More Space, Less Privacy? Measuring the Effectiveness of IP-based Website Fingerprinting in IPv6

Sumeer Ahmad[⋆†] Michalis Polychronakis[⋆] Theophilus A. Benson[§] Nguyen Phong Hoang[†]
[⋆]Stony Brook University [§]Carnegie Mellon University [†]University of British Columbia

Abstract

Despite the widespread adoption of domain name encryption protocols (e.g., DoH, DoT, and ECH), website fingerprinting attacks remain a significant threat to online privacy due to the visibility of IP connections that can be observed by network-level adversaries. This problem has been investigated in IPv4 thoroughly but remains largely unexplored in IPv6, where it is even more pronounced. From a design perspective, IPv6 offers a vastly larger address space, eliminating the need for virtual hosting and domain co-location—practices prevalent in IPv4. These practices obscure several domains behind a single IPv4 address. The adoption of IPv6 can theoretically lead to more accurate fingerprinting based on IPv6 connections because each domain is now more likely to be resolved to a unique globally routable IPv6 address unlike IPv4.

In this study, we systematically investigate the feasibility, accuracy, and privacy implications of IP-based website fingerprinting in IPv6 environments. Utilizing empirical data collected via active DNS measurements, Web crawling, and entropy-based analyses across half a million dual-stack websites, we conduct the largest evaluation of website fingerprinting in IPv6. We find that dual-stack websites that still have some IPv4-only dependencies are easier to fingerprint, achieving close to 94% accuracy on both IPv4 and IPv6. However, fingerprinting pure dual-stack websites is significantly harder: accuracy drops to 56% over IPv4 and 45% over IPv6. These results reveal distinct trends in hosting infrastructures and indicate that IPv6 itself does not inherently diminish privacy, contrary to existing concerns in the community. Rather, fingerprinting risk is shaped by how hosting providers deploy IPv6 and their choices regarding domain co-location.

Keywords

IPv6, website fingerprinting, web privacy, domain name encryption, traffic analysis.

1 Introduction

With the growing deployment of domain name encryption protocols, e.g., DNS over TLS (DoT) [38], DNS over HTTPS (DoH) [35], and Encrypted Client Hello (ECH) [68], plaintext domain names in DNS queries and the TLS ClientHello SNI extension field [67] are no longer visible to network-level observers. Despite concealing domain names, these mechanisms do not hide the IPs or other transport-level metadata. During a typical website visit, a client communicates with multiple servers to retrieve various resources [94], and the destination IPs of these outgoing packets are still visible. This visibility enables network-level observers to fingerprint the websites users visit based on the IP-level characteristics of encrypted network connections [87].

Internet Service Providers (ISPs) can exploit these IP-based fingerprints to reconstruct users' browsing habits, generate behavioral profiles [43, 57, 73], and monetize this information through advertising or data-sharing practices. Network administrators (e.g., enterprise, campus, or public networks) can similarly monitor users' activity, enforce restrictive usage policies, or engage in surveillance that extends beyond legitimate operational needs. Government censors can use the same fingerprinting techniques to detect connections to politically sensitive content, social media, or foreign websites, and subsequently block, throttle, or interfere with access [11, 33, 62, 92].

The effectiveness of IP-based fingerprints primarily depends on the degree of IP sharing across domains. If an IP address hosts a single domain, an adversary can easily correlate that IP with that particular domain. Conversely, if an IP address hosts multiple domains, it becomes more challenging to accurately infer the specific domain being contacted based solely on the hosting IP address [32]. The limited address space of IPv4, along with the widespread use of virtual hosting and server-side domain co-location for sharing IPs among multiple services, makes website fingerprinting inherently non-trivial [5].

In principle, IPv6's vastly larger address space removes the scarcity constraints that shaped IPv4 deployment. Unlike IPv4, where address exhaustion incentivized extensive IP address sharing across domains, IPv6 allows providers to allocate addresses more liberally [17, 53]. Whether IPv6 deployments actually translate into more discriminative IP-based fingerprints remains an open empirical question. As IPv6 adoption continues to expand [56, 64], it is therefore critical to evaluate how real-world IPv6 deployment practices shape the risk of IP-based fingerprinting.

In this paper, we explore the practicality of IP-based website fingerprinting in the context of IPv6-enabled Web browsing. We conduct a large-scale measurement study of half a million websites—the largest of its kind at the time of writing—to assess the effectiveness of fingerprinting in the realm of IPv6. Analyzing data and metrics collected from active DNS resolutions, Web crawls, and Shannon entropy-based analysis [74], our study answers the following interrelated research questions:

(1) *What are the trends in IPv6 deployment across different website categories, and how do these affect domain-to-IP mappings?* (§4.1 and §5.2)
We investigate whether IPv6-enabled website(s) are more likely to have unique, persistent IPs (IPv4+IPv6), and analyze IPv6 usage across hosting infrastructures, CDNs, and secondary resources (e.g., advertisements, extensions, plug-ins). Around 56%

Proceedings on Privacy Enhancing Technologies 2026(4), 36–52

https://doi.org/10.56553/popets-2026-0109

of the websites in our first measurement batch are capable of operating in a fully IPv6 environment ("dual-stack complete"). As the measurement progressed, some domains became unavailable over IPv6 due to misconfigurations or maintenance; the longitudinally stable figure is approximately 32% (Figure 3).

(2) *How does IPv6 adoption influence the effectiveness of IP-based fingerprinting compared to IPv4?* (§5.3 and §6.1)
We assess whether IPv6 enhances or weakens fingerprinting under real-world conditions, especially in environments with domain name encryption in place. Combining both dual-stack complete and incomplete sites, we correctly identify 72% of all websites using IPv4-based fingerprints, and 68% using IPv6-based fingerprints.

(3) *Can stable, scalable, and computationally efficient fingerprinting techniques be used to test large IPv6 fingerprinting datasets over time?* (§6.2 and §6.4)
We apply entropy-based analysis [74] to quantify fingerprinting effectiveness across IPv6's vast address space, through a large-scale measurement of about 520K IPv6-enabled websites. Our longitudinal study also tracks both the stability of domain-IP mappings and churn rates over a period of three months.

(4) *What are the patterns of IPv6 deployment across different hosting providers compared to IPv4? How do they affect IP-based fingerprinting?* (§7.1)
To identify the underlying factors of IPv6 deployment, we analyze domain co-location in IPv4 and IPv6 across major hosting providers, and examine how their deployment influences fingerprintability. We observe trends in how domains are clustered on shared IP addresses across both protocols. Several top providers (e.g., Cloudflare, Google, Fastly) exhibit notably high levels of domain co-location in IPv6, in contrast to other providers such as Amazon.

By answering these questions with empirical evidence, we quantify IPv6's real-world impact on website fingerprinting, assess its potential privacy implications, and examine whether providers are leveraging the full potential of IPv6. To support reproducibility, our code and dataset are publicly available at https://github.com/sumeer2602/ipv6-web-fingerprinting-pipeline.

2 Background and Motivation

This section provides an overview of the key developments in domain name encryption, the continued relevance of website fingerprinting in the face of these technologies, and the transition from IPv4 to IPv6, a change that may reshape the deployment characteristics influencing IP-based distinguishability.

2.1 Website Fingerprinting in IPv4

Website fingerprinting remains a relevant technique for inferring the website(s) a user visits by analyzing observable network metadata. Before the deployment of domain name encryption, adversaries could directly observe plaintext DNS queries and the TLS SNI field, making it trivial to identify user browsing activity [76]. The introduction of DoT, DoH, and ECH, prevents adversaries from relying on the domain name as an identifier for website identification. However, even with domain names concealed, significant portions of traffic metadata remain visible. Adversaries can still leverage features, i.e., packet timing and size patterns [61], website oracle [65], and the sequence of IP connections a client makes to different servers [33, 47] to infer the visited website.

Among these techniques, we focus on IP-based fingerprinting: identifying website(s) based solely on the *set* of observed IP addresses. By omitting temporal statistics, packet-level characteristics, and connection sequences, this abstraction isolates the inherent address-level distinguishability of a service. Consequently, our analysis remains independent of transient network conditions, browser-specific optimizations, or caching artifacts [39, 42]. This focus aligns with a conservative and operationally realistic threat model, where adversaries rely only on readily observable network-layer metadata. In this context, IP addresses serve as persistent network-layer identifiers that can rival the efficacy of browser cookies [25, 30, 55].

While richer ML/DL-based fingerprinting techniques can achieve higher accuracy, they depend on detailed side-channel features, traffic timing analysis, and continuous model retraining [7, 78]. Such approaches may introduce computational and operational overheads that are unnecessary to scale [70] for adversaries who already collect destination IP metadata through mechanisms such as IPFIX [2] or NetFlow [36]. Moreover, destination IP addresses will remain visible even as future protocols encrypt additional metadata. Studying IP-based fingerprints therefore enables us to evaluate privacy risks that persist under minimal attacker capabilities and within increasingly encrypted web ecosystems.

Hoang et al. [33] demonstrated that, despite challenges introduced by virtual hosting, IP-based fingerprinting remains feasible in IPv4 networks. Their study was able to correctly identify 84% of 200K top websites, highlighting the effectiveness of using only IP-level information for fingerprinting attacks. However, this prior work was limited to IPv4 only, and the feasibility of IP-based fingerprinting in IPv6 environments remains unexplored. Inspired by this prior work, we seek to assess whether and how the deployment of IPv6 affects the effectiveness of IP-based fingerprinting.

2.2 The IPv6 Landscape

The transition from IPv4 to IPv6 has been driven primarily by the exhaustion of IPv4 addresses and the need for a protocol that can accommodate the rapidly growing number of connected devices. IPv6 solves this by offering a vastly larger address space, improved routing efficiency, and native end-to-end connectivity [41]. Under IPv4, address scarcity encouraged extensive IP sharing across domains via virtual hosting and shared infrastructure, resulting in large anonymity sets for a domain at IP level.

Theoretically, IPv6 allows devices, domains, or services to be assigned globally routable addresses without the scarcity constraints of IPv4, increasing the potential granularity of domain-IP mappings. Prior measurements report that a substantial fraction of IPv6 addresses observed in web hosting environments are associated with exactly one domain, in contrast to IPv4 addresses, where address scarcity has led to widespread domain co-location [32, 34]. Our measurements exhibit a similar pattern across 850K domains. As shown in Figure 1, IPv6 has a higher number of single-domain addresses (≈250K) compared to IPv4 (≈150K). While IPv6 address abundance reduces the structural pressure for large-scale IP sharing, it does

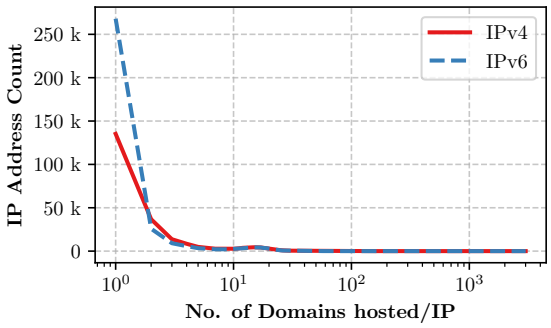

Figure 1: Distribution of number of domains hosted on each IPv4 & IPv6 address in our data.

not eliminate co-location entirely; rather, it reshapes the size and distribution of anonymity sets relative to IPv4 deployments.

The structure of IPv6 deployment itself has raised concerns within the community. Major hosting providers deploy IPv6 in ways that can further amplify these risks. Providers and CDNs such as Cloudflare, Google, and Fastly often co-locate large clusters of domains behind a small set of IPv6 addresses, while others (i.e., Amazon) operate vast IPv6 prefixes that respond on every address within a block, creating distinctive and easily recognizable provider-level patterns [27, 95]. Combined with the well-documented stability of IPv6 prefixes [31], these deployment behaviors create distinctive provider-level patterns that may influence distinguishability. These factors have fueled a growing concern in the community that IPv6 may inadvertently strengthen IP-based tracking.

These concerns are particularly significant for IP-based fingerprinting, where the visibility of unique destination IP addresses in network traffic may allow an adversary to infer the websites a user visits, profile browsing behavior, and compromise user privacy. The trends observed in IPv6 deployment suggest that IPv6 may either strengthen or weaken fingerprinting depending on how addresses are assigned and managed in practice. These contrasting dynamics set the stage for our work and raise two unresolved questions in the community, which we address in this paper:

- *Uniqueness vs. Privacy.* If a large proportion of IPv6 addresses map to a single domain as compared to IPv4, does the higher prevalence of single-domain IPv6 addresses translate into increased fingerprintability?
- *Abundance vs. Stability.* If the vast IPv6 address space encourages frequent churn or rotation, does address abundance affect the stability and persistence of domain-IP mappings over time?

