Chain-Sawing: A Longitudinal Analysis of Certificate Chains

Marcus Döberl*, York Freiherr von Wangenheim*, Carl Magnus Bruhner*,
David Hasselquist*†, Martin Arlitt‡, Niklas Carlsson*
*Linköping University, Sweden
†Sectra Communications, Sweden
‡University of Calgary, Canada

***Abstract*—The security and integrity of TLS certificates are essential for ensuring secure transmission over the Internet and protecting millions of people from man-in-the-middle attacks. Certificate Authorities (CAs) play a crucial role in issuing and managing these certificates. This paper presents a longitudinal analysis of certificate chains for popular domains, examining their evolution over time and across different categories. Using publicly available certificate data, primarily from *crt.sh*, we created a longitudinal dataset of certificate chains for domains from the Tranco top-1M list. After categorizing the certificates based on their type and service category, we analyze a selected set of domains over time and identify the patterns and trends that emerge in their certificate chains. Our analysis reveals several noteworthy trends, including a trend towards shorter certificate chains and fewer paths from domains to root certificates. This implies that the certificate process is becoming more simplified and streamlined. Combined with our observations that there is an increasing use of new CAs and a shift in the types of certificates used that we observe, we expect part of this to be an effect of individual choices made by some popular CAs (e.g., less cross-signings). In general, the observed trends, patterns, and findings capture tradeoffs in overhead, backward compatibility, and security. The quick shifts in some of the observed metrics (e.g., chain lengths) therefore also highlight the importance of continued monitoring and analysis of certificate chains.**

I. Introduction

The trust underpinning the Internet rests upon the fragile shoulders of X.509 certificates, issued by Certificate Authorities (CAs) and validated during each HTTPS handshake. However, not all certificates are the same.

While prior research has shown reductions in the number of weak or misissued certificates [1], as well as improvements in the management of certificates [2], less research has studied the trust chains associated with each certificate and how they have changed over time. In addition to having security implications, the lengths of these chains are important since they implicitly impact validation times, with the extra validation times associated with each additional link silently adding precious milliseconds to connection times. Living in an era where speed reigns supreme, it is clear that the chain lengths are important to both understand and optimize.

Certificate hierarchy: The validation of digital certificates occurs within a hierarchical structure known as a certificate chain. Each certificate in the chain is signed by the one above it, ultimately leading back to a trusted root CA. Leaf certificates, issued to specific entities like websites or servers, reside at the bottom of the chain, while root CAs sit at the top, with their self-signed certificates pre-installed and trusted by client systems. As a security measure, root CAs rarely sign leaf certificates directly. Instead, intermediate CAs are used as intermediaries, signing leaf certificates and linking them to the trusted root. This hierarchical structure verifies the authenticity of each certificate through a chain of signings, linking back to a trusted root, ultimately reassuring the client that the public key in the certificate is authentic and can be trusted.

Chain of trust: These certificate chains play a crucial role in the X.509 protocol as they establish a chain of trust between leaf certificates and trusted self-signed root certificates [3]. However, the landscape of certificate validation chains is dynamic, complex, and shaped by evolving technologies and practices. Further compounding this complexity is the use of multiple intermediate CAs, each possessing variable trust levels. By cross-signing each other, these intermediate CAs create a complex network of validation paths and lengths [4]. Furthermore, the potential compromise of CAs poses a serious threat, leading to fraudulent certificate issuance or hijacking of legitimate chains. With both security and performance implications, it is clear that it is important to understand the evolving certificate chain dynamics across domains.

Contributions: In this paper, we present a comprehensive 10-year longitudinal analysis of certificate chains for different domains, focusing on their evolution over time across different categories. Leveraging publicly available certificate transparency (CT) logs, we collect a longitudinal dataset containing over 50 million unique certificates for domains listed on the Tranco top-1M list [5]. By focusing on different domain aspects (rank, TLD, and category), as well as their certificate chain length and number of paths to root, we identify patterns and trends in certificate chains. Our findings provide valuable insights into the factors that shape the use and evolution of certificate validation chains. Most notably, our analysis reveals a trend toward shorter certificate chains and fewer paths from the domain to the root certificate, indicative of streamlined trust chains. While these results imply that the trust landscape is becoming more simplified and streamlined, we also observe some quick shifts in some of the observed metrics (e.g., chain lengths and number of paths), capturing the impact of individual choices made by some popular CAs

The definitive version was published in *Proc. 2024 IFIP Networking Conference (IFIP Networking)*, https://doi.org/10.23919/IFIPNetworking62109.2024.10619717

TABLE I: The most used intermediate CAs in the dataset, ranked by share of certificates they have signed.

Rank	Intermediate CA	Leaf certs valid (from/to)		Share
1	R3 Let's Encrypt	2020-12-02	2023-07-16	40.30%
2	Let's Encrypt Authority X3	2016-03-25	2021-03-02	20.53%
3	COMODO ECC Domain Validation Secure Server CA 2	2014-09-30	2024-01-18	11.05%
4	GlobalSign CloudSSL CA - SHA256 - G3	2015-10-06	2024-04-27	4.75%
5	cPanel, Inc. Certification Authority	2016-03-01	2024-04-11	2.30%
6	Amazon	2015-11-10	2024-04-29	1.47%
7	GlobalSign Organization Validation CA - G2	2011-06-27	2019-03-28	1.32%
8	COMODO RSA Domain Validation Secure Server CA 2	2015-01-13	2024-01-18	1.26%
9	GlobalSign Organization Validation CA - SHA256 - G2	2014-03-31	2024-02-20	1.22%
10	Cloudflare Inc ECC CA-3	2020-05-14	2024-04-16	1.16%
			Total share of top 10	85.36%

(e.g., less cross-signings) or an increasing use of new CAs and a shift in the types of certificates used.

Outline: After describing our dataset and how we perform chain extraction (Section II), we present our analyses based on domain ranks (Section III), top-level domain (Section IV), domain category (Section V), and validation type (Section VI). Finally, we discuss our observations (Section VII), present related works (Section VIII), and conclude (Section IX).

II. Dataset and chain extraction

A. Data collection and categorization

For each of the domains listed on the Tranco top-1M list [5], we query *crt.sh* [6] using Certwatch [7] and obtain certificates belonging to the domain from the years 2013 to 2023 (indicated by the *notBefore* and *notAfter* date fields). This resulted in a dataset of over 50 million unique certificates from 884,312 distinct domains, averaging 56.5 certificates per domain.[1]

For the analysis, we categorize domains along three dimensions: (1) domain popularity (rank), (2) popular generic TLD (gTLD) and sponsored TLD (sTLD), and (3) domain category.

