On IP Addresses as Identifiers of Internet Users and Services

RUMAISA HABIB, Stanford University, USA
SUDHEESH SINGANAMALLA, Independent, USA
MARWAN FAYED, Cloudflare, Inc., UK
ZAKIR DURUMERIC, Stanford University, USA

The Internet is increasingly attracting attention from legislators, regulators, and policymakers, who seek to balance identity and accountability online with safety and privacy for their constituents. Many policy decisions and operational practices around data handling and privacy, content blocking, abuse mitigation, and threat attribution, treat IP addresses as identifiers of Internet users and services. This assumption may have been reasonable in the past, but only owing to the Internet's early deployment models rather than to its design. The Internet Protocol (IP) specifications explicitly state that non-network attributes must be represented by other means for the Internet to succeed. Indeed, the Internet has evolved substantially because of those designs, and scales because of address sharing, content delivery networks and modern web hosting, as well as large scale privacy proxies, all of which raise questions about IP addresses functioning as meaningful identifiers in practice. In this paper, we provide a holistic empirical assessment of the ability of IP addresses to represent identity, for both clients (the users) and servers. Using data collected from a global CDN, historical DNS records, and active DNS measurements, we investigate the extent to which IP addresses uniquely identify users and services on today's Internet. We find that IP sharing has steadily increased over the past decade, with fewer than 0.2% of domains using a unique IPv4 address and up to millions of domains co-located on single IP addresses. On the client side, traffic is similarly skewed: 5% of client IP addresses account for over half of observed web requests worldwide. We further show that commonly cited explanations for regional differences do not fully explain the observed variation. Our findings have direct implications for Internet policies and practice that depend on IP addresses and risk disproportionate effects on both services and their users. Our results also provide empirical grounding for policymakers and operators to foster precise, accountable, proportionate, and *private* mechanisms for notions of identity.

1 INTRODUCTION

Internet Protocol (IP) addresses are used to identify endpoints and to route traffic across networks. However, they are often mistakenly assumed to uniquely identify specific clients and services by both operators and policymakers. For example, in 2022, Austrian courts directed Internet Service Providers (ISPs) in the country to prevent citizens from accessing several websites by blocking connections to a set of server IP addresses that hosted those websites. Unfortunately, the IP addresses belonged to Cloudflare, a large content distribution network, and the policy resulted in citizens losing access to thousands of unrelated websites [29, 39, 40].

The conceptualization of IP addresses as uniquely identifying users and services likely stems from past practices. IP addresses were originally envisioned to serve as both a network *locator* for routing traffic and an *identifier* for uniquely identifying endpoints. However, as IPv4 addresses became increasingly scarce and expensive, operators began to multiplex large numbers of users behind a single client IP address using Network Address Translation (NAT). Content delivery also evolved. Rather than host a single Internet service like a website on a single IP address, operators began to use higher-layer technologies to colocate multiple websites and services onto the same IP address. Today, many cloud providers and content distribution networks route traffic to data centers and servers using IP addresses, but route traffic to specific websites using HTTPS-based Server Name Indication (SNI) identifiers. As a result, IP addresses often fail to serve as reliable identifiers for clients or servers.

Competing Interests Disclosure: Marwan Fayed is employed by Cloudflare, Inc, a major content distribution network whose operations have previously been affected by IP-based policies (e.g., [29, 39, 40]). Sudheesh Singanamalla was previously employed by Cloudflare, Inc. While this research was performed at Stanford University, Zakir Durumeric is concurrently employed at Censys, Inc., a cybersecurity company that collects data about both IP- and name-based Internet services.

Authors' addresses: Rumaisa Habib, Stanford University, USA; Sudheesh Singanamalla, Independent, USA; Marwan Fayed, Cloudflare, Inc., UK; Zakir Durumeric, Stanford University, USA.

Despite broad appreciation that users and services share IP addresses [4, 20, 24, 34, 42], there is relatively little understanding of the extent to which this happens in practice—understanding that operators and policymakers need to design more precise and proportionate alternatives. In this paper, we present a holistic view of how well IP addresses serve as identifiers as seen by both clients and servers. First, using data from OpenINTEL (a large public DNS repository) [41] and our own active DNS measurements, we investigate the distribution of names across IP addresses. We show that sharing has steadily increased over the past ten years. Today, fewer than 0.2% of domains in the .com, .net, and .org top level domains (TLDs) use a unique IP address and only 9% of the top million domains have unique IP addresses. Not only do websites share IP addresses, but there is extreme concentration with a small number of addresses each hosting millions of domains.

Second, we analzye traffic seen at Cloudflare, a large CDN, to understand the distribution of web requests across client IP addresses in 198 countries. We find that, while the extent varies across countries, there is significant concentration of client traffic on a subset of IPs: 5% of client IP addresses seen sending any requests account for over half (55%) of the (non-abusive) web requests in our dataset. Investigating regions where client traffic is more concentrated, we find that there are no obvious patterns that predict client concentration and that some prior held beliefs (e.g., regions of low historical IP allocation have higher traffic concentration) may be misconceptions. Future work is needed to better understand what leads to varying levels of client concentration on IP addresses.

Grounded in our analysis, we argue that IP addresses should be regarded as the online equivalents of street addresses. They are necessarily public so that they can be found, but little else is known about them. Looking at a street address does not reveal the number of suites, business, or people who use it, nor its purpose; IP addresses share the same form. The use of street addresses, however, is protected. IP addresses are necessarily public and not inherently personally identifiable information but their use can be abused and so, just like street addresses, IP addresses deserve similar protections.

2 BACKGROUND

In this section, we formalize how IP addresses serve two distinct roles, first as a *locator* and second as an *identifier*. Then, we discuss how the evolution of IP allocation, service naming, and routing on the public Internet have allowed the two roles to diverge. In the next section, we empirically measure the extent to which IP addresses serve as unique identifiers of users and services on the public Internet today.

2.1 Identifiers and Locators

An IP address—in its original conception—was envisioned to be both an *identifier* and a *locator* [31]. As an *identifier*, an IP address is a bit string used throughout the lifetime of communication session to uniquely identify each endpoint, while a *locator* is a bit string used to simply indicate where packets should be delivered [31]. As an example of the subtle difference, consider anycast addresses that locate one of many nodes performing the same service. Since an anycast address does not uniquely identify a host, an anycast address is a locator but not an identifier. The dual role has been the source of much confusion, but the Internet community has generally recognized the use of IP addresses as locators more important than their use as identifiers [31]:

> It has been considered more useful to deliver the packet, then worry about how to identify the end points, than to provide identity in a packet that cannot be delivered. [RFC 2101]

There is a long history of arguments in favor of separating identity and location, and it is now generally recognized that the current approach to use IP addresses as both an identifier and a locator was a poor design choice [5]. In fact, the IETF (Internet Engineering Task Force) has a Locator/ID Separation Protocol (LISP) Working Group that aims to support "a routing architecture which decouples the routing locators and identifiers" [21]. While, in many situations, IP addresses are used as both an identifier and locator, we note that one reason a separate

identifier header was not added to the packet header as part of the original IP specification was that it was deemed unnecessary for routing the packet and was considered private to the end-points.

