Prefix Top Lists Reloaded: A Temporal Prefix Ranking Dataset

Savvas Kastanakis[1], Rick Fontein[1], Shyam Krishna Khadka[1], Ebrima Jaw[1], Cristian Hesselman[1,2], and Mattijs Jonker[1]

[1] University of Twente, Enschede, The Netherlands {`s.kastanakis, r.l.h.fontein, s.k.khadka, e.jaw, c.e.w.hesselman, m.jonker`}`@utwente.nl`
[2] SIDN Labs, Arnhem, The Netherlands `cristian.hesselman@sidn.nl`

Abstract. Accurate Internet measurements depend on well-defined targets. A popular mechanism for target selection is domain-based top lists, e.g., the Tranco or Cisco Umbrella lists. Such lists have a few shortcomings such as the lack of aggregation across related domain names and high volatility over time. Prefix Top Lists (*PTL*) were introduced in 2019 to address these issues, by aggregating domain names into IP prefixes and applying a Zipf-based ranking model to improve stability and representativeness, nonetheless, the original *PTL* resource was discontinued, leaving a gap in publicly available prefix-level data.
In this replication study, we revive and enhance the *PTL* resource by incorporating a broader range of domain-based top lists. Our approach involves mapping domain names to IP prefixes using DNS resolution and BGP routing data, ranking prefixes through a Zipf-based weighting system, and conducting three use-case studies to promote the applicability of *PTLs*. We release the complete *PTL* toolchain as open-source software and publish weekly *PTL* snapshots under `https://openintel.nl/data/prefix-top-lists`, ensuring sustained, versioned and publicly accessible prefix-level rankings for the measurement community.

1 Introduction

Internet measurements typically begin with careful target selection [21]. Researchers often rely on Domain-based Top Lists (DTLs), such as Cloudflare Radar [14], Cisco Umbrella [13] and Google CrUX [18] to identify key Internet properties. However, these lists exhibit rank fluctuations, do not group related infrastructure under a single entity, and lack a consistent weighting mechanism, making it challenging to interpret their significance [26]. Specifically, DTLs fluctuate significantly over time and treat separate domain names under the same Autonomous System (AS) or organization (e.g., google.com and google.co.uk) as distinct entities, despite serving the same function [21]. Additionally, these lists do not reflect the relative importance of ranked entries, making it unclear how much more significant one domain name is compared to another. For example, a low-traffic but persistent domain name might be ranked similarly to a widely used, high-impact one, despite their vastly different influence on the Internet ecosystem.

To address these challenges, the ***Prefix Top Lists (PTL)*** method introduced in [21] aggregated domain names into IP prefixes, and employed a Zipf-based ranking [27] to reflect relative importance, and reduce temporal volatility. *PTLs* complement DTLs by shifting focus to the infrastructure-level, capturing how services are actually deployed across IP space. By exposing this infrastructure, *PTLs* reveal patterns of concentration and operational dependencies that domain-based top lists cannot capture. This makes *PTLs* a complementary resource that enables network measurement studies [21, 9, 25].

PTLs are operationally important because key Internet processes, e.g., routing, hijack propagation, DNS and CDN deployment and security enforcement, occur at the prefix and AS level. This view is essential for understanding patterns of centralization, exposure to routing risks, and operational diversity. Since prefixes are the units of routing and control in BGP, *PTLs* provide a stable and interpretable lens into the Internet's structural hubs, making them well-suited for measurement tasks that require network-aware target selection. While the initial study relied on DTLs from Cisco Umbrella, Majestic and the since retired Alexa, we expand this scope to include additional sources that offer varied perspectives on domain name importance, including: Tranco, Google CrUX and Cloudflare Radar [13, 19, 18, 23, 14].

However, despite its value, the original *PTL* resource [5] was discontinued, creating a void in publicly available prefix-level data. This absence limits the ability of researchers and operators to conduct longitudinal studies and reproduce previous analyses, highlighting the need for a renewed and expanded approach to prefix-based ranking methodologies. In this work, we aim to *revive* and *enhance* this valuable resource, and commit to sustaining it for the research community.

Specifically, we replicate and extend the *PTL* methodology by reconstructing and improving prefix-level aggregation, while adhering to the respective ACM replication guidelines [10]. Our replication approach is as follows:

1. We gather recent DTLs, i.e.,: *Umbrella, Majestic, CrUX, Tranco, and Radar.*
2. We leverage public *DNS* and *BGP data* through the OpenINTEL platform [24] to map domain names to their respective IP prefixes and ASes.
3. We implement both a *Zipf-based ranking model* and a *presence-based frequency model*, allowing flexible aggregation strategies based on domain name rank or multi-list presence.
4. We explore the applicability of *PTLs* across three study domains: a) DNS name server deployment, b) BGP prefix hijacks prevalence, and c) Post-Quantum Cryptography (PQC) adoption.
5. We provide the complete *PTL* toolchain as open-source code under `https://github.com/kastanakis/prefix-top-lists` and publish weekly *PTL* snapshots via the long-standing OpenINTEL platform at `https://openintel.nl/data/prefix-top-lists`, ensuring stable, versioned, and long-term availability of both the methodology and the datasets.