3 Threat Model

We consider a threat model in which an *on-path* adversary—typically an ISP or similar entity capable of monitoring network traffic—seeks to infer the websites a user visits and reconstruct their browsing history. This adversary is regionally co-located with the victim and operates in an environment where domain names are encrypted (e.g., via DoH/DoT/ECH), thus cannot access plaintext DNS queries or TLS SNI fields. However, the adversary can observe destination IP addresses contacted during each browsing session.

We consider the realistic case in which users typically start browsing by loading the landing page of *one website at a time*. Prior web-performance and caching studies [48] show that a site's landing page is responsible for downloading and caching the bulk of its static resources (scripts, styles, fonts, images). Since most pages within a site share the same static resource bundle, later in-site navigation tends to produce relatively small increments of traffic compared to the first page load. The landing page concentrates most of the stable, useful domains for fingerprinting [9, 61].

While multi-tab browsing is possible, empirical studies and user behavior suggest that most interactions still occur sequentially, particularly on mobile or single-screen devices. Even when multiple tabs are open, switching between them generally involves a temporal gap, allowing the adversary to segment the traffic into distinct sessions [93]. There are proposed mechanisms to disaggregate mixed-tab traces into individual website visits using time and flow-based heuristics [20, 22]. We assume the adversary leverages such techniques and operates on clean, isolated traffic segments.

The adversary does not rely on any flow-level metadata (e.g., packet size, frequency, protocol), TLS fingerprints, or any auxiliary identifiers. The adversary only sees the *set of destination IPs* for each website session and uses these to match against pre-built IP fingerprints using the matching algorithm described in §5.3 to predict the website being visited.

We intentionally scope our study to destination IPs alone to establish a *conservative lower bound* on fingerprintability. Any adversary incorporating additional features (port numbers, packet timing, protocol metadata) would only improve upon the accuracy we report. Similarly, this study focuses exclusively on web browsing; non-web protocols (e.g., RTP, VoIP) are outside our threat model by design, as the adversary's goal is identifying visited websites.

For our IPv6 fingerprints, when a domain has both A and AAAA records, we use only the IPv6 address (§5.2). IPv4 is included only as a fallback for domains that lack AAAA records entirely. This design isolates the contribution of IPv6 addressing and measures fingerprinting in the direction of a future where IPv6 deployment is complete. Under RFC 8305 (Happy Eyeballs [91]), a browser attempts both IPv4 and IPv6 connections in parallel, so an on-path adversary would in practice observe addresses from both protocols for dual-stack domains—which would only *increase* accuracy over what we report.

We model the adversary as observing a *single, isolated browsing session* at a time. While an AS-level observer would see aggregated traffic from multiple devices, prior work [85] shows that users can be distinguished behind NAT with high accuracy using only NetFlow-style metadata: distinct source ports and time-based heuristics enable per-session flow segmentation. Our setup reflects the common scenario of one website visit per session.

4 Experimental Setup

Next, we outline the experimental setup for our fingerprinting approach, including the selection and categorization of website(s), configuration of the measurement environment, and data collection procedures to ensure reliability and accuracy. Our primary goal is to ensure scalability, maintain up-to-date IP information, and reproducibility of our results.

4.1 Target Websites Dataset

To build a comprehensive and representative dataset of popular and sensitive website(s), we merged two rankings: Tranco[1] [44] and Cloudflare Radar[2] [15], generated on January 20, 2024. By combining the top one million domains from each list and removing duplicates, we obtained a unified dataset of approximately 1.3 million unique websites, capturing a wide range of high-traffic and widely visited website(s) across different categories.

From this dataset, we select only those websites whose effective top-level domain plus one label (eTLD+1), resolves to at least one IPv6 address. We filtered for domains with a valid AAAA record on public DNS servers (indicating IPv6 support) and identified approximately 520K such websites. These domains are accessible over both IPv4 and IPv6, allowing us to perform a direct comparison of fingerprinting risk between the two protocols. It is important to note that the vast majority of domains offering IPv6 support also maintain IPv4 connectivity, reflecting the operational norm of dual-stack deployment to ensure backward compatibility and broader reach [13].

We further partition our 520K dual-stack websites into two disjoint categories based on the IPv6 availability of the domains involved in loading the webpage. For every website, we collect all domains that a browser contacts when fetching external resources (e.g., ads, APIs, CDNs, images). For a given website, if *all* of its dependency domains are accessible over IPv6, we classify it as *dual-stack complete* website. Conversely, if *at least one* domain is not available over IPv6, we classify the website as *dual-stack incomplete* (more discussion in §5.2). This results in approximately 295K dual-stack complete and 228K dual-stack incomplete websites. These numbers fluctuate throughout our measurement period due to factors such as website downtime, maintenance and unavailability over IPv6 due to changes in network configurations.

Since the majority of websites on the Internet still operate in dual stack mode [4], it is essential to account for these real-world deployment conditions in our analysis. This distinction enables us to isolate trends across both dual-stack incomplete and emerging dual-stack complete deployments, while simultaneously identifying how IPv4-only third-party dependencies drive fingerprinting risk.

4.2 Measurement Environment

While prior work has shown that geographic proximity and load balancing can affect DNS resolutions [32, 34, 60], as discussed in §3 our threat model assumes that the adversary is local. To align with our threat model, we deployed our measurement setup within an academic network in a single autonomous system (AS) in the US, ensuring a consistent resolver view and routing perspective. Each crawler machine used a distinct, static public IP address throughout the entire measurement period. All crawls were conducted inside isolated Docker containers to eliminate any DNS or browser cache interference between successive website visits, ensuring consistency across crawl batches. All measurements were collected from April to August 2024.

[1]Tranco list: https://tranco-list.eu/list/YXQ4G

[2]Cloudflare Radar domain rankings: https://radar.cloudflare.com/domains?dateStart=2024-01-20&dateEnd=2024-01-21

As our measurements originate from a single geographic vantage point, provider-level conclusions reflect local CDN IP mappings. This is consistent with our local on-path adversary model: such an adversary (e.g., an ISP, IXP, or campus network admin) observes only localized IP assignments. Our methodology is directly replicable by any regional adversary to generate region-specific fingerprints, and prior work [34] has confirmed that co-hosting degrees are similar across different locations.

We used 3 machines to parallelize crawling and each machine used Google Chrome as the browser to crawl websites and record browsing sessions. We chose Chrome due to its dominant market share (68%) and its popularity among ordinary users, who are generally more vulnerable to privacy threats [66]. However, the observed trends are not exclusive to Chrome, as they stem from how network protocols and infrastructure are implemented, rather than browser-specific behavior.

During each website visit, we capture HTTP Archive (.har) files to record all domain connections initiated during the browsing session for every website in our dataset. HAR files provide structured, comprehensive logs of all web requests, responses, and associated metadata, enabling precise identification of both primary and secondary domains (see §5.1). Measurement data was collected over a three-month period in *five* **crawl batches**, each batch completing a full crawl of the dataset over approximately one month. Successive batches were initiated 15 days apart, allowing us to observe temporal variations in the set of domains a website contacts due to dynamic content, CDN migrations, advertising and analytics APIs, etc. [25, 45, 84]. These temporal snapshots enable us to assess how stable our IP-based fingerprints are across each crawl batch.

In order to keep up-to-date IP data, we resolve all observed domains into their A and AAAA records (if available), utilizing ZDNS [40], a fast DNS toolkit that is optimized for conducting large-scale active DNS measurements. This tool's speed and scalability enabled us to query both A and AAAA records for all domains within short timeframes. As new domains are discovered during crawling, they are promptly added to the DNS query batch. For an extended coverage, we utilized six popular public DNS resolvers that provide domain name encryption services: Google DNS [46], Cloudflare DNS [16], Quad9 [81], OpenDNS [58], Comodo Secure DNS [18], and CleanBrowsing DNS [14].

Using multiple resolvers mitigates bias introduced by resolver specific caching, geo-distribution, or provider driven DNS steering [1]. We maintained a rolling resolution schedule, refreshing A and AAAA records at hourly intervals for domains observed in recent crawl batches. Maintaining timely domain-IP mappings is critical, as DNS records fluctuate due to dynamic server allocations or load balancing [72]. While DNS resolutions may vary across geographic locations, maintaining a single vantage point aligns with our local adversary model. Our experimental setup mirrors the visibility available to a passive ISP-level observer within a single network vantage point.

5 Methodology

We now describe the pipeline used to construct IP-based fingerprints and evaluate their effectiveness under the threat model defined in §3. Figure 2 illustrates the end-to-end workflow.

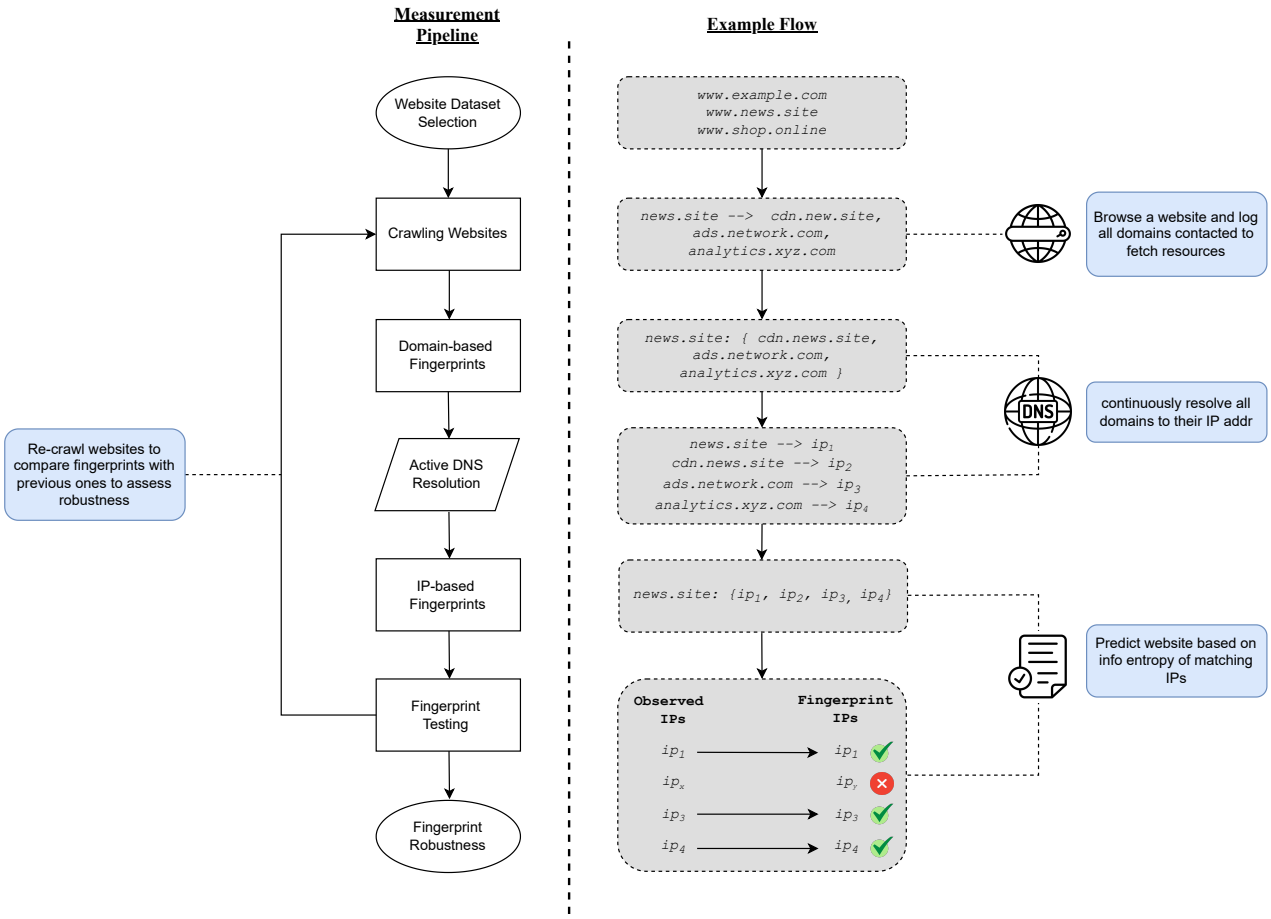

Figure 2: Pipeline for constructing IP-based fingerprints and evaluating their effectiveness for both IPv4 and IPv6.

5.1 Domain-based Fingerprints

After finalizing the target websites dataset and crawling them (§4), we extract the primary and secondary domains from the HAR files. The *primary domain* is the domain name of the website being visited, defined at the eTLD+1 level using the Public Suffix List (PSL), while *secondary domains* are any additional domains contacted to fetch resources (images, scripts, stylesheets, advertisements, etc.) during the visit to the primary domain.

Using this information, we construct a high-level, domain-based fingerprint for each website. Each fingerprint consists of two components: the primary domain and a set of associated secondary domains. Formally, the fingerprint is represented as:

$$\mathcal{D}_w = \{d_p\} \cup \{d_{s_1}, d_{s_2}, \ldots, d_{s_n}\}$$

where d_p is the primary domain and d_{s_i} are the secondary domains observed during the session. We represent fingerprints as sets to eliminate redundancy and to disregard the order in which domains are contacted during page load.