2.2 Addresses, Names, and Routes

Since their original introduction, IP addresses have had significant evolution in function and form.

2.2.1 Private IP Space and NAT. To preserve limited IPv4 address space [2], "private" address space [11] was allocated in 1994, and Network Address Translator (NAT) [15] was introduced to ensure that hosts on private enterprise networks could access "public hosts of different enterprises" [11]. With the adoption of NAT, public IP addresses ceased to act as a unique client *identifier* for some networks, but continued to act as a global routing *locator*. Beyond NAT being used within individual subscriber networks (e.g., a single home network), Internet Service Providers (ISPs), in particular, many cellular carriers, began to deploy Carrier-Grade NAT (CG-NAT) to multiplex a large number of subscribers across a single IP address [24, 34].

2.2.2 IPv6. The continuing growth of the Internet and the limited IPv4 address space led to the IPng effort, which was eventually standardized as IPv6. IPv6 sought to solve the address exhaustion issue plaguing IPv4 by using a 128-bit address. However, the surprising effectiveness and resilience of the "short-term" fixes such as NAT and private IP addresses to conserve IPv4 addresses has significantly reduced the incentive to convert to IPv6.

2.2.3 Autonomous Systems. With the emergence of regions with different routing policies, the Autonomous System (AS) [18] construct was introduced to be the unit of routing policy for protocols like BGP (Border Gateway Protocol) [25]. Consequently, the Internet has evolved into a nested routing structure. BGP is used to deliver a packet to the correct AS. Every AS advertises its own prefixes (CIDR addressing), which are then used as *flat identifiers* for routing computation by BGP. On delivery to the AS, an *intra*-AS routing protocol (an AS can choose any protocol it wants) is then used to deliver the packet to the intended end-point within that AS. Frequently, the end-point could itself be a NAT box which then appropriately rewrites to deliver to the correct host within its own region. It is worth noting that the original Internet architecture did not include ASes; they were added later as needs evolved. As such, through evolution of the Internet, primary elements of the original design such as the single global address space turned out not to be a crucial constraint, while aspects like ASes which were not considered at all have turned out to be crucial for the functioning of the Internet.

2.2.4 Service Identifiers. Since the early days of the Internet, a distinction was sought to be made between names, addresses, and routes [28, 38]. While an IP address indicates which network a node resides in and how to get there, it does not inherently indicate what services are on the IP address [38]. Instead, we have relied on human-readable *names* to indicate what service is sought by the user. A binding between a name and an IP address is provided by DNS, and a binding between the *resolved* IP address and the route is provided by BGP [37]. However, while we might have considered the name to be an identifier and IP address the locator in this situation, after a name was resolved to an IP address, the IP address inadvertently became both the identifier and locator of the expected service since the IP layer provided no opportunity to present a secondary identifier. For example, historically, only one TLS-enabled website could be hosted on a given IP address.

The TLS Server Name Indication (SNI) Extension explicitly separates the role of identity and locator for TLS-based services. Clients indicate the identity of the service they seek in a TLS header field, and use IP as only a locator to route packets to a web server. TLS SNI has seen widespread adoption over the past 10–15 years [13, 42]. Recently, Cloudflare has argued that SNI and ECMP allow for IP to be split from services, returning IP's role to merely that of reachability [16].

Yet, operational and policy practice has not kept pace with this technical evolution. Many of the systems that block, rate-limit, or attribute traffic by IP address continue to treat IPs as an identifier long after its locator role became primary—a mismatch we quantify for both clients and servers in the remainder of this paper.

3 NAMES AND SERVICES

HTTP(S)-based Internet services regularly share IP addresses, routing requests to the appropriate backend service using mechanisms like TLS Server Name Indication (SNI) [42]. This architecture has direct consequences for enforcement actions that target an IP address rather than the service it hosts: because a single address can serve an enormous and unrelated set of domains, blocking, filtering, or rate-limiting decisions made at the IP layer routinely produce effects far broader than intended. The 2022 Austrian court order discussed in Section 1 is one documented instance of this dynamic, but the same structural risk applies wherever regulators, ISPs, or network operators rely on IP addresses to identify or restrict a specific service [3, 7, 8, 30].

In this section, we measure the extent to which domains share public IP addresses across two datasets, providing an empirical basis for assessing how often such collateral effects are likely to occur. First, we analyze domains in the `.com`, `.net` and `.org` TLDs from 2015–2025 using data provided by OpenINTEL [41]. Second, we conduct active DNS queries for popular domains in the Cloudflare Radar top sites list using ZDNS [23] in April 2025. Because we investigate domains rather than subdomains, we emphasize that our results *underestimate* IP address sharing, but better account for unrelated entities sharing IPs (e.g., websites on an IP do not share organizational ownership).

As of February 2025, only 0.2% of domains in the .com, .net, and .org TLDs in the OpenINTEL dataset map to a unique IP address in the dataset (Figure 1). Conversely, only 16% of IP addresses that host at least one domain in one of the three TLDs host only one domain (from the three TLDs). To understand whether the websites that users frequent in practice use shared IPs, we analyzed the IP addresses serving the top million most popular domains in the Cloudflare Radar Domain Rankings [1]. (Ruth et al. previously showed that the top million websites account for about 95% of user traffic [35].) Looking at the top million domains, we see that only 9% have a dedicated IP address amongst the set (Figure 2c). Even at the high end, only 19% of the Top 100 domains have dedicated IP addresses and only 19% of the IP addresses serving a Top 100 domain serve only one domain (Figure 2b).

Not only do most domains use shared IP addresses, but a relatively small number of IP addresses serve most domains in our three popular TLDs: 50 IPs serve 55% of domains, 1K IPs 72% of domains, and 5M IPs 99% of domains. Much of this concentration is due to several outlier IP addresses. For example, as can be seen in Figure 3a, a single Google IP address (34.102.136.180) hosted over 43 million domains starting in 2020, which then shifted to Amazon (3.33.130.190) in October 2023 after Squarespace acquired Google Domains. Manual inspection reveals that this IP is used to host content for parked domains that are currently not in use. While the top four IPs with the most domains primarily host parked domains, many other IPs have millions of live domains. For example, 198.185.159.144 (Squarespace) serves nearly 10M domains and Amazon's global accelerator accounts for four out of the top ten IP addresses by domain count (Table 1).