By reviving and enhancing *PTLs*, we aim to create a robust resource that will benefit the Internet measurement ecosystem. We commit to continually updating the *PTL* data and maintaining accessibility.

1.1 Enhancements Over Original PTL Study

While our work is rooted in the original *PTL* methodology introduced at IMC 2019 [21], we introduce several important advancements that significantly improve both the quality and the scope of prefix-level Internet measurements.

First, rather than, as in the original study, relying on in-house DNS resolution, we ingest public DNS data from the OpenINTEL project [24]. This relieves us and anyone seeking to reproduce our results from having to perform DNS queries, and is designed so others can access the same underlying data.

Second, we expand the range of DTLs used to construct *PTLs*. Whereas the original work relied on Alexa, Majestic, and Umbrella, our study incorporates a more diverse set of DTLs, including Tranco, Google CrUX, and Cloudflare Radar, which capture a more diverse and modern view of popular Internet infrastructure.

Since the original *PTL* work, several shortcomings of DTLs have been addressed by efforts such as Tranco [23], which stabilizes domain name rankings via aggregation across multiple sources, and CrUX [18], which offers accurate, user-centric web performance telemetry. While these advances reduce volatility and increase representativeness at the domain level, they do not resolve a distinct challenge: the disconnect between domain names and network infrastructure. As such, *PTLs* and modern DTLs offer complementary perspectives on Internet measurement.

In addition to the methodological improvements, we also broaden the scope of the use-case analyses. While the original *PTL* work focused solely on DNS name server deployment, we further demonstrate the applicability of *PTLs* to two additional contemporary challenges [20, 12]: the measurement of prefix hijack incidents and the analysis of Post-Quantum Cryptography (PQC) adoption trends. These use cases illustrate the continued relevance of prefix-based aggregation for both operational and security-focused network measurement.

Furthermore, we are committed to the ongoing public availability of our data and tooling to promote reproducibility and longitudinal research. We plan to release updates on a weekly basis, aligned with the release cycles of source datasets such as Umbrella and Tranco, and archive historical snapshots to support longitudinal analysis. Collectively, these enhancements position our replication as a robust, flexible, and sustainable resource for the measurement community.

2 Prefix Top Lists (*PTLs*)

This section describes our approach for generating *PTLs*, from input DTLs and DNS/BGP data to the final prefix rankings, as illustrated in Figure 1.

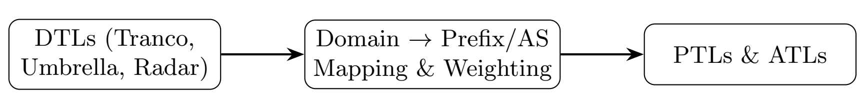

Fig. 1: Overview of the PTL/ATL generation pipeline.

Tranco Domain	**Weight**	**Top Rank**	**Bottom Rank**	**Umbrella Domain**	**Weight**	**Top Rank**	**Bottom Rank**
google.com	0.0693	1	1	**google.com**	0.0510	1	10
microsoft.com	0.0346	2	2	**microsoft.com**	0.0238	1	375
mail.ru	0.0231	3	3	data.microsoft.com	0.0108	3	223025
facebook.com	0.0173	4	4	events.data.microsoft.com	0.0086	4	42786
root-servers.net	0.0138	5	5	**apple.com**	0.0063	6	288
dzen.ru	0.0115	6	6	office.com	0.0061	6	3458
amazonaws.com	0.0099	7	7	clientservices.googleapis.com	0.0053	14	31
apple.com	0.0086	8	8	live.com	0.0049	8	89958
youtube.com	0.0077	9	9	e2ro.com	0.0043	5	61
googleapis.com	0.0069	10	10	windowsupdate.com	0.0042	8	121494

Table 1: Top 10 domain names from Tranco and Umbrella (Apr 1–7, 2025), with Zipf-based weights averaged over a 7-day rolling window. Rank ranges reflect Tranco's long-term aggregation and Umbrella's real-time DNS activity; limited overlap (in bold) shows how source choice affects observed domains.

2.1 Domain Top List Generation

Existing domain-based top lists [13, 19, 18, 23, 14] have the following properties. They: a) generally rank domain names based on proprietary metrics; b) differ in methodology and underlying data; and c) are prone to short-term fluctuations [26, 21]. To improve their stability, we follow the conventions of the original work and apply a Zipf-based weighting scheme [27] to each list individually.

