishing – one of the most evenly distributed malicious categories – shows 99% of phishing domains concentrated at 92 TLDs.

Total		Malicious		Phishing		Malware		Grayware		C2		Sensitive	
TLD	CD	TLD	CD	TLD	CD	TLD	CD	TLD	CD	TLD	CD	TLD	CD
com	0.47	com	0.49	com	0.41	com	0.51	xyz	0.38	com	0.58	com	0.39
net	0.51	icu	0.53	xyz	0.48	icu	0.56	com	0.68	net	0.66	tk	0.48
de	0.55	xyz	0.58	tk	0.52	cn	0.61	tokyo	0.71	tk	0.72	icu	0.54
org	0.58	cn	0.62	ml	0.55	net	0.66	club	0.73	cn	0.76	ga	0.59
tk	0.61	net	0.66	cf	0.59	ml	0.70	net	0.75	info	0.79	cf	0.63
uk	0.63	ml	0.70	icu	0.61	org	0.72	work	0.76	cf	0.82	gq	0.67

cn	0.65	tk	0.73	ga	0.64	tk	0.75	ru	0.77	ml	0.85	ml	0.71
ru	0.67	org	0.75	top	0.66	cf	0.77	co	0.79	ga	0.88	cn	0.74
icu	0.68	cf	0.77	pw	0.68	xyz	0.79	info	0.80	gq	0.91	xyz	0.77
xyz	0.70	ga	0.79	net	0.70	ga	0.80	org	0.81	top	0.92	net	0.80

Table 1. The biggest TLDs and their cumulative distribution (CD) for various categories.

As mentioned earlier, the `.com` TLD is responsible for nearly half of all domains registered and subsequently also for nearly half of all malicious domains. This does not mean that the .com TLD is malicious, yet the operator of the `.com` TLD is uniquely positioned to help with clearing up malicious domain registrations. This is also true for other large TLDs such as `.net`, `.org`, `.tk`, `.cn`, `.icu` and `.xyz`.

In Table 1, we can also observe that TLDs like `.pw`, `.ml`, `.club`, `.cf` and `.top` are in the top 10 for certain types of abuse but are not among the top 10 largest TLDs, clearly signaling criminal preference. While the `.xyz` TLD is barely among the top 10 TLDs by total size, it is second only to the `.com` TLD in the number of phishing domains and has the highest number of grayware domains accommodated. Some TLD operators try to combat cybercrime, for example, in the case of `.xyz`, whose anti-abuse policy allows them to investigate and take actions on malicious domain names. As pinpointing domains that can be suspended or taken down can be difficult, certain TLD operators often make reporting of such domains easy (e.g. `.xyz's` reporting page).

Conversely, some large TLDs such as `.de` and `.uk` are not present in the top 10 list for any malicious categories. Both of these TLDs have significantly below the average number of malicious domains, showcasing that a dutiful TLD registry can help curb abuse.

Half of the top 10 phishing TLDs are not in the top 10 TLDs by total size, providing evidence that phishers prefer some TLDs over others. By randomly sampling phishing domains at these five TLDs, we observe that they often target the brands of the largest tech companies, such as social networks, payment solutions, secure messaging apps and webmail providers. Phishing domains targeting brands fall into different domain squatting categories that we discussed in our cybersquatting blog post. We also find more generic phishing domains containing words like login, support and account. The third category of phishing domains is related to specific trending topics such as COVID-19. We did a deep dive into COVID-19 related domains when the pandemic was still relatively new.

Furthermore, some ccTLDs such as `.pw` and `.tk` have a glaringly high number of malicious domains registered, comparable to the population of these regions – or even higher. The `.tk` TLD also has more phishing domains registered than the population of Tokelau. If we compare the malicious domain to population ratio (the malicious domain per capita count) of these ccTLDs to Germany's `.de` ccTLD, we find ratios that are hundreds or even hundreds of thousands times higher. For example, this ratio is 271,768, 2,373 and 610 times larger, respectively for `.tk`, `.pw` and `.ws`, compared to `.de`. Considering only phishing domains, the same ratio is 462,098 and 20,286 times larger for `.tk` and `.pw` than for `.de`.

ccTLD	Malicious domains per capita compared to .de	Phishing domains per capita compared to .de
.tk	271,768	462,098
.pw	2,373	20,286
.ws	610	11.44

Table 2. Comparing malicious domains and phishing domains per capita to .de, which has a relatively low incidence of malicious domains, to the more commonly abused TLDs .tk, .pw and .ws.

One of the most fascinating stories in the domain name world is how `.tk`, the ccTLD of a small Pacific island called Tokelau, became one of the most populous TLDs in the world. Domain registrations contributed at one point one-sixth of Tokelau's income. Their TLD

became popular by providing free domain registrations, where the source of income for the TLD operator is through advertisement rather than domain registration fees. Unfortunately, their domain registration policy also invites abuse, spam and a large amount of sensitive content, as we can observe in Table 1.

In fact, the TLDs `.tk`, `.ga`, `.cf` and `.ml`, all run by Freenom, appear on our list of top TLDs hosting phishing, and some of them also appear on our lists of top TLDs for other malicious categories. Freenom's fifth TLD, `.gq`, also appears on our top sensitive category list and barely missed the top 10 for malicious categories. Of note, all these ccTLDs are owned by developing countries where the income from registered domains might outweigh the issues arising with malicious registrations. This is in contrast with a TLD like `.de`, where monetary loss from cybercrime dwarfs the revenue from domain registrations.