Excluding IP addresses that solely host parked domains, the top 50 IPs serve over 182M domains (48%). To better understand why this occurs, we analyze the top five service providers by domain name. These five providers serve approximately 66% of domains in the .com, .net, and .org TLDs as well as 49% of domains in the top million. Figure 4 shows the distribution of domain count for the top 5 ASes ranked by maximum number of domains under a single IP address. Squarespace and Wix exhibit a sharp decline: although they peak at over 9M and 5M domains on a single IP, respectively, their median IP hosts only 3 domains (Squarespace) and 4 domains (Wix). In contrast, Amazon, Cloudflare, and Google Cloud (all of which provide CDN services) consistently host thousands of domains under a single IP address.

IP Address	Domains	Owner
198.185.159.144	9,780,432	Squarespace
198.185.159.145	9,733,980	Squarespace
198.49.23.145	9,728,856	Squarespace
198.49.23.144	9,728,127	Squarespace
15.197.225.128	9,355,756	Amazon
3.33.251.168	9,355,731	Amazon
13.248.243.5	5,894,632	Amazon
76.223.105.230	5,870,872	Amazon
34.149.87.45	5,842,294	Google Cloud
185.230.63.107	5,552,351	Wix

Table 1. IP addresses that serve the most live (non-parked) domains—The top 10 IP addresses serving live domains account for over 80M (21%) of the domains in the .com, .net, and .org TLDs.

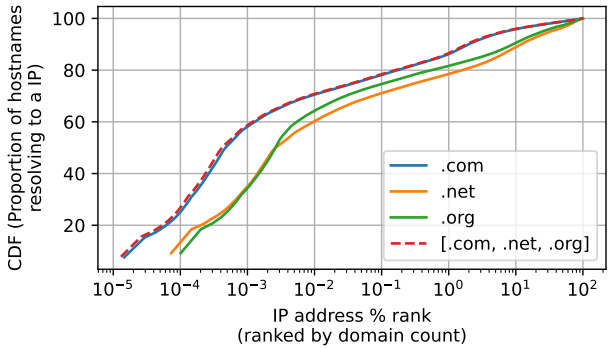


Fig. 1. CDF of domains by hosting IP addresses—1% of IP addresses in our dataset host 87% of the domains under the .com, .net, and .org top level domains (TLDs).

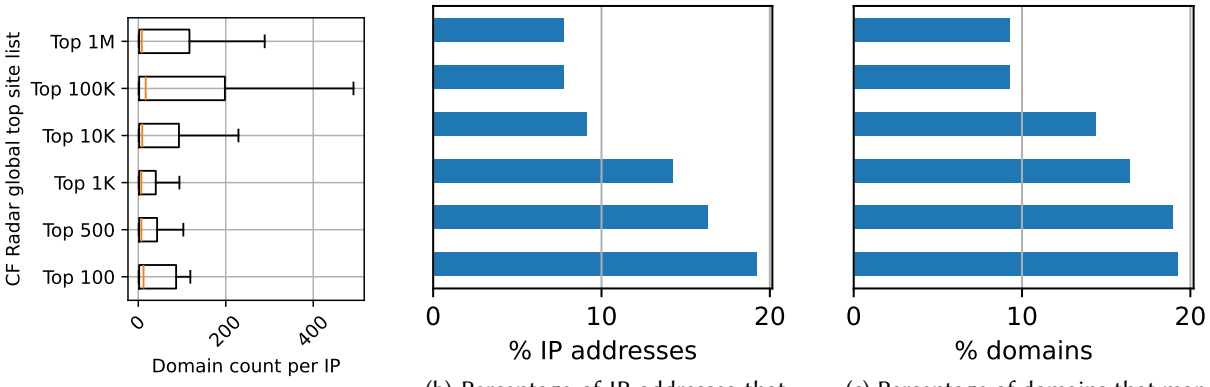
If history is a predictor of the future, we can expect IP addresses to continue to lose their role as a service identifier. As can be seen in Figure 3b, domains have become increasingly concentrated on a small number of IP addresses every year for the past decade. In 2015, 0.9% of IP addresses accounted for 80% of domains while in 2024, 0.2% of IP addresses account for the same percentage of domains.

4 USERS AND CLIENT IPS

Client IP addresses raise a distinct but related concern: rather than determining what content is delivered, they are frequently used to determine *who* may access a service, at what rate, or under what reputation [7, 8]. Because CGNATs, VPNs, and other proxies place many users behind a single address, such IP-based policies risk penalizing every client who shares that address for the actions of one. We next consider the distribution of traffic across client IP addresses as seen by a global content distribution network to quantify the scale of this risk.

4.1 Dataset

Our client dataset consists of anonymized per-IP measurements of web requests from 0.1% sampled HTTP request logs of Cloudflare, a global CDN that handles an average of 78M requests per second from POPs in 330 cities

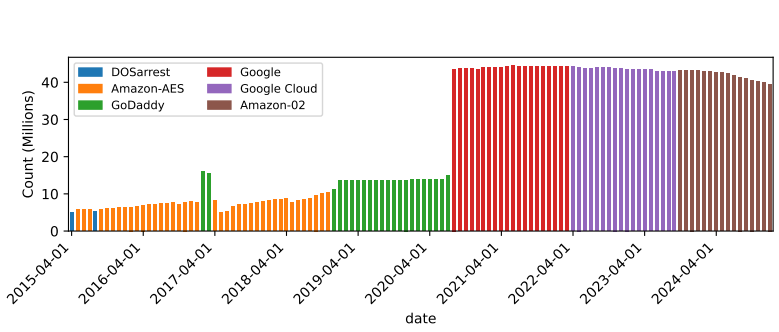


(a) Distribution of domain counts.

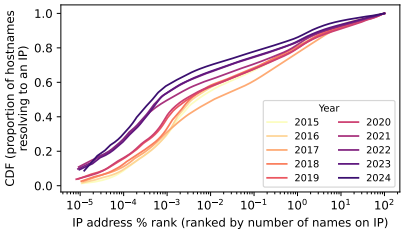
(b) Percentage of IP addresses that map to a single domain.

(c) Percentage of domains that map to a single IP address.