Zipf's law has been used to model the popularity of web content and domain access patterns [17, 11, 16, 8], where a few top-ranked domain names receive disproportionately high attention while the majority receive very little. By assigning Zipfian weights, we ensure that higher-ranked domain names have a stronger influence on the final ranking, while still accounting for long-tail entries. To formalize this intuition, we compute a normalized Zipfian weight for each domain name based on its rank position. For a given rank k in a list of size N, the Zipf weight w_k (using $s = 1$, consistent with the original study [21]) is:

$$w_k = \frac{1/k^s}{\sum_{n=1}^{N} 1/n^s}$$

Here, the numerator $1/k^s$ reflects that higher-ranked domains receive more weight than lower-ranked ones, while the denominator $\sum_{n=1}^{N} 1/n^s$ acts as the normalization constant that scales these values across the entire list. This normalization ensures that the weights sum to one and that lower-ranked domain names contribute progressively less to the overall distribution.

To further stabilize the data, we aggregate weights using a seven-day sliding window (consistent with the original PTL methodology [5]). For each domain name, we compute the average Zipf weight over the preceding week. Domain names that do not appear in a given list on a particular day are assigned a weight of zero for that day. This smoothing process reduces the impact of daily fluctuations and ensures that transient spikes or drops in popularity do not dominate the resulting prefix rankings. The final averaged weights form the input to our DNS resolution and prefix aggregation steps.

The DTLs for April 1-7, 2025, based on the Tranco and Umbrella datasets [23, 13], are shown in Table 1. Tranco provides stable domain name rankings by aggregating data from multiple sources over a rolling window, effectively smoothing out short-term fluctuations. In contrast, Cisco Umbrella derives its rankings from real-time DNS query data observed at recursive resolvers, which captures a broader spectrum of activity and as a result, Umbrella rankings exhibit significantly higher variance, reflecting the dynamic DNS nature at the edge.

2.2 DNS Resolution and Routing Data

Unlike the original *PTL* study, which performed custom DNS resolution in-house, we use publicly available DNS datasets. This approach saves us and those seeking to generate their own *PTLs* with our tooling from the operational complexity of deploying and managing a custom resolution pipeline. We use data from the OpenINTEL project, which makes various types of DNS data available for academic research [24]. We specifically rely on forward DNS (fDNS) measurement data, which OpenINTEL collects daily by querying sizable lists of domain names for address records (i.e., `A` and `AAAA`) and others. Among its public data are top-list based measurements for Tranco, Majestic, CrUX, Umbrella, and Radar. OpenINTEL also embeds BGP-related data such as the covering prefix and origin AS of the resolved IP addresses making it particularly suitable for our replication study.

2.3 Prefix Top List Generation

We construct a Prefix Top List (*PTL*) by augmenting the DTL from Sec. 2.1 with DNS data obtained from OpenINTEL (see Sec. 2.2) over the same seven-day sliding window. Each domain name d is assigned a normalized weight w_d according to its Zipfian rank in the input *DTL*. Domain names frequently resolve to multiple IP addresses—e.g., due to global load balancing, CDN distribution, or DNS-based failover. We denote the set of addresses associated with a domain name d as:

$$I(d) = \{i_1, i_2, \ldots, i_m\}$$

We assume that each IP address in $I(d)$ shares equal responsibility for serving d, and thus evenly divide the domain name's weight across its resolved addresses. Each address $i \in I(d)$ receives a proportional share of $\frac{w_d}{|I(d)|}$. OpenINTEL maps DNS address records to one or more BGP prefixes $P(i)$ based on publicly available BGP routing data. Typically, an address will map to a single, most-specific prefix, but multi-origin entries can occur as well. We use these metadata to compute the weight of each prefix p by summing the contributions of all addresses that map to it:

$$W(p) = \sum_{\substack{d \\ i \in I(d) \\ P(i)=p}} \frac{w_d}{|I(d)|} \tag{1}$$

(a) **Prefix Top List (Zipf)**

Prefix	AS	Weight	# of Domains
2a00:1450:400e::/48	15169	0.0705	20473
2a02:26f0:1180::/48	20940	0.0264	34724
142.250.0.0/15	15169	0.0193	18853
142.251.36.0/24	15169	0.0186	7502
2603:1000::/25	8075	0.0166	14324
2606:4700::/44	13335	0.0141	47681
162.159.128.0/19	13335	0.0110	14101
172.217.23.0/24	15169	0.0107	8405
151.101.204.0/22	54113	0.0102	18848
172.217.0.0/16	15169	0.0090	10008

