Detecting and Characterizing Massively Shared IP Addresses

AMANDA HSU*, University of Chicago, USA
PAUL PEARCE, University of California, Irvine, USA
FRANK LI, Georgia Institute of Technology, USA
ARTHUR BERGER†, Akamai, Massachusetts Institute of Technology, USA
PHILIPP RICHTER, Akamai, USA

IP addresses are commonly shared across devices and users for a variety of reasons, including NAT and proxies. These technologies operate at different scales, from residential NATs that share an IP address across devices in a home to large-scale Carrier Grade NATs that share hundreds or thousands of users on a single IP. Cases of large-scale IP sharing are distinct as they have significant implications for IP-based mechanisms such as attribution, blocklisting, and rate-limiting, where the consequences of mishandling affect a large quantity of end-users and organizations.

In this work, we detect and characterize IP addresses shared at large scales, which we coin massively shared. Leveraging diurnal patterns in traffic shape, we use data from a large CDN to characterize these IPs globally. We broadly find that massive IP sharing is responsible for a large fraction of IPv4 traffic, concentrated in a small fraction of address space, with over 40% of total traffic coming from less than 2% of active IP addresses. We observe distinct patterns in deployment geographically, with particularly high rates of massively shared traffic from some smaller countries. Comparatively, in IPv6, we find far fewer massively shared addresses with some surprising exceptions among mobile providers. We additionally contextualize these addresses by other network characteristics, including identifying cellular connectivity and dual-stack capabilities, and identifying several instances of massively shared IPs in proxy services hosted on cloud networks. Finally, we find that rates of massively shared traffic are increasing over time, predicting future reliance on these technologies. Our work contextualizes the state of IP sharing, providing a uniquely broad perspective globally.

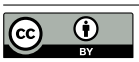
CCS Concepts: • **Networks → Naming and addressing**; **Network layer protocols**; **Network measurement**.

Additional Key Words and Phrases: Network measurement; IP addressing; Shared IP Addresses

ACM Reference Format:
Amanda Hsu, Paul Pearce, Frank Li, Arthur Berger, and Philipp Richter. 2026. Detecting and Characterizing Massively Shared IP Addresses. *Proc. ACM Netw.* 4, CoNEXT4, Article 50 (December 2026), 24 pages. https://doi.org/10.1145/3856961

*This work was conducted while the author was a Ph.D. student at Georgia Institute of Technology.
†Our colleague, mentor, and friend, Arthur, passed away on July 12, 2026. He reviewed the content of this paper prior to his passing.

Authors' Contact Information: Amanda Hsu, amandahsu@uchicago.edu, University of Chicago, USA; Paul Pearce, pearce@gatech.edu, University of California, Irvine, USA; Frank Li, frankli@gatech.edu, Georgia Institute of Technology, USA; Arthur Berger, awberger@akamai.com, Akamai, Massachusetts Institute of Technology, USA; Philipp Richter, prichter@akamai.com, Akamai, USA.



This work is licensed under a Creative Commons Attribution 4.0 International License.
© 2026 Copyright held by the owner/author(s).
ACM 2834-5509/2026/12-ART50
https://doi.org/10.1145/3856961

1 Introduction

IP addresses are commonly shared across multiple devices and users, frequently due to IPv4 address exhaustion. In ISPs, this has included Network Address Translation (NAT) and Carrier-Grade NAT (CGN). However, IP sharing extends to other scenarios, including VPNs, proxies, enterprise gateways, and other browse-out systems, whether for anonymity, security, or facilitating access to geo-restricted content. While such IP sharing supports a variety of functional and operational purposes, it also introduces problems, such as reducing the effectiveness of IP-based attribution, blocklisting, rate limiting, and geolocation, and broadly increases the complexity of network topologies [60, 62]. Furthermore, widespread CGN use can reinforce IPv4 dependence, slow IPv6 adoption, and add operational complexity, with uneven effects on regions with limited IPv4 resources.