Importantly, we disregard the sequence, frequency, and order in which these domain requests occur, as these features are inherently unstable and non-deterministic. Prior work shows that fetch order can vary widely across visits due to network conditions, browser scheduling, resource prioritization, and operating system differences [90]. Modern browsers further exacerbate this variability through speculative parsing, script prioritization, and asynchronous resource loading, leading to nondeterministic patterns even for the same webpage [54]. Thus, we opt to not use request order as part of our fingerprint design.

In short, fingerprints are constructed independently per crawl batch and each fingerprint is modeled as an unordered set of domains to focus on identifiers that persistently characterize the site. Of course, such domain-based fingerprints become ineffective if domain names are encrypted via DoH/DoT or TLS 1.3 ECH, which is the case according to our threat model. These fingerprints serve as an intermediate representation. In the next step (§5.2), we resolve each domain to its corresponding A and AAAA records to derive IP-based fingerprints consistent with our adversarial visibility.

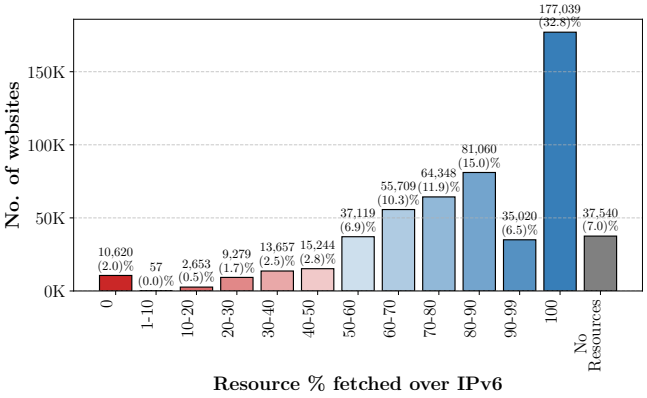

Figure 3: Percentage of resources fetched over IPv6 by each website (consistently across the measurement period) to recognize the extent of IPv6 adoption across websites in terms of third party resources.

5.2 IP-based Fingerprints

We model the adversary as observing the set of destination IP addresses actually contacted during the session. To model this scenario, we convert each domain-based fingerprint into an IP-based fingerprint using our collected DNS data (§4.2). For evaluation within each crawl batch, fingerprints are constructed using IPs observed during that batch. When analyzing stability across batches, we compare batch-specific fingerprints over time. For each domain in a domain-based fingerprint, we resolve the domain to both its A and AAAA records to construct:

5.2.1 IPv4 Fingerprints. The IPv4 fingerprint of a website is constructed by resolving the primary and secondary domains to all of their corresponding IPv4 addresses observed during the crawl window. Formally, we represent it as:

$$\mathcal{F}_{v4} = \{ip^A_{p,1}, ip^A_{p,2}, \ldots, ip^A_{p,m}\} \cup \bigcup_{i=1}^{n} \{ip^A_{s_i,1}, ip^A_{s_i,2}, \ldots ip^A_{s_i,k}\}$$

where $ip^A_{x,j}$ denotes the j-th IPv4 address obtained by resolving domain x via an A record. Here, p is the primary domain and s_i is the i-th secondary domain. The parameters m and k correspond to the number of unique IPv4 addresses observed for the primary and each secondary domain, respectively.

In essence, $\mathcal{F}_{v4}$ is the set of all distinct IPv4 addresses associated with the domains contacted during the site's loading process. If a domain resolves to multiple IPv4 addresses either simultaneously or at different times within the crawl window, all such addresses are included in $\mathcal{F}_{v4}$. For websites that are IPv6-enabled (dual-stack), their IPv4 fingerprint contains only the IPv4 addresses to compare it against their IPv6 counterpart.

5.2.2 IPv6 Fingerprints. IPv6 fingerprints are constructed similarly, but with a preference for IPv6 addresses. Since our dataset only includes websites whose primary domain has a valid IPv6 address, we attempt to resolve each secondary domain to its IPv6 addresses. If an IPv6 address is unavailable for a given domain, we fall back to that domain's IPv4 address. We represent this fingerprint as:

$$\mathcal{F}_{v6} = \{ip^{AAAA}_{p,1}, ip^{AAAA}_{p,2} \ldots\} \cup \bigcup_{i=1}^{n} \{ip^*_{s_i,1}, ip^*_{s_i,2}, \ldots, ip^*_{s_i,k}\}$$

where $ip^{AAAA}_{p,j}$ are IPv6 addresses for the primary domain, and $ip^*_{s_i,j}$ denotes an address for secondary domain s_i that is IPv6 (ip^{AAAA}) or otherwise IPv4 (ip^A).

In other words, $\mathcal{F}_{v6}$ is the set of all IPv6 addresses for each resource domain, using IPv4 only as a fallback. This approach approximates modern client behavior (e.g., Happy Eyeballs [91]), where IPv6 is preferred but IPv4 is used when IPv6 is unavailable. As a result, the fingerprint of a dual-stack complete site contains only IPv6 addresses, whereas dual-stack incomplete sites include both IPv4 and IPv6 addresses in their fingerprints as discussed in §4.1.

This behavior reflects the transitional nature of today's Internet, where many websites are IPv6-capable but still rely on IPv4 for portions of their content because a substantial fraction of resource domains lack IPv6 support. We construct these two separate fingerprints to help us draw direct comparison of effectiveness between IPv4- and IPv6-based fingerprinting as shown in Table 1.

The distribution of IPv6 usage in our dataset is illustrated in Figure 3: approximately 32% of sites resolved all dependency domains to IPv6 addresses throughout the measurement period, requiring no IPv4 fallback, whereas 58% of sites had a mix (1–99%) of IPv6 and IPv4 (dual-stack incomplete). This finding is consistent with other recent measurements showing that only a small portion of IPv6-enabled websites are fully reachable over IPv6 without any IPv4 dependencies. For example, one study of the top 100k sites found that nearly three-quarters of sites that support IPv6 still pull in some resources via IPv4 in practice [82].

We observed about 10K websites in our dataset where, despite the primary domain being accessible over IPv6, all resources were fetched via IPv4, effectively negating any IPv6 usage beyond the landing page. We label such sites as dual-stack incomplete since they require both IP versions for a full page load. Meanwhile, websites that fetch no secondary resources at all (approximately 37K), but whose primary domain is accessible over IPv6, are still classified as dual-stack complete.

5.3 Fingerprint Matching

As discussed earlier, the effectiveness of a fingerprint depends on how uniquely a given domain or IP address appears across websites in our dataset. If a domain is contacted during nearly every site visit, it provides little discrimination, making it difficult to infer which specific site is being accessed. In contrast, a domain that appears exclusively when visiting a single website becomes a strong identifier—its presence in network traffic allows a confident prediction of the corresponding website. To quantify this uniqueness, we use the concept of self-information (*information gain/entropy*) [74], measured in bits, defined as:

$$\textit{Information Gain} = -\log_2 P(d) \tag{1}$$

where $P(d)$ is the probability of a domain d being contacted when visiting a given website from our dataset. A high entropy value for a domain indicates it is contacted by relatively few sites, making it a strong identifier. Conversely, a low entropy value suggests that

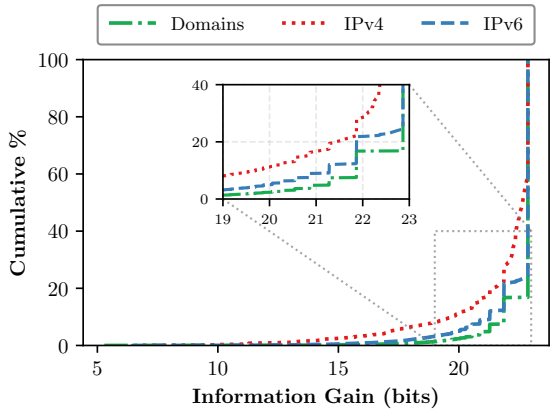

Figure 4: CDF of Information Gain provided by each domain, IPv4 and IPv6 address in the scope of our dataset.

Algorithm 1 IP-Based Fingerprint Matching

Require: Observed IP set O, fingerprint set $\{\mathcal{F}_w\}$, information entropy values $H(ip)$
Ensure: Predicted website $\hat{w}$

```
Extract primary IP ip_p from O
Candidate set C ← {w | ip_p ∈ primary(F_w)}
for all w ∈ C do
    Score[w] ← 0
    for all ip ∈ O do
        if ip ∈ F_w then
            Score[w] ← Score[w] + H(ip)
        end if
    end for
end for
ŵ ← arg max_{w∈C} Score[w]
return ŵ
```

the domain is common to many sites, reducing its usefulness in a fingerprint and making it a weak identifier.

To compute an information gain value for an IP address, we take the average of the entropy values of all domains that were observed to be hosted on that IP during the crawl window. The resulting value provides a measure of the IP's uniqueness as a fingerprinting feature in the absence of plaintext domain names in the traffic.

5.3.1 Matching Algorithm. Given an observed browsing session, the adversary obtains the set of destination IP addresses $O = \{ip_1, ip_2, \ldots, ip_k\}$ contacted during the landing-page load. Let $\{\mathcal{F}_w\}$ denote the set of precomputed fingerprints for all websites w, and let $H(ip)$ represent the entropy value associated with IP address ip. As shown in Algorithm 1, for each candidate website w, we compute a cumulative score $Score[w]$ by summing the entropy values of all IP addresses that appear in both the observed set O and the fingerprint $\mathcal{F}_w$. Intuitively, IP addresses that are rare across the dataset (i.e., have higher $H(ip)$) contribute more strongly to the score than widely shared IPs. The predicted website $\hat{w}$ is the one whose fingerprint yields the maximum score across all candidates. We report top-1 identification accuracy, defined as the fraction of sessions in which $\hat{w}$ matches the ground-truth website.

Figure 4 plots the information gain (entropy) provided by each domain, IPv4 address, and IPv6 address in our dataset, shown as a cumulative distribution. As expected, domain names offer the highest distinctiveness: approximately 84% of domains have the maximum entropy (appearing on only a single website). The curve for IPv6 addresses lies just below the domain curve, with about 78% of all IPv6 addresses associated with only one site.

In contrast, the IPv4 address curve lies significantly towards lower entropy and diverges much earlier, reflecting that IPv4 addresses are more frequently shared among multiple sites with only 43% highly distinctive IPv4 addresses. This aligns with findings of another study [26], that found that out of 250 million active domains, 40% of all domains were sharing an IP with two or more domains. More strikingly, about 20% of all live domain names were pointing to only 100 IPv4 addresses.

5.3.2 Operational Feasibility. IP-based fingerprinting requires no machine learning and is inexpensive to deploy: fingerprints are compact (≈100 KB per site, ≈50–60 GB for 520K sites), and identification takes seconds. Our stability analysis (§6.4) shows fingerprints remain accurate for months; periodic DNS re-resolution suffices to keep the database current. For ISPs already collecting NetFlow data, the additional overhead is negligible. Further details on mitigation strategies are discussed in Appendix §D.

6 Results

In this section, we present the key findings from our methodology (§5) and the metrics collected during our study, including accuracy, uniqueness, and stability of fingerprints.

6.1 Fingerprint Testing

Once the fingerprints are constructed, the next step is to simulate a testing environment that aligns with our threat model. It is crucial to accurately reflect both the capabilities and limitations of the adversary, as well as the relevance of the observable data. In our scenario, we assume a network-level adversary who can observe the destination IPs of outgoing HTTPS packets from a user.

IP Connections. Similarly to §5, we generate traces of IP connections that resemble the Happy Eyeballs algorithm [91]. Extracting from each crawl's HAR file, we determine the sequence of domain connections initiated by the browser. For each domain in these sequences, we resolve it to an IP address using our active DNS data. We choose the *first* IP address returned by a DNS resolver that was closest in time to the crawl, simulating realistic resolution behavior. For IPv4 fingerprints, every contacted domain in a website's trace is mapped to its corresponding IPv4 address. For IPv6 fingerprints, we resolve each domain to IPv6 and fall back to its A record if no IPv6 record is available. The resulting set of contacted IPs make the observation O used in Algorithm 1 for fingerprint prediction.

As a baseline, we first evaluate how accurately websites can be identified using *only* the primary domain's IP address. This represents the minimal adversarial capability. The accuracy is low

across both categories: for dual-stack complete sites, primary-IP-only achieves 5.98% (IPv4) and 7.01% (IPv6); for dual-stack incomplete sites, 4.71% (IPv4) and 7.67% (IPv6). Including the IPs of all sub-resources dramatically increases accuracy to the levels shown in Table 1 (up to 94% for dual-stack incomplete sites), demonstrating that the discriminating power comes from the combination of site-specific resource IPs rather than the primary domain's IP alone.

