If you cite this paper, please use the TMA reference: J. Zirngibl, S. Deusch, P. Sattler, J. Aulbach, G. Carle, M.Jonker. 2022. Domain Parking: Largely Present, Rarely Considered!. *In Proc. of Network Traffic Measurement and Analysis Conference (TMA '22).* IFIP.

Domain Parking: Largely Present, Rarely Considered!

Johannes Zirngibl*, Steffen Deusch*, Patrick Sattler*, Juliane Aulbach*, Georg Carle*, Mattijs Jonker†
Technical University of Munich, Munich, Germany
{zirngibl,deusch,sattler,aulbach,carle}@net.in.tum.de
†*University of Twente, Enschede, The Netherlands*
m.jonker@utwente.nl

***Abstract*—Domain parking typically involves leveraging advertisements to generate revenue on otherwise inactive domain names. Their content is rarely of real value to users and tends to be highly similar across parked domains. They have commonalities beyond content alone: parked domains can share hosting and DNS infrastructure. Parking rarely receives special treatment in existing studies (e.g., content analyses or infrastructure concentration studies). While the presence and possible bias introduced by parked pages is sometimes acknowledged in studies, the studies still treat parked domains as any other, either because differentiation is infeasible, or because doing so is considered out-of-scope.**

We argue that the impact of parked domains on analyses regarding the current state and future development of the Internet should not be overlooked. In this paper, we motivate this argument through quantification, and take steps towards helping other researchers identify parked domains.

We systematically collect a list of 82 parking services and develop DNS-based indicators to help identify parked domains. We next quantify the presence of parked domains, using large-scale DNS data containing hundreds of millions of registered domain names, representative for a significant part of the global DNS namespace. Overall, we pinpoint 60 M parked domains, which is a significant percentage of all names under consideration (23 %) and identify up to 4 % of domains from top lists to be parked. These findings demonstrate that the effect of parked pages is potentially pronounced. We also break down into the various parking services and DNS zones. This helps us demonstrate and further discuss the effect that domain parking can have on research and Internet consolidation.

***Index Terms*—DNS, Domain parking, Internet measurements**

I. Introduction

The Internet is becoming increasingly centralized. Over the past years, the development towards centralization and consolidation has emerged as an important subject of discussion among research and network operator communities.

The upsurge of Content Delivery Networks (CDNs) has added to Internet centralization. A number of recent studies evaluate properties of CDNs and consolidation – related in particular to the web and Domain Name System (DNS) ecosystems – and offer a basis on which to discuss and evaluate its impacts [1]–[6]. To evaluate trends and possible impacts, studies often rely on domain names and an analysis of the infrastructure with which names are associated. In such studies, domain names are usually treated the same, even though not all names are equal. In particular, domains can be parked. Domain parking is a concept to generate revenue from registered but otherwise unused domains, for example via advertisements. So-called domain parking providers offer the means for such monetization.

978-3-903176-47-8 ©2022 IFIP

Parking pages differ from user-centric web content. Their content is of little use and importance to users, yet similar across parked domains. Hosting and DNS infrastructure can also be in common. As a result, parked domains can introduce *bias* in centralization studies. While some studies do mention this limitation, differentiating parked domains is then either considered infeasible or out of scope.

We argue that parked domains require consideration when evaluating and discussing Internet consolidation. To this end, we quantify the prevalence of parked domains at scale across multiple Top Level Domains (TLDs) but also top lists. We further evaluate the impact of domain parking on DNS and hosting providers, *e.g.*, CDNs, and reason about effects of lack of consideration in related studies. To the best of our knowledge, we are the first to offer this perspective. This work does not evaluate the monetization schemes of domain parking services, individually hosted domains or potential wrongdoing (*e.g.*, typo-squatting) and vulnerabilities.

Our contributions in this work are:

(i) We systematically collect 82 parking, marketplace and placeholder services. We analyze their modus operandi and develop DNS-based indicators that enable identifying parked domains. We share these with the community to lower the barriers to considering parked domains;

(ii) We study the prevalence of domain parking using multiple sources of large-scale DNS data. These data are representative for a sizable part of the global DNS namespace, containing hundreds of millions of registered domain names from well over 1 k TLDs, including country-code TLDs (ccTLDs), legacy generic TLDs (gTLDs) such as *.com*, and newer gTLDs such as *.tokyo*;

(iii) With our new-found insights into parked domain name prevalence, we discuss findings from existing research regarding the development and consolidation of the Internet and of CDNs.

We identify domain name parking services in Section II. In Section III, we outline the data sources on the basis of which we study the prevalence of domain parking. We present our analyses on prevalence and the impact on effective TLDs (eTLDs) and infrastructure providers in Section IV. Section V contains an outline of related work. Building on our insights, we discuss the impact of domain parking on previous findings in Section VI. Finally, we conclude in Section VII.

II. Domain Parking Services

Monetizing unused domains through advertisements or sales is referred to as domain parking. This concept has been generalized and professionalized by dedicated domain parking services, that administer and monetize registered but unused domains. These services host websites on parked domains including advertisements and sales banners, manage visibility and cash flows. Domain owners can register their domains with these services and park their assets using DNS in one of several ways. First, domains can be parked by delegating authority for a domain name to the name servers of a parking service. Under this approach the name server delegation (*i.e.*, NS records) will point to the service-specific name servers. Second, domain name owners can use their own name servers but configure an IP address record (*i.e.*, A or AAAA) or canonical name record (CNAME) and point it to the infrastructure of the parking service. Both approaches can be inferred from DNS data, for example by actively querying for records.

We systematically collected a list of 82 parking services and established if they require delegation or involve canonical names or IP address records. These DNS-based indicators are input to our methodology to identify parked domains similarly to Vissers *et al.* [7]. Our list contains providers of varying size. We started our collection with a web search for prominent parking services and published configurations, *e.g.*, GoDaddy (Free Parking) [8]. Based on these insights, we searched for indicators of parking in domains of our data, *e.g.*, ParkingCrew uses the name servers `nsX.parkingcrew.net`. Lastly, we rely on our DNS data to analyze frequently used name servers and IP addresses to identify the most impactful services. This follows our assumption that parking services rely on few name servers or IP addresses for many domains. We selected a random set of parked domains for each service and used visual confirmation to verify parking services (cf Section IV-C). We exclude services without a clear identification possibility based on DNS indicators, *e.g.*, Namecheap mixes services on the same infrastructure and relies on HTTP redirects. For a more detailed description of our collection process we refer to [9]. We did not investigate infrastructure associated with less than $10\,\mathrm{k}$ domains and might miss specialized services for TLDs not in our data.