Domain rank: To facilitate comparisons between a set of domains of different popularity, we select the last 1,000 domains (ordered by popularity) in each magnitude sample except for (1) the last magnitude interval on the top-1M, for which we use 10,000 domains to improve collection rate, and (2) the top-1,000 which we break into a top-100 and the rest, as the top-100 are of particular interest. In summary, we include the following domain rank sets: (0, 100], (100, 1K], (9K, 10K], (99K, 100K], and (990K, 1M]. The collection rate is 82–93% of domains for each interval.

Top-level domain (TLD): For our TLD analysis, we compare the certificate chain based on the domain suffix. To identify the most commonly used gTLDs not associated with a specific country, we select the top seven most popular gTLDs on Google Domains [8] (.app, .co, .com, .eu, .info, .net, and .org) and the three most common sTLDs (.edu, .gov, and .mil).

Domain category: For our domain category analysis, we select domains from Cloudflare Radar [9] and the Website Categorization API [10]. Considering the domains overlapping with our dataset and keeping only those categories that contain at least 50 domains, we end up with the categories: Society & Lifestyle (361), Entertainment (176), Technology (143), Shopping & Auction (127), and Business & Economy (62). While these domain sets capture only a small share of the total domains, the size of each set still provides some statistical assurance and the analysis provides several complementing perspectives on the evolution of certificate chains.

Table I shows the most used intermediate CAs in the dataset and the share of certificates signed by each of them. We note that Let's Encrypt has signed most certificates and is responsible for the top-2 most used intermediate CAs, together responsible for over 60% of all observed certificates.

B. Chain extraction

To efficiently traverse the chain of trust for multiple certificates, we developed and applied a heuristic, described next.

Identifying CA links: Due to the large amount of data, we first divide the dataset into groups of months containing all the certificates valid during that month. This allows us to create a network graph containing every possible path from a leaf certificate to a root CA specific for each month. Fig. 1 shows such an example of a network graph based on certificates valid during March 2023. Here, we illustrate the intermediate CAs (blue dots), certificate chain (arrows), and root CAs (red dots). The network graph also contains several sub-graphs that can easily be analyzed separately. As the figure shows, cross-signing and long certificate chains can create more complex graphs while there are also examples of much simpler graphs (however, the many examples of graphs with $\leq$2 intermediate CAs are omitted from the figure).

Identifying certificate chain links and chain of trust: To establish the chain of trust from a domain to a root certificate, we first obtain the domain certificate and identify the CA that signed it using the CA ID field. Then, we check all valid certificates of the CA during the validity period of the domain certificate, repeating the process until we reach a root certificate. However, due to cross-signing, some CAs may have multiple distinct chains leading to a root certificate, complicating the determination of the chain of trust. To account for this, we extract several chain paths for each CA.

Fig. 2 shows an example of a certificate chain. Here, the leaf certificate issued for domain *example.com* is signed by a certificate belonging to the *GeoTrust Global TLS* intermediate CA. The leaf certificate is also indirectly signed by two root CAs (included in the root stores, at the top). Here, the number of paths from *example.com* to a root is 2. The first one

[1]Dataset: https://www.ida.liu.se/~nikca89/papers/networking24.html

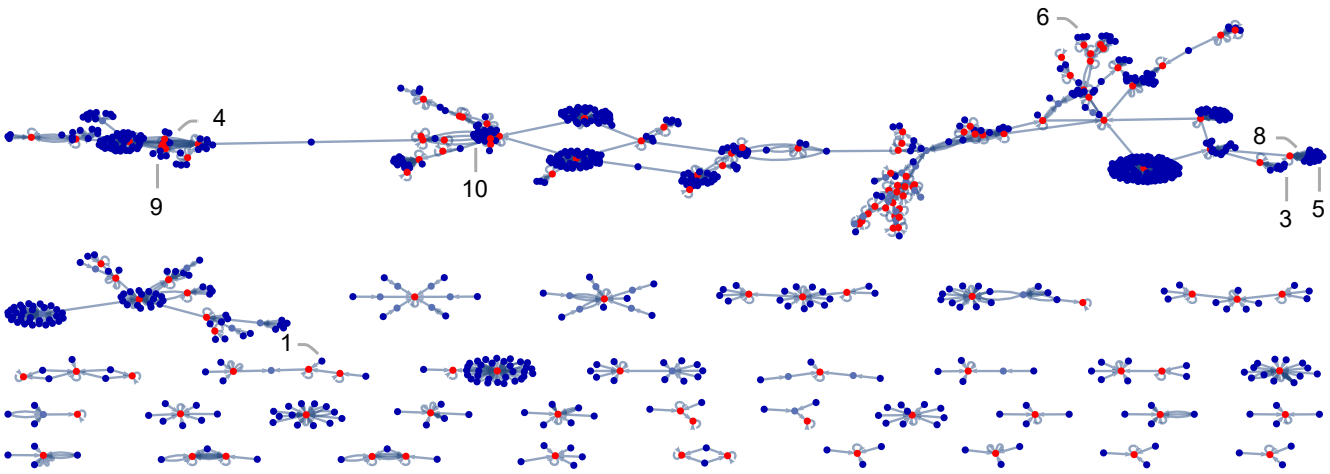


Fig. 1: Example network graph linking intermediate CAs (blue) to root CAs (red) for certificates valid March 2023. The top 10 most used CAs in the collected dataset (Table I) are labeled 1–10, with 2 and 7 missing due to not being valid at the time.

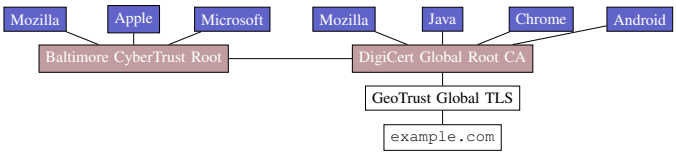


Fig. 2: Example of a certificate chain for the leaf certificate of *example.com* with one intermediate CA and two root CAs.

has the length 3 (*example.com* → *GeoTrust Global TLS* → *DigiCert Global Root CA*), and the second path has length 4 (*example.com* → *GeoTrust Global TLS* → *DigiCert Global Root CA* → *Baltimore CyberTrust Root*), resulting in a mean path length of 3.5. Previous work (see Section VIII) has focused on the shortest path length, but in this work, we specifically look at the mean path length as it allows for a more probable path length due to clients using different chains of trust depending on available intermediate and root certificates.