Table 1 presents the fingerprinting accuracy for IPv4 and IPv6 across dual-stack incomplete and complete websites, grouped by their Tranco rankings. Several trends emerge: **(1)** dual-stack incomplete websites consistently show high fingerprinting accuracy on both IPv4 and IPv6 (around 89-94%), with minimal differences between the two protocols. **(2)** dual-stack complete websites exhibit significantly lower accuracy: only 56% are correctly identified using IPv4 fingerprints and 45% using IPv6 fingerprints. **(3)** For dual-stack complete sites, the gap between IPv4 and IPv6 accuracy is sizable (about 12 % points), indicating that IPv6-only websites (with no IPv4 dependency) are substantially harder to fingerprint.

The discrepancy in accuracy between dual-stack incomplete and dual-stack complete sites is initially counter-intuitive. Earlier analyses showed that IPv6 addresses exhibit higher marginal entropy than IPv4 addresses (Figure 4) and are more frequently mapped to individual domains (Figure 1). Under a naïve interpretation, this would suggest that IPv6 fingerprints should be more discriminative.

However, fingerprint accuracy depends not only on the distinctiveness of individual IP addresses, but on the structural overlap between entire fingerprint sets. Dual-stack complete sites often rely on shared IPv6-enabled hosting infrastructures, leading to substantial overlap among their IP sets. This reduces separability at the classifier level despite high per-address entropy. We examine this structural similarity explicitly in the next subsection (§6.2).

In contrast, some dual-stack incomplete sites may have at least one IPv4-only dependency with a distinctive address, which contributes a strong discriminative signal and explains their consistently higher accuracy. Interestingly, the gap between IPv4 and IPv6 accuracy for dual-stack incomplete sites is negligible (< 1%) across all ranking tiers. This consistency suggests similar deployment patterns across the two protocols, despite IPv6's vastly larger address space. The underlying reasons become clearer when examining the hosting and deployment practices of major providers. We analyze these provider practices and their contribution to fingerprinting characteristics in §7.1.

Impact of Ad Blocking. Ad blocking shortens the IP fingerprint by removing third-party resource loads. However, Hoang et al. [33] found that accuracy drops only marginally (91% to 80%) with aggressive ad blocking, since tracker and ad domains are contacted across many sites and therefore carry low entropy. Their removal has limited impact on overall fingerprinting accuracy.

6.2 Fingerprint Similarity

To further investigate the observations in §6.1 and Table 1, we compute pairwise similarity between IP-based fingerprints of all websites across our dataset to quantify the overlap between the IP sets of one website with every other website. Jaccard similarity between website i and j is defined as:

$$\mathbf{J}_{i,j} = \frac{|\mathcal{F}_i \cap \mathcal{F}_j|}{|\mathcal{F}_i \cup \mathcal{F}_j|} \tag{2}$$

We compute similarities separately for each protocol (IPv4 & IPv6) within each dual-stack category (complete & incomplete) to ensure consistent structural comparison. Across categories, the proportion of zero-similarity pairs differs. For dual-stack incomplete sites, only 17.8% (IPv4) and 19.1% (IPv6) of pairs have zero overlap. In contrast, for dual-stack complete sites, approximately 75% of pairs share no common IP under both IPv4 and IPv6 (Appendix §C)

The non-zero pairwise similarities are visualized in Figure 5, which shows kernel density estimates (KDEs) of non-zero similarity scores for both IPv4 and IPv6 fingerprints. Zero-similarity pairs are omitted from both website categories to improve readability and highlight meaningful overlap patterns. Although zero-valued pairs are excluded, the KDE curves may still appear to start near zero because kernel smoothing spreads probability mass toward the lower boundary when many values are very close to zero. Similarly, the y-axis represents probability density rather than raw counts.

The left panel illustrates that dual-stack incomplete websites share almost no similarity with each other: both the IPv4 and IPv6 curves are sharply concentrated near zero, with virtually no area beyond 0.45. This indicates that dual-stack incomplete deployments are highly heterogeneous. Different domains typically use distinct providers, diverse allocation schemes, and unshared address blocks. As a consequence, their IP-based fingerprints are extremely sparse and dispersed, resulting in high accuracy.

The right panel reveals a contrasting pattern for dual-stack complete websites, with a clear multi-modal structure. Here, both IPv4 and IPv6 similarity distributions show substantial mass at moderate similarity interval $(0, 0.6]$ and also retain some density near the upper end of the scale, including values approaching one. Although most dual-stack complete pairs are disjoint, the non-zero similarity distribution exhibits a clear multi-modal structure. This indicates that overlap, when it occurs, is concentrated within specific clusters of websites, rather than uniformly distributed across the dataset.

The IPv6 similarity distribution peaks sharply near zero, indicating that most unrelated websites have minimal overlap, and it is highly structured rather than uniformly distributed. When similarity occurs, it is concentrated within clusters of similarly deployed websites hosted by the same provider. Such clustered overlap increases confusion among websites sharing infrastructure, reducing global separability despite low average similarity [27, 59]. These structured clusters reduce global separability and ultimately lower IPv6 classification accuracy (see §7.1).

6.3 Realized Entropy

While similarity analysis explains structural overlap between websites, entropy analysis examines the distinctiveness of individual IP addresses. At first glance, the entropy CDF in Figure 4 appears to reinforce the intuition that IPv6 addresses carry substantially more *information* than IPv4, hinting at potentially better fingerprinting performance. Yet this intuition breaks when compared against the actual fingerprinting results in Table 1 and similarity distributions in Figure 5 (§6.2).

Table 1: Fingerprinting accuracy in IPv4 & IPv6 across dual-stack complete & incomplete sites

Ranking	Dual-stack incomplete sites			Dual-stack complete sites		
	# Sites	IPv4 %	IPv6 %	# Sites	IPv4 %	IPv6 %
Top 100	39	89.74	89.74	61	29.51	24.59
Top 1000	540	90.37	90.19	460	56.96	49.35
Top 10K	6038	94.10	94.17	3962	64.84	54.72
Top 50K	29742	94.81	94.96	20258	65.33	56.64
Top 100K	56491	94.68	94.86	43509	62.24	53.29
Top 250K	119997	93.06	93.31	130003	56.37	44.94
Top 500K	219501	93.79	94.16	280499	56.09	44.80
All Websites	228108	93.78	94.20	295315	56.02	44.82

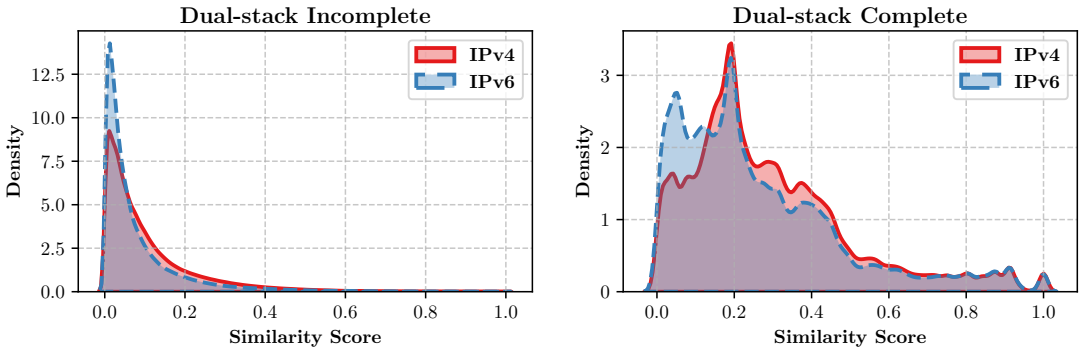

Figure 5: KDE distributions of non-zero pairwise similarity scores between website fingerprints for IPv4 & IPv6.

However, these raw entropy measurements can be misleading because they do not account for address-space scale or temporal behavior. In IPv6, the enormous address space enables providers to allocate large pools of addresses to domains, leading to continuous discovery of previously unseen IPv6 addresses across crawl batches. High entropy values therefore arise not only from structural distinctiveness across websites, but also from the growth of the observed IPv6 address pool over time. As a result, Figure 4 reflects entropy as cumulative distribution over 1.04M distinct IPv6 versus 552K IPv4 addresses, reflecting the substantially larger observed IPv6 address pool. Many of these short-lived IPv6 addresses are never used for fingerprint testing.

Figure 6 illustrates this perfectly: the number of newly observed IPv6 addresses continues to grow throughout the measurement period, while IPv4 quickly saturates and flattens out. This persistent influx of fresh IPv6 addresses causes fingerprints to accumulate large sets of ephemeral addresses, inflating naive entropy and conflating genuine diversity with simple growth in the number of observations of new IPv6 addresses.

To disentangle true structural distinctiveness from address-space expansion, we introduce a *realized entropy* metric. Rather than computing entropy over the union of all IP addresses observed across the entire crawl window—which inflates IPv6 uniqueness because the IPv6 address pool grows continuously throughout the study—realized entropy is computed using a snapshot of domain-to-IP mappings at a single point in time (one crawl batch). Each IP address is scored based only on how many websites share it

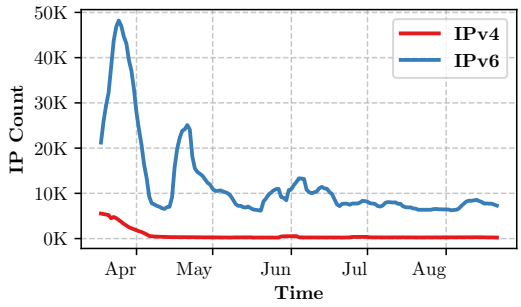

Figure 6: New IPs found across measurement period.

within that snapshot, not across the full longitudinal dataset. This snapshot-based approach isolates the structural distinctiveness that actually aids fingerprint classification, excluding short-lived IPv6 addresses that are never seen during testing.

Figure 7 shows the cumulative distribution of realized entropy across websites for IPv4 and IPv6. Once address-scale inflation is removed, we find that the entropy gap between IPv4 and IPv6 becomes far smaller. In fact, more than 20% of the IPv6 addresses provide less information gain than their IPv4 counterparts, challenging the assumption that higher raw entropy directly translates to stronger fingerprinting capability. This reduced effective entropy helps explain the relatively lower IPv6 fingerprinting accuracy of dual-stack complete websites compared to their IPv4 counterparts.

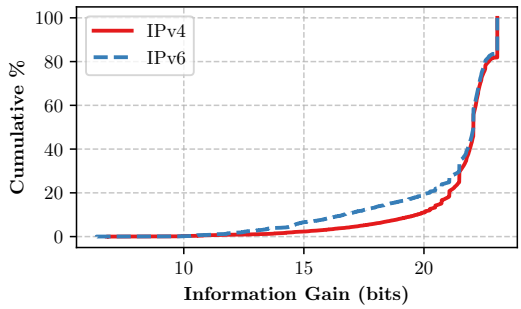

Figure 7: Realized entropy of IP connections of each website used for testing respective fingerprints.

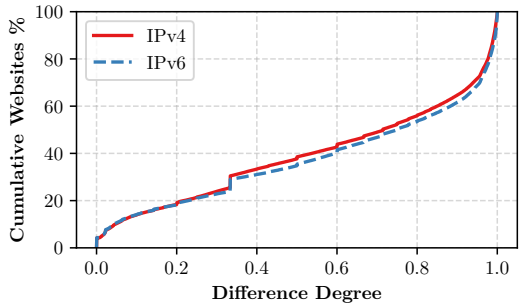

Figure 8: Average difference degree of IP-based fingerprints for all websites within subsequent crawl batches.

6.4 Fingerprint Stability

The stability of our IP-based fingerprints largely depends on two factors. (1) the temporal evolution of domain- and IP-based fingerprints, and (2) the rate at which associated IP addresses change over time. To evaluate this, we analyze the temporal stability of website fingerprints across different crawling batches and examine the IP rotation/churn behavior of both IPv4 and IPv6 addresses over the course of our study.

6.4.1 Difference Degree. Since IP-based fingerprints are derived from domain-based fingerprints, it is essential to understand how both evolve over time. As described in §4.2, we repeatedly crawl each website and compute the difference degree between fingerprints across crawl batches. For a website i, the difference degree between fingerprints at time t_0 and t_1 yields a value between 0 (no change) and 1 (complete change), defined as:

$$\textit{Difference degree} = \frac{|(\mathcal{D}_{t_0} \cup \mathcal{D}_{t_1}) - (\mathcal{D}_{t_0} \cap \mathcal{D}_{t_1})|}{|(\mathcal{D}_{t_0} \cup \mathcal{D}_{t_1})|}$$

Prior work has shown that domain-based fingerprints are relatively stable. Hoang et al. [33] report that over 90% of domains exhibit a difference degree below 0.5 over two months. Our results show a similar trend, with dual-stack incomplete sites showing slightly higher stability than dual-stack complete sites (§A). We confirm domain-level stability, and then examine whether this stability persists once domain fingerprints are translated into IP space.