Depending on the specific parking service, different types of indicators allow for parked domains to be identified. For example, GoDaddy (Free Parking) uses specific IP addresses (`34.102.136.180` and `34.98.99.30`), while AfterNic relies on a set of name servers (*ns*.afternic.com.*). We identified the reference points that are explicitly used for parking and use these values to identify parked domains using DNS measurement data (see Section IV). To lower the barriers for other researchers to identify and consider parked domains separately, we publish our list of services including the reference points alongside this paper.

https://tma22-parking.github.io

We divide the 82 identified services into four categories based on the content they display on parked pages.

Advertisements These services use domains to host a web page focusing on advertisements which generate money if a client accesses the page. Monetization is either click based (Pay-Per-Click (PPC)) or based on redirects (Pay-Per-Redirect (PPR)). We identified 25 services in this category including well known companies in the segment of hosting like GoDaddy.

Domain marketplaces These services mainly sell domain names that are considered valuable. To advertise the domain itself, they host a plain web page with a banner indicating the domain is for sale. These services are not necessarily open for use by others and can try to only sell their own domain name assets. This applies, for example, to HugeDomains.com. While not strictly acting as domain parking services as mentioned by Halvorson *et al.* [10], they show similar behavior to parking services, namely centralized infrastructure, similar content, and limited importance to most users. Out of the 82 identified services, 30 are of this category. They are marked in the published list of services accordingly.

Placeholders The third category has some similarities with the previous two, but notably comes without apparent monetization attempts. This category involves landing pages of the "this domain is taken but not yet in use" kind. These domains can also share dedicated infrastructure and the hosted pages thus only display placeholders. In some cases, these services are operated by registrars or CDNs. Our list includes 23 services of this category.

Mixed Four identified services use the same infrastructure for advertisement focused parking but also domain name sales or simple placeholders, *e.g.*, Sedo and Uniregistry. These services are still parking providers but can not be mapped to one of the introduced categories. We add them to the list of services but indicate them as mixed category.

III. DNS Data Sources

To quantify the prevalence of parked domains, we need DNS data. We rely on three different DNS data sources to this end. All three data sources involve independent collection efforts. One collection effort is implemented at the Technical University of Munich (TUM). The other two involve independent projects that collect and share DNS data with researchers.

TUM scans We conduct weekly DNS scans from TUM, targeting more than $325\,\mathrm{M}$ domain names each run. The total input of domains remains relatively even with a slight increase by $3\,\%$ throughout one year. Each scan takes $24\,\mathrm{h}$ to $48\,\mathrm{h}$ to complete, responsibly distributing measurement load and impact on name servers. This measurement is seeded with domains from:

- Well over $1\,\mathrm{k}$ available zone files from the Centralized Zone Data Service (CZDS), which includes legacy gTLDs and newer gTLDs;

- Names on the Alexa [11], Majestic [12] and Umbrella [13] Top 1M lists;
- A static collection of 98.1 M domains from 52 ccTLDs (partial zones, *e.g.*, 22 M *.tk* and 13 M *.de* domains).

The data set we use includes scans performed between Jan 1st, 2021 and Jan 28th, 2022. As such, we cover roughly 13 months.[1] We resolve A and AAAA records during the complete period and started to explicitly resolve and collect NS records in May 2021. We cover the ethical considerations for this measurement in Section VI.

OpenINTEL We use data from the OpenINTEL project [14]. This project collects, among others, the A, AAAA and NS records of domain names, through active querying. OpenINTEL primarily seeds its measurement on the basis of full TLD zone files and covers approximately 65 % of the global DNS namespace. The OpenINTEL measurement is seeded with domains from:

- Well over 1 k available zone files from the CZDS;
- Names under 16 ccTLDs (full zone);
- Names on the Alexa [11] and Umbrella [13] Top 1M lists;

We used aggregate data: statistics regarding parked services for one scan each week. The data from OpenINTEL allows us to analyze equally longitudinal and complementary data from the same time period as the TUM scans, but from another vantage point.

Rapid7 We obtain and use a single snapshot of *forward* DNS measurement data from Rapid7 [15] for the date Jan 28th, 2022. The Rapid7 measurement is done from a US-based vantage point and also includes some domain names that the other sources do not include, which is mostly due to differences in ccTLD coverage and special zones, *e.g.*, *.blogspot.com*. We use the Rapid7 snapshot to verify that inferences made on the basis of data collected from a US-based vantage point results in similar observations about the numbers of parked domains.

Public Suffix List We use the Public Suffix List [16] (PSL) to account for eTLD specific statistics. To see why PSL data is needed for this, consider that *example.com.br* and *example.edu.br* are both registered names under the TLD *.br*, but have different eTLDs.

In summary, two of the DNS data sets involve longitudinal data, measured at regular intervals, and covering 13 months. By combining the three sources we can compare results from different vantage points. In general, sources seed the active measurement from multiple input sources and resolve at least the A, AAAA and NS records of domains. While data sets overlap in their seed, each data set includes unique domains extending our view. For example, one project primarily uses full TLD zones as seed, the other uses partial seeds which extends coverage into other ccTLDs. It is worth noting that some collection efforts include, to some extent, fully qualified domain names. To increase comparability of the input data, we focus on registered domains only. In doing so, our analysis will produce a comparable lower bound of parked domains.

[1]Due to a system change, no data is available for week 25 & 26 of 2021. This does not influence later weeks due to the independence of scans.