(b) **Prefix Top List (Presence)**

Prefix	AS	Weight	# of Domains
2606:4700:3030::/48	13335	0.0215	167066
23.227.38.0/23	13335	0.0173	72047
2606:4700:20::/44	13335	0.0156	73091
104.26.0.0/20	13335	0.0112	75479
2606:4700::/44	13335	0.0105	62653
141.193.213.0/24	209242	0.0088	23516
2620:127:f00f::/48	13335	0.0081	58288
188.114.96.0/24	13335	0.0080	116787
188.114.97.0/24	13335	0.0080	116782
2a06:98c1:3120::/48	13335	0.0074	112756

(c) **AS Top List (Zipf)**

ASN	Weight	# of Prefixes
15169	0.1576	390
13335	0.1550	442
16509	0.1161	3389
8075	0.0646	208
20940	0.0636	829
54113	0.0438	200
32934	0.0366	352
14618	0.0265	291
16625	0.0259	580
396982	0.0214	1995

(d) **AS Top List (Presence)**

ASN	Weight	# of Prefixes
13335	0.2654	540
16509	0.1271	3522
8075	0.0393	222
14618	0.0274	299
396982	0.0207	2050
20940	0.0186	879
24940	0.0182	60
16276	0.0178	170
209242	0.0174	254
15169	0.0171	396

Fig. 2: Top 10 entries from the Prefix Top Lists (PTL) and AS Top Lists (ATL) generated using two methods: Zipf-based ranking and presence-based frequency.

This yields a weighted ranking of BGP prefixes based on the volume of popular domain names they serve, defined as the Prefix Top List (PTL). We then aggregate these prefix-level weights to the AS level. Let $P(AS)$ denote the set of prefixes originated by AS AS. The total weight assigned to AS is:

$$W(AS) = \sum_{p \in P(AS)} W(p) \tag{2}$$

This yields a weighted ranking of ASes based on the volume of popular prefixes they serve, defined as the AS Top List (ATL). Unlike traditional domain-centric lists, these rankings (i.e., PTL and ATL) highlight which prefixes and ASes play central roles in hosting or distributing the most frequently accessed web services.

Our approach generates two types of Top lists: ranked (Zipf-based) and presence-based. For the ranked $PTLs$ and $ATLs$, we apply a Zipfian weighting scheme to domain-based top lists from Tranco, Umbrella, and Majestic, assigning greater weight to higher-ranked domain names. In the presence-based variant, we ignore domain ranks and instead assign weights based only on whether a domain name appears in each input list. The resulting value reflects the domain's normalized appearance frequency across sources. We then aggregate these weights to prefixes and ASes following the same method used for the Zipf-weighted lists.

These variants are constructed using a broader set of sources, including CrUX and Cloudflare Radar in addition to Tranco, Umbrella, and Majestic, to better reflect raw domain appearance frequency across diverse datasets.

Figure 2 shows the top 10 prefixes and ASes under each ranking variant, illustrating the differences in outcome between Zipf-based and presence-based methods. Zipf-based rankings (subfigure a) are dominated by Google prefixes, making Google (AS15169) the top AS in the corresponding ATL (subfigure c), due to a few highly ranked services like YouTube and Google Search. In contrast, the presence-based PTL (subfigure b) surfaces more Cloudflare-operated prefixes, with Cloudflare (AS13335) topping the presence-based ATL (subfigure d), reflecting its support for many moderately popular or regional domain names. These patterns illustrate how aggregation shapes AS-level rankings and highlight differing strategies: centralized hosting vs. widespread coverage.

To comply with the original methodology and due to space constraints, we focus our analysis on the Zipf-based ranked lists. However, we release the presence-based variants to support research that prioritizes breadth of coverage over popularity [6], e.g., assessing geographic diversity, evaluating the spread of services across hosting providers, or tracking technology adoption in the long tail.

2.4 Considerations for DNS Resolution

OpenINTEL provides broad DNS and BGP coverage across multiple top lists, but like any passive measurement platform, it is not exhaustive. Some domain names may remain unresolved, and not all resolved IP addresses map cleanly to BGP prefixes, particularly in cases involving sparse or short-lived announcements. Although we do not supplement the dataset with active DNS resolution in this study, our methodology is compatible with such extensions. We consider this a pragmatic trade-off between reproducibility and completeness.

As a result, our mapping approach requires additional care in handling incomplete data. Some domain names may be delegated (i.e., have an `NS` record) and be included in a DTL due to query volume even though no address records exist, or failed to resolve due to transient errors. We disregard such names, consistent with the original study [5]. To preserve ranking integrity, we normalize weights after filtering. Once non-resolvable domain names are removed, the remaining weights are rescaled to sum to 1.0, preventing bias toward any specific prefix or AS.

3 Temporal Analysis of Prefix Top Lists

In this section, we analyze the growth of *PTLs* over an one-month period, using four weekly snapshots. Our analysis begins in mid-March, 2025, coinciding with the start of our data collection. For each weekly snapshot, we identify newly discovered BGP prefixes (i.e., those not present in earlier lists) and quantify their relative importance by accumulating their Zipf weights.