We now examine IP-based fingerprints. Figure 8 shows the CDF of the average difference degree for IPv4 and IPv6 across successive crawl batches. The two distributions are strikingly similar: roughly 40% of websites have a difference degree below 0.5 after 15 days for both IPv4 and IPv6. This indicates that IPv6 fingerprints are not substantially more volatile than their IPv4 counterparts; its operational churn largely mirrors IPv4, reflecting similar hosting and deployment practices (§7.1). In practice, IPv6 churn more often manifests as the appearance of previously unseen addresses, consistent with the larger observed IPv6 address pool.

6.4.2 IP Churn. While domain-based fingerprints are relatively stable, the associated IP addresses may change over time, directly affecting the robustness of IP fingerprints. To quantify this, we analyze the churn behavior of both IPv4 and IPv6 by tracking timestamps for each domain at which its IP set changes (i.e., IP is added, removed, or replaced) across the measurement period

Let a domain be observed over a set of timestamps T and let $C = \{t_{c_1}, t_{c_2}, \ldots, t_{c_k}\} \subseteq T$ denote the ordered set of churn events. The average *IP churn interval* for a domain is:

$$\textit{IP-Churn Interval} = \frac{1}{k-1} \sum_{i=1}^{k-1} \left(t_{c_{i+1}} - t_{c_i}\right)$$

Figure 9 shows the average IP churn rate (right y-axis) and the number of IP changes observed (left y-axis) for both IPv4 and IPv6 for all domains in our dataset ($\approx 850K$). Red and blue dotted lines show the number of IP changes for IPv4 and IPv6 respectively. The solid red and blue lines show the average time interval after which the IP address of a domain changes for IPv4 and IPv6 respectively.

Roughly 60% of domains show no IP changes at all and remain fully stable in both IPv4 and IPv6. Among the remaining 40%, IPv6 addresses churn slightly more often than IPv4. Approximately 38% of these domains exhibit churn intervals under 10 hours; of these, 28% rotate IPv4 addresses within an hour on average, compared to 35% for IPv6. At the other extreme, over 50% of domains with any churn have intervals exceeding 100 hours (≈4 days), and around 10% exceed 1000 hours (≈42 days). These findings indicate that IPv6 does not introduce fundamentally different operational volatility compared to IPv4. Importantly, while the aggregate IPv6 address pool grows more rapidly over time (§6.3), per-domain churn behavior remains broadly comparable to IPv4.

7 Discussion

In this section, we discuss our findings from §6 and examine their implications for the future of IP-based fingerprinting in an increasingly IPv6-enabled Internet. In particular, we focus on how hosting infrastructure and provider deployment choices shape fingerprintability in dual-stack incomplete and complete environments.

7.1 Hosting Infrastructure

Our analysis indicates that fingerprintability is primarily shaped by provider-level address allocation strategies rather than by protocol alone. Address pool granularity, domain co-location patterns, and CDN infrastructure collectively determine how distinctive a website's IP footprint appears under IPv4 and IPv6.

To better understand these trends, we examine provider-level domain co-location. We define the *co-location degree* of a provider

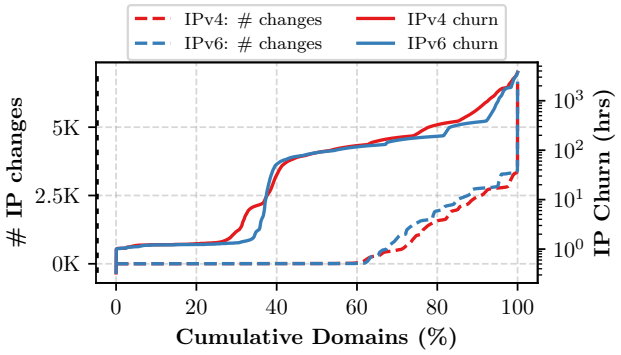

Figure 9: Average IP churn rate (solid, right y-axis) & the number of IP changes (dotted, left y-axis) observed for each domain in our dataset.

as the average number of domains hosted on each IP address associated with that provider. Higher co-location reduces per-address uniqueness and can weaken fingerprinting signals.

Figure 11 (Appendix §B) shows the cumulative distribution of IPv4 and IPv6 co-location degrees across all providers. Contrary to expectation, the two distributions are strikingly similar: approximately 45% of providers host only one domain per IPv4 address, and around 40% do so for IPv6. By a co-location degree of two, both curves reach roughly 80%, and by degree ten, approximately 95% of providers fall below that threshold for both protocols. These results indicate that the majority of providers exhibit low co-location under both IPv4 and IPv6, with only a small fraction accounting for heavily shared addresses.

Although co-location degrees are broadly comparable across protocols, co-location alone does not fully determine fingerprintability. Many providers allocate domains within structured IPv6 address pools, causing clusters of websites to share overlapping address subsets even when per-IP co-location remains modest. This pool-level clustering arises from performance incentives: large CDNs concentrate traffic into dense anycast clusters [10] for routing efficiency and DDoS resilience, which incidentally pools IPv6 addresses across unrelated domains [6]. Consequently, separability is governed not only by how many domains share an IP, but by how address pools are reused across sites—a structural pressure that mirrors IPv4 co-location dynamics despite IPv6's abundant address space [32, 34].

7.2 Provider Specific Trends

We next examine provider-level trends to understand how address allocation strategies translate into measurable fingerprinting performance. Using MaxMind GeoLite2 [49], we attribute each website's IP addresses to its hosting provider (AS) and compare fingerprinting accuracy across providers.

Table 3 reveals substantial variation across providers. For example, Cloudflarenet dual-stack incomplete sites achieve 97% fingerprinting accuracy, whereas Cloudflarenet dual-stack complete sites drop to 43.6%. Even more pronounced, Google-GCP and Google-hosted dual-stack complete sites exhibit extremely low accuracy

Table 2: Top IPv6 hosting providers and their IPv6 co-location degree (average number of domains per IPv6 address).

ASN	Provider Name	Colocation
AS15169	GOOGLE	65.74
AS54113	FASTLY	59.41
AS13335	CLOUDFLARENET	12.37
AS9123	JSC Timeweb	9.76
AS8560	IONOS SE	7.40
AS396982	GOOGLE-GCP	6.88
AS209242	Cloudflare London LLC	6.14
AS16276	OVH SAS	5.80
AS20940	Akamai Int'l B.V.	5.43
AS14061	DIGITALOCEAN	4.83
AS24940	Hetzner Online GmbH	2.30
AS47583	Hostinger Int'l Ltd	1.16
AS16509	AMAZON	1.13

(9.8% and 16.3%, respectively), indicating heavy IPv6 address sharing within these infrastructures.

In the dual-stack complete category, providers with higher IPv6 co-location degrees (Table 2) generally exhibit lower fingerprinting accuracy. When many domains share common IPv6 pools, per-address distinctiveness diminishes and inter-site overlap increases, reducing separability. Conversely, providers with co-location degrees close to one often show higher fingerprinting accuracy, as their IPs are more uniquely associated with specific domains.

However, in the dual-stack incomplete category, the relationship between co-location and fingerprinting accuracy is less consistent. While the trend holds for some providers, there are notable cases where, even with high IPv6 co-location, a large number of websites remain accurately fingerprintable. This suggests that factors beyond IPv6 contribute to fingerprinting in dual-stack incomplete environments —for example, highly distinctive IPv4 addresses. Even when the IPv6 portion of a fingerprint is heavily co-located, a single high-entropy IPv4 address can still suffice to identify a website.

These observations indicate that fingerprintability depends strongly on provider-level allocation policies. Large CDNs that allocate domains within shared IPv6 address pools induce clustered overlap and lower separability [86] (e.g., Cloudflare, Fastly, Google), while providers that assign near-unique addresses per domain yield higher distinctiveness (e.g., Amazon). As IPv6 deployment practices remain heterogeneous across providers, fingerprintability continues to vary widely across the hosting ecosystem [80].

Structural Differences. Dual-stack incomplete sites are disproportionately fingerprintable because they embed third-party adtech, analytics, and widget resources that currently lack IPv6 support. These IPv4-only dependencies are sourced from diverse providers with unique IP footprints, producing a highly distinctive per-site fingerprint. Dual-stack complete sites, by contrast, serve all resources through a small set of large CDNs with shared IPv6 address pools, causing substantial fingerprint overlap. If adtech providers adopt IPv6 via shared anycast infrastructure, fingerprintability for incomplete sites will likely decrease; adoption in the style of Amazon would maintain or increase it.

Table 3: IPv6 fingerprinting accuracy of the top 10 hosting providers for dual-stack complete and incomplete sites.

Rank	Dual-stack incomplete sites			Dual-stack complete sites		
	Provider	# Sites	WF %	Provider	# Sites	WF %
1	CLOUDFLARENET	146115	97.0	CLOUDFLARENET	212195	43.6
2	Hostinger Int'l Ltd	4255	97.6	Hostinger Int'l Ltd	4775	79.5
3	AMAZON	4006	84.2	Hetzner Online GmbH	3892	48.5
4	FASTLY	3344	71.4	Cloudflare London, LLC	3626	25.6
5	DIGITALOCEAN	3195	14.6	OVH SAS	2865	49.6
6	OVH SAS	1985	86.1	AMAZON	2508	62.8
7	Hetzner Online GmbH	1896	93.4	IONOS SE	2321	50.7
8	Akamai Int'l B.V.	1848	78.9	FASTLY	1706	22.5
9	IONOS SE	1717	93.8	GOOGLE-GCP	1629	9.8
10	JSC Timeweb	1607	88.1	GOOGLE	1601	16.3

8 Related Work

8.1 Website Fingerprinting Against Domain Name Encryption

Website fingerprinting has been extensively studied in the context of encrypted traffic, but much of the early work focused on anonymity networks like Tor, where destination IPs are hidden and attackers rely on packet timing and size patterns [12, 79, 88, 89]. In contrast, fingerprinting attacks against web traffic on the clearnet (i.e., without the use of network relays like Tor) have received less attention until recently.

Historically, a network eavesdropper could simply harvest domain name information from plaintext DNS resolutions and the TLS SNI extension to identify which website is being visited, making sophisticated fingerprinting unnecessary. However, with the advent of domain name encryption protocols (e.g., DoH, DoT, and ECH) that conceal plaintext domain names, the focus has shifted to identifying websites based solely on observable IP addresses and traffic patterns. This has prompted researchers to ask whether and how a network observer can still infer visited websites when only IP addresses and network metadata (e.g., timing and packet size) remain visible [8, 21, 23, 37, 50, 71, 75, 77, 83].

Hoang et al. [33] presented an IP-based website fingerprinting attack assuming an idealized future of universal DNS/SNI encryption. They showed that the set of destination IPs contacted during a page load often serves as a unique fingerprint for that page. In their experiments (primarily IPv4), 84% of 200k websites could be identified by IP address sequences alone. This success rate climbed above 90% for popular or sensitive sites, highlighting that even widely shared CDN infrastructure was not enough to prevent site identification. Notably, the constructed IP fingerprints remained about 70% effective after two months, despite the dynamic nature of the web and Internet infrastructure.

Our work is inspired by this insight and directly extends this prior work along four key dimensions. (1) We provide an explicit IPv4/IPv6 comparison on the same dataset, enabling a direct assessment of how the two protocols differ in fingerprintability. (2) We introduce the dual-stack complete versus dual-stack incomplete distinction, which reveals that fingerprinting risk is not uniform across IPv6-enabled websites but depends critically on whether all sub-resources are reachable over IPv6. (3) Our dataset is 2.4× larger (520K vs. 220K websites), improving generalizability to the broader Web. (4) We add a provider-level co-location analysis that identifies which hosting infrastructure choices drive or suppress fingerprintability—a dimension entirely missed by prior work.

More recently, Mazzuz et al. [50] further advance IPv4-based fingerprinting by employing a Siamese neural network to learn similarity metrics between sets of destination IP addresses. Their approach enables fingerprinting even with limited training data (as few as five samples). However, their methodology incorporates flow-level features and relies on learned embeddings, whereas our work is a purely symbolic and set-based analysis of IP addresses. Our design choice of using a more lightweight approach that does not rely on computationally expensive neural networks allows us to scale our analysis to a much larger dataset of more than half a million IPv6-enabled websites. This is crucial for generalizing our findings to the broader Web and understanding the broader fingerprinting landscape, especially when it comes to the practicality of deploying such techniques in real-world scenarios such as nation-state adversaries.