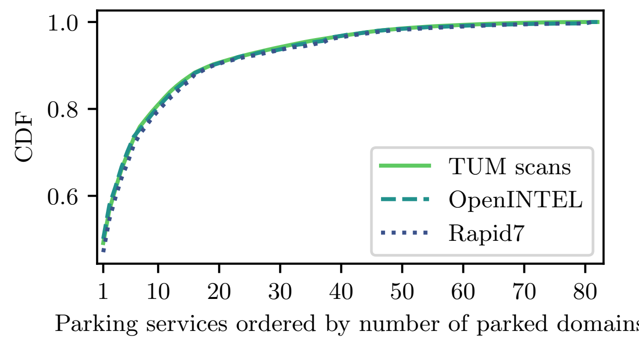

Figure 1: Distribution of ∼60 M parked domains across services on Jan 28th, 2022. Note the y-axis does not start at 0.

IV. Analysis of Domain Parking

We initially use snapshots from all three data sources for Jan 28th, 2022 and evaluate the prevalence and impact of parking services on resolved domains. In the TUM scan data, 60.7 M of 267 M successfully resolved domains (23 %) can be mapped to parking services. Using OpenINTEL data, we infer 59.7 M of 228 M (26 %) domains as parked. Finally, Rapid7 data reveals 61 M out of 332 M domains parked (18 %). Rapid7 contains more domain names because of its coverage (see Section III). However, the number of parked domains we infer is on a par. The additional domains involve 35 M AWS domains and 11 M *.blogspot.com*, among others. These can not be used freely in combination with external infrastructure and thus cannot be parked. We stress that the number of parked domains we infer forms a lower bound, because our list of parking services does not encase every possible provider, and the DNS data is not without end. However, we provide a present-day look at prevalence and report nearly eight times more parked domains than Vissers *et al.* [7], who inferred 8 M parked domains for 15 services in 2015. Note that their work relied on a historic, cumulative data set covering registered domains from two years and the number of total registered domains during their validation of identified parked domains is not available. Thus, we can not argue about the fraction of parked domains in this comparison. Also, the total number of existing domains is expected to have increased since 2015.

Figure 1 shows the presence and distribution of parked domains under the 82 services considered. Minor input differences aside, the results are largely similar for all three DNS data sources, which shows that our inferences are consistent among data collected at different vantage points.

Table I shows the 10 most prevalent parking services, along with their category, number of parked domains, number of eTLDs (*i.e.*, public suffixes), and the hosting organization. GoDaddy is the predominant service with more than 32 M domains. We bisect GoDaddy into free and paid *CashParking* services, managing 30 M and 2 M parked domains respectively. GoDaddy increased their parking operations over the last years by acquiring other services, *e.g.*, AfterNic and SmartName [17]. While AfterNic is operated on discernible infrastructure,

Table I: Top 10 parking services with the number of parked domains and covered eTLDs based on all three input sources on January 28th, 2022. The web hosting locations are based on A/AAAA records for parked domains.

		TUM Scans		OpenINTEL		Rapid7		Web Hosting	
Service	Category	Domains	eTLDs	Domains	eTLDs	Domains	eTLDs	ASN	Organization
GoDaddy (Free Parking)	Parking	29.95 M	571	29.89 M	510	28.92 M	792	15169	Google
HugeDomains.com	Marketplace	4.62 M	6	4.61 M	21	4.61 M	6	16509	Amazon
Sedo	Mixed	2.97 M	671	2.51 M	573	3.09 M	877	47846	Sedo
Skenzo	Parking	2.76 M	574	2.73 M	502	2.77 M	832	40034	Confluence Networks
GoDaddy (CashParking)	Parking	2.19 M	521	2.17 M	469	2.16 M	587	15169	Google
dan.com	Marketplace	2.14 M	640	1.97 M	552	2.47 M	879	16509	Amazon
ParkingCrew	Parking	1.62 M	725	1.04 M	576	1.56 M	1031	61969	Team Internet
Bodis	Parking	1.08 M	638	1.00 M	552	1.13 M	795	16509	Amazon
survey-smiles.com	Parking	1.04 M	352	1.04 M	310	1.17 M	376	60781[1]	Leaseweb
AfterNic	Marketplace	0.96 M	438	0.95 M	401	0.99 M	477	16509	Amazon

[1] Leaseweb hosts only 23 % of parked domains. 15 further ASes host at least 10 k domains see Section IV-B.

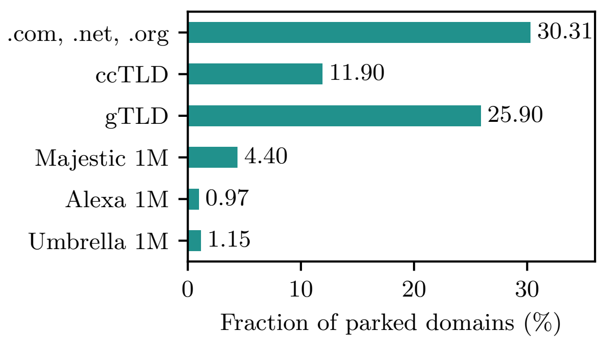

Figure 2: Fraction of parked in relation to resolved domains per input on January 28th, 2022.

SmartName is run on the same infrastructure as GoDaddy (CashParking).

The second rank, *i.e.*, HugeDomains.com is a domain marketplace but does not provide selling services to customers and only sells its own names (see Section II). We observe that they hold more than 4.6 M domains from a comparably small set of eTLDs. The third-largest service, Sedo, a German domain parking service, is categorized as mixed. Besides general, advertisement-focused parking, it also offers a marketplace and domain brokerage service. Both run on the same infrastructure, which hinders a clear distinction made solely based on DNS information. The leading services mainly showing placeholders (*i.e.*, no monetization) are Alibaba with 390 k domains followed by 123 Reg with 270 k domains (ranked 19 and 21). While 15 services already cover 85 % of parked domains, differences between our services and the ones identified by Vissers *et al.* [7] show a significant shift in the domain parking ecosystem, *e.g.*, GoDaddy is not included in their list while Namedrive was bought by and integrated into ParkingCrew since 2015 [18].

