The Evolution of Top-Level Domains: A Comparative Study of .org and .dev

Florian Pfisterer, Johannes Zirngibl*, Patrick Sattler*
*Chair of Network Architectures and Services
School of Computation, Information and Technology, Technical University of Munich, Germany
Email: florian@pfisterer.dev, zirngibl@net.in.tum.de, sattler@net.in.tum.de

***Abstract*—The Domain Name System (DNS) is key to the modern Internet. Many studies investigate the changes of domain names over time. We argue that the the development of top level domains (TLDs) as a whole can introduce bias into such DNS measurements, leading researchers to draw conclusions about patterns in the data that actually stem from how the TLD changes over time. Additionally, one needs to take the churn of a given TLD into account when assessing the validity of DNS data given its age.**

To investigate how DNS data change varies across TLDs, we analyze the recent development of two very different TLDs between 2019 and 2023. By studying the zone files, we compare the almost 40 year old .org TLD with the much newer .dev TLD first made available only in 2019.

We aggregate zone file data over time and show that .org and .dev differ in a variety of aspects – .org is almost 30 times larger, grows only about three percent as fast, and exhibits 29 % less churn. We introduce the idea of the "core" of a zone (the domains that are registered for the full investigated period) and find that .org's core is 14 % larger than .dev's core, suggesting that we can consider .org a more stable TLD.

***Index Terms*—top-level domains, DNS, CZDS, zone files**

1. Introduction

Centralized Zone Data Service (CZDS) [1] is an online portal provided by the Internet Corporation for Assigned Names and Numbers (ICANN) that provides access to zone files of a variety of TLDs (including .com, .net and .org). With daily updates to the zone files, it offers the opportunity to study the Domain Name System (DNS) as a whole and to do DNS measurements on a large scale.

It is important for researchers doing such studies (e.g., [2]–[4]) to assess the validity of their measurements, given the age of their data. The correlation between validity and age can vary based on the corresponding TLD: if the TLD has very high churn, more recent data is to be preferred. Futhermore, the development of a TLD as a whole can introduce bias into the patterns that researchers study in their DNS measurement data.

Therefore, it is important to consider the TLD-specific changes within DNS data. In order to examine these differences, we study the development of the old .org TLD (which has been available for almost 40 years [5]) and compare it to the development of a much more recent TLD, .dev (which has only been made available in 2019 [6]). We analyze how both of these TLDs changed from 2019 until 2023 by investigating the zone files for each day provided by the CZDS.

The key research questions we seek to answer are:

- How did the number of resolvable domains in the .org and .dev TLDs develop over this time span?
- How stable are the TLDs, i.e., how many domains stay present in the zone file vs. how much churn is there?
- What is the typical lifetime of a .org domain vs. a .dev domain?
- If domains are not resolvable during the full considered timespan, how many of them are re-added and how long are the periods in between?

2. Background

The DNS is used for resolving readable domain names to IP addresses on the Internet [5]. It uses TLDs such as .org, .com or .dev to partition the name space of domain names [7]. A name server stores different kinds of records for the set of domain names it is responsible for and replies to DNS queries with this information [7].

TLD name servers maintain information on which second level domains (SLDs) (e.g., tum.de) are registered within the TLD and how to resolve them. The collection of all such records stored on a name server is called a zone file.

Following the introduction of many new TLDs and the associated security risks, easier access to such zone files has been required. Thus, the ICANN introduced the CZDS, which is a service that provides easy access to the zone file of all participating TLDs on a daily basis [8].

Domains are usually registered in the corresponding TLD's registry under contract with ICANN [9]. On behalf of the buyer, the registry then communicates the name servers of the domain to the name server maintaining the zone file of the TLD. This allows DNS clients to find the name servers responsible for the domain, which can then provide information to correctly resolve the domain to the corresponding IP address. In this paper, when we say that a domain is *registered* or *resolvable*, we mean that the domain's name servers are present in the TLD's zone file.

3. Related Work

Past surveys have analyzed the kinds of TLDs participating in the CZDS as well as the results of access requests to the zone files [1], [8].

 doi: 10.2313/NET-2023-11-1_17

While most domains are registered for at least one year, Foremski et. al. [10] analyze domains with shorter lifetimes, which they claim are used for abusive purposes. They investigate the time it takes from the registration of new domains until they "die" as well as the most common "causes of death". They do separate analyses of domain mortality per TLD and find that old TLDs such as `.org` or `.com` experience much less "domain deaths" than the newer TLDs like `.party` or `.work` [10].