Fig. 2. Domains from the .com, .net, and .org zone files that resolve to IPs of domains in the Cloudflare Radar Domain Rankings buckets— Less popular websites are less likely to have dedicated IP addresses.



(a) Maximum domains per IP over 10 years.



(b) Domains per IP over time. We show the number of domains per IP address as seen by OpenINTEL over the past ten years for the .com, .net, and .org TLDs. IP sharing has increased in each subsequent year.

Fig. 3. Longitudinal evolution of IP sharing over the past decade.

and 125 countries. Entries in the log are randomly sampled at each server in the fleet, before being recorded in a global log. Aggregates are tabulated over a 24-hour interval on February 26, 2025, annotated with device type (desktop, mobile, tablet, and other), origin autonomous system, and country (as determined by MaxmindDB [26]), filtered to remove bots and other automated requests using Cloudflare's proprietary bot detection system [9], and then anonymized. Each entry in the dataset contains the total number of requests, number of distinct user-agents (UAs), and number of distinct domains visited per (anonymized) IP address.

To respect user privacy and ensure sufficient data points in our 0.1% sample, we exclude countries with fewer than 100 client IPs and fewer than 10K distinct user agents from our analysis. This excludes 0.2% of the total requests and 48 ISO 3166-1-defined countries. In total, we analyze requests from about 195.8M distinct source IPs.

Limitations and Bias. Our client dataset represents the passive view of Cloudflare, a popular CDN service, excluding bots. The quality of the data is shaped by the diversity of users who visit the websites hosted by Cloudflare. Third-parties estimate ~20% of all websites use this CDN [43], which helps mitigate bias from any

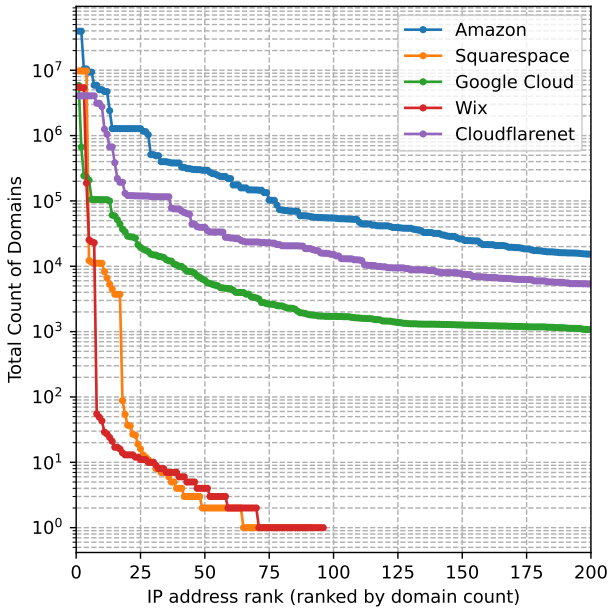


Fig. 4. Distribution of domain count for top 5 organizations, ranked by number of total domains. Only the top 200 IP addresses are shown—Squarespace and Wix show a sharp decline in IP sharing over their fewer IP addresses, while Amazon, Cloudflare, and Google have a consistent spread of thousands of domains per IP address across their larger IP pool.

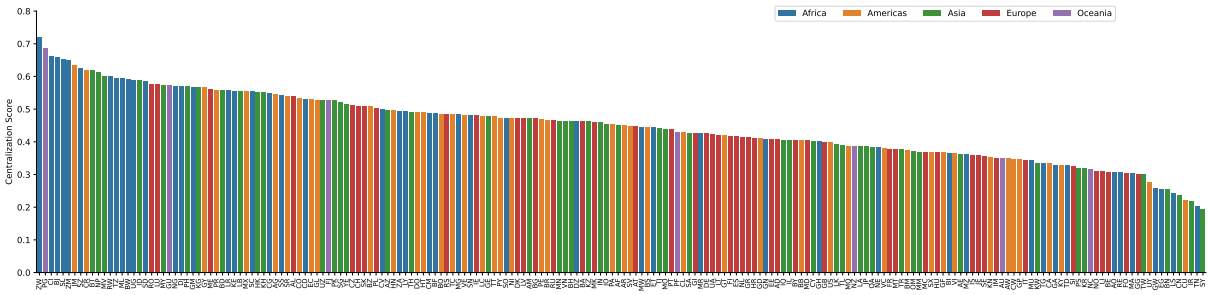


Fig. 5. Earth Mover's Distance (EMD) for the distribution of User Agent counts, normalized as a fraction of the maximum EMD (in which all distinct user agents have the same IP address). A higher score indicates a greater extent of IP sharing.

one service of any popularity. Separately, potential biases associated with population distributions are mitigated by sampling close to the sources. For example, servers in Africa sample their requests independently from servers in North America, so populations from each region are sampled separately. Finally, the sensitive and proprietary nature of the data precludes the identification of ASNs and network names. However, we do describe network types (e.g., broadband, hosting, mobile) where useful.

4.2 Metrics

There exists no externally visible indicator for the number of users behind a client IP address. Instead, we analyze the concentration of requests and unique user agents across client IP addresses to understand the extent to which

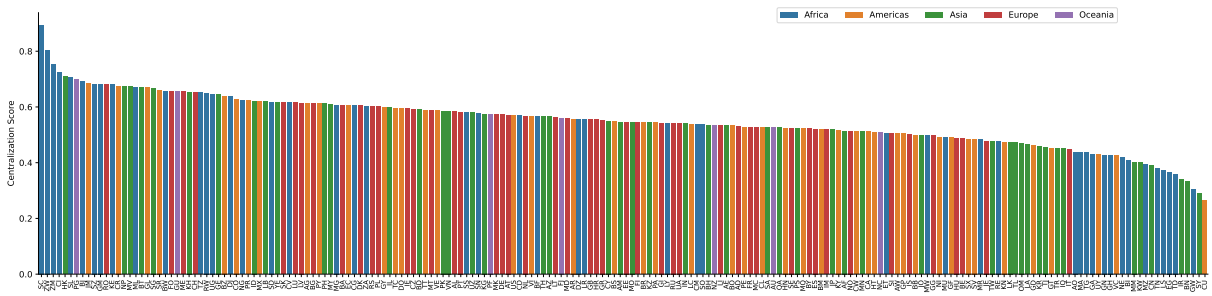


Fig. 6. Earth Mover's Distance (EMD) for the distribution of number of requests, normalized as a fraction of the maximum EMD (in which all requests come from the same IP address). A higher score indicates a greater extent of IP sharing.

client IPs are likely shared. While past work has shown high correlation between varying client metrics as seen from a large CDN [36], we acknowledge the limitations of these two measures. User-agents are perceived as estimators of real users behind IPs [33], but the relationship between users and user-agents behind any single IP address is poorly understood. Similarly, request counts likely correlate with numbers of users, but are also affected by other factors like website complexity and the subrequests generated by a webpage. Here, too, the signal quality for individual IPs is not well understood.