We also investigated the impact of domain parking on different scan seeds and eTLDs. Figure 2 shows the fraction of parked domains in relation to all domains accounted for per input of the TUM scans as described in Section III. In general, parking is most prevalent under the legacy gTLDs *.com*, *.net* and *.org* which contain roughly 30 % parked domains, followed by other gTLDs and then ccTLDs. Interestingly, even top lists include parked domains. On Jan 28th, 2022, the Majestic Top 1M [12] contained 40 k (4.4 %) parked domains, 15 of which had Top 10 k ranks. In the Alexa Top 1M [11], 5.3 k domains were parked. In general, most parked domains in top lists are on a rank above 100 k. A large fraction of parked domains listed on the Majestic Top 1M is parked throughout most of our measurement period and simultaneously listed. A higher churn can be seen in combination with other top lists. We assume the Majestic list to be more broadly impacted due to differences in the generation of top lists. The Majestic Top 1M relies on web crawls and rates pages based on link metrics. This might be influenced due to efforts to showcase parked domains and increase revenue from page visits. Advertising a parked domain by placing a link onto a large variety of web pages to attract visits can result in a ranking, especially with the Majestic Top 1M. In contrast, the Alexa Top 1M requires active page visits by users recorded by its toolbar and the Umbrella Top 1M is based on visible DNS requests and includes automatically generated API requests This shows that domain parking even affects the large amount of studies solely relying on top lists.

Table II: Top 10 eTLDs with more than 500 k resolving domains based on the percentage of parked domains on Jan 28th, 2022.

		Parking	
eTLD	Total Domains	Domains	%
app	0.54 M	0.29 M	54.24
co	1.26 M	0.59 M	47.34
us	1.13 M	0.45 M	39.95
vip	0.52 M	0.19 M	37.10
club	0.77 M	0.27 M	35.25
info	3.23 M	1.09 M	33.89
in	0.87 M	0.27 M	31.37
com	138.58 M	42.37 M	30.57
me	0.51 M	0.15 M	29.76
org	9.06 M	2.62 M	29.00

The domains in TUM input data involve 4095 eTLDs in total. We observe one or more parked domains under 901. We study the prevalence of parked domains under eTLDs with more than 500 k domains in total and rank by percentage. Table II shows

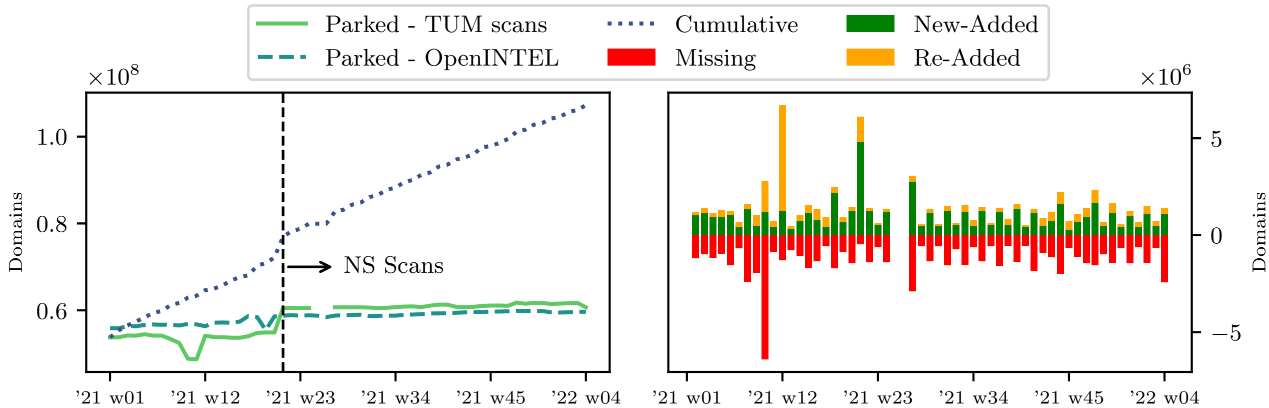

Figure 3: Development of parked domain names over time (Jan 2021 through Jan 2022). The gap in week 25 and 26 is due to scan system changes (see Section III).

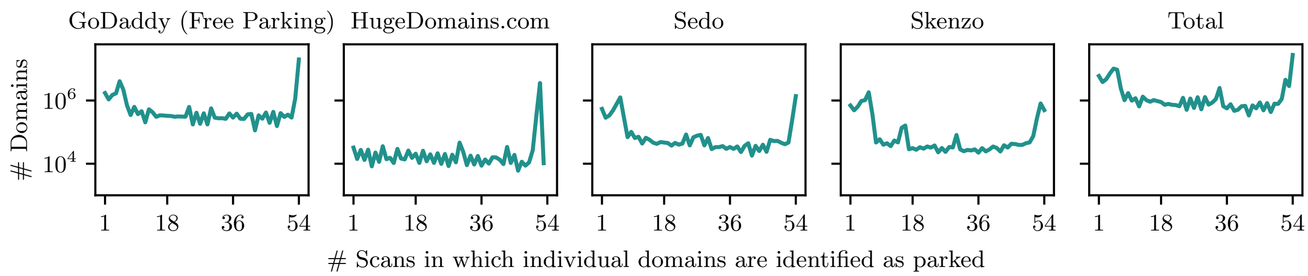

Figure 4: Lifetime of parked domains in number of scans for the Top 4 services and all 82 services combined from Jan 2021 until Jan 2022. All results are based on the TUM scans. Note the y-axis does not start at 0.

the results. *.app* contains the highest percentage of parked domains: 54.23 % (298 k of 550 k). *.com* accounts for most parked domains overall with 42.4 M. The eTLD *.realty*, which is not tabulated because its size is 13.2 k, even contains 91.4 % (12 k) parked domains. A majority of these is registered and parked by a single service, *i.e.*, epik.