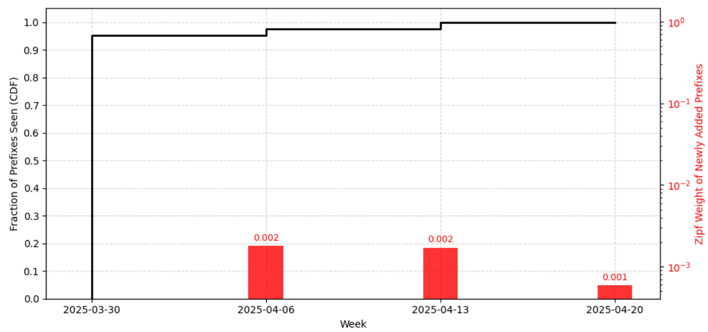

Fig. 3: Growth of Unique Prefix Coverage by *PTL*. The left y-axis shows the CDF of discovered prefixes on a linear scale, representing the proportion of total unique prefixes observed over time. The right y-axis shows the cumulative Zipf weight of newly added prefixes on a logarithmic scale, emphasizing the diminishing contribution of lower-ranked domain names.

Fig. 3 shows the cumulative prefix discovery across the four weeks. The black step line (left y-axis) indicates the fraction of total prefixes observed over time, while the red bars (right y-axis, log scale) show the Zipf weight of newly added prefixes. Prefix coverage increases rapidly in the first snapshot (March 30), capturing over 90% of all observed prefixes. Subsequent weeks (April 6, 13, and 20) add only a few new prefixes, contributing marginal Zipf weights of 0.002, 0.002, and 0.001, respectively, demonstrating the long-tail of domain name popularity.

Prefix weights also remain stable across time. Although we do not calculate explicit weight deltas (as in the original study), the minimal changes in cumulative Zipf weight indicate low volatility across snapshots. This behavior mirrors the original findings [5] and reflects that prefixes associated with highly ranked domains tend to be discovered early and persist across measurement periods. As a result, most of the PTL's structure converges quickly, which is consistent with the intended stability of prefix- and AS-level rankings. A more extensive, longer-term stability analysis is beyond the scope of this paper but represents a natural direction for future work.

4 Applications of Prefix Top Lists

To demonstrate the practicability of the revived *PTL* resource, we evaluate its utility in three network measurement domains: a) Border Gateway Protocol (BGP) security, b) Post-Quantum Cryptography (PQC) compliance, and c) Domain Name System (DNS) resilience.

It is important to note that *PTLs* are not intended as direct replacements for DTLs. The use cases in this section illustrate analyses that benefit specifically from a prefix- or AS-level perspective, which cannot be reliably inferred from domain-level rankings alone.

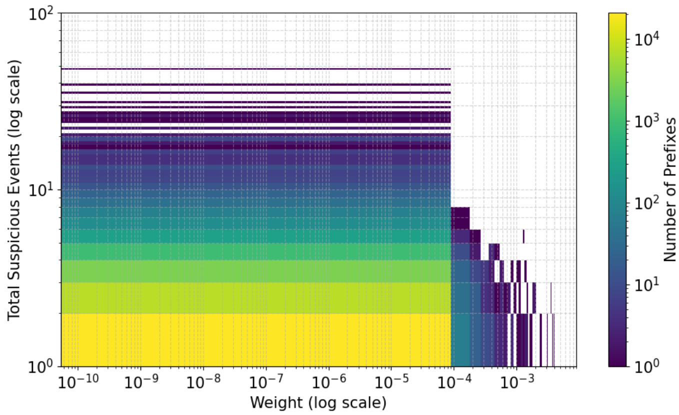

Fig. 4: Heatmap showing the density of prefixes based on their weight and the total number of suspicious events observed in 2024. Both axes are log-scaled, and color reflects the number of prefixes in each bin.

4.1 Exposure to Suspicious Routing Events

To evaluate the relationship between prefix popularity and routing security incidents [20], we analyzed suspicious routing events using the Global Routing Intelligence Platform (GRIP) from Georgia Tech's INET-Intel platform [1].

We queried GRIP's API for suspicious events through the whole duration of 2024 involving each prefix in our Zipf-weighted PTL for April 1-7, 2025 (compiled in Section 2.3). For each prefix, we retrieved the total number of Multiple Origin AS (MOAS) events flagged with high suspicion scores (> 80) occurring between January 1 and December 31, 2024. Although there is a temporal offset between the PTL snapshot and the MOAS event data, Section 3 shows that popular prefixes remain stable across neighboring periods, making it reasonable to relate the 2025 snapshot to 2024 routing events. If a prefix had no associated events, it was recorded with a count of zero. The result is a mapping between PTL*-ranked* prefixes and their corresponding number of suspicious events.