Instead of attempting to determine the exact number of users behind an IP address, we instead investigate how concentrated user requests and user agents are across client IP addresses across different regions and device types. We specifically borrow inspiration from the centralization metric introduced by Habib et al. [17] and measure request concentration using Earth Mover's Distance (EMD): the distance from a theoretical, perfectly uniform distribution—one where each IP has an equal number of user agents or requests. We normalize the EMD by the maximum possible value to get a score between 0 and 1. We calculate EMD in two contexts for each country:

User Agent EMD, E_u. We define E_u as the work required (normalized by the maximum possible work) to shift the distribution of user-agents per IP in a country to a fully uniform distribution where every IP address has an equal number of user agents.

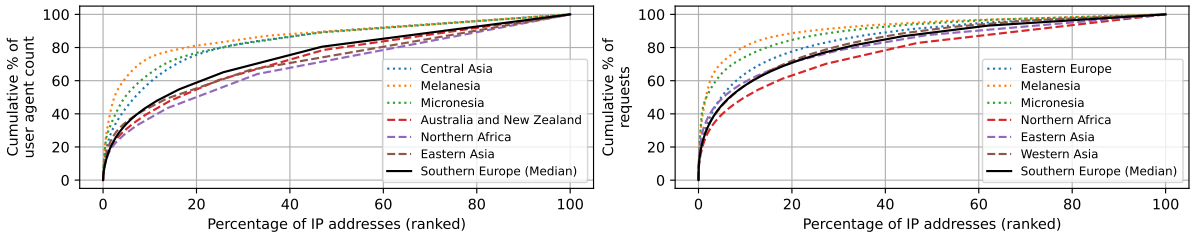
Request EMD, E_r. We define E_r as the normalized work required to shift the distribution of *web requests* in a country to a fully uniform distribution where every IP address makes an equal number of requests.

We emphasize that we do not expect the distribution to be fully uniform (*i.e.*, $E = 0$); rather this is a theoretical distribution against which to compare such that we can understand how concentration of requests and user agents compares between countries and regions.

4.3 Results

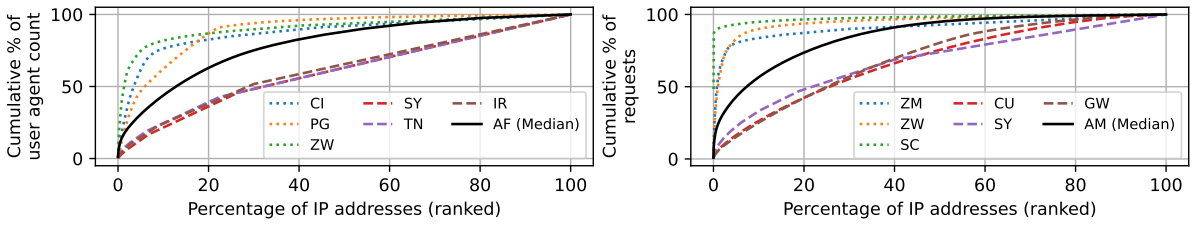
Globally, 5% of IP addresses account for over half (55%) of web requests in our dataset. However, as can be seen in Figures 5 and 6, concentration varies significantly between countries. Globally, the distribution of requests per IP ($\bar{E_r} = 0.55$) is more concentrated than user agents ($\bar{E_u} = 0.45$), but they are highly correlated with a Pearson correlation of $(r) = 0.85, p \ll 0.05$. Across all source IPs, 44% of IPs have at least two distinct user agents and, at the extreme, one IP address has requests from more than 7,000 UAs. Within individual countries, the most dense IP address accounts for an average of 0.36% of all user-agents, ranging from 0.003% in Italy to 16% in Seychelles.

The variation across countries has several notable outliers (Figure 8). For example, requests and user-agents are concentrated on notably few IP addresses on the islands of Seychelles ($E_u = 0.55, E_r = 0.89$) and Papua New Guinea ($E_u = 0.69, E_r = 0.70$). In both cases, a large fraction of requests originate from incumbent telecommunication companies. In Seychelles, 30% of total UAs originate from one AS. Elsewhere, in Zimbabwe ($E_u = 0.72$,$E_r = 0.80$),



(a) User Agents (b) Requests

Fig. 7. Percentage of IP addresses against the percentage of (a) distinct user agents and (b) requests they account for, for 3 subregions with the highest, lowest and median average (a) E_u and (b) E_r.



(a) User Agents (b) Requests

Fig. 8. Percentage of IP addresses against the percentage of (a) distinct user agents and (b) requests they account for, for 3 countries with the highest, lowest and median (a) E_u and (b) E_r.

5% of IP addresses account for 83% of all requests, and 10% of IP addresses account for 82% of user agents. The top five client IP addresses also originate from a single ISP. In contrast, traffic from Syria is much less concentrated ($E_u = 0.19$,$E_r = 0.29$), with 26% of IP addresses showing more than one user agent. However, we are mindful that the one unique UA behind the remaining 74% of addresses may be a software client used by many and an underestimation given the 0.1% sample. Unsurprisingly, the number of IP addresses in the United States makes it the country that appears most often in the dataset. Concentration is middle of the range ($E_u = 0.40$, $E_r = 0.57$), with 5% of IP addresses accounting for over half (57%) of its requests and 39% of its UAs.

Focusing on subregions, we see more variation (Figure 7). Micronesia and Melanesia, characterized by many islands, have the highest E_u concentration (0.65 and 0.59, respectively). In Micronesia, 20% of IP addresses account for 76% of user agents. Africa (not shown) appears polarized. Sub-Saharan Africa contains four of the top five countries with greatest concentration values by both user-agents and requests. Northern Africa has lower concentration and yet, still, 20% of IP addresses account for 50% of the total of distinct user-agents behind each IP. Asia also has lower concentration on average ($E_u = 0.44$,$E_r = 0.53$) but large variation across individual countries.

4.4 Potential Misconceptions

We test three hypotheses for why the concentration of clients across IPs varies dramatically across countries. We find that none adequately explain the differences, suggesting that other forces may be driving the variation.