C. Dataset limitations

While our dataset offers a comprehensive view of certificate chains from 2013 to 2023, certain limitations should be acknowledged. First, our dataset relies on *crt.sh* for certificate retrieval, which may not encompass all certificates issued within the specified time frame due to potential variations in data collection methods or updates to their database.[2] Additionally, the categorization of domains is based on sources like Cloudflare Radar and Website Categorization API, which may not fully represent the diversity of domains on the Internet. Finally, while we aim to provide insights into certificate chain evolution, our dataset may not capture every aspect of this dynamic process, especially regarding the intricacies of cross-signing and complex chain paths. In light of these dynamics and complexities, we limit our analysis to using relatively simple metrics (i.e., chain lengths and number of paths to root) and incorporate statistical tests to support key findings.

[2]However, we note that they monitor most (if not all) active CT logs available: https://crt.sh/monitored-logs

III. Rank-based analysis

We begin our analysis by looking at the impact of domain popularity. Table II shows the most commonly used intermediate CA in each popularity range. For each range, we show (1) the intermediate that has signed the greatest number of certificates (column *Intermediate CA*), (2) the percentage of certificates in the group signed by this intermediate (*Share*), (3) the mean length of the chain of trust, from a domain to a root CA (*Chain length*), and (4) the mean number of paths from a domain to a root CA (*Paths*). The last two are an average for the last ten years. There is an evident difference between the groups of ranks, where Google's intermediate is the most used in the top, 1–100 and 101–1,000, while Let's Encrypt's intermediate is the most used for the five lowest groups. Only in the group 9,001–10,000 is a different CA (i.e., GlobalSign) the most used intermediate. Fig. 3 shows the distribution of CA popularity (grouping same-CA intermediates) per domain rank interval, supporting the above findings.

Chain length: We calculated the chain length from a domain leaf certificate to a root certificate and extracted the mean and the 95th percentile of the length. The calculation was done once per month from January 1, 2013, to March 1, 2023. Fig. 4 presents the results where the *mean* length of the chain (blue solid line) is based on all domains belonging to the subcategory and the *95th percentile* (blue dashed line) is the chain length of which 95% of the domains in the subcategory have shorter chain length than. A corresponding figure of the intermediate ranks (249K, 250K], (499K, 500K], and (749K, 750K] can be found in the Appendix.

Looking at the mean chain lengths (and mean paths to root, presented next) it is primarily in the last 3–5 years that there are notable changes with decreasing chain lengths and number of paths (especially the latter) making it more interesting to analyze this period. Before that, there was a different landscape with lower HTTPS adoption rate [11]. Going forward, we will look both at trends for the past 10 years and look closer at the last three years, capturing the most recent trends.

TABLE II: The most used intermediate CA and its share of each domain ranks group, and the mean chain length and mean number of paths to root for certificates in the group.

Ranks	Intermediate CA	Share	Chain length	Paths
(0, 100]	Google [...] G2	30.90%	3.613	3.350
(100, 1K]	Google [...] G2	16.84%	3.480	2.939
(9K, 10K]	GlobalSign [...] G3	15.81%	3.424	3.113
(99K, 100K]	R3 Let's Encrypt	32.08%	3.442	3.149
(249K, 250K]	R3 Let's Encrypt	38.88%	3.514	3.228
(499K, 500K]	R3 Let's Encrypt	41.92%	3.507	3.240
(749K, 750K]	R3 Let's Encrypt	50.34%	3.449	3.149
(990K, 1M]	R3 Let's Encrypt	45.01%	3.418	3.064

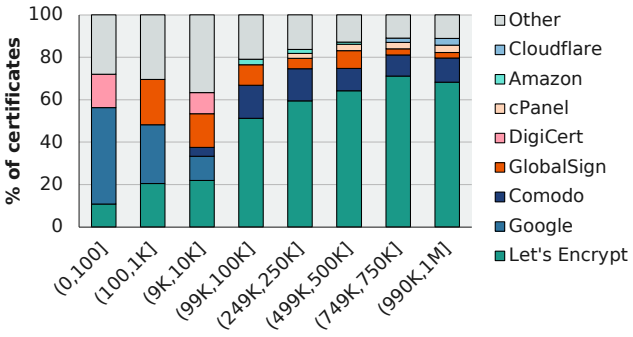
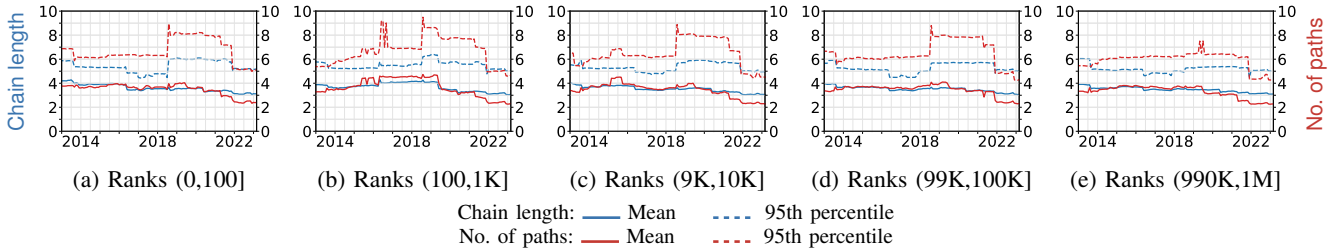


Fig. 3: Use of top CAs for 8 domain rank intervals.



(a) Ranks (0,100] (b) Ranks (100,1K] (c) Ranks (9K,10K] (d) Ranks (99K,100K] (e) Ranks (990K,1M]

Fig. 4: Chain length and number of paths to root for studied domain ranks from January 2013 to March 2023.

TABLE III: The mean chain length and mean number of paths to root per group of domain ranks.

	Mean chain length			Mean paths to root		
Ranks	**2020**	**2023**	**Change**	**2020**	**2023**	**Change**
(0, 100]	3.604	3.137	−0.467*	3.666	2.413	−1.253**
(100, 1K]	3.456	3.084	−0.372*	3.327	2.323	−1.004**
(9K, 10K]	3.566	3.050	−0.516**	3.327	2.323	−1.004**
(99K, 100K]	3.602	3.067	−0.535**	3.564	2.314	−1.250**
(249K, 250K]	3.652	3.130	−0.522**	3.565	2.347	−1.218**
(499K, 500K]	3.712	3.169	−0.543**	3.565	2.347	−1.218**
(749K, 750K]	3.463	3.166	−0.297*	3.268	2.458	−0.810*
(990K, 1M]	3.472	3.092	−0.381*	3.181	2.275	−0.906*

* 95%, ** 99.9% statistical significance

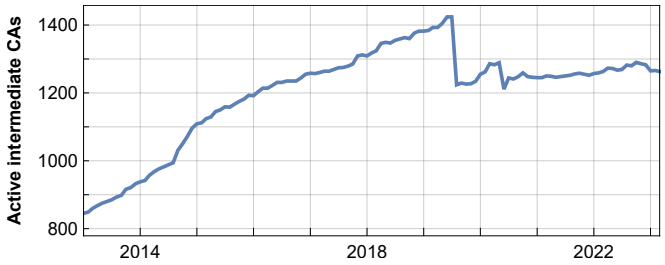


Fig. 5: Number of active intermediate CAs over time.