Key take-away: *Parked domains are present in large numbers (∼60 M) and as significant fractions of a variety of eTLDs (e.g., 31 % of .com). Furthermore, they are administered by a few services, which can drastically influence the appearance of the Internet.*

A. Development of Domain Parking over Time

We analyze the development of parking services and parked domains over a 13-month period (Jan 2021 through Jan 2022). Figure 3 shows the total number of parked domains in data from TUM and OpenINTEL for each week, and the cumulative number of distinct parked domains seen in between Jan 2021 and each scan. The second part shows the differences between two consecutive scans, *i.e.*, how many domains are not parked anymore or are newly parked. The latter is further divided into domains parked before based on TUM data and new, previously unknown parked domains. No general reason why domains are not parked anymore can be given. They are evenly distributed across three reasons: *(i)* they are not part of zone files used as input anymore; *(ii)* they do not resolve during that respective scan, and *(iii)* they resolve but are classified as not parked. The daily number of parked domains remains relatively stable (1-2 % change between scans similar to the overall change in input domains), but the overall parked names (cumulative) learned keeps growing. The trends observed in TUM scan data are confirmed by OpenINTEL data (dashed line).[2]

While the total number of parked domains remains relatively stable, individual services can involve noticeable change. Figure 4 shows the number of scans during which individual domains are parked with a service. In general, most domain names can be identified as parked throughout many scans up to the complete measurement period (54 scans), thus they remain parked with their parking provider for extensive periods. For example, most domains from HugeDomains.com are parked throughout all but two scans. Similarly, 19.9 M (66.6 %) of domains parked with GoDaddy (Free Parking) are visible throughout the 13-month period. However, we observe occasional changes in the portfolios of parking providers, resulting in domains only identified as parked in few scans. For GoDaddy (Free Parking), these account for 12.3 M distinct domains, identified as parked in at most six scans.

[2]The rise in daily parked domains in May 2021 is due to improvements to TUM scans (see Section III) that increase the lower bound inference.

Table III: Top-10 ASes with web hosting based on number of parked domains. The general rank is based on the total number of domains relying on web hosting in the Autonomous System (AS). R_c is a cleaned rank if parked domains are excluded.

		General		Parking		
ASN	Organization	Rank	Domains	Domains	%	R_c
15169	Google	1	40.73 M	32.33 M	79.4	3
16509	Amazon	2	24.71 M	8.76 M	35.5	1
14618	Amazon AES	7	5.53 M	2.94 M	53.1	12
40034	Confluence N.	14	2.91 M	2.88 M	98.9	427
19324	DOSArrest	43	0.84 M	0.84 M	99.6	1140
29873	Newfold	15	2.29 M	0.76 M	33.2	19
46606	Unified Layer	5	5.74 M	0.72 M	12.7	8
20857	Trans IP	37	0.96 M	0.51 M	53.7	61
24940	Hetzner	11	3.23 M	0.47 M	14.7	11
63949	Linode	28	1.22 M	0.41 M	33.8	38

Table IV: Top-10 ASes containing name servers based on number of parked domains. The general rank is based on the total number of domains delegated to a name server in the AS. R_c is a cleaned rank if parked domains are excluded.

		General		Parking		
ASN	Organization	Rank	Domains	Domains	%	R_c
44273	GoDaddy DNS	1	57.88 M	33.08 M	57.1	2
16509	Amazon	3	17.76 M	8.25 M	46.4	8
14618	Amazon AES	13	8.63 M	6.44 M	74.6	24
47846	Sedo	29	1.98 M	1.87 M	94.4	204
13335	Cloudflare	2	27.09 M	1.59 M	5.8	1
40034	Confluence N.	51	1.02 M	0.96 M	94.2	310
33438	Highwinds	52	0.97 M	0.93 M	96.2	408
397238	NeuStar	9	9.99 M	0.68 M	6.8	9
397220	NeuStar	10	9.95 M	0.68 M	6.8	10
397213	NeuStar	11	9.94 M	0.68 M	6.8	11

Domains, identified as parked only for a short amount of time, are often parked by drop-catch services (cf. Lauinger *et al.* [19]) shortly after expiration or by registrars themselves after cancellation by a customer until final expiration. This effect is more clearly visible for, *e.g.*, Sedo and Skenzo compared to HugeDomains.com, resulting in a lower stability of parked domains (cf. Figure 4). Our published list of services allows to analyze the life cycle of domains in more detail in the future.

As a side note: we take the overall stability of parked domains over time as a sign that the DNS-based indicators that we developed (see Section II) remain valid throughout. This also suggests that the list that we share will not quickly lose its value. It can of course be amended in the future.

Key take-away: *The overall number of parked domains remains stable in both long-running data sources and the majority of domains is parked throughout the complete period. However, a frequent change of 1-2 % of domains per week is visible.*

B. Service Infrastructure

We posit that domain parking mainly affects three infrastructure components: *(i)* Web hosting; *(ii)* the DNS; and *(iii)* monetization means (*e.g.*, advertisement systems). We focus on the first two components. The latter was analyzed by Alrwais *et al.* [20].

Note, that due to the TUM scan spanning more than 24h, distributing resolved domains and records evenly, some domains, identified using specific records might not be parked during the scan of all remaining records anymore/yet. However, as shown in Section IV-A, most domains remain stable over time and are seen in multiple scans.

Web hosting Parking services can use their own infrastructure to host parked pages or rely on external, large providers such as Google or Amazon. Table III shows the Top 10 ASes ranked by the number of hosted parking pages, and the total number of hosted domains and associated ranking.

Multiple prominent hosting locations are extensively used by domain parking services. All domains parked with GoDaddy (Free Parking) are hosted within Google and all of dan.com are within Amazon. In fact, for 72 services at least 95 % of the parked pages are hosted in a single AS. Evaluating the effects of these ASes on Internet centralization without taking parking into account can bias results. For example, consider that Google (AS15169) is the most-used AS based on DNS resolutions in total and by parked domains. In TUM data, 40.7 M domains in total resolve to a Google IP address, but GoDaddy relies on Google Cloud (see Table I) and as such 32.1 M (78.9 %) are in fact parked. Such effects were recognized by Zembruzki *et al.* [4] who found that in recent years GoDaddy (Free Parking) switched from self-hosting to Google Cloud, which drastically changed their view on hosting centralization.

Amazon (AS16509) is second in rank and is used by multiple services, including HugeDomains.com, dan.com and Bodis (see Table I). However, only 35.5 % domains in Amazon (8.8 M out of 24.7 M) are parked. Confluence Networks, ranked 14 in terms of overal hosting, contains 98.9 % parked domains. This is comparable to DOSArrest with 99.6 %.