While many IPs are shared amongst a small number of users or devices (e.g., a NAT within a residential network or small business), some are shared at large scales (e.g., hundreds, thousands, or more users). These cases of large-scale sharing are particularly important to identify and handle distinctly, especially with IP-based mechanisms and policies, as they can exhibit unique behavior and the consequences of IP-based actions can be substantially different. As a simple example, IP rate limiting must account for the large number of users behind such addresses when thresholding traffic rates, and mishandled decisions can cause widespread collateral damage, interfering with many legitimate users and organizations. Identifying these IPs based on traffic volume alone is unreliable, as other highly active hosts, e.g., bots and web crawlers, also generate high traffic volumes, as can malicious hosts. Using transport- or application-layer attributes (e.g., port number distributions, User-Agent strings, or device identifiers [11, 20, 61]) is limited to only certain vantage points and offers mixed reliability, such as with minimal attribute diversity exhibited by proxies that only relay specific applications or protocols [3, 15, 52]. Overall, the manifestation of address sharing can differ substantially depending on the underlying technology or deployment mechanism.

Although prior work has studied the extent of general IP address sharing on the Internet [38, 42, 43, 62], no work has characterized the phenomenon of these particularly highly shared IP addresses across IPv4 and IPv6 networks at a global scale. Existing detection methods [38, 42, 43, 62] rely on client-based measurements or properties of Peer-to-Peer traffic, yielding valuable insights but lacking the global coverage needed to characterize address sharing at scale.

In this work, we present a lightweight and more widely applicable approach to detect IPs that are shared at large scales across many users, which we call **massively shared IPs**. Our approach builds on an intuitive observation that the temporal properties of network traffic volumes from such addresses become increasingly predictable once they reflect the aggregated activity of a sufficiently large population of end users and devices. In particular, traffic from an address shared among many subscribers smooths out individual usage patterns, producing organic diurnal patterns. Building on this intuition, we present a method to formally detect these characteristics using Fourier analysis and curve-fitting evaluation. Since our method only requires traffic volumes per IP over time, it is lightweight and does not rely on any additional client-side information such as accessed content, packet-level specifics, or device identifiers that are otherwise used to identify sharing [20]. Thus, it can be used by a wide set of stakeholders in network operations, such as transit Internet providers, those with encrypted traffic or flow data, and samples from high-volume links.

We deploy our method across access logs of a major global Content Delivery Network (CDN), providing an unprecedented characterization of massively shared IP addressing across IPv4 and IPv6, spanning geographic regions, access networks, and cloud environments.

Our key contributions include:

(1) We develop a method to identify massively shared IP addresses based on diurnal traffic and smoothness, informed by Fourier analysis of empirical patterns, and evaluate it on both synthesized and real-world shared traffic. We evaluate and characterize our method, including heuristics to estimate how many subscribers are behind massively shared IPs, proper parameters, and suitable timeframes.
(2) We present a large-scale characterization of client IPv4 massive sharing on the Internet, quantifying it globally and regionally. We find that while only 1-3% of active IPs are massively shared, they originate between 25-54% of all IPv4 traffic, depending on the country. This demonstrates a stark concentration of traffic from massively shared IPv4 addresses.
(3) Broadly, we find that 82% of all massively shared traffic originates from end-user ISPs. When we observe massive sharing in IPv4, we ask whether the networks also have IPv6 connectivity, finding that only 55% of top providers with massively shared IPv4 deploy IPv6. We find that IPv6 sharing in end-user ISPs is rare, but sometimes deployed in cellular setups.
(4) Notably, we find that 14% of massively shared traffic comes from cloud providers. We find that massively shared cloud traffic is largely driven by proxy services [3, 14].
(5) Finally, we evaluate massively shared traffic over time, finding that the fraction of traffic from massively shared IPv4 IPs steadily increases.

2 On Shared Address Detection

Related Work. Although no work has considered the traffic quantity of shared IPs, prior work has leveraged traffic features to identify and quantify users behind home NAT deployments, including using IP TTLs, HTTP User-Agent strings [40], BitTorrent client identifiers [38], and BitTorrent client identifiers and RIPE Atlas logs [60]. Others have used IP leakage from BitTorrent [62] and on-client measurements to identify and characterize NAT deployments [42, 43, 62]. Others have used diversity in User-Agent strings [11] and source ports [13] to identify shared IP addresses. A CDN has extended this to identify CGN using ML classifiers over a variety of unspecified signals, including client diversity and traffic volume, then validated using PTR and WHOIS records that contain hints about address deployment [20].