Similarly, Affinito et. al. [9] analyze the registration periods of domains for different TLDs over a time frame of ten years. They find that in ten prominent TLDs, 95 % of registration periods are exactly the minimum period of one year. They further identify that the share of malicious domain registrations with shorter periods correlates to the corresponding kind of TLD. The data they use comes from zone files in the Domain Zone Database (DZDB) by CAIDA, which provides a historical record of TLD zone files [9].

This paper focuses instead on only two TLDs in detail and does an in-depth analysis and comparison of the stability and churn of domain names in those TLDs.

4. Methodology

This section describes the zone files and the aggregations we perform with them.

4.1. CZDS Zone Files

The zone file that CZDS provides on a daily basis is a compressed, tab-separated file with one entry per line, sorted by domain. Each entry is comprised of the domain name, the cache TTL, the class (usually "IN" for Internet), the type of record (especially relevant for this paper "NS" records for the name servers of a domain) as well as the value of the record (for "NS", this is the domain name of the name server responsible for the domain).

An uncompressed `.org` zone file for a single day in May 2023 is about 1.8 GB large and contains almost 30 M rows. An uncompressed `.dev` zone file for the same day spans 2.1 M entries at 234 MB. For comparison: a zone file of the largest TLD (`.com`) on the same day is 25 GB large and contains over 410 M rows.

Thus, the raw data of the `.org` TLDs for the investigated timespan from November 2019 till May 2023 (1270 days in total) amounts to over 2 TB alone. Efficiently processing and analyzing such large amounts of data in a reasonable amount of time and with reasonable computing resources poses a technical challenge, which is discussed in more detail in the following sections.

4.2. Data Aggregation

To answer the research questions of this paper, computationally intensive aggregations of the raw zone file data are needed. In principle, one could execute all required queries on the raw data itself. Our approach however was to first transform the zone files into a more compact and aggregated form, materialize that form, and then perform our analyses on the aggregates instead. This allowed us to iterate on the downstream analysis tasks more quickly, without constantly perfoming the low-level aggregations.

The two types of aggregates we materialized are: 1) Distinct domains: We de-duplicate the NS records in the zone file by domain and then count the unique number of domains present in the zone file on each day.
2) Registration periods: For each time frame of consecutive days on which a domain was present in the zone file, we store the domain as well as the first and last date of the time frame.

While the first aggregation allows us to investigate the development of the size of the two TLDs as a whole, the second gives us the opportunity to do more fine-grained analyses such as studying the periods between registrations, the registration periods and the churn.

4.3. Data storage

Operating on the raw compressed files would require custom optimized programs to perform the aforementioned aggregations with reasonable computing resources. Thus, we decided to use ClickHouse, an open-source, column-oriented database management system suited for analytical workloads such as our aggregations [11]. It allowed us to simply access the data using SQL.

We imported the raw and compressed zone files provided by CZDS using the `clickhouse-client` CLI to benefit from the built-in parallelization options it offers. With this setup, the import took around 25 seconds per `.org` zone file. Since only NS records are relevant for the research questions of this paper, we filtered the DNS records during the insert process already. By storing the data ordered by domain, we can efficiently partition the full name space and thus make aggregation queries more efficient.

4.4. Challenges

While implementing the pipeline from raw CZDS zone files to final aggregates, we faced several challenges, which are discussed below.

4.4.1. Missing Days. Between November 2019 and May 2023, there have been seven days on which our automated CZDS download process did not successfully acquire the current zone files. This led to incorrect registration periods in the corresponding aggregation (one additional period for each missing day). The query assumed that a missing (domain, date)-entry for such a day meant the domain was not resolvable on that day. To solve this, we instead explicitly determined the missing dates and adjusted the query to not consider registration gaps on those dates.

4.4.2. Registration Periods Aggregation. If one considers only one registration period per domain, the aggregation becomes a simple minimum and maximum of the (domain, date)-entries grouped by domain. However, it can happen that a domain is registered for some period, then it is not present in the zone file for a while, and then comes back later.

Since we need to distinguish separate registration periods, our approach was to group the (domain, date)-entries by domain and collect the sorted dates in a list. We then compare the i-th date with the (i+1)-th date in that list and append a running sum of how often the two dates were

not apart by exactly one day (and the day(s) in between are not part of the missing dates, cf. Section 4.4.1). This running sum provides an ID of the consecutive registration period each (domain, date)-entry corresponds to. Finally, we calculate the minimum and maximum date for each domain and period.