Dominance of ISPs. Seychelles has the highest concentration of client requests on a small number of IPs with 84% of requests originating from a single IP, consistent with the dominance of one large provider. However, we find the association untrue elsewhere. For example, 99.5% of requests appearing from Syria and 99.7% from Cuba

come from a single provider in each country, and yet both countries' distributions are closest to uniform ($E_u = 0.19$, $E_r = 0.29$ and $E_u = 0.22$, $E_r = 0.26$, respectively). In further contrast, Zimbabwe has high concentration even though the top ISP in the dataset generates a much lower 32% of requests. Our data suggests that concentration on a small number of IPs is not always governed by a singular ISP. However, in some cases, the decisions of large providers can shape traffic patterns.

Mobile Networks. We considered the extent of mobile networks and their use of carrier-grade NATs as a possible explanation, where we expect mobile devices are most often connected. We do find that lower desktop usage corresponds to a higher E_u ($r = -0.42, p \ll 0.05$), but also find no statistically significant correlation with mobile devices, which is not the strict complement of desktop share given the presence of tablet and other device classes. If large-scale IP sharing behind CGNATs, VPNs, and proxies were causal, then we would expect correlation between IP sharing and mobile usage.

RIR IP Allocations. We also considered that historically fewer IP addresses are allocated to regions where we see high EMD values. We took IP allocations as reported by IANA and regional registries [22], which we normalized by the population of the respective owning countries. Surprisingly, E_r and E_u are poorly correlated with the number of allocated IP addresses per person per country ($r = 0.19$ and $r = -0.15$, respectively). This further suggests that IP usage is largely influenced by individual ISPs' decisions rather than simply IP allocation.

Overall, many of the hypotheses where we see the highest concentration did not play out as expected, leaving open questions for future work as to what leads to increased concentration, or if this is simply due to ad-hoc ISP decisions. This unpredictability itself has policy relevance: because concentration appears to follow arbitrary infrastructure choices rather than any deliberate or documented policy, the clients most exposed to collateral damage from IP-based enforcement cannot be identified in advance, and bear no responsibility for that exposure.

5 DISCUSSION

Our results suggest that we need to rethink what client and server identifiers we should use on the Internet today. IP addresses have—at least in the case of HTTP(S)-based services—ceased to act as identifiers for servers. In some environments, IP addresses continue to serve as a valid client identifier, but, in many cases they do not, creating inequitable Internet experiences for much of the world due to historic allocation decisions. We contend that the issue of identifiers and locators is not merely a semantic one, but one that has significant impact on the Internet and its use. In this section, we raise salient issues that are rooted in this identifier/locator tussle, issues that the community as a whole must deliberate on and consciously chart a path forward.

5.1 On Separate Client and Service Identifiers

IP addresses today are asked to serve as proxies for a wide range of requirements that cannot be ignored: they are used both to rate-limit abusive users despite collateral effects, and to geolocate users to provide or restrict region-specific content—two fundamentally different uses of a single label designed for neither. Identifiers should instead be restricted to their needs: specific and narrowly scoped to prevent side-uses, and fair, so that no entities or individuals suffer collateral effects. Our study makes clear that IP addresses are neither specific nor fair, given their varying allocations and usage patterns.

We contend that a uniform, symmetrical end-point identifier (like an IP address) for both services and clients does not fit the needs of the Internet. Client and server identity needs differ significantly, and a single shared identifier that can be used by both clients and servers —as most proposed alternative Internet architectures do — may very well not support either case in full. A single identifier also requires significant Internet re-architecture, whereas we need better solutions *today*. Servers, particularly on the web, have already gravitated towards using names, cryptography bound to public keys and certificates, rather than IP addresses. We suggest that other

cryptographic Internet protocols beyond TLS consider how server identities can be baked into their authentication mechanisms such that IP address's purpose can be reduced to being used for only routing as a locator. We note that while SNI adoption has enabled name-based identification of TLS flows, there remains an open question of whether Encrypted Client Hello (ECH) will cause server filtering to continue to occur at the IP-layer and whether any server indicator should be publicly included in a flow.

We explicitly *do not want* client identifiers to be public or be meaningful to entities beyond the destination service. In fact, for client identifiers privacy is a key driving factor [10, 19]: such identifiers must not be public or hold meaning beyond a specific use, and are likely ephemeral. Privacy Pass [10] already safeguards access to Apple's Private Relay service, and a variant of it being considered for standardization at IETF [19] extends this model to let clients obtain stable and unique identifiers from networks more generally. We argue that these client tokens should be considered more broadly than where initial conversations began (e.g., Tor users) such that we can migrate away from IP addresses being used as a client identifier altogether.

Finally, we argue that the Internet has *nearly* evolved to this split naturally. The SNI represents the culmination of an ideal service identifier. It embeds an identifier in the packet at a fixed location (though not in the header) and is not only meaningful but guarantees authenticity (WebPKI). At the same time, the modern DNS with its dynamicity has pretty much solved the problem of keeping location information of names up-to-date [16]. Overall, we contend that acknowledging SNI as an identifier will allow us to apply decades of thinking on identifiers to reason about SNI's evolution in a more structured way.

5.2 On IPv6

Since most of our discussion so far has revolved around IPv4, the adoption of IPv6 may appear to be the most obvious solution to all the problems that plague IPv4, such as address exhaustion and the need to share addresses as a result. We argue that IPv6 does not, in fact, fundamentally solve the identifier/locator conflict that plagues IP addresses, nor is it a reliable identifier in its own right. Hyperscalers and CDNs are already agile about how they allocate and reuse IP addresses, a practice unlikely to change under IPv6 [16]. Client operating systems, meanwhile, rotate IPv6 addresses at most every 24 hours for privacy by default, and prior work shows that scanners and bots use IPv6 in different ways than legitimate clients [32]. If anything, we contend that IPv6's larger address space only increases the temptation to continue using addresses as some form of end-point identifier, and we believe that this temptation must be resisted.

The explicit assumption at the heart of IPv6 is the assumption that the Internet needs a single, global address space. While there is no doubt that complete adoption of IPv6 will be a net positive since it fixes the inequities that are implicit in IPv4 (*e.g.*, unequal address space distribution), whether the Internet truly needs a single, global address space remains an open question. Moreover, even with IPv6, we find that providers choose to host multiple domains at the same IP though not at the same scale as IPv4. Thus, even with IPv6 the need for SNI (or an alternate service identifier) remains. On the client side, IPv6's larger address space does not translate into stable identifiers either: beyond the default 24-hour rotation noted above, clients who wish to circumvent censorship or other restrictions may also use a VPN or Tor, which share IPv6 addresses just as they do IPv4 addresses today, leading us back to the same issue we face with IPv4.