8.2 Co-location and Fingerprinting Risk

In another earlier study by Hoang et al. [32], the theoretical privacy gains from encrypting domain names was quantified by assessing the *co-location of domains* on shared IP addresses. They found that for about 20% of domains in their sample, encryption would yield no privacy benefit at all since each of those domains maps to a unique IP. An eavesdropper thus could still identify them just by relying on the IP address. On the other hand, less popular domains often benefit from large anonymity sets, ranging from dozens to hundreds of domains co-located on the same IP address, yielding significantly improved privacy. Their work, which used metrics like k-anonymity, revealed a crucial limitation: IP address information remains a privacy bottleneck. Our study investigates whether this limitation persists in the context of IPv6, where the address space is vastly larger and the dynamics of domain co-location may differ. Our work is different in that we move beyond static anonymity analysis to actively fingerprint websites based on their IP addresses, and we specifically evaluate entropy and classifier performance for IPv6 address distributions. More importantly, the previous study

was conducted in 2019, when domain name encryption was still in its early stages. In contrast, our study reflects the current state of the Web, where protocols like DoH and ECH are starting to see wider adoption. This shift allows us to assess the practical implications of domain co-location in today's Web environment, where many websites are now actively using encrypted DNS protocols.

We also incorporate findings from another study by Hoang et al. [34] and evaluate their web co-location evaluations in context of both IPv4 and IPv6. Interestingly, their measurement in 2019 showed that IPv6 has only slightly lower co-location than IPv4, e.g., 61.7% of IPv4 addresses host a single domain vs 70.5% for IPv6. This, together with our findings, indicate that simply having more addresses (IPv6) has not necessarily led to less co-location. These findings justify treating IPv6 with equal concern: if a majority of IPv6 endpoints are unique per site, the fingerprinting risk remains high. We explicitly compare our IPv6 fingerprinting results to such prior IPv4 measurements to highlight any changes (or lack thereof). In our case, we found that dual-stack complete websites are less fingerprintable than mixed-IP (dual-stack incomplete) websites, mainly because of how hosting providers deploy services in these two environments. Indeed, the comparative results across our test website categories follow a similar trend. The co-location patterns we observe for IPv6 are consistent with their findings, suggesting that underlying hosting practices have not fundamentally changed despite increasing IPv6 adoption over time.

8.3 Traffic Patterns and Other Side Channels

Another dimension of web fingerprinting research examines how encrypted traffic patterns themselves can leak identifying information. Patil and Borisov [63] first demonstrated that page load sequences of IP addresses (page load fingerprints), are overwhelmingly unique per site (over 95% uniqueness in the Alexa top-1M sites). They concluded that effective web privacy protection will require changes beyond protocol encryption, notably changes in how websites are hosted (e.g., increasing the number of domains sharing each IP address to dilute fingerprints). This finding aligns with our motivation: even in an encrypted-by-default Web, network-level observables have enough entropy to identify pages. We compare our IPv6 results to their IPv4-era conclusions and find that many of the same issues persist. In fact, we observe substantial multihosting among dual-stack complete websites compared to dual-stack incomplete websites, indicating that simply moving to IPv6 has not eliminated the fingerprinting challenges.

Similarly, researchers have explored side channels beyond IP addresses. For example, Shao et al. [75] showed that the pattern of encrypted DNS queries in DNS-over-HTTPS (DoH) traffic can fingerprint websites with high accuracy. While their threat model differs from ours (their observer sees DNS query traffic but not the web connection IPs), both studies underscore a common theme: encrypting domain names and messages is not sufficient if some other part of the communication remains distinctive. In our case, the distinguishing feature is the set of server IP addresses (especially IPv6), whereas in theirs it is the sequence of DNS lookups. These prior investigations [51], together with our insights, reinforce that comprehensive web privacy requires considering all layers: domain encryption, resistance to traffic analysis, and rethinking how resources are fetched or cached.

To that end, our work explicitly focuses on the IPv6 aspect of website fingerprinting at scale. To the best of our knowledge, no prior work has conducted such a large-scale evaluation of IPv6-enabled fingerprinting. Previous fingerprinting studies largely evaluated IPv4 environments or did not isolate IPv6 behavior. By filling this gap, we reveal how the ongoing transition to IPv6 might affect privacy, and we show that the dual-stack mode of the Internet can be even more vulnerable to original IP-based fingerprinting attacks. For instance, we examine whether IPv6-specific practices, such as address rotation via privacy extensions or provider-specific address allocations influence information entropy in ways that enhance or mitigate fingerprinting, an aspect that prior work did not consider.

9 Conclusion

We present the largest empirical study of IP-based website fingerprinting in IPv6, analyzing over 520K dual-stack websites through active DNS measurements, repeated browser crawls, and entropy-based evaluation. Our work addresses an open question in the community: does IPv6's vastly larger address space translate into greater fingerprintability and therefore a new privacy threat?

Our findings reveal a nuanced picture that challenges prevailing assumptions. Contrary to widespread concern, IPv6 does not inherently amplify fingerprinting risk. Dual-stack complete websites are substantially harder to fingerprint than their dual-stack incomplete counterparts, achieving only 44.8% accuracy under IPv6 compared to 94.2% for incomplete sites. This gap arises not from protocol-level properties, but from structural differences in how hosting providers deploy IPv6. Providers such as Cloudflare, Google, and Fastly co-locate large clusters of domains behind shared IPv6 address pools. Conversely, Amazon and other providers assign near-unique addresses per domain, yielding high fingerprinting accuracy even in IPv6 environments.

We further demonstrate that raw entropy metrics can be misleading in the IPv6 context. The continuous growth of the observed IPv6 address pool inflates naive entropy measurements, and our introduced realized entropy metric reveals that over 20% of IPv6 addresses contribute less discriminative information than their IPv4 counterparts once address-pool expansion is controlled for. Additionally, our longitudinal analysis over three months shows that IPv6 fingerprint stability largely mirrors IPv4: roughly 60% of domains exhibit no IP churn at all, and per-domain rotation intervals are broadly comparable across protocols.

Taken together, these results demonstrate that IPv6's privacy impact is deployment-contingent rather than protocol-determined. The transition presents an opportunity—not a guarantee—for improved privacy. Shared address pools reduce fingerprintability, while sparse, unique addresses per domain increase it. As domain name encryption becomes prevalent, IP-level observables remain a critical privacy bottleneck, and provider address allocation strategies will decisively shape Web fingerprintability. We hope our findings will foster further research into IPv6 hosting trends and their privacy implications, informing privacy-preserving designs for future IPv6 adoption.

Acknowledgments

We thank the Internet Society for partially funding this study through their Pulse IPv6 Fellowship, and the Natural Sciences and Engineering Research Council of Canada (NSERC) for support under awards RGPIN-2025-04929 and DGECR-2025-00173. This research was enabled in part by computational resources and services provided by Advanced Research Computing at the University of British Columbia (UBC ARC). We also thank the PoPETs reviewers for their constructive feedback, as well as all other reviewers who provided valuable feedback at various stages of this project. We are grateful to our colleagues at Hexlab for their informal guidance and contributions to this work. We also acknowledge the use of generative AI tools (ChatGPT, Claude, Gemini) in a limited and supervised capacity for writing and editing assistance, LaTeX formatting, and plotting scripts. The authors used these tools to revise the text, improve flow and correct any typos, grammatical errors, and awkward phrasing.

References

[1] Bernhard Ager, Wolfgang Mühlbauer, Georgios Smaragdakis, and Steve Uhlig. 2010. Comparing DNS resolvers in the wild. In *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*. 15–21.

[2] Paul Aitken, Benoît Claise, and Brian Trammell. 2013. Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of Flow Information. RFC 7011. https://doi.org/10.17487/RFC7011

[3] Michael Bailey, David Dittrich, Erin Kenneally, and Doug Maughan. 2012. The menlo report. *IEEE Security & Privacy* 10, 2 (2012), 71–75.

[4] Vaibhav Bajpai and Jürgen Schönwälder. 2016. Measuring the effects of happy eyeballs. In *Proceedings of the 2016 Applied Networking Research Workshop*. 38–44.

[5] Steven M Bellovin. 2002. A technique for counting NATted hosts. In *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment*. 267–272.

[6] Ignacio N Bermudez, Marco Mellia, Maurizio M Munafo, Ram Keralapura, and Antonio Nucci. 2012. Dns to the rescue: Discerning content and services in a tangled web. In *Proceedings of the 2012 Internet Measurement Conference*. 413–426.

[7] Sanjit Bhat, David Lu, Albert Kwon, and Srinivas Devadas. 2018. Var-CNN: A data-efficient website fingerprinting attack based on deep learning. *arXiv preprint arXiv:1802.10215* (2018).

[8] J. Bushart and C. Rossow. 2020. Padding Ain't Enough: Assessing the Privacy Guarantees of Encrypted DNS. In *FOCI '20*.

[9] Xiang Cai, Xin Cheng Zhang, Brijesh Joshi, and Rob Johnson. 2012. Touching from a distance: Website fingerprinting attacks and defenses. In *Proceedings of the 2012 ACM conference on Computer and communications security*. 605–616.

[10] Matt Calder, Ashley Flavel, Ethan Katz-Bassett, Ratul Mahajan, and Jitendra Padhye. 2015. Analyzing the Performance of an Anycast CDN. In *Proceedings of the 2015 Internet Measurement Conference*. 531–537.

[11] Zimo Chai, Amirhossein Ghafari, and Amir Houmansadr. 2019. On the Importance of Encrypted-SNI (ESNI) to Censorship Circumvention. In *FOCI @ USENIX Security Symposium*.

[12] Giovanni Cherubin, Rob Jansen, and Carmela Troncoso. 2022. Online website fingerprinting: Evaluating website fingerprinting attacks on tor in the real world. In *31st USENIX Security Symposium (USENIX Security 22)*. 753–770.

[13] Kenjiro Cho, Matthew Luckie, and Bradley Huffaker. 2004. Identifying IPv6 network problems in the dual-stack world. In *Proceedings of the ACM SIGCOMM workshop on Network troubleshooting: research, theory and operations practice meet malfunctioning reality*. 283–288.

[14] CleanBrowsing, Inc. 2024. Free DNS Filtering. https://cleanbrowsing.org/filters/.

[15] Cloudflare, Inc. 2024. Cloudflare Radar: Domain Rankings. https://radar.cloudflare.com/domains.

[16] Cloudflare, Inc. 2024. What is 1.1.1.1? https://www.cloudflare.com/learning/dns/what-is-1.1.1.1/.

[17] L Colitti, V Cerf, S Cheshire, and D Schinazi. 2016. RFC 7934: Host Address Availability Recommendations.

[18] Comodo SecureDNS 2.0. 2024. An open DNS recursive service for free security and high privacy. https://securedns.dnsbycomodo.com/.

[19] Alissa Cooper, Fernando Gont, and Dave Thaler. 2016. Security and Privacy Considerations for IPv6 Address Generation Mechanisms. RFC 7721. https://doi.org/10.17487/RFC7721

[20] Scott E Coull, Michael P Collins, Charles V Wright, Fabian Monrose, Michael K Reiter, et al. 2007. On Web Browsing Privacy in Anonymized NetFlows.. In *USENIX Security Symposium*. 339–352.

[21] Levente Csikor, Himanshu Singh, Min Suk Kang, and Dinil Mon Divakaran. 2021. Privacy of DNS-over-HTTPS: Requiem for a Dream? *2021 IEEE European Symposium on Security and Privacy (EuroS&P)* (2021).

[22] Weiqi Cui, Tao Chen, Christian Fields, Julianna Chen, Anthony Sierra, and Eric Chan-Tin. 2019. Revisiting assumptions for website fingerprinting attacks. In *Proceedings of the 2019 ACM asia conference on computer and communications security*. 328–339.

[23] Thilini Dahanayaka, Zhiyi Wang, Guillaume Jourjon, and Suranga Seneviratne. 2022. Inline Traffic Analysis Attacks on DNS over HTTPS. *2022 IEEE 47th Conference on Local Computer Networks (LCN)* (2022).

[24] Matthew Dunlop, Stephen Groat, William Urbanski, Randy Marchany, and Joseph Tront. 2011. Mt6d: A moving target ipv6 defense. In *MILCOM Military Communications Conference*. IEEE, 1321–1326.

[25] Steven Englehardt and Arvind Narayanan. 2016. Online tracking: A 1-million-site measurement and analysis. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*. 1388–1401.

[26] George G. 2020. The most crowded ipv4 addresses on the internet. https://hostingchecker.com/research/the-most-crowded-ipv4-addresses-on-the-internet/

[27] Oliver Gasser, Quirin Scheitle, Pawel Foremski, Qasim Lone, Maciej Korczyński, Stephen D Strowes, Luuk Hendriks, and Georg Carle. 2018. Clusters in the expanse: Understanding and unbiasing IPv6 hitlists. In *Proceedings of the Internet Measurement Conference 2018*. 364–378.

[28] Fernando Gont and Tim Chown. 2016. Network Reconnaissance in IPv6 Networks. RFC 7707. https://doi.org/10.17487/RFC7707

[29] Fernando Gont, Suresh Krishnan, Dr. Thomas Narten, and Richard P. Draves. 2021. Temporary Address Extensions for Stateless Address Autoconfiguration in IPv6. RFC 8981. https://doi.org/10.17487/RFC8981

[30] Seungyeop Han, Vincent Liu, Qifan Pu, Simon Peter, Thomas Anderson, Arvind Krishnamurthy, and David Wetherall. 2013. Expressive privacy control with pseudonyms. *ACM SIGCOMM Computer Communication Review* 43, 4 (2013), 291–302.