By partitioning the `.org` and `.dev` name spaces based on the domain and by indexing entries by the domain (cf. Section 4.3), we were able to restrict the query to only consider domains in pre-computed disjoint partitions. That way, computing the aggregation query for one out of 1000 `.org` partitions only took 24 seconds on average and was well within the resource constraints.

5. Analysis

Using our two exported aggregations, we examine at the size of the `.org` and `.dev` TLDs as a whole and investigate the domain churn.

5.1. Size Of The TLD

As one can see in Figure 1, the total size of the `.org` TLD grew linearly by about 7 % in the investigated 3.4-year timespan. At the same time, `.dev` more than doubled its size. Notably however, in absolute numbers, `.org` grew by around 700 k domains, while `.dev` added less than 200 k dinstinct domains. In May of 2023, `.dev` still only contained a fraction of the domains of `.org` (368 k `.dev` domains vs. 10.7 M `.org` domains). This shows the large difference both in size and growth of these two TLDs.

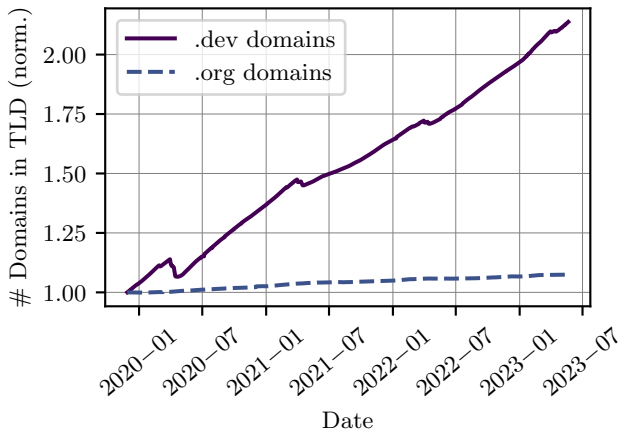

Figure 1: The number of distinct `.org` and `.dev` domains between November 2019 and May 2023. (Both are normalized to the number on November 28, 2019, which is 10 M for `.org` and 172 k for `.dev`)

One further pattern can be observed in the figure: the number of registered `.dev` domains slightly drops in periodic intervals of around one year, each time approximately in April. The magnitude of these drops decreases over the years and is barely visible in 2023. These drops supposedly come from the large amount of initial `.dev` domains registered when the TLD first became available in March 2019 [6]. Domains registrations are mostly sold in one-year increments [9], which is why we observe the small drops exactly one, two, three and four years after the TLD's inception. For `.org`, we do not see this behavior, as it was first available almost 40 years ago [5].

While the number of simultaneously registered `.org` domains increased only by around 7 %, the total number of domains that were present in the `.org` zone file for at least one day during the timespan is around 40 % larger. For `.dev`, the size of the zone file increased by 114 %, but the total number of seen domains is 61 % over that number. This large difference is caused by churn – domains that are registered at some point, and then later again removed from the zone file.

In fact, only 40.4 % of all `.org` domains and 16.1 % of all `.dev` domains ever seen in our data were registered for the full duration. The other 59.6 % of `.org` domains and 83.9 % of `.dev` domains were not present in the zone file on at least one day. We can already see that `.org` displays less churn than `.dev`. Below, we analyze churning domains and the periods in which they were not registered in more detail.

5.2. Periods Between Registrations

Of particular interest are the domains that are registered for some period, are not present in the zone file for a while, and then come back. These domains allow us to study the typical number of days a previously registered domain was not registered for, before it came back. This period represents an attack window for malicious users to publish different name servers and temporarily hijack traffick to a domain [12].

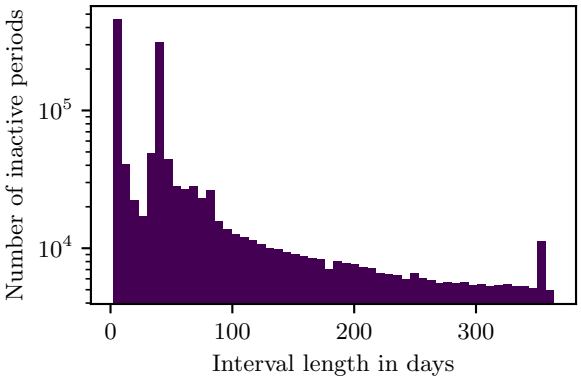

Figure 2: Histogram of the number of days `.org` domains are not present in the zone file after being registered for some duration and before coming back again (logarithmic scale). One vertical bar represents a bucket of 1 week and its height describes the number of periods between registrations we saw whose lengths was this particular number of weeks.