While the move to IPv6 is a net positive, we are at risk of repeating the same mistakes made with IPv4. The temptation to use IPv6 as an identifier only delays, and perhaps exacerbates, the trends we document, and the community as a whole should resist that temptation while deliberating what problems we expect IPv6 to fix.

5.3 On Blocklists

On the Internet today, blocklists represent an expression of a policy. The policy they represent may vary, ranging from parental control to copyright strikes to outright censorship. The perception of IP addresses as identifiers,

leads blocklists to implement these policies, at times, by blocking IP addresses. However, this systemic blocking of IP addresses may result in significant collateral damage as seen before [7, 8, 27, 29, 30, 40]. Recognizing this collateral damage, there has been a recent shift towards using SNI and DNS to block services.

More generally, we note that any identifier embedded in a packet may be used to block traffic. Notably, the original design of the Internet assumed that the interests of the entities on the Internet were always *aligned* (*i.e.*, the interest of sender matched with those of the receiver (or anything on the route)) [6]. Unfortunately, attacks such as Denial of Service (DoS) demonstrate that this assumption was incorrect. As a result, users want blocking to succeed when they are being attacked, while they may want it to fail when content and services are being censored by repressive regimes. As such, to introduce an "improved identifier" is to improve the granularity of blocking traffic though it may not always serve interests aligned with our own.

Overall, the issue of blocklists (*i.e.*, blocking traffic towards a policy objective) is a tricky one. The presence of granular identifiers prevents collateral damage but may also compromise privacy. Overall, we argue that the community needs to deliberate on how we balance the need for privacy, and the legal mandates to take down services, while at the same time preventing collateral damage.

5.4 On Ownership and Responsibility

Another question that is raised by the evolving use of IP addresses is how we attribute ownership, and consequently responsibility, of services and resources. One model is to assign ownership of services and resources to the underlying "owners" of the IP addresses. However, in the case of CDNs which serve content and not host it, this model may fall short. Another model would be to attribute ownership to the domain owners of the service identifiers, but it is not always easy to find this information readily. As such, another aspect to consider is whether we as a community want identifiers that allow us to attribute ownership and responsibility.

6 CONCLUSION

In this paper, we presented an empirical assessment of whether IP addresses still serve as reliable identifiers for clients and servers on the Internet. Combining a decade of DNS records with a global CDN's traffic logs, we find that they largely do not: server IP addresses are shared at massive scale, and client traffic is concentrated on a small, unpredictable subset of addresses in every region we examined. Closing this gap requires the networking community to rethink Internet identity around principles of authenticity, transparency, and privacy, rather than continuing to lean on an address whose original purpose was to route packets, not identify the people and services behind them.

ETHICS

For the active measurements performed on our server, we adhered to the active scanning guidelines outlined by Durumeric et. al [12, 14]. We did not receive any opt-out requests. The data from the large CDN provider was anonymized and aggregated before being shared with us to ensure privacy and confidentiality. As such, our work does not constitute human subjects research per our institutional policy and does not require IRB approval.

ACKNOWLEDGMENTS

We thank Gautam Akiwate, Gerry Wan, Liz Izhikevich, and members of the Empirical Security Research Group. This work was supported in part by the National Science Foundation under Grant Numbers #2319080 and #2540803, as well a Sloan Research Fellowship and Stanford Graduate Fellowship. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding organizations.

REFERENCES

[1] [n. d.]. https://radar.cloudflare.com/domains Cloudflare Domain Rankings.

[2] Philip Almquist and Phillip G. Gross. 1992. IESG Deliberations on Routing and Addressing. RFC 1380. https://doi.org/10.17487/RFC1380 https://www.rfc-editor.org/info/rfc1380.

[3] Mohamed Boucadair, Mat Ford, Phil Roberts, Alain Durand, and Pierre Levis. 2011. Issues with IP Address Sharing. RFC 6269. https://doi.org/10.17487/RFC6269 https://www.rfc-editor.org/info/rfc6269.

[4] Frank Cangialosi, Taejoong Chung, David Choffnes, Dave Levin, Bruce M Maggs, Alan Mislove, and Christo Wilson. 2016. Measurement and analysis of private key sharing in the https ecosystem. In *ACM Conference on Computer and Communications Security*.

[5] David D Clark. 2018. *Designing an Internet*. The MIT Press.

[6] David D Clark, John Wroclawski, Karen R Sollins, and Robert Braden. 2002. Tussle in cyberspace: defining tomorrow's internet. In *Proceedings of the 2002 conference on Applications, technologies, architectures, and protocols for computer communications*. 347–356.

[7] Cloudflare. 2018. Getting Cloudflare capcha on almost every website I visit for my home network. Help! Forum. https://community.cloudflare.com/t/getting-cloudflare-capcha-on-almost-every-website-i-visit-for-my-home-network-help/42534.

[8] Cloudflare. 2019. Cloudflare blocking my IP? Forum. https://community.cloudflare.com/t/cloudflare-blocking-my-ip/65453.

[9] Cloudflare. 2025. Improved Bot Management flexibility and visibility with new high-precision heuristics. https://blog.cloudflare.com/bots-heuristics/. Accessed: 2025-03-20.

[10] Alex Davidson, Ian Goldberg, Nick Sullivan, George Tankersley, and Filippo Valsorda. 2018. Privacy Pass: Bypassing Internet Challenges Anonymously. In *Proceedings of Privacy Enhancing Technologies 2018 (PoPETs '18)*. De Gruyter, Barcelona, Spain, 164–180.

[11] Geert Jan de Groot, Daniel Karrenberg, Yakov Rekhter, and Robert Moskowitz. 1994. Address Allocation for Private Internets. RFC 1597. https://doi.org/10.17487/RFC1597 https://www.rfc-editor.org/info/rfc1597.

[12] Zakir Durumeric, David Adrian, Phillip Stephens, Eric Wustrow, and J Alex Halderman. 2024. Ten years of zmap. In *ACM Internet Measurement Conference*.

[13] Zakir Durumeric, James Kasten, Michael Bailey, and J Alex Halderman. 2013. Analysis of the HTTPS certificate ecosystem. In *ACM Internet measurement conference*.

[14] Zakir Durumeric, Eric Wustrow, and J Alex Halderman. 2013. ZMap: Fast Internet-wide scanning and its security applications. In *22nd USENIX Security Symposium*.

[15] Kjeld Borch Egevang and Paul Francis. 1994. The IP Network Address Translator (NAT). RFC 1631. https://doi.org/10.17487/RFC1631 https://www.rfc-editor.org/info/rfc1631.