While most services rely on a single hosting location, some services distribute domains over several ASes. Among all parking services, survey-smiles.com shows the most distribution. They have 1 M parked domains and host at least 10 k domains using 16 different providers.

DNS provider The mode of operation of a parking service determines how parked domains need to configure their DNS, which for example involves name server delegation (see Section II). Some services use their name servers not only for parked domains but also other resources. GoDaddy (Free Parking) is an example. We can use the NS records of parked domains to analyze the DNS infrastructure used by parked domains and how this infrastructure relates to parking services. Table IV shows the Top 10 ASes in which authoritative name servers for parked domains are located, as well as the total number of domains that use this DNS infrastructure and the associated rank.

GoDaddy DNS is authoritative for the highest number of domains in the TUM data set in total with 57.8 M domains out of 260 M. However, 33.1 M (57.1 %) out of these domains are parked. The AS is mostly used by domains parked with GoDaddy itself, but also, *e.g.*, by 960 k domains of AfterNic, which GoDaddy owns. In contrast, Cloudflare ranks second and is authoritative for 27 M domains out of which *only* 1.6 M (5.8 %) are parked. Similarly to web hosting, Amazon is used by a variety of parking services, *e.g.*, dan.com (2 M parked domains), Sedo (1.7 M) and ParkingCrew (1.4 M). Sedo (mostly used by Sedo itself), Confluence Networks (mostly Skenzo) and Highwinds (mostly Bodis) are in the Top 60 of name server hosting ASes and are used almost exclusively by parked domains (94% – 96%).

Key take-away: *Domain parking accounts for large fractions of domains hosted within large providers and of domains delegated to authoritative name servers within well known DNS providers, e.g., Google, Amazon or Linode. Therefore, they directly influence the analysis of commonly selected, important organizations.*

C. Verification and Content Similarity

We take a number of steps towards validating how well the parking service reference points that we developed (in Section II) fare in terms of enabling the identification of parked domain names.

Visual identification First, we sample 1 k parked domains per service and visit them to visually confirm correctness. We automated browsing using Puppeteer[3], set it to visit all sampled names, and take a screenshot of the landing page if available. Due to timeouts or rate limiting, some services resulted in fewer screenshots. For most services, we were able to capture more than 80 % of sampled websites. The remaining services resulted in a smaller success rate most likely due to stricter rate limits. No correlation between timeouts and a specific service category (see Section II) can be seen. All successful samples showed expected parking pages containing advertisements, sales banners or placeholders. Due to the limited content on parking pages and similarities of pages within each service, manual verification of this set was possible by us. Evidently, visiting 60 M parked domains, taking screenshots, and visually inspecting them is not feasible.

Common Crawl While visual identification allowed us to validate the inferences to a reasonable extent, we wanted to upscale validation. Therefore, we consider Common Crawl (CC) data [21] and calculate the similarity of pages. Common Crawl is an open repository of web crawl data, collected at monthly intervals, accounting for hundreds of millions of unique domain names, and many more URLs.

We consider CC data for Jan 2022 and the $\sim$60 M parked domains that we identify on Jan 28th, 2022. We extract the HTML content of parked pages from CC data, only considering URLs that contain *exactly* the registered domain. Furthermore, we require the crawl target to have been the landing page (*i.e.*, the path of the URL is /) and also to have resulted in a useful response (*i.e.*, HTTP status code of 200). Given these filters, $\sim$1.29 M HTML rich responses can be obtained. We extract visible text and tokenize it into words, remove stop words, apply lemmatization, and create a vector for the most-frequently used words for each page. Using these vectors and the MinHash algorithm, we estimate the Jaccard similarity between each and every pair of pages. We then look at the average similarity (ranging from 0 to 1) for pairs of parked domains under the same provider, as well as inter-provider pairs. For the vast majority of the parking services, the intra-provider similarity is significantly higher (often nearing 1) than the inter-similarity (often near 0). Combined with visual identification, it becomes apparent that all are indeed correctly identified. Some services send an HTTP redirect and thus not a status code of 200. Domains parked with HugeDomains.com, *e.g.*, redirect to `www.hugedomains.com/domain\_profile.cfm?d=...`. For these cases, we looked at the *fetch_redirect* value in the CC data. By counting common occurrences of domains in redirect URLs and presuming that this leads to similar pages, we also confirmed identification.

Key take-away: *The verification shows, that our collected list of services and DNS indicators results in correctly identified parked domain names. It highlights the marginal importance to users due to the high similarity of content within each service.*

[3]https://github.com/puppeteer/puppeteer

V. Related Work

Domain parking as a service was previously analyzed by Alrwais *et al.* [20] focusing on the customer perspective, analyzing monetization chains and potential malpractice. They used passive DNS data and a hitlist of name server records to find monetization chains. Vissers *et al.* [7] in turn focused on the user perspective. They also infer the use of parking services on the basis of DNS indicators. Based on the DNS Census data covering 106 M registered domains collected throughout two years, they identify 7.5 M parked domains for a collection of 15 parking services. They further analyzed the identified domains with respect to typo-squatting and malicious behavior. They propose a set of features derived from identified parking pages which could be used as basis for a browser-based classifier. We offer an extended and updated list of providers, resulting in a number of parked domains nearly an order of magnitude larger. We also focus on the impact of ignoring the special role of parked domains can have on research, in particular as it relates to centralization.

Kührer *et al.* [22] identified domain parking as reason for up to 10 % of blocklist entries. Similar to Vissers *et al.* [7], they design a classifier based on websites to identify parked domain names. However, they focus on blocklists and top lists and train their classifier with a relatively small input set. We focus on a general quantification of domain parking and its impact on further areas.