Figure 2 shows a histogram of the duration of these periods in which `.org` domains are not present in the zone file. In total, 1.57 M of such periods from 1.31 M re-appearing `.org` domains were analyzed. Notably, these domains only represent a small fraction of all 10 M `.org` domains. For illustration purposes, the histogram was cut off at a period of 1 year, which represents 88 % of the total data. For `.dev`, we see a very similar pattern, suggesting that the two TLDs do not differ a lot in this regard.

 doi: 10.2313/NET-2023-11-1_17

About a third of periods are less than a week long (most of these – 60 % – are 3 days long). We suspect these do not come from intentional de-registrations of domains, but instead from temporary configuration issues or name server changes.

A second peak occurs at around 35 days. When a domain is not renewed, it is in the redemption grace period status for 30 days, in which it may be re-registered by the previous owner [9]. After that, it takes up to five days until the domain is available to the public again [9]. We assume the second peak comes from domains that were re-registered as soon as they were available again, which is exactly 30 to 35 days after de-registration.

After that, we consistently only see a lower number of period lengths, suggesting that the reason for different inactive durations is just that someone registers the same domain again later. Overall, we can conclude that most attack windows are very short for both `.org` and `.dev`.

5.3. Registration Periods

After analyzing the periods during which domains are not present in a zone file, we study the periods during which domains *are* registered. Figure 3 and Figure 4 show the color-coded fraction of domains still active after a given number of months, differentiated by when the domain was first seen. The first row represents the domains that were present on the first day of our investigated time frame. The other rows represent the domains that were first registered during the month shown on the left (the cohort), while the last row displays an average of all rows. The color in each cell indicates the fraction of domains (between 0 and 1) that are still present in the zone file after the number of months indicated on the x-axis.

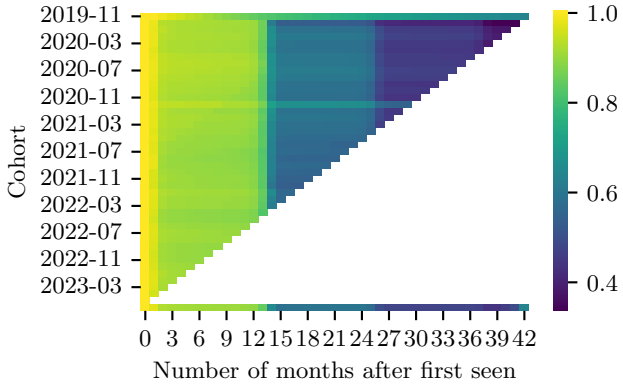

Figure 3: Cohort analysis of `.org` domains from Nov 2019 until May 2023. The last row displays the average over all cohorts.

For both `.dev` and `.org`, one can clearly observe the significant drops in the fraction of registered domains after one, two, and three years for all cohorts except the first – which is unsurprising, given that domain registrations are mostly sold in one year increments [9]. The one- to two-month deviations from the full 12 –month cycle that can be observed in Figure 3 come from auto-renew grace periods during which domains may still be in the zone file.

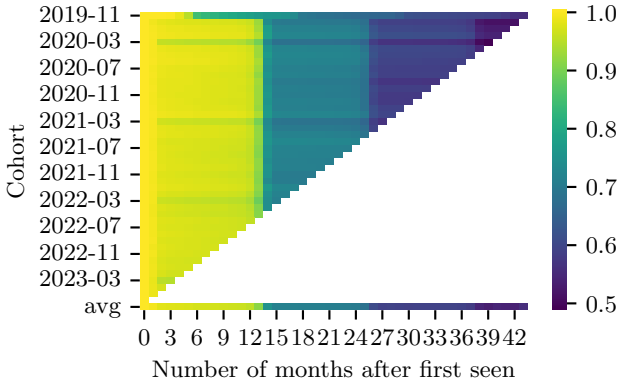

Figure 4: Cohort analysis of `.dev` domains from Nov 2019 until May 2023. The last row displays the average over all cohorts.