[31] Dominik Herrmann, Christine Arndt, and Hannes Federrath. 2012. Ipv6 prefix alteration: An opportunity to improve online privacy. *arXiv preprint arXiv:1211.4704* (2012).

[32] Nguyen Phong Hoang, Arian Akhavan Niaki, Nikita Borisov, Phillipa Gill, and Michalis Polychronakis. 2020. Assessing the privacy benefits of domain name encryption. In *Proceedings of the 15th ACM Asia Conference on Computer and Communications Security*. 290–304.

[33] Nguyen Phong Hoang, Arian Akhavan Niaki, Phillipa Gill, and Michalis Polychronakis. 2021. Domain name encryption is not enough: privacy leakage via IP-based website fingerprinting. *Proceedings on Privacy Enhancing Technologies* 4 (2021), 420–440.

[34] Nguyen Phong Hoang, Arian Akhavan Niaki, Michalis Polychronakis, and Phillipa Gill. 2020. The web is still small after more than a decade. *ACM SIGCOMM Computer Communication Review* 50, 2 (2020), 24–31.

[35] Paul E. Hoffman and Patrick McManus. 2018. DNS Queries over HTTPS (DoH). RFC 8484. https://doi.org/10.17487/RFC8484

[36] Rick Hofstede, Pavel Čeleda, Brian Trammell, Idilio Drago, Ramin Sadre, Anna Sperotto, and Aiko Pras. 2014. Flow monitoring explained: From packet capture to data analysis with netflow and ipfix. *IEEE Communications Surveys & Tutorials* 16, 4 (2014), 2037–2064.

[37] Rebekah Houser, Zhou Li, Chase Cotton, and Haining Wang. 2019. An Investigation on Information Leakage of DNS over TLS. In *ACM CoNEXT*.

[38] Zi Hu, Liang Zhu, John Heidemann, Allison Mankin, Duane Wessels, and Paul E. Hoffman. 2016. Specification for DNS over Transport Layer Security (TLS). RFC 7858. https://doi.org/10.17487/RFC7858

[39] Sunghwan Ihm and Vivek S Pai. 2011. Towards understanding modern web traffic. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*. 295–312.

[40] Liz Izhikevich, Gautam Akiwate, Briana Berger, Spencer Drakontaidis, Anna Ascheman, Paul Pearce, David Adrian, and Zakir Durumeric. 2022. ZDNS: a fast DNS toolkit for internet measurement. In *Proceedings of the 22nd ACM Internet Measurement Conference*. 33–43.

[41] Antonio J Jara, Latif Ladid, and Antonio Fernandez Gómez-Skarmeta. 2013. The Internet of Everything through IPv6: An Analysis of Challenges, Solutions and Opportunities. *J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl.* 4, 3 (2013), 97–118.

[42] Marc Juarez, Sadia Afroz, Gunes Acar, Claudia Diaz, and Rachel Greenstadt. 2014. A critical evaluation of website fingerprinting attacks. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*. 263–274.

[43] Srinivas Krishnan and Fabian Monrose. 2010. DNS prefetching and its privacy implications: when good things go bad. In *Proceedings of the 3rd USENIX conference on Large-scale exploits and emergent threats: botnets, spyware, worms, and more*. 10–10.

[44] Victor Le Pochat, Tom Van Goethem, Samaneh Tajalizadehkhoob, Maciej Korczyński, and Wouter Joosen. 2019. Tranco: A Research-Oriented Top Sites Ranking Hardened Against Manipulation. In *Proceedings of the 26th Annual Network and Distributed System Security Symposium (NDSS 2019)*. https:

//doi.org/10.14722/ndss.2019.23386
[45] Ada Lerner, Anna Kornfeld Simpson, Tadayoshi Kohno, and Franziska Roesner. 2016. Internet jones and the raiders of the lost trackers: An archaeological study of web tracking from 1996 to 2016. In *25th USENIX Security Symposium (USENIX Security 16)*.
[46] Google LLC. 2024. Google Public DNS: 8.8.8.8. https://developers.google.com/speed/public-dns.
[47] Liming Lu, Ee-Chien Chang, and Mun Choon Chan. 2010. Website fingerprinting and identification using ordered feature sequences. In *European Symposium on Research in Computer Security*. Springer, 199–214.
[48] Yun Ma, Xuanzhe Liu, Shuhui Zhang, Ruirui Xiang, Yunxin Liu, and Tao Xie. 2015. Measurement and analysis of mobile web cache performance. In *Proceedings of the 24th International Conference on World Wide Web*. 691–701.
[49] MaxMind. 2025. GeoLite Databases and Web Services. https://dev.maxmind.com/geoip/geolite2-free-geolocation-data/ Accessed: July 31, 2025.
[50] Neriya Mazzuz and Asaf Shabtai. 2025. Domain Name Encryption Does Not Ensure Privacy: Website Fingerprinting Attack With Only a Few Samples Using Siamese Network. In *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*. Springer, 26–45.
[51] Muhammad Muzammil, Oleksii Starov, Zane Ma, and Nick Nikiforakis. 2026. New Trains, Same Rails: Fingerprinting Cryptocurrency Software via Network Requests. In *IFIP Networking Conference*.
[52] Dr. Thomas Narten, Richard P. Draves, and Suresh Krishnan. 2007. Privacy Extensions for Stateless Address Autoconfiguration in IPv6. RFC 4941. https://doi.org/10.17487/RFC4941
[53] Dr. Thomas Narten, Tatsuya Jinmei, and Dr. Susan Thomson. 2007. IPv6 Stateless Address Autoconfiguration. RFC 4862. https://doi.org/10.17487/RFC4862
[54] Javad Nejati and Aruna Balasubramanian. 2016. An in-depth study of mobile browser performance. In *Proceedings of the 25th International Conference on World Wide Web*. 1305–1315.
[55] Nick Nikiforakis, Alexandros Kapravelos, Wouter Joosen, Christopher Kruegel, Frank Piessens, and Giovanni Vigna. 2013. Cookieless monster: Exploring the ecosystem of web-based device fingerprinting. In *2013 IEEE Symposium on Security and Privacy*. IEEE, 541–555.
[56] Mehdi Nikkhah and Roch Guérin. 2015. Migrating the internet to IPv6: An exploration of the when and why. *IEEE/ACM Transactions on Networking* 24, 4 (2015), 2291–2304.
[57] Paul Ohm. 2009. The rise and fall of invasive ISP surveillance. *U. Ill. L. Rev.* (2009), 1417.
[58] OpenDNS, Inc. 2024. Use OpenDNS. https://use.opendns.com/.
[59] Fariba Osali, Khwaja Zubair Sediqi, and Oliver Gasser. 2025. Sibling Prefixes: Identifying Similarities in IPv4 and IPv6 Prefixes. In *Proceedings of the 2025 ACM Internet Measurement Conference*. 100–119.
[60] John S Otto, Mario A Sánchez, John P Rula, and Fabián E Bustamante. 2012. Content delivery and the natural evolution of DNS: remote DNS trends, performance issues and alternative solutions. In *Proceedings of the 2012 Internet Measurement Conference*. 523–536.
[61] Andriy Panchenko, Fabian Lanze, Jan Pennekamp, Thomas Engel, Andreas Zinnen, Martin Henze, and Klaus Wehrle. 2016. Website Fingerprinting at Internet Scale.. In *NDSS*, Vol. 1. 23477.
[62] Andriy Panchenko, Lukas Niessen, Andreas Zinnen, and Thomas Engel. 2011. Website fingerprinting in onion routing based anonymization networks. In *Proceedings of the 10th annual ACM workshop on Privacy in the electronic society*. 103–114.
[63] Simran Patil and Nikita Borisov. 2019. What can you learn from an IP?. In *Proceedings of the 2019 Applied Networking Research Workshop*. 45–51.
[64] John Pickard, John Southworth, and Dale Drummond. 2017. The IPv6 Internet: An assessment of adoption and quality of services. *Journal of International Technology and Information Management* 26, 2 (2017), 48–64.
[65] Tobias Pulls and Rasmus Dahlberg. 2020. Website fingerprinting with website oracles. *Proceedings on Privacy Enhancing Technologies* (2020).
[66] Digvijaysinh Rathod. 2017. Web browser forensics: google chrome. *International Journal of Advanced Research in Computer Science* 8, 7 (2017), 896–899.
[67] Eric Rescorla and Tim Dierks. 2008. The Transport Layer Security (TLS) Protocol Version 1.2. RFC 5246. https://doi.org/10.17487/RFC5246
[68] Eric Rescorla, Kazuho Oku, Nick Sullivan, and Christopher A. Wood. 2025. *TLS Encrypted Client Hello*. Internet-Draft draft-ietf-tls-esni-25. Internet Engineering Task Force. https://datatracker.ietf.org/doc/draft-ietf-tls-esni/25/ Work in Progress.
[69] Philipp Richter, Oliver Gasser, and Arthur Berger. 2022. Illuminating large-scale ipv6 scanning in the internet. In *Proceedings of the 22nd ACM Internet Measurement Conference*. 410–418.
[70] Vera Rimmer, Davy Preuveneers, Marc Juarez, Tom Van Goethem, and Wouter Joosen. 2017. Automated website fingerprinting through deep learning. *arXiv preprint arXiv:1708.06376* (2017).
[71] Mohammad Amir Salari, Abhinav Kumar, Federico Rinaudi, Reza Tourani, Alessio Sacco, and Flavio Esposito. 2025. Privacy Analysis of Oblivious DNS over HTTPS: a Website Fingerprinting Study. *2025 55th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)* (2025).
[72] Kyle Schomp, Tom Callahan, Michael Rabinovich, and Mark Allman. 2013. On measuring the client-side DNS infrastructure. In *Proceedings of the 2013 conference on Internet measurement conference*. 77–90.
[73] Faisal Shaman, Bogdan Ghita, Nathan Clarke, and Abdulrahman Alruban. 2019. User profiling based on application-level using network metadata. In *2019 7th International Symposium on Digital Forensics and Security (ISDFS)*. IEEE, 1–8.
[74] C. E. Shannon. 2001. A Mathematical Theory of Communication. *SIGMOBILE Mob. Comput. Commun. Rev.* (2001).
[75] Yong Shao, Kenneth Hernandez, Kia Yang, Eric Chan-Tin, and Mohammed Abuhamad. 2023. Lightweight and Effective Website Fingerprinting over Encrypted DNS. In *2023 Silicon Valley Cybersecurity Conference (SVCC)*.
[76] Haya Shulman. 2014. Pretty bad privacy: Pitfalls of DNS encryption. In *Proceedings of the 13th Workshop on Privacy in the Electronic Society*. 191–200.
[77] Sandra Siby, Marc Juarez, Claudia Diaz, Narseo Vallina-Rodriguez, and Carmela Troncoso. 2020. Encrypted DNS => Privacy? A Traffic Analysis Perspective. In *NDSS '20*.
[78] Payap Sirinam, Mohsen Imani, Marc Juarez, and Matthew Wright. 2018. Deep fingerprinting: Undermining website fingerprinting defenses with deep learning. In *Proceedings of the 2018 ACM SIGSAC conference on computer and communications security*. 1928–1943.
[79] Payap Sirinam, Mohsen Imani, Marc Juarez, and Matthew Wright. 2018. Deep fingerprinting: Undermining website fingerprinting defenses with deep learning. In *Proceedings of the 2018 ACM SIGSAC conference on computer and communications security*. 1928–1943.
[80] Florian Streibelt, Patrick Sattler, Franziska Lichtblau, Carlos H Ganán, Anja Feldmann, Oliver Gasser, and Tobias Fiebig. 2023. How ready is DNS for an IPv6-only world?. In *International Conference on Passive and Active Network Measurement*. Springer, 525–549.
[81] The Quad9 Foundation. 2024. An open DNS recursive service for free security and high privacy. https://quad9.net/.
[82] Sulyab Thottungal Valapu and John Heidemann. 2025. Towards a Non-Binary View of IPv6 Adoption. *arXiv e-prints* (2025), arXiv–2507.
[83] Martino Trevisan, Francesca Soro, Marco Mellia, Idilio Drago, and Ricardo Santos Morla. 2022. Attacking DoH and ECH: Does Server Name Encryption Protect Users' Privacy? *ACM Transactions on Internet Technology* (2022).
[84] Tobias Urban, Martin Degeling, Thorsten Holz, and Norbert Pohlmann. 2020. Beyond the front page: Measuring third party dynamics in the field. In *Proceedings of the web conference 2020*. 1275–1286.
[85] Nino Vincenzo Verde, Giuseppe Ateniese, Emanuele Gabrielli, Luigi Vincenzo Mancini, and Angelo Spognardi. 2014. No NAT'd user left behind: Fingerprinting users behind NAT from NetFlow records alone. In *2014 IEEE 34th International Conference on Distributed Computing Systems*. IEEE, 218–227.
[86] Kai Wang, Liyun Chen, and Xingkai Chen. 2018. Website Fingerprinting Attack Method Based on DNS Resolution Sequence. In *International Conference on Applications and Techniques in Cyber Security and Intelligence*. Springer, 1227–1233.
[87] Tao Wang, Xiang Cai, Rishab Nithyanand, Rob Johnson, and Ian Goldberg. 2014. Effective attacks and provable defenses for website fingerprinting. In *23rd USENIX Security Symposium (USENIX Security 14)*. 143–157.
[88] Tao Wang and Ian Goldberg. 2013. Improved website fingerprinting on tor. In *Proceedings of the 12th ACM workshop on Workshop on privacy in the electronic society*. 201–212.
[89] Tao Wang and Ian Goldberg. 2016. On realistically attacking Tor with website fingerprinting. *Proceedings on Privacy Enhancing Technologies* (2016).
[90] Maarten Wijnants, Robin Marx, Peter Quax, and Wim Lamotte. 2018. HTTP/2 prioritization and its impact on web performance. In *Proceedings of the 2018 World Wide Web Conference*. 1755–1764.
[91] Dan Wing and Andrew Yourtchenko. 2012. Happy Eyeballs: Success with Dual-Stack Hosts. RFC 6555. https://doi.org/10.17487/RFC6555
[92] Eric Wustrow, Scott Wolchok, Ian Goldberg, and J Alex Halderman. 2011. Telex: Anticensorship in the network infrastructure. In *20th USENIX Security Symposium (USENIX Security 11)*.
[93] Yixiao Xu, Tao Wang, Qi Li, Qingyuan Gong, Yang Chen, and Yong Jiang. 2018. A multi-tab website fingerprinting attack. In *Proceedings of the 34th Annual Computer Security Applications Conference*. 327–341.
[94] Sami Zhioua. 2015. The web browser factor in traffic analysis attacks. *Security and Communication Networks* 8, 18 (2015), 4227–4241.
[95] Johannes Zirngibl, Lion Steger, Patrick Sattler, Oliver Gasser, and Georg Carle. 2022. Rusty clusters? Dusting an IPv6 research foundation. In *Proceedings of the 22nd ACM Internet Measurement Conference*. 395–409.