Fig. 4 presents the results as a log-log heatmap. The x-axis shows the popularity as derived from the PTL rankings, and the y-axis shows the total number of suspicious events associated with each prefix. The color of each bin indicates how many prefixes fall into that weight–event range, using a logarithmic color scale: brighter areas (yellow) represent higher concentrations of prefixes, while darker areas (purple) indicate sparsely populated regions. While most high-weight prefixes (right) experience few or no suspicious events, lower-weight prefixes (lower left quadrant) exhibit higher counts. This exploratory view suggests that less prominent infrastructure may be more prone to routing anomalies or targeted hijacks. In contrast, prefixes with both high weight and high event counts are rare but may signal critical infrastructure under persistent targeting.

Table 2: PQC compliance across popularity tiers. A prefix is classified as compliant if at least one domain within it successfully completes a PQC handshake.

Prefix popularity tier	**PQC-compliant prefixes**
Top 100 prefixes	71%
Top 1,000 prefixes	48%
Top 10,000 prefixes	29%
All prefixes	6%

4.2 Investigation of PQC Deployment

To assess the early adoption of post-quantum cryptography (PQC) [12], we integrate TLS handshake testing into our generated *PTL* for April 1-7. Our methodology aligns with recent community-driven scanning efforts [7, 15] and serves as a practical baseline for monitoring PQC adoption at the prefix level.

Building on the Open Quantum Safe (OQS) project [3], we deploy a custom TLS 1.3 scanner based on OpenSSL [4] with the oqsprovider module enabled. This setup allows us to attempt handshakes with hybrid key exchange groups. Specifically, we test the following popular [2, 7] groups: `mlkem768`, `X25519MLKEM768`, `SecP256r1MLKEM768`, `x25519kyber768`. For each domain name in the *PTL* dataset, the scanner attempts to negotiate a connection using each hybrid group. To capture whether a prefix exhibits any evidence of PQC readiness, we apply a binary rule: if at least one domain name within a prefix successfully completes a PQC handshake, we classify the entire prefix as compliant. This consideration reflects the prefix-oriented nature of our study, where the presence of a single PQC-capable domain name demonstrates that the underlying infrastructure can support PQC, even if deployment is not uniform across all hosted domain names.

Table 2 summarizes PQC compliance rates across prefix popularity tiers, from the top 100 to the full set of observed prefixes. We observe a strong decreasing trend: while 71% of prefixes in the top 100 tier host at least one PQC-capable domain name, this rate drops sharply with rank. The compliance rate falls below 30% for prefixes ranked beyond 10,000 and below 6% overall. This suggests that PQC adoption is currently concentrated among prominent prefixes, indicating an uneven and early-stage deployment pattern. These findings offer a prefix-level view of PQC adoption, showing that support is emerging but concentrated in popular, likely well-resourced prefixes.

4.3 Evaluation of DNS Resilience

To replicate the DNS analysis from the original *PTL* study [5], we evaluate name server placement compliance with RFC 2182 [22] across the April 1-7 *PTL*. As the RFC requires name servers to span multiple prefixes, we identify each zone's hosting prefixes and treat zones located within a single prefix as non-compliant.

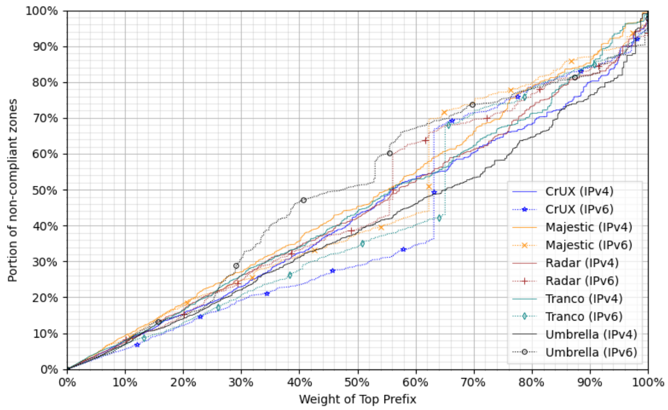

Fig. 5: CDF of non-compliant DNS zones sorted by prefix weight.

Fig. 5 shows the cumulative distribution of non-compliant zones across our ranked *PTLs*, broken down by *DTL* and IP version. On the x-axis, prefixes are sorted by their cumulative weight, from most popular (left) to least popular (right). The y-axis represents the fraction of non-compliant zones. The IPv4 distribution is relatively smooth, suggesting that non-compliance is more evenly spread across prefixes. In contrast, IPv6 displays step-like jumps, indicating that misconfigurations are concentrated within a small number of top-ranked prefixes.