Table III includes a comparison of how the chain length has evolved for the different domain rank intervals in the last three years, showing the mean length in 2020, in 2023, and the change between. If the change is statistically significant, it is marked with asterisk(s). Over this time period, the average chain length decreased with 0.454 (standard deviation 0.092), while the number of intermediate CAs remained relatively stable, as shown in Fig. 5.

Paths to root: In addition to calculating the path length, we calculated the number of valid paths one could take from a certificate belonging to a domain to a root certificate, extracting the mean and the 95th percentile of the month. Also, this calculation was done once per month from January 1, 2013, to March 1, 2023. Fig. 4 shows how the paths to roots have evolved in the last 10 years, where the *mean* number of paths to root (red solid line) is based on all domains belonging to the subcategory and the *95th percentile* (red dashed line) is the number of paths that 95% of the domains of the subcategory have fewer paths than. A corresponding figure of the intermediate ranks can be found in the Appendix.

Referring to the three right-most columns in Table III, we also observe significant reductions in the mean number of paths over the last three years, with an average decrease of 1.083 (standard deviation 0.174) across all domains.

Analysis: When looking at our results, we see that the mean chain length of all domain rank ranges has decreased between 2020 and 2023, with an average length of 3.426 in 2020 vs. 3.112 in 2023. The number of paths has decreased from an average of 3.433 in 2020 to 2.350 in 2023. All changes shown in Table III are statistically significant.

A significant decrease in the chain length and number of paths can be seen in the first three domain rank ranges, where Google CAs have a dominant market share. Notably, larger companies like Google have had their own root certificate in all major root stores. Before August 2020, Google's GTS CA 1C3 CA was additionally cross-signed by GlobalSign R3/R2 root certificate. This no longer being the case corresponds with the decreased chain length and can especially be seen

TABLE IV: The most used intermediate CA and its share of each studied TLD, sorted by domain count, and the mean chain length and mean number of paths to root.

TLD	Intermediate CA	Share	Chain length	Paths	Certificate count	Domain count	Avg. num. cert./dom.
.com	R3 Let's Encrypt	37.80%	3.380	2.761	26,586,330	416,194	55.78
.org	R3 Let's Encrypt	37.20%	3.646	3.157	3,386,242	45,577	69.65
.net	R3 Let's Encrypt	39.87%	3.348	3.026	2,457,046	44,325	53.77
.co	R3 Let's Encrypt	35.30%	3.713	4.150	424,740	6,901	60.46
.info	R3 Let's Encrypt	44.44%	3.488	3.258	350,856	6,761	50.99
.eu	R3 Let's Encrypt	46.16%	3.925	4.794	212,786	3,752	55.77
.edu	R3 Let's Encrypt	23.59%	3.515	3.301	290,857	3,524	80.33
.gov	R3 Let's Encrypt	19.07%	3.763	3.812	127,066	1,832	7.32
.app	R3 Let's Encrypt	54.06%	3.219	3.121	72,880	1,732	42.04
.mil	Entrust [...] - L1K	39.16%	8.972	15.257	711	54	12.61

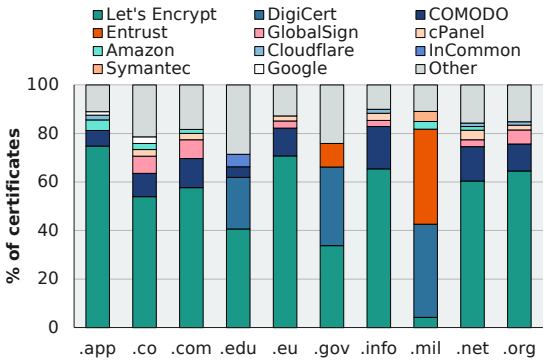


Fig. 6: CA popularity among the 10 TLDs.

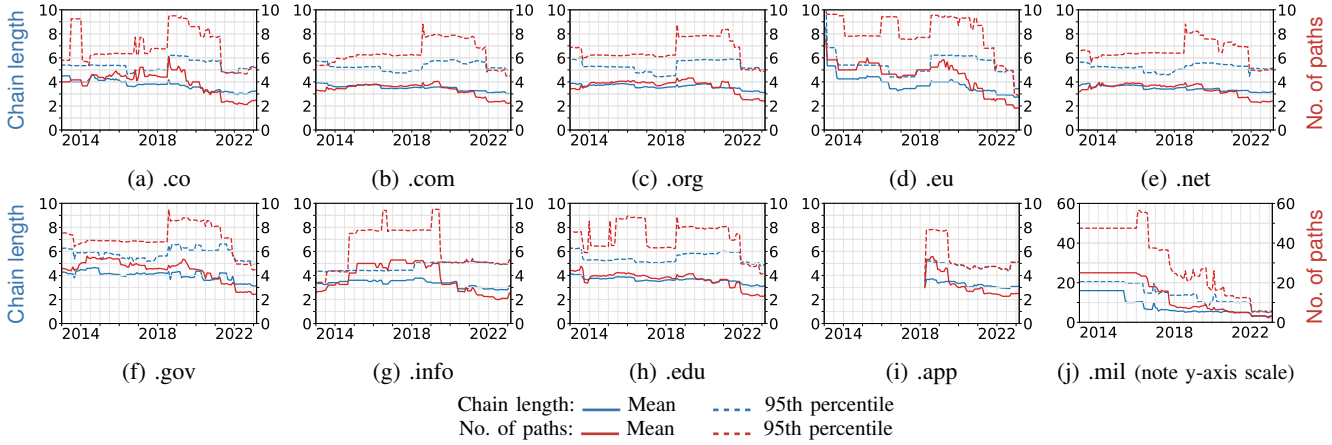


Fig. 7: Chain length and number of paths to root for studied TLDs from January 2013 to March 2023.

in the number of paths to root in Fig. 4a and 4b. The same effects can be observed with Let's Encrypt, launched in 2016 and initially cross-signed but recently phasing out its cross-signed root, contributing to fewer number of paths to root and a shorter chain length [12]. Finally, we note that the reduction in number of paths to root (red curves) has seen a relatively bigger reduction than the chain lengths (blue curves), suggesting a streamlining of the chains.