A Domain Stability

To complement our analysis of fingerprint stability in the main body, we also examine the temporal behavior of domain-based fingerprints, since changes in domain dependencies directly affect

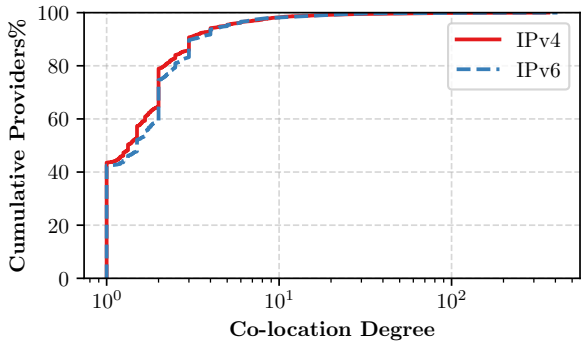

Figure 11: Co-location degrees of providers across all websites, shown separately for IPv4 & IPv6

Table 4: Summary statistics of pairwise Jaccard similarity scores between IP-based fingerprints. Mean and median are computed over non-zero similarity pairs only.

Dual-stack	Protocol	#Pairs	Non-zero	Median	Mean
Incomplete	IPv4	8.4 M	82.23 %	0.0656	0.1043
	IPv6	8.4 M	80.87 %	0.0428	0.0799
Complete	IPv4	10 M	25.05 %	0.2373	0.2900
	IPv6	10 M	24.99 %	0.2010	0.2638

the stability of the resulting IP sets. For each website, we compute the avg. difference degree between its domain-based fingerprints across all subsequent crawls.

Figure 10 presents the cumulative distribution of these difference degrees for dual-stack incomplete and dual-stack complete websites. The figure provides a comparative view of how consistently each category contacts the same set of domains over time. Lower difference degrees indicate higher domain-level consistency across batches, whereas higher values reflect greater temporal variability in the domains contacted. This appendix figure therefore serves as a reference for understanding the domain-level dynamics that underlie the IP-based stability results presented in §6.4.

B Provider Co-location Degrees

Figure 11 shows the cumulative distribution of co-location degrees for all providers in our dataset, plotted separately for IPv4 and IPv6. For each provider, we compute the co-location degree by counting how many domains resolve to each of the provider's IP addresses and then aggregating these counts across the provider's address pool. The x-axis is shown on a logarithmic scale to capture the large variation in domain density, ranging from single-domain mappings to heavily shared IPs used by major hosting platforms. The IPv4 and IPv6 curves are derived from the same set of domains and providers, ensuring a directly comparable view of address-sharing behavior across both protocols.

The two distributions are strikingly similar. Approximately 45% of providers host only one domain per IPv4 address, and around 40% do so for IPv6. By a co-location degree of two, both curves reach roughly 80%, and by degree ten, approximately 95% of providers fall below that threshold for both protocols. This supplements the results presented in the main body by providing the full distributional view that underlies the summary statistics discussed in §7.1.

C Pairwise Similarity Statistics

To complement the kernel density visualizations presented in §6.2, we provide detailed descriptive statistics of the pairwise Jaccard similarity scores between website IP-based fingerprints.

Table 4 reports results separately for dual-stack incomplete and dual-stack complete websites under both IPv4 and IPv6. We emphasize that summary statistics (minimum, median, and mean) are

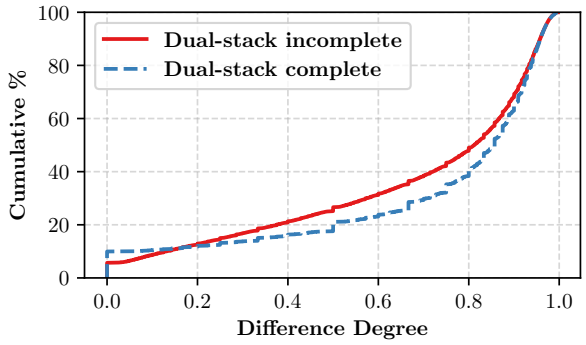

Figure 10: Cumulative distribution of difference degrees of websites' domain-based fingerprints.

computed only over non-zero similarity pairs in order to capture meaningful structural overlap between fingerprints.

For dual-stack incomplete websites, the majority of website pairs share some IP overlap (over 80% non-zero pairs), but similarity values remain low in magnitude (median below 0.07 for IPv4 and below 0.05 for IPv6). This confirms that while minor overlap exists, fingerprints are largely heterogeneous and weakly clustered.

In contrast, DSC websites exhibit a markedly different structure. Approximately 75% of website pairs share no overlap at all, yet the remaining non-zero pairs show substantially higher similarity (median between 0.20 and 0.24). This indicates the presence of clustered hosting patterns: most sites are disjoint, but subsets of sites share significant IP overlap within provider-controlled pools.

These statistics reinforce the structural explanation developed in §6.1 and §6.2: fingerprintability is governed not merely by per-address entropy, but by how IP addresses are distributed and shared across websites within hosting infrastructures.

D Mitigation and Evasion Strategies

The effectiveness of IP-based website fingerprinting ultimately hinges on the stability and predictability of address mappings. When a domain consistently resolves to the same IPs over time, adversaries can exploit this consistency to build robust fingerprints. Therefore, mitigation strategies must focus on introducing uncertainty or variability into IP mappings to degrade fingerprint.

One such strategy is enabled directly by IPv6's design: the use of temporary addresses and prefix rotation [19] to protect user privacy. At the client side, many operating systems implement privacy extensions that randomize the interface identifier, and some ISPs

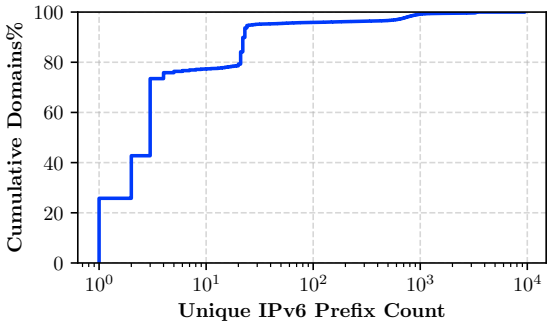

Figure 12: Number of unique IPv6 prefixes observed for each domain across our measurement period.

periodically rotate the delegated /64 prefix assigned to customer networks. These features help disrupt long-term tracking by forcing adversaries to relearn identifiers as they change over time [28].

It is important to distinguish between client-side privacy extensions and the server-side behavior relevant to IP-based website fingerprinting. IPv6 privacy extensions [29, 52] primarily randomize the client's interface identifier [19] to prevent long-term tracking of end users. However, our threat model assumes an adversary observing destination IP addresses of web connections, not the client's source address. Consequently, client-side privacy extensions do not mitigate the core risk studied in this work. From the perspective of IP-based fingerprinting, the relevant question is whether servers and hosting providers introduce sufficient variability in their address assignments.

Importantly, our study reveals that similar dynamics are emerging server-side. As shown in Figure 12, domains in our dataset exhibited varying degrees of IPv6 prefix variability: 25% retained a static prefix throughout the measurement period, 55% rotated among fewer than 10 prefixes, and 20% exhibited high variability, cycling through dozens or even hundreds of unique prefixes.

While some degree of server-side prefix variability is observable in practice, it is neither universal nor consistently deployed as a privacy mechanism [69]. Many domains retain static prefixes throughout our measurement, and even among rotating deployments, changes often occur at operational rather than adversarially disruptive timescales. As a result, current IPv6 privacy mechanisms only partially mitigate fingerprinting risk and do not eliminate the structural separability introduced by hosting allocation policies.

This behavior mirrors moving target defense (MTD) techniques proposed in prior security research, such as MT6D [24], where periodic IP changes are used to confuse adversaries. In our context, frequent server-side prefix changes can invalidate previously learned fingerprints. While the mechanisms behind this rotation may differ, the effect remains the same: adversaries must constantly adapt, increasing their cost and reducing precision.

Beyond rotation, another impactful mitigation lies in limiting co-location of closely related domains on shared IPs. High degrees of colocation (common in shared hosting environments) create large overlaps in fingerprints and enable adversaries to use a single IP as a signal for multiple domains. IPv6's vast address space provides an opportunity to move away from this practice: hosting providers could assign distinct addresses to domains or smaller groups, injecting more entropy into the IP footprint of each site.

However, these techniques do not come without trade-offs. Server-side rotation introduces operational complexity in DNS, load balancing, and configurations, incurring management overhead. Dedicated addressing may be constrained by legacy architectures or cost considerations. Yet, our findings suggest that some providers are already embracing these models, either intentionally for privacy or incidentally through modern orchestration proving its feasibility.

Based on our findings, we offer the following recommendations. Hosting providers can reduce user exposure by maximizing IPv6 domain co-location within shared address pools, as Cloudflare and Google already do naturally through their CDN infrastructure. Website developers should ensure that all third-party dependencies—analytics platforms, CDN-hosted assets, advertising networks, and font services—are reachable over IPv6. IPv4-only sub-resources are the primary driver of high fingerprintability in dual-stack incomplete sites; migrating or replacing these dependencies is the most direct action developers can take to reduce exposure.

E Ethics

In conducting this study on IPv6 WF, we have carefully considered the ethical implications of our methodology and findings, guided by the principles outlined in the Menlo Report [3]. Our study aims to contribute to the understanding of privacy risks in the evolving IPv6 landscape, potentially benefiting the internet community by encouraging the development of privacy-preserving mechanisms. However, we acknowledge the dual-use nature of our work and have taken steps to mitigate potential misuse.

Our data collection and analysis processes were designed to respect individual privacy and autonomy. We exclusively utilized publicly available data, including domain names and DNS data, avoiding the collection or processing of any PII. This approach aligns with the ethical considerations regarding network measurement research. Our automated website crawling process was implemented to mimic normal user behavior, minimizing any potential disruption to the targeted websites. Additionally, we utilized public DNS resolvers, querying once every hour, to ensure our research does not exploit public resources. We curated a diverse dataset that includes a wide range of websites, ensuring a representative sample across various categories and regions. This approach aims to distribute the benefits and risks of our research fairly across the internet ecosystem. Furthermore, by making our dataset and code publicly available, we adhere to the principles of open science, allowing for verification and extension of our work by the broader scientific community.

Throughout our study, we maintained strict compliance with applicable laws and regulations, including adherence to the terms of service of all public services used in our research. We remain committed to the responsible disclosure of any potential risks that may arise from our research. We believe that the potential benefits of increased awareness and improved privacy protections outweigh the potential risks associated with our research. By openly discussing these ethical considerations, we aim to contribute to the ongoing dialogue about responsible research practices in the field of network security and privacy.