In addition to the TLDs offering free registrations, TLDs like `.xyz` and `.icu` follow another strategy by offering cheap domains – typically only a couple of dollars. Their pricing strategy has allowed them to become two of the most popular TLDs among benign and malicious users alike. Their popularity can make it challenging for them to combat offending domains, even if they are determined to do so, as in the case of `.xyz`.

We confirm previous research showing that free or cheap domain registration leads to a high level of abuse. Additionally, the same paper found that restricted registration leads to a lower abuse rate. Other researchers have tried to devise registration policy strategies to decrease malicious registrations. Unfortunately, the domain naming ecosystem is complex, and far-reaching changes are not likely to occur in the near future.

TLDs With the Highest Rate of Malicious Domains

Malicious		Phishing		Malware		Grayware		C2	
TLD	MAD	TLD	MAD	TLD	MAD	TLD	MAD	TLD	MAD
zw	30.37	pw	43.48	zw	38.05	sbs	89.66	cyou	7.95
bd	26.18	quest	32.00	bd	30.98	tokyo	66.08	pw	6.72
ke	25.38	ke	17.28	ke	28.69	xyz	40.94	ws	4.25
am	18.48	date	15.47	am	24.46	cam	21.21	gq	4.03
sbs	17.58	cyou	13.80	cd	16.07	date	18.56	cf	3.84
date	15.38	support	11.38	date	13.12	cm	16.21	ml	3.81
pw	13.35	win	8.55	bid	12.81	casa	15.78	ga	3.36
quest	11.92	rest	7.14	ml	12.00	uno	11.77	info	2.93
cd	11.88	casa	6.45	ws	10.68	email	8.39	su	2.74
bid	10.96	help	5.47	icu	9.08	stream	7.38	best	2.44

Table 3. The TLDs hosting the highest rate of malicious domains..

The MAD score is the median of the absolute deviation from the median. As shown in Table 3, we calculate the MAD score for the ratio of malicious to all domains in a TLD. We use the MAD score as a malicious reputation to compare TLDs to the median. MAD score is better to use than standard deviation when large outliers bias the average. For example, the `.com` TLD appears near average malicious when considering -0.06 standard deviation. However, the MAD score is 0.81, telling us that the `.com` TLD is more malicious than the median TLD.

In Table 3, we can find that some TLDs have a high proportion of malice, and they are rarely the same as top TLDs by count . A few TLDs have an extremely high ratio of malicious domains, such as `.zw`, `.bd` and `.ke`. The high rate of malice is unexpected for these TLDs, as registering a domain in them is four to 14 times more expensive than registering in the `.com` TLD. To our surprise, we found that a significant portion of the malicious domains in these TLDs are not registered with malicious intent but are compromised instead. The per

capita GDP of these regions is 60 to 30 times lower than that of the U.S., which suggests a lower budget and less expertise for setting up websites in these TLDs. Consequently, we expect lower-quality websites – confirmed by inspecting random samples of domain names – and hence more security holes.

One interesting case is `.cm`, a top TLD for phishing (No. 12) and grayware (No. 6). We conjecture that criminals prefer `.cm` for phishing attacks due to its similarity to `.com`, making domains registered at this TLD less conspicuous in phishing attacks.

Sensitive TLD Categories

TLD	Ratio	MAD
casino	0.88	54.05
xxx	0.86	52.67
poker	0.84	51.22
porn	0.81	49.88
bet	0.66	40.35
sex	0.60	35.97
sexy	0.54	32.39
adult	0.39	22.74
gq	0.29	16.63
webcam	0.26	14.57

Table 4. Top category-specific TLDs for sensitive categories.

Next to malicious content, we at Palo Alto Networks track sensitive or risky categories organizations might want to block. For example, both educational and government institutions often prefer to block adult and gambling sites. Our customers can block entire TLDs using our External Dynamic List feature, such as the sensitive-category-specific TLDs `.xxx` and `.casino`.

In Table 4, we can note that several TLDs have a high proportion of domains in sensitive categories. Even though we track many sensitive categories, mostly adult and gambling TLDs made it to our top list of sensitive-category-specific TLDs. Another category frequent at certain TLDs is “Proxy Avoidance and Anonymizers,” common at TLDs such as `.gq`, `.ga`, `.ml` and `.tk`.

Conclusion

We observe that the vast majority of malicious domains can be found at a handful of TLDs, providing an opportunity to them to help with fighting cybercrime. We also find TLDs many MAD scores above the median, suggesting that we can devise a TLD reputation to help in classifying domain names as malicious or benign. We also confirmed TLDs specific to sensitive categories that our customers can block using our External Dynamic List.

Palo Alto Networks offers multiple security subscriptions, including Advanced URL Filtering and DNS Security, that can be used to block malicious and sensitive category domains, including those discussed in this blog.

Acknowledgements

We want to thank Arun Kumar, Daiping Liu, Erica Naone, Laura Novak and Oleksii Starov for their invaluable input on this blog post.