IV. TLD-based analysis

We next look at the selected TLDs. Table IV shows the most commonly used intermediate CA for each included TLD. All but one have Let's Encrypt as the most used CA, with .gov just below 20% and .app with over 50%. The only TLD studied with a different dominating CA certificate is .mil, using Entrust instead, which seems reasonable given that the US military might not use free certificates with simpler validation. Fig. 6 shows the distribution of top CAs for each TLD, confirming an almost negligible use of Let's Encrypt in .mil.

Chain length: Fig. 7 shows how the chain lengths have evolved for each TLD, like presented for the domain ranks. We note that .app is a fairly new TLD, thus lacking some older data. Also, .mil suggests a more complex history regarding certificates given the historically long chain lengths, which only recently have leveled out similar to other TLDs.

Similarly, Table V shows a comparison of how the mean chain lengths have evolved over the past three years. Given the notable change for .mil, this is the only TLD with a statistically significant decrease over the last years. The average decrease of chain length was 0.723 (standard deviation 0.623).

Paths to root: Considering the mean number of paths to root, we again observe bigger and more significant reductions than for the chain lengths. For example, referring to Fig. 7, showing how the mean number of paths has evolved, most classes have seen substantial reductions in the mean number of paths over the last 3-to-5 years. The main exception is .mil, which saw the by far biggest reduction between 2016 and 2018, and has since continued to see a reduction.

Examining the changes over the last three years (Table V), the average decrease in paths to root was 1.551 (standard deviation 0.866) across all domains, with three classes (.eu, .gov, and .mil) all having statistically significant decreases. The big drops in both mean and 95th percentile for these domains suggest a conscious effort to streamline their trust chains.

Analysis: While the decrease in mean chain length is only statistically significant for .mil, all TLDs have seen decreases. The same trend goes for the number of paths, but with the decreases of .eu and .gov also being statistically significant. With every single TLD studied having a decrease, both in mean chain length and in mean paths to root, we can conclude that the overall trend is a decrease for both.

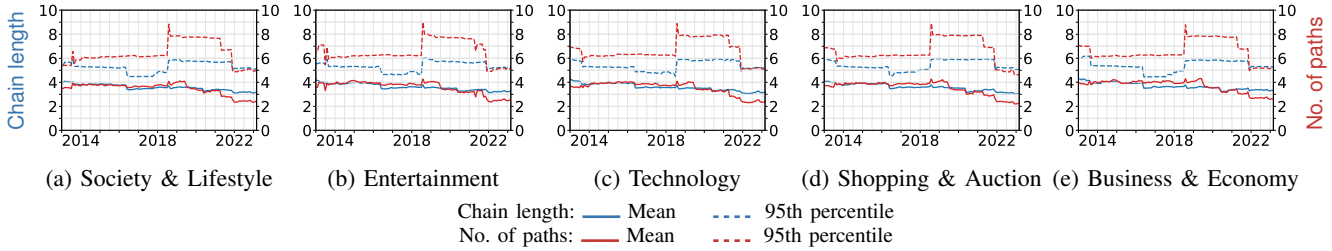
TABLE V: The mean chain length and mean number of paths to root per TLD.

TLD	Mean chain length			Mean paths to root		
	2020	2023	Change	2020	2023	Change
.co	3.571	3.300	−0.271	4.000	2.556	−1.444
.com	3.534	3.033	−0.501	3.404	2.256	−1.148
.org	3.808	3.143	−0.665	3.865	2.449	−1.416
.eu	3.769	2.750	−1.019	4.692	1.833	−2.859*
.net	3.448	3.212	−0.236	3.293	2.439	−0.854
.gov	4.347	3.140	−1.207	4.449	2.453	−1.996*
.info	3.444	3.154	−0.290	3.333	2.538	−0.795
.edu	3.654	3.146	−0.508	3.615	2.341	−1.274
.app	3.385	3.083	−0.302	3.076	2.500	−0.576
.mil	5.500	3.272	−2.228*	6.333	3.181	−3.152*

* 95%, ** 99.9% statistical significance

TABLE VI: The most used intermediate CA and its share of each studied domain category, and the mean chain length and mean number of paths to root for certificates in the category.

Category	Intermediate CA	Share	Chain length	Paths
Society & Lifestyle	GlobalSign [...] G3	33.26%	3.555	3.276
Entertainment	GlobalSign [...] G3	29.79%	3.588	3.339
Technology	GlobalSign [...] G3	33.12%	3.543	3.288
Shopping & Auction	GlobalSign [...] G3	31.29%	2.518	3.351
Business & Economy	Google Internet Authority G2	18.31%	3.643	3.667



(a) Society & Lifestyle (b) Entertainment (c) Technology (d) Shopping & Auction (e) Business & Economy

Chain length: Mean 95th percentile
No. of paths: Mean 95th percentile

Fig. 8: Chain length and number of paths to root for studied domain categories from January 2013 to March 2023.

V. Domain Category-Based Analysis

We now look at the selected domain categories. Table VI show the most commonly used intermediate CA for each domain category. These groups are primarily dominated by GlobalSign, used by close to one third of the cases. Only the last category stands out by having Google as the most used intermediate CA. As the number of domains and certificates studied in this section is smaller, we do not go into the same details, however, noting that the general trends of previous categorizations apply here as well.

Chain length: Fig. 8 shows how the mean chain length has evolved for each domain category. We observe the appearance to be very similar across all categories, with only minor changes and a slow but steady decrease in length.

Paths to root: Fig. 8 shows how the mean number of paths has evolved. The similarity between each category continues and overall, we see steady decrease in mean length.

Analysis: When looking at our results, we see that both the mean chain length and mean number of paths to root of all domain categories has decreased steadily over the period, with number of paths being the most notable. This follows the trends we have already seen in previous categorizations, possibly with a little less variation.

VI. Certificate Validation Type-Based Analysis

In addition to studying the different subsets of domains, we also look at the different validation types of certificates and how the structure of the chains has evolved over the last three years in the dataset using the previously defined domain ranks. With validation types, we mean the three most popular validation types as standardized by the CA/Browser Forum [2], [13]: Domain (DV), Organization (OV), and Extended (EV). Focusing on the recent trends, we present how the chain length and the number of paths to the root have evolved for these in each popularity range.

Chain length: Table VII shows an overview of the mean chain length changes across various certificate types over the past three years. Here, the average decrease (and standard deviation) for DV, OV, and EV was 0.411 (0.051), 0.350 (0.137), and 0.485 (0.191), respectively.