Notably, Umbrella (IPv6) shows the earliest and most pronounced jump, starting around the 30% mark, whereas the other top-listed zones demonstrate similar but smaller spikes that occur later, roughly after the 55% weight mark for Radar and after 60% for Majestic, Crux and Tranco. These findings align with the original *PTL* study and confirm that many popular domain names, do not adhere to robust DNS deployment practices, especially in the IPv6 space. An additional breakdown by domain name rank is provided in Appendix B.

5 Conclusion

In this study, we revived and enhanced the *PTL* resource originally introduced at IMC'19. Our findings corroborate key insights from the original work, including the early discovery of high-impact prefixes and the long-tail distribution of domain name popularity. Through three applied use cases, we demonstrate that *PTLs* offer a meaningful lens into the structure and stability of the Internet's edge. We publicly release our tooling and data to support further reproducibility and future research, and commit to continually updating the *PTL* dataset.

Acknowledgements

This research received funding from the Dutch Research Council (NWO) under the projects UPIN and CATRIN.

References

1. Grip: Global routing intelligence platform. https://grip.inetintel.cc.gatech.edu/ (2024), accessed: 2025-07-31
2. Cloudflare research: Post-quantum key agreement. https://pq.cloudflareresearch.com/ (2025), accessed: 2025-08-01
3. Open quantum safe. https://openquantumsafe.org/ (2025), accessed: 2025-08-07
4. Openssl. https://www.openssl.org/ (2025), accessed: 2025-08-07
5. Prefix top lists (2025), https://prefixtoplists.net.in.tum.de/, accessed: 2025-07-30
6. Prefix top lists datasets and code. To be released upon publication. (2025)
7. Sites using pqc (march 2025). https://www.netmeister.org/blog/pqc-use-2025-03.html (2025), accessed: 2025-08-01
8. Adamic, L.A.: Zipf, power-laws, and pareto-a ranking tutorial. Xerox Palo Alto Research Center, Palo Alto, CA, http://ginger. hpl. hp. com/shl/papers/ranking/ranking. html (2000)
9. Amos, R., Acar, G., Lucherini, E., Kshirsagar, M., Narayanan, A., Mayer, J.: Privacy policies over time: Curation and analysis of a million-document dataset. In: Proceedings of the Web Conference 2021. p. 2165–2176. WWW '21, Association for Computing Machinery, New York, NY, USA (2021). https://doi.org/10.1145/3442381.3450048, https://doi.org/10.1145/3442381.3450048
10. Association for Computing Machinery: Artifact Review and Badging (2025), https://www.acm.org/publications/policies/artifact-review-and-badging-current, accessed: 2025-07-30
11. Berners-Lee, T.: The fractal nature of the web. https://edshare.soton.ac.uk/392/3/DesignIssues/Fractal.html (1998), accessed: 2025-08-01
12. Bernstein, D.J., Lange, T.: Post-quantum cryptography. Nature **549**(7671), 188–194 (2017)
13. Cisco Systems, I.: Cisco umbrella popularity list. Website (2025), https://umbrella.cisco.com/, accessed: 2025-07-30
14. Cloudflare, I.: Cloudflare radar: Internet traffic trends. Website (2025), https://radar.cloudflare.com/, accessed: 2025-07-30
15. Fabrizio, G., Sperotto, A., Van Rijswijk-Deij, R.: Measuring the impact of post-quantum cryptography on complex applications: A case study on federated identity management. In: 2025 9th Network Traffic Measurement and Analysis Conference (TMA). pp. 1–10 (2025). https://doi.org/10.23919/TMA66427.2025.11097002
16. Kepner, J., Cho, K., Claffy, K.C., Gadepally, V., McGuire, S., Milechin, L., Arcand, W., Bestor, D., Bergeron, W., Byun, C., et al.: New phenomena in large-scale internet traffic. In: Massive Graph Analytics, pp. 241–285. Chapman and Hall/CRC (2022)
17. Krashakov, S.A., Teslyuk, A.B., Shchur, L.N.: On the universality of rank distributions of website popularity. Computer Networks **50**(11), 1769–1780 (2006). https://doi.org/https://doi.org/10.1016/j.comnet.2005.07.009, https://www.sciencedirect.com/science/article/pii/S1389128605002513
18. LLC, G.: Chrome user experience report (crux). Website (2025), https://developers.google.com/web/tools/chrome-user-experience-report, accessed: 2025-07-30
19. Majestic: Majestic million: The top 1 million websites. Website (2025), https://majestic.com/reports/majestic-million, accessed: 2025-07-30