[16] Marwan Fayed, Lorenz Bauer, Vasileios Giotsas, Sami Kerola, Marek Majkowski, Pavel Odintsov, Jakub Sitnicki, Taejoong Chung, Dave Levin, Alan Mislove, Christopher A. Wood, and Nick Sullivan. 2021. The Ties That Un-Bind: Decoupling IP from Web Services and Sockets for Robust Addressing Agility at CDN-Scale. In *Proceedings of the 2021 ACM SIGCOMM 2021 Conference* (Virtual Event, USA) *(SIGCOMM '21)*. Association for Computing Machinery, New York, NY, USA, 433–446. https://doi.org/10.1145/3452296.3472922

[17] Rumaisa Habib, Kimberly Ruth, Gautam Akiwate, and Zakir Durumeric. 2025. Formalizing Dependence of Web Infrastructure. In *ACM SIGCOMM Conference*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3718958.3750492

[18] John A. Hawkinson and Tony J. Bates. 1996. Guidelines for creation, selection, and registration of an Autonomous System (AS). RFC 1930. https://doi.org/10.17487/RFC1930 https://www.rfc-editor.org/info/rfc1930.

[19] Scott Hendrickson, Jana Iyengar, Tommy Pauly, Steven Valdez, and Christopher A. Wood. 2023. Rate-Limited Token Issuance Protocol. Work in Progress.

[20] 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. *PETS* (2021).

[21] IETF. [n. d.]. Locator/ID Separation Protocol (lisp) Working Group. (Accessed 2023-06-20).

[22] IP2Location. 2025. Internet IP Address Report 2025. https://www.ip2location.com/reports/internet-ip-address-2025-report. Accessed: 2025-05-11.

[23] 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* (Nice, France) *(IMC '22)*. Association for Computing Machinery, New York, NY, USA, 33–43. https://doi.org/10.1145/3517745.3561434

[24] Ioana Livadariu, Karyn Benson, Ahmed Elmokashfi, Amogh Dhamdhere, and Alberto Dainotti. 2018. Inferring carrier-grade NAT deployment in the wild. In *IEEE Conference on Computer Communications*.

[25] K. Lougheed and Y. Rekhter. 1990. Border Gateway Protocol (BGP). RFC 1163. https://doi.org/10.17487/RFC1163 https://www.rfc-editor.org/info/rfc1163.

[26] MaxMind, Inc. 2025. MaxMind: Industry Leading IP Geolocation and Online Fraud Prevention. https://www.maxmind.com/ Accessed: 2025-05-15.

[27] Open Observatory of Network Interference (OONI). 2026. Collateral Damage of IP-Based Blocking During LALIGA Football Streaming in Spain: Evidence from OONI Measurements. https://ooni.org/post/2026-laliga-collateral/. Accessed: 2026-08-01.

[28] Jon Postel. 1981. Internet Protocol. RFC 791. https://www.rfc-editor.org/info/rfc791.
[29] Andreas Proschofsky. 2022. Überzogene Netzsperre sorgt für Probleme im Österreichischen Internet. Der Standard Austria. https://www.derstandard.at/story/2000138619757/ueberzogene-netzsperre-sorgt-fuer-probleme-im-oesterreichischen-internet.
[30] Sivaramakrishnan Ramanathan, Anushah Hossain, Jelena Mirkovic, Minlan Yu, and Sadia Afroz. 2020. Quantifying the Impact of Blocklisting in the Age of Address Reuse. In *Proceedings of the ACM Internet Measurement Conference* (Virtual Event, USA) *(IMC '20)*. Association for Computing Machinery, New York, NY, USA, 360–369. https://doi.org/10.1145/3419394.3423657
[31] Yakov Rekhter, Jon Crowcroft, and Brian E. Carpenter. 1997. IPv4 Address Behaviour Today. RFC 2101. https://doi.org/10.17487/RFC2101 https://www.rfc-editor.org/info/rfc2101.
[32] 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 (IMC)* (Nice, France) *(IMC '22)*. 410–418. https://doi.org/10.1145/3517745.3561452
[33] Philipp Richter, Georgios Smaragdakis, David Plonka, and Arthur Berger. 2016. Beyond Counting: New Perspectives on the Active IPv4 Address Space. In *Proceedings of the 2016 Internet Measurement Conference* (Santa Monica, California, USA) *(IMC '16)*. Association for Computing Machinery, New York, NY, USA, 135–149. https://doi.org/10.1145/2987443.2987473
[34] Philipp Richter, Florian Wohlfart, Narseo Vallina-Rodriguez, Mark Allman, Randy Bush, Anja Feldmann, Christian Kreibich, Nicholas Weaver, and Vern Paxson. 2016. A multi-perspective analysis of carrier-grade NAT deployment. In *ACM Internet Measurement Conference*.
[35] Kimberly Ruth, Aurore Fass, Jonathan Azose, Mark Pearson, Emma Thomas, Caitlin Sadowski, and Zakir Durumeric. 2022. A world wide view of browsing the world wide web. In *ACM Internet Measurement Conference*.
[36] Kimberly Ruth, Deepak Kumar, Brandon Wang, Luke Valenta, and Zakir Durumeric. 2022. Toppling top lists: Evaluating the accuracy of popular website lists. In *ACM Internet Measurement Conference*.
[37] Jerome H. Saltzer. 1993. On the Naming and Binding of Network Destinations. RFC 1498. https://doi.org/10.17487/RFC1498 https://www.rfc-editor.org/info/rfc1498.
[38] John F. Shoch. 1978. Inter-Network Naming, Addressing, and Routing. Internet Experiment Note 19.
[39] Alissa Starzak and Marwan Fayed. [n. d.]. The unintended consequences of blocking IP addresses. https://blog.cloudflare.com/consequences-of-ip-blocking/.
[40] Bill Toulas. 2022. Pirate sites ban in Austria took down Cloudflare CDNs by mistake. Bleeping Computer. https://www.bleepingcomputer.com/news/security/pirate-sites-ban-in-austria-took-down-cloudflare-cdns-by-mistake/.
[41] Roland van Rijswijk-Deij, Mattijs Jonker, and Anna Sperotto. [n. d.]. The OpenINTEL Open Access Portal. ([n. d.]).
[42] Benjamin VanderSloot, Johanna Amann, Matthew Bernhard, Zakir Durumeric, Michael Bailey, and J Alex Halderman. 2016. Towards a complete view of the certificate ecosystem. In *ACM Internet Measurement Conference*.
[43] W3Techs. 2025. Usage statistics and market share of Cloudflare. https://w3techs.com/technologies/details/cn-cloudflare. Accessed: 2025-05-14.