Paths to root: Table VIII shows an overview of how the mean number of paths to root has changed in the last three years for each certificate type. The average decrease (with standard deviation) of DV was 1.034 (0.080), OV was 0.959 (0.276), and EV was 2.823 (0.951).

Analysis: The results show that the mean number of paths from a leaf to a root certificate has decreased in all combinations of ranks and validation types over the past three years. (To give some context, with 24 cases in total, observing decreases for all 24 cases is statistically significant with a p-value of $6 \cdot 10^{-8}$ if performing a binomial test.) In terms of mean chain length in 2023, OV certificates exhibit the shortest chains, with a mean length of 3.076. Comparatively, EV certificates have a mean length of 3.204, while DV certificates have a slightly higher mean length of 3.216. It is worth noting that OV certificates also display the smallest deviation from their 2020 mean chain length.

When considering the number of paths to a root, OV certificates continue to exhibit the lowest change and the lowest number of paths in 2023. Of particular interest is the number of paths to a root for DV certificates, which experienced a decrease from a mean of 3.487 to 2.432, representing a change of nearly 1, and statistically significant changes for all 8 popularity ranges. Diving further into the data, the change

TABLE VII: The mean chain length from a DV/OV/EV certificate to a root March 1, 2020 vs. March 1, 2023.

	DV			OV			EV		
Ranks	**2020**	**2023**	**Change**	**2020**	**2023**	**Change**	**2020**	**2023**	**Change**
(0, 100]	3.667	3.273	−0.394**	3.407	3.074	−0.333*	3.667	3.273	−0.394
(100, 1K]	3.632	3.157	−0.475**	3.469	3.138	−0.330*	3.632	3.157	−0.475*
(9K, 10K]	3.639	3.151	−0.488**	3.583	3.107	−0.476*	3.639	3.151	−0.488*
(99K, 100K]	3.618	3.273	−0.345**	3.405	3.000	−0.405*	3.618	3.273	−0.345
(249K, 250K]	3.629	3.209	−0.419**	3.571	3.037	−0.534*	3.629	3.209	−0.419
(499K, 500K]	3.647	3.238	−0.409**	3.393	3.115	−0.277	3.647	3.238	−0.409
(749K, 750K]	3.536	3.129	−0.407**	3.200	3.118	−0.082	3.536	3.129	−0.407
(990K, 1M]	3.597	3.242	−0.354**	3.377	3.019	−0.359*	4.125	3.182	−0.943*

* 95%, ** 99.9% statistical significance

TABLE VIII: The mean number of paths to root from DV/OV/EV certificates March 1, 2020 vs. March 1, 2023.

	DV			OV			EV		
Ranks	**2020**	**2023**	**Change**	**2020**	**2023**	**Change**	**2020**	**2023**	**Change**
(0, 100]	3.583	2.545	−1.038**	3.333	2.370	−0.963*	5.500	2.000	−3.500*
(100, 1K]	3.500	2.353	−1.147**	3.391	2.369	−1.021*	5.211	2.100	−3.111*
(9K, 10K]	3.472	2.377	−1.095**	3.542	2.339	−1.202*	5.087	2.333	−2.754*
(99K, 100K]	3.500	2.500	−1.000**	3.324	2.200	−1.124*	4.318	2.300	−2.018
(249K, 250K]	3.486	2.442	−1.044**	3.464	2.148	−1.316*	6.333	2.500	−3.833*
(499K, 500K]	3.500	2.405	−1.095**	3.250	2.538	−0.712*	6.200	2.500	−3.700*
(749K, 750K]	3.357	2.419	−0.938**	2.933	2.471	−0.463	3.500	2.500	−1.000
(990K, 1M]	3.323	2.409	−0.913**	3.019	2.148	−0.871*	5.125	2.455	−2.670*

* 95%, ** 99.9% statistical significance

is particularly evident with a notable dip in September 2021, coinciding with when the initially cross-signed [14] certificates of Let's Encrypt ceased to be cross-signed thus reducing the number of paths with 1. Being the dominant CA among DV certificates, this clearly contributes to the relatively consistent decrease in number of paths for all DV certificates.

To illustrate the change, Fig. 9 shows the chain of trust for Let's Encrypt CA certificates, illustrating how they have transitioned away from having a cross-signed root certificate.

Another interesting observation is the very large decreases in mean number of paths to root for EV certificates compared to the other types for all eight rank-intervals. While the smaller number of samples in each category limit the number of significant changes on a per subset level, this observation in itself is significant as we see it for eight out of eight subsets (p-value of 0.0039 with binomial test). Finally, we again see decreases for all 24 cases, regardless of type and rank subset, strengthening our observation of reduced number of paths to root across the board.

VII. Discussion

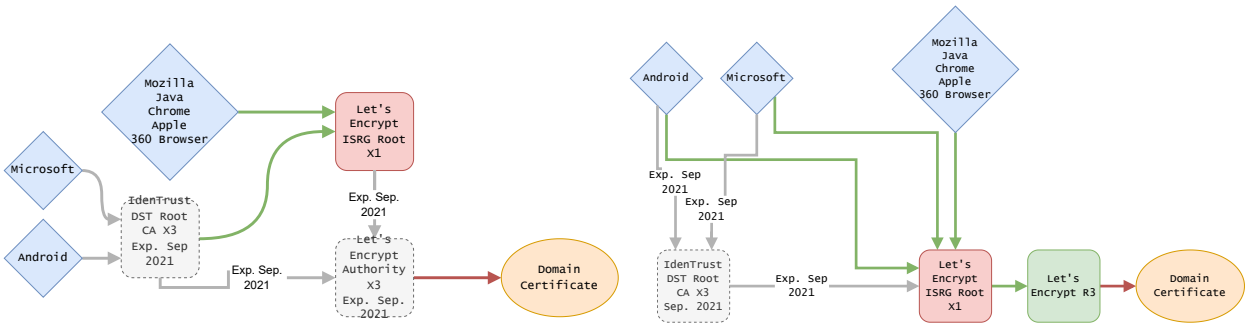
In 2016, Let's Encrypt officially launched as the first free and automated CA, achieved using the then-new Automatic Certificate Management Environment (ACME) protocol simplifying the certificate issuance process [14], [15].

In 2017, Google launched Google Trust Service, acquiring two existing root CAs to be able to launch quickly [16]. Subsequently, in 2018, a substantial milestone was reached as the Internet traffic across Google that employ HTTPS finally reached 95% [11]. Parallel to this, the trend has been towards shorter validity periods as stipulated by CA/Browser Forum [13]. In line with this timeline, capturing a combination of both an increased HTTPS adoption and shorter validity periods of certificates, our dataset includes a notable surge in the number of certificates with expiry dates in 2018–2021.