20. Mitseva, A., Panchenko, A., Engel, T.: The state of affairs in bgp security: A survey of attacks and defenses. Computer Communications **124**, 45–60 (2018)
21. Naab, J., Sattler, P., Jelten, J., Gasser, O., Carle, G.: Prefix top lists: Gaining insights with prefixes from domain-based top lists on dns deployment. In: Proceedings of the Internet Measurement Conference. p. 351–357. IMC '19, Association for Computing Machinery, New York, NY, USA (2019). https://doi.org/10.1145/3355369.3355598, `https://doi.org/10.1145/3355369.3355598`
22. Patton, M.A., Bradner, S.O., Elz, R., Bush, R.: Selection and Operation of Secondary DNS Servers. RFC 2182 (Jul 1997). https://doi.org/10.17487/RFC2182, `https://www.rfc-editor.org/info/rfc2182`
23. Pochat, V.L., Van Goethem, T., Tajalizadehkhoob, S., Korczyński, M., Joosen, W.: Tranco: A research-oriented top sites ranking hardened against manipulation. arXiv preprint arXiv:1806.01156 (2018)
24. van Rijswijk-Deij, R., Jonker, M., Sperotto, A., Pras, A.: A high-performance, scalable infrastructure for large-scale active dns measurements. IEEE Journal on Selected Areas in Communications **34**(6), 1877–1888 (2016). https://doi.org/10.1109/JSAC.2016.2558918
25. Ruth, K., Kumar, D., Wang, B., Valenta, L., Durumeric, Z.: Toppling top lists: evaluating the accuracy of popular website lists. In: Proceedings of the 22nd ACM Internet Measurement Conference. p. 374–387. IMC '22, Association for Computing Machinery, New York, NY, USA (2022). https://doi.org/10.1145/3517745.3561444, `https://doi.org/10.1145/3517745.3561444`
26. Scheitle, Q., Hohlfeld, O., Gamba, J., Jelten, J., Zimmermann, T., Strowes, S.D., Vallina-Rodriguez, N.: A long way to the top: Significance, structure, and stability of internet top lists. In: Proceedings of the Internet Measurement Conference 2018. pp. 478–493 (2018)
27. Wikipedia contributors: Zipf's law — wikipedia, the free encyclopedia. Website (2025), `\url{https://en.wikipedia.org/wiki/Zipf\%27s_law}`, accessed: 2025-07-30

A Ethics

This work does not raise any ethical issues. We consulted the original Prefix Top Lists authors, who acknowledged our replication effort without objections. All TLS handshakes were performed using non-intrusive probes with rate limiting in place to avoid overwhelming any individual host. The measurements were conducted over an extended period to distribute load, and no content was fetched beyond the handshake itself.

B Breakdown by Domain Rank

This appendix complements Section 4.3 by showing the distribution of DNS non-compliance across domain ranks. Fig. 6 shows the cumulative distribution of non-compliant zones across domain-based top lists. As in the original study [21], the curves are mostly linear, indicating that non-compliance to RFC2182 [22] is

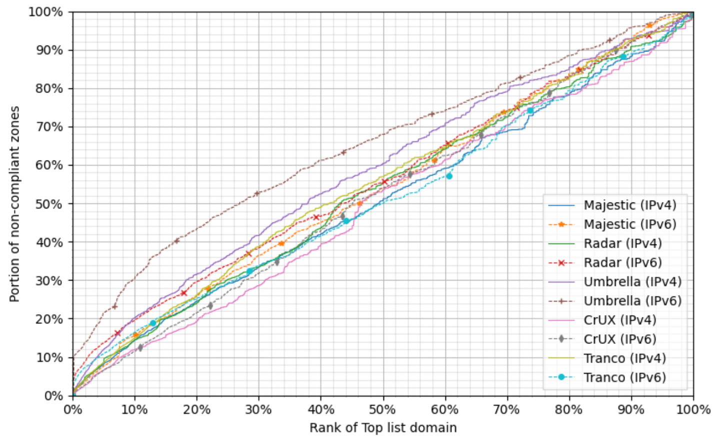

Fig. 6: Share of DNS zones with non-compliant NS placement, sorted by DTL-ranked domains.

fairly evenly distributed across domain ranks. However, meaningful differences emerge between sources and IP versions.

IPv6 consistently shows higher rates of non-compliance than IPv4, with Umbrella (IPv6) exhibiting the steepest slope reflecting its inclusion of many ephemeral or operational domains with poor NS redundancy. In contrast, lists like Tranco and Majestic show slightly lower non-compliance at the top, likely due to their emphasis on stable, long-lived web domains. These patterns confirm that poor topological diversity in NS placement is not confined to the tail of domain rankings but remains a widespread issue, especially in IPv6 environments.

Crucially, the variability across lists also reinforces the motivation for prefix-based aggregation. While domain-level rankings reflect differing views of popularity and usage, they do not account for shared infrastructure or hosting patterns. *PTLs* address this by elevating the analysis to the prefix level, smoothing out list-specific biases and surfacing systemic configuration issues tied to underlying infrastructure regardless of which domains are affected.