Domain parking was also observed as part of analyses of specific (*i.e.*, *.xxx* and *.biz*) and newly introduced (starting in 2013) gTLDs [10], [23], [24], the practice of domain registrars [19], [25], and the analysis of hosting providers [4]. In various works, Halvorson *et al.* [10], [23], [24] have shown that around 30 % of domains from gTLDs are parked and even 23 % of *.com* is parked. We identify a larger fraction for the established TLDs *com*, *net*, *org* (30 %) but slightly less parked domains on average for all remaining, available gTLDs (25 %). Tammy *et al.* [25] and Lauinger *et al.* [19] identified parking as common practice of registrars as well, promoting their portfolio and collecting expired domains. We include dedicated parking services but also services from registrars into our analysis, and extend evaluation further, to domains from ccTLDs and all available gTLDs, including traditional gTLDs, *e.g.*, *.com* but also newly registered gTLDs. Our DNS scans cover more than $1\,\mathrm{k}$ gTLDs compared to 502 investigated by Halvorson *et al.* [10]. Zembruzki *et al.* [4], saw domain parking as influential factor on the analysis of hosting centralization but did not investigate its impact.

VI. Discussion

Our results underpin the prevalence of parked domains. Taking *.com* as example, up to $31\,\%$ of names are parked. Evidently, analyses that indiscriminately rely on domain names to identify infrastructure (*e.g.*, hosting or DNS) can inadvertently be biased by parked domains. So even though most parked domains are part of the web, we argue that their special characteristics need to be taken into account.

The hosting of advertisements and sales banners on parked pages results in large numbers of similar pages (see Section IV) that are centralized on the same infrastructure. For example, GoDaddy (Free Parking) uses only two IP addresses to host more than $29\,\mathrm{M}$ domains. The content of these pages only has circumstantial benefits for visitors. Furthermore, our analysis of used DNS infrastructure (see Section IV-B) highlights potential impact. While more domains would be affected in total in case of a service disruption of GoDaddy DNS, $57.1\,\%$ of these are parked. This, we argue, would not be as consequential as a Cloudflare disruption, which would affect considerably more non-parked (and thus user-centric) domains.

Due to their economical value and prevalence, parking services and parked domains should of course still be the subject of future research. However, we propose that, based on their specific appearance and client value, they need to be classified as a specific asset and evaluated as such.

A re-evaluation of findings based on our work can help to better understand the ecosystem and support future discussions regarding Internet consolidation. Studies, *e.g.*, on web [1], [5] and DNS [3], [26] consolidation often already rely on large DNS data sources but do not consider parked domains, leaving their role for future work, if recognized. Our collected parking indicators can be used by research to filter own DNS resolutions or in combination with available DNS data, *e.g.*, from OpenINTEL and Rapid7. Future research utilizing top lists should filter parked domains using our DNS indicators.

While results are not completely overturned if domain parking is considered properly, its effect on results is clearly visible, as seen in changes of ranks in Table III and Table IV. Even for top providers, changes are visible. As an example, Zembruzki *et al.* [4] saw a massive shift of hosted domains from GoDaddy to Google Cloud in between 2020 and 2021, and pinpoint this event to domains parked with GoDaddy. This finding is supported by our results in Section IV. However, they miss further effects of domain parking, *e.g.*, classifying Confluence Networks (AS40034) similar to other hosting providers. As shown in Table III, its relevance drops significantly if parked domains are excluded.

Besides the evaluation of infrastructure and hosting providers, domain parking can bias analyses of new protocol deployments, *e.g.*, regarding TLS 1.3 [27] or QUIC [28]. Both studies show that mainly large providers drive the deployment of new protocols. However, they do not consider the effect of domain parking. While the most influential service GoDaddy does neither support TLS 1.3 nor QUIC at the moment, it can change the deployment status drastically if these protocols are supported. Zirngibl *et al.* [28] identified around $30\,\mathrm{M}$ domains supporting QUIC in May 2021. The domains parked with GoDaddy (Free Parking) alone could nearly double these findings, if GoDaddy starts to support QUIC.

Ethical Considerations: For scans, we strictly followed a set of ethical measures, *i.e.*, informed consent [29] and community best practices [30]. To not burden network operators unnecessarily, we used independent and (semi)publicly available data sources to provide other vantage points, rather than conducting unnecessary scan campaigns ourselves. All our scans are conducted with a limited rate and use a request based blocklist. Furthermore, our measurement vantage point is clearly identified based on reverse DNS, WHOIS information and a hosted website indicated as measurement. We did not receive any inquiries related to our scans during this work. Our study does not relate to users, personally identifiable information, or otherwise privacy-sensitive data. We focus on publicly reachable and available services. While sensitive data may occasionally be present in such source (*e.g.*, the DNS), this is in no way in focus.

VII. Conclusion

In this paper, we analyzed the prevalence of domain parking on the Internet. We systematically collected 82 parking services and used large-scale and longitudinal DNS data sets, representative for a sizable part of the global DNS namespace and collected at different vantage points, to identify $60\,\mathrm{M}$ parked domains. Domain parking accounts for $20\,\%$ to $30\,\%$ of registered domains in most available gTLDs, including *.com*, *.net* and *.org*, and does not show any sign of decline throughout our 13-month observation period.

Our findings show that most parked domains are concentrated on a few services, often consistently relying on single hosting and DNS infrastructure locations. Furthermore, because parking services largely focus on advertisements and domain sales, the content of parked pages is highly similar, while only circumstantially relevant for visitors. This highlights our initial proposition, that while domain parking is of low importance for most users and the Internet in general, it represents a large part of the DNS and web ecosystem, and can introduce bias in analyses that treat parked domains the same as any other. Major shares of domains hosted within, *e.g.*, Google Cloud or Amazon AWS or served by name servers within GoDaddy are only parked domains. Analyzing the impact of these on the Internet in general and its consolidation towards certain providers is drastically influenced by domain parking.

Acknowledgments

The authors would like to thank the anonymous reviewers and our shepherd Matt Calder for their valuable feedback. This work was partially funded by the German Federal Ministry of Education and Research under the project PRIMEnet (grant 16KIS1370), and by the EU H2020 CONCORDIA project (830927). This research used data from OpenINTEL, a project of the University of Twente, SURF, SIDN, and NLnet Labs.

References

[1] N. P. Hoang, A. A. Niaki, M. Polychronakis, and P. Gill, "The Web is Still Small after More than a Decade," *ACM SIGCOMM Computer Communication Review*, 2020.