This finding highlights that companies like Google with a substantial part of the Internet traffic have managed to streamline their certificate infrastructure, sometimes by establishing their own trusted root certificates and certificate infrastructure. This means that they have found an effective move to decrease and optimize the chain of trust. This also highlights the influence of market share and the role of the leading companies in the industry in shaping the evolution of certificate chains. The ability of prominent organizations to streamline their certificate chains demonstrates their commitment to enhancing security measures and ensuring efficient validation processes.

We can see in all popularity ranges that cross-signing has become less popular. In Fig. 4, we see that the chain length is trending towards a chain length of 3 where the chain of trust is Root → Intermediate → Leaf certificate. The likelihood of certificate chains shrinking to less than three in the future is highly unlikely due to the common practice in certificate management of keeping the root certificate of CAs offline for security reasons. Instead, intermediates are used to sign certificates for domains or organizations, creating a hierarchical structure. This approach ensures robust security and minimizes the risk of compromising the root certificate. Thus, the use of intermediary certificates is expected to continue, maintaining a minimum chain length of three.

As discussed at the end of the last section, Let's Encrypt is a dominant player and the decreases in chain lengths towards the end of 2021 might to a large extent be explained by Let's Encrypt ceasing to be cross-signed at that time. Fig. 9 shows how the chain of trust has evolved for Let's Encrypt before and after September 2021, when the cross-signing from *DST Root X1* had ended. What is not shown is that Let's Encrypt had their root certificate cross-signed to enable older Android devices to continue having a valid certificate chain [17]. However, this will end in September 2024 (which



(a) Let's Encrypt X3. Mean length/paths to root: 3.5/2 (2021). Valid chain of trust between Mar. 2016 and Sept. 2021.

(b) Let's Encrypt R3: Mean length/paths to root: 3/1 (2023). Valid chain of trust between Sept. 2020 and Sept. 2025.

Fig. 9: The chains of trust for the two most used intermediate CAs (both Let's Encrypt) in May 2023 (see Table I). The blue rhombuses are root stores, red squares are root CAs, green squares are intermediate CAs, gray squares are CA certificates that have expired, green arrows are valid relations in May 2023, and gray arrows are expired relations.

has made Cloudflare go with other default alternatives [18]) as Let's Encrypt continues their efforts to provide smaller and lighter chains [19]. Fig. 3 illustrates the distribution of market shares among the top-10 entities in the 99,001–1M range. Notably, it reveals that *Let's Encrypt's R3 CA* commands a market share exceeding 50%. The decline in chain lengths at the end of 2021 is evident in nearly all figures displaying the chain length due to the extensive use of Let's Encrypt in all categories. This dip is not as evident in .mil, in Fig. 7j, due to Let's Encrypt low market share for the TLD.

In conclusion, we see that in all the different ways we look at the data (i.e., domain ranks, TLDs, domain categories, and certificate validation types), we see a decrease in both mean chain length and mean number of paths to root. The change is statistically significant even at the level of individual subcategories, and all things combined the trend is evident.

VIII. Related Work

There have been a number of studies on TLS/SSL certificate characterization, including chain of trust and longitudinal analyses on certificates and certificate management.

Holz et al. [20] used various active and passive measurements on TLS/SSL traffic from November 2009 to April 2011 to analyze the web PKI at that time, including previous work by Electronic Frontier Foundation. Looking at the length of certification chains, the vast majority of certificate chains were found to have a length of ≤ 3, excluding root and leaf certificates (meaning ≤ 5 in this paper). Half of the certificates had a chain length of 0, meaning issued from the root (length of 2 in this paper) or self-signed (length of 1), noting that a large fraction of certificates was found to be self-signed. The maximum chain length was close to 20.

Durumeric et al. [21] used Internet-wide scans over 14 months (June 2012 to August 2013) to describe the HTTPS certificate ecosystem, identifying top-CAs and looking at aspects such as root stores and distribution of trust. They found almost all (98%) certificates being issued by an intermediate CA one step away from a root (path length of 3). However, they did not consider alternative cross-signing(s) with multiple paths to root as we have done in this paper. They did note that 62.6% of leaf certificates had "multiple parents" (cross-signed), which suggests that the average chain length might have been longer considering all paths. 38.7% of certificates had two parents, 12.3% had three, 11.3% four, and the rest 5–9 parents (meaning number of paths to root in this paper).

Hiller et al. [4] used data of 7 years to study benefits and challenges with cross-signing, finding examples of mismanagement as well as highlighting opportunities in terms of helping to bootstrap new CAs like Let's Encrypt.

Bruhner et al. [2] used data of 7 years to study certificate replacement relationships and certificate management practices, finding that OV and EV certificates seemingly tend to be better managed, whereas certificates issued through automation tend to be replaced at the most regular intervals. Cerenius et al. [22] built on this work to study the effects of revocation on certificate replacements, finding notable shortcomings and further suggesting improvements to automation.

VanderSloot et al. [23] investigated ways to improve the view of the certificate ecosystem through combinations of various collection techniques. Among other results, they found CT logs to have good coverage of certificates for web content in contrast to other TLS-based services (e.g., webmail).

Kumar et al. [1] introduced a certificate linter, ZLint, analyzing how well CAs construct their certificates. In this, they studied how intermediates contribute to CAs overall misissuance, finding that in 80% of CA organizations issuing 10K+ certificates, one intermediate is responsible for the majority of misissued certificates.

In a systematization of knowledge paper, Chuat et al. [24] looked at issues with delegations and revocations of certificates, suggesting an alteration to the chain of trust with proxy certificates and delegated credentials as solutions to address current shortcomings. Contrary to the trend shown in this paper, such solutions would expand the chain of trust with one level below the current chain, giving the domain owner the possibility/responsibility to issue proxy certificates or short-lived credentials to serve in connections instead.

Hasselquist et al. [25] presented a 10-year longitudinal study

of wildcard certificates, highlighting substantial differences in wildcard and multi-domain certificate practices. Using Google's CT logs, *crt.sh*, and Rapid7, they conclude that differences in wildcard practices cannot be attributed to only individual CAs or to policy suggestions. However, they do not consider certificate chain aspects.

This paper presents an analysis of the historical certificate chains of a large-scale dataset of certificate validation across a diverse range of domains. Specifically, we examine the evolution of the length and number of paths associated with each domain and its certificates over time. To the best of our knowledge, this study represents the first longitudinal perspective on the evolution of certificate validation chains for popular domains using such a large dataset. Previous work [20], [21] have looked at the (shortest) certificate chain lengths and the number of paths to root, but only over a shorter period of time and not including the combined perspectives to see average chain lengths and number of paths.