The domains represented by the rightmost square in the first row can be considered the "core" of the TLD – these have been in the zone file without interruption since the start of our investigated time frame in 2019. Most notably, 63.6 % of `.org` domains present in the zone file at the start of the investigated time frame are still present at the end. Factoring in that additionally some of the domains that were present at the start but not for the full period might be so because of configuration errors (and the many short 3-day periods between registrations, cf. Section 5.2) we can conclude that the core of `.org` domains is quite stable.

For `.dev`, slightly less domains are there for the full time frame (55.6 %), suggesting that it has a smaller stable core. In the first row of the cohort analysis of `.dev` domains (Figure 4), we can additionally observe the periodic drops in April of each year for the core. This behavior supposedly comes from the considerable share of `.dev` domains that were initially registered when the TLD first became available (as previously discussed in Section 5.1).

6. Conclusion and Future Work

In this paper, we analyzed the development of the old `.org` TLD between November 2019 and May 2023 and compared it to the development of the new `.dev` TLD during the same time span.

We found that the development of `.dev` and `.org` differs in a number of factors. Firstly, `.org` is a much larger TLD, containing 10.7 M domains vs. `.dev`'s 368 k domains in May 2023. Secondly, `.org` shows much less relative growth - it only increased its size by 7 % in the investigated time frame, while `.dev` more than doubled its size. Thirdly, while analyzing the registration periods, we recognized several patterns – the yearly registration period increments as well as `.dev`'s initial registrations in March of 2019. Finally, we showed that `.org` has a larger core of domains that have been registered without interruption since 2019, shows significantly less churn and thus can be considered more stable than `.dev`.

Future work may analyze other large TLDs such as `.com` or a regional TLD such as `.bayern` and investigate differences to `.org` and `.dev`.

 doi: 10.2313/NET-2023-11-1_17

References

[1] A. R. Kang and A. Mohaisen, "Transparency of the new gTLD's centralized zone data service: A measurement study," in *2016 IEEE Conference on Communications and Network Security (CNS)*, 2016, pp. 354–355.

[2] J. Zirngibl, S. Deusch, P. Sattler, J. Aulbach, G. Carle, and M. Jonker, "Domain Parking: Largely Present, Rarely Considered!" in *Proceedings of the 6th edition of the network traffic measurement and analysis conference (TMA conference 2022). International Federation for Information Processing (IFIP). https://tma. ifip. org*, 2022.

[3] 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, pp. 3–15, 2020.

[4] 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 *Proceedings of the 21st ACM Internet Measurement Conference*, 2021, pp. 261–275.

[5] J. Postel, "Domain name system structure and delegation," Internet Requests for Comments, RFC Editor, RFC 1591, 3 1994. [Online]. Available: https://www.rfc-editor.org/rfc/rfc1591.html

[6] ICANNWIki, ".dev — icannwiki,," 2019, [Online; accessed 15-June-2023]. [Online]. Available: https://icannwiki.org/index.php?title=.dev&oldid=1170428

[7] P. Mockapetris and K. J. Dunlap, "Development of the Domain Name System," in *Symposium Proceedings on Communications Architectures and Protocols*, ser. SIGCOMM '88. New York, NY, USA: Association for Computing Machinery, 1988, p. 123–133. [Online]. Available: https://doi.org/10.1145/52324.52338

[8] J. Park, J. Choi, D. Nyang, and A. Mohaisen, "Transparency in the New gTLD Era: Evaluating the DNS Centralized Zone Data Service," *IEEE Transactions on Network and Service Management*, vol. 16, no. 4, pp. 1782–1796, 2019.

[9] A. Affinito, R. Sommese, G. Akiwate, S. Savage, K. Claffy, G. M. Voelker, A. Botta, and M. Jonker, "Domain Name Lifetimes: Baseline and Threats," in *6th Network Traffic Measurement and Analysis Conference, TMA 2022*. International Federation for Information Processing (IFIP), 2022.

[10] P. Foremski and P. Vixie, "The modality of mortality in domain names," *Virus*, p. 1, 2018.

[11] I. ClickHouse. ClickHouse - fast open-source OLAP DBMS. [Online]. Available: https://clickhouse.com/

[12] G. Akiwate, R. Sommese, M. Jonker, Z. Durumeric, K. Claffy, G. M. Voelker, and S. Savage, "Retroactive identification of targeted DNS infrastructure hijacking," in *Proceedings of the 22nd ACM Internet Measurement Conference*, 2022, pp. 14–32.