[2] P. Gigis, M. Calder, L. Manassakis, G. Nomikos, V. Kotronis, X. Dimitropoulos, E. Katz-Bassett, and G. Smaragdakis, "Seven Years in the Life of Hypergiants' off-Nets," in *Proc. ACM SIGCOMM*, Virtual Event, USA, 2021.

[3] S. Wang, K. MacMillan, B. Schaffner, N. Feamster, and M. Chetty, "A First Look at the Consolidation of DNS and Web Hosting Providers," 2021. [Online]. Available: https://arxiv.org/abs/2110.15345

[4] L. Zembruzki, R. Sommese, L. Z. Granville, A. S. Jacobs, M. Jonker, and G. C. M. Moura, "Hosting Industry Centralization and Consolidation," 2022. [Online]. Available: https://arxiv.org/abs/2109.01187

[5] T. V. Doan, R. van Rijswijk-Deij, O. Hohlfeld, and V. Bajpai, "An Empirical View on Consolidation of the Web," *ACM Trans. Internet Technol.*, vol. 22, no. 3, feb 2022.

[6] I. Society. (2019) Consolidation in the Internet Economy. [Online]. Available: https://future.internetsociety.org/2019/

[7] T. Vissers, W. Joosen, and N. Nikiforakis, "Parking Sensors: Analyzing and Detecting Parked Domains," in *Proc. Network and Distributed System Security Symposium (NDSS)*, San Diego, CA, 2015.

[8] GoDaddy. (2022) Park a domain registered with GoDaddy. [Online]. Available: https://www.godaddy.com/help/park-a-domain-registered-with-godaddy-23936

[9] S. Deusch, "Analyzing the Effect of Domain Parking on DNS Based Research," Bachelor's Thesis, Technical University of Munich, 2022.

[10] 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 *Proc. ACM Int. Measurement Conference (IMC)*, Tokyo, Japan, 2015.

[11] Alexa. (2022) Top 1M sites. [Online]. Available: https://www.alexa.com/topsites

[12] Majestic. (2022) The Majestic Million. [Online]. Available: https://majestic.com/reports/majestic-million/

[13] Cisco. (2022) Umbrella Top 1M List. [Online]. Available: https://umbrella.cisco.com/blog/cisco-umbrella-1-million

[14] R. van Rijswijk-Deij, M. Jonker, A. Sperotto, and A. Pras, "A High-Performance, Scalable Infrastructure for Large-Scale Active DNS Measurements," *IEEE Journal on Selected Areas in Communications*, 2016.

[15] Rapid7 Labs. (2022) Forward dns (fdns). [Online]. Available: https://opendata.rapid7.com/sonar.fdns_v2/

[16] Mozilla Foundation. (2022) Public Suffic List. [Online]. Available: https://publicsuffix.org/

[17] GoDaddy. (2013) GoDaddy Acquires Afternic - Primes Domain Aftermarket for New TLDs Move Also Delivers Improved Service, Selection & Speed to More Aftermarket Customers. [Online]. Available: https://aboutus.godaddy.net/newsroom/press-releases/press-release-details/2013/GoDaddy-Acquires-Afternic---Primes-Domain-Aftermarket-for-New-TLDs-Move-Also-Delivers-Improved-Service-Selection--Speed-to-More-Aftermarket-Customers/default.aspx

[18] D. N. Wire. (2015) ParkingCrew acquires NameDrive. [Online]. Available: https://domainnamewire.com/2015/07/16/parkingcrew-acquires-namedrive/

[19] T. Lauinger, A. Chaabane, A. S. Buyukkayhan, K. Onarlioglu, and W. Robertson, "Game of Registrars: An Empirical Analysis of Post-Expiration Domain Name Takeovers," in *Proc. USENIX Security Symposium*, Vancouver, BC, 2017.

[20] S. Alrwais, K. Yuan, E. Alowaisheq, Z. Li, and X. Wang, "Understanding the Dark Side of Domain Parking," in *Proc. USENIX Security Symposium*, San Diego, CA, 2014.

[21] C. Crawl. (2022) The Common Crawl Corpus. [Online]. Available: https://commoncrawl.org/

[22] M. Kührer, C. Rossow, and T. Holz, "Paint It Black: Evaluating the Effectiveness of Malware Blacklists," in *Research in Attacks, Intrusions and Defenses*, Cham, 2014.

[23] T. Halvorson, K. Levchenko, S. Savage, and G. M. Voelker, "XXXtortion? Inferring Registration Intent in the .XXX TLD," in *Proceedings of the 23rd International Conference on World Wide Web*, Seoul, Korea, 2014.

[24] T. Halvorson, J. Szurdi, G. Maier, M. Felegyhazi, C. Kreibich, N. Weaver, K. Levchenko, and V. Paxson, "The BIZ Top-Level Domain: Ten Years Later," in *Passive and Active Measurement*, Berlin, Heidelberg, 2012.

[25] T. Alexander, "Domain Name Registrars: Are They Part of the Domain Name Fraud Problem?" in *Proceedings of the 3rd Annual Conference on Information Security Curriculum Development*, Kennesaw, Georgia, 2006.

[26] M. Allman, "Comments on DNS Robustness," in *Proc. ACM Int. Measurement Conference (IMC)*, Boston, MA, USA, 2018.

[27] 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," *ACM SIGCOMM Computer Communication Review*, vol. 50, no. 3, 2020.

[28] J. Zirngibl, P. Buschmann, P. Sattler, B. Jaeger, J. Aulbach, and G. Carle, "It's over 9000: Analyzing early QUIC Deployments with the Standardization on the Horizon," in *Proc. ACM Int. Measurement Conference (IMC)*, Virtual Event, USA, 2021.

[29] D. Dittrich, E. Kenneally *et al.*, "The Menlo Report: Ethical principles guiding information and communication technology research," *US Department of Homeland Security*, 2012.

[30] C. Partridge and M. Allman, "Addressing Ethical Considerations in Network Measurement Papers," *Communications of the ACM*, vol. 59, no. 10, Oct. 2016.