IX. Conclusion

In this paper, we have presented a novel analysis of longitudinal trends of certificate chain lengths and number of paths to root. The certificate data of the Tranco top-1M websites was collected from the *crt.sh* database, dating back 10 years. We then calculated the chain lengths and the number of paths to root for each domain and performed statistical analysis of recent years to validate the observed changes. Our main findings are that the mean chain length has decreased significantly in the last two years, reaching an average of 3.142 in 2023 (an average decrease of 0.482 since 2020), getting closer to the suggested equilibrium state of 3 based on current certificate management practices. We can also see that the number of paths has dropped from 3.789 in 2020 to around 2.401 in 2023. By analyzing the results from various different perspectives, we see that the results are consistent with the changes being unambiguous. With fewer intermediates, this can help prevent misissuance and ensure consistent revocations based on challenges identified in previous studies. These results further suggest that the certificate ecosystem is not only becoming more streamlined and efficient, but also more centralized and dependent on fewer CAs.

Acknowledgment

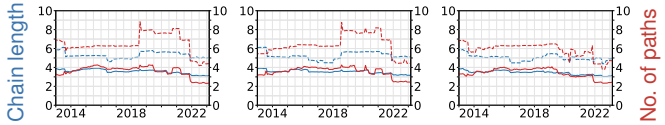
This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

References

[1] D. Kumar, Z. Wang, M. Hyder, J. Dickinson, G. Beck, D. Adrian, J. Mason, Z. Durumeric, J. A. Halderman, and M. Bailey, "Tracking certificate misissuance in the wild," in *Proc. IEEE S&P*, 2018.
[2] C. M. Bruhner, O. Linnarsson, M. Nemec, M. Arlitt, and N. Carlsson, "Changing of the guards: Certificate and public key management on the internet," in *Proc. Passive and Active Measurement (PAM)*, 2022.
[3] D. Cooper, S. Santesson, *et al.*, "Internet X.509 public key infrastructure certificate and certificate revocation list (CRL) profile," 2008.
[4] J. Hiller, J. Amann, and O. Hohlfeld, "The boon and bane of cross-signing: Shedding light on a common practice in public key infrastructures," in *Proc. ACM CCS*, 2020.
[5] V. Le Pochat, T. Van Goethem, S. Tajalizadehkhoob, M. Korczyński, and W. Joosen, "Tranco: A research-oriented top sites ranking hardened against manipulation," in *Proc. NDSS*, 2019.
[6] Sectigo, "crt.sh certificate search," https://crt.sh, 2024.
[7] crt.sh, "certwatch_db," https://github.com/crtsh/certwatch_db, 2024.
[8] Google Domains. Find your perfect ending. https://domains.google/tld/. Accessed 2023-04-17.
[9] Cloudflare Radar, https://radar.cloudflare.com/domains, 2023.
[10] Website Categorization API, "URL categorization database," https://www.websitecategorizationapi.com/url_database.php, 2023.
[11] Google Transparency Report, "HTTPS encryption on the web," https://transparencyreport.google.com/https/overview, 2024.
[12] Let's Encrypt, "Shortening the Let's Encrypt chain of trust," https://letsencrypt.org/2023/07/10/cross-sign-expiration, 2023.
[13] CA/Browser Forum, "Baseline requirements for the issuance and management of publicly-trusted certificates," https://cabforum.org/uploads/CA-Browser-Forum-BR-v2.0.1.pdf, 2023.
[14] J. Aas, R. Barnes *et al.*, "Let's Encrypt: An automated certificate authority to encrypt the entire web," in *Proc. ACM CCS*, 2019.
[15] Let's Encrypt, "Leaving beta, new sponsors," https://letsencrypt.org/2016/04/12/leaving-beta-new-sponsors, 2016.
[16] Google, "The foundation of a more secure web," https://security.google blog.com/2017/01/the-foundation-of-more-secure-web.html, 2017.
[17] Let's Encrypt, "Extending Android device compatibility for Let's Encrypt certificates," https://letsencrypt.org/2020/12/21/extending-android-compatibility, 2020.
[18] Cloudflare, "Upcoming Let's Encrypt certificate chain change and impact [...]," https://blog.cloudflare.com/upcoming-lets-encrypt-certificate-chain-change-and-impact-for-cloudflare-customers, 2024.
[19] Let's Encrypt, "New intermediate certificates," https://letsencrypt.org/2024/03/19/new-intermediate-certificates, 2024.
[20] R. Holz, L. Braun, N. Kammenhuber, and G. Carle, "The SSL landscape: A thorough analysis of the X.509 PKI using active and passive measurements," in *Proc. Internet Measurement Conference (IMC)*, 2011.
[21] Z. Durumeric, J. Kasten, M. Bailey, and J. A. Halderman, "Analysis of the HTTPS certificate ecosystem," in *Proc. IMC*, 2013.
[22] D. Cerenius, M. Kaller, C. M. Bruhner, M. Arlitt, and N. Carlsson, "Trust issue(r)s: Certificate revocation and replacement practices in the wild," in *Proc. Passive and Active Measurement (PAM)*, 2024.
[23] B. VanderSloot, J. Amann, M. Bernhard, Z. Durumeric, M. Bailey, and J. A. Halderman, "Towards a complete view of the certificate ecosystem," in *Proc. Internet Measurement Conference (IMC)*, 2016.
[24] L. Chuat, A. Abdou, R. Sasse, C. Sprenger, D. Basin, and A. Perrig, "SoK: Delegation and revocation, the missing links in the web's chain of trust," in *Proc. IEEE EuroS&P*, 2020.
[25] D. Hasselquist, L. Bolin, E. Carlsson, A. Hylander, M. Larsson, E. Voldstad, and N. Carlsson, "Longitudinal analysis of wildcard certificates in the WebPKI," in *Proc. IFIP Networking*, 2023.

Appendix

Fig. 10 shows the chain length (blue curves) and number of paths to root (red curves) for three additional domain rank ranges that complement those shown in Fig. 4. We note that there appears to be a relatively smooth transition in the patterns observed with the patterns observed for (249K, 250K] most resembling those for (99K, 100K], and the patterns of (749, 750K] most resembling those of (990K, 1M].



(a) (249K,250K] (b) (499K,500K] (c) (749K,750K]

Chain length: Mean 95th percentile
No. of paths: Mean 95th percentile

Fig. 10: Chain length and number of paths to root for additional domain ranks from January 2013 to March 2023.