Most similarly, Giostas et al. identified IPs with large numbers of users behind them (denoted "large multi-user IPs") [21] from the request logs of a large CDN. While this work has similar goals, the detection method leverages a multi-class classification model of IPs and their prefixes, then reproduces their labeling with public speed tests, and, finally, focuses its Internet-wide analysis specifically on CGN deployment. Our method only relies on diurnal patterns in request volume, and we analyze VPNs, proxies, and other shared technologies at Internet scale.

Challenges in Sharing Detection. Detection of massive IP sharing remains challenging because the necessary information about address sharing is rarely exposed, especially at scale. Approaches that rely on transport- or application-layer attributes to identify sharing require vantage points with such data available. While this data may be available to individual web platforms, transit providers cannot observe these attributes if they are encrypted. However, smaller web platforms cannot observe broad attributes such as port distributions, whereas an ISP or transit provider could. Moreover, different types of IP sharing setups may or may not result in detectable evidence. For example, proxy services such as Apple Private Relay only work for select applications [3, 15, 52], so it may not result in a large number of distinct User-Agent strings [13]. Carrier-Grade NAT setups can use a plethora of configurations [62], making detection based on transport-layer attributes, such as source ports, only partially effective. Thus, methods relying on these attributes can only observe a subset of IP sharing depending on the technology behind it. Finally, as we focus on massively shared IPs, one may assume that particularly high volumes of traffic (e.g., client requests)

may be sufficient to detect them. However, high volume alone cannot discriminate high numbers of end users from automated network traffic, such as crawling or infrastructure, as we will study in Section 8. Although an RFC published in May 2026 proposes a way for operators to report shared prefixes [19], it is not clear if and how widely it will be adopted.

Overcoming Challenges Towards Global Characterization. We overcome these challenges by detecting the traffic shape produced by *many* users sharing an IP address: smooth and diurnal request patterns. Our method only requires traffic volume over time and is agnostic to the actual implementations or technologies used to share IP addresses across users. We elaborate on the specifics of our method in Section 4. Indeed, while our method is limited by user behavior and measurement vantage, other methods are limited by the technology responsible for IP sharing. As we will show, the CDN access logs provide us with a global view, allowing us to characterize massively shared IPs, agnostic of the technology behind them.

3 Data Sources and Processing Steps

In this section, we introduce our vantage point and the corresponding datasets we use.

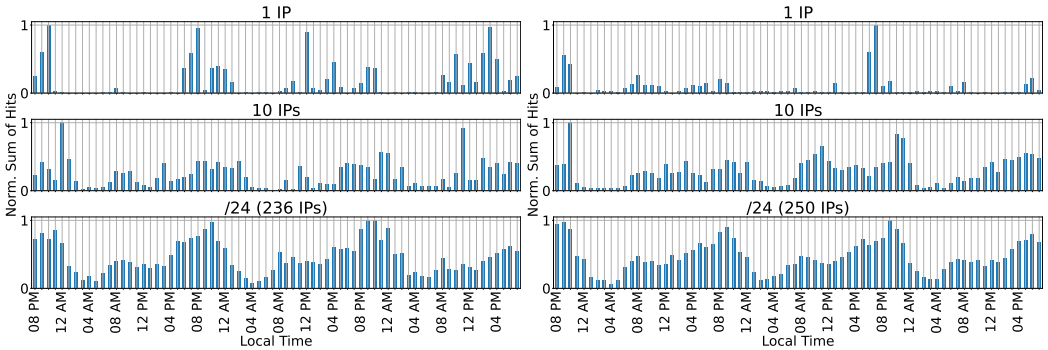
3.1 Primary CDN Log Data

Hourly Request Counts. The primary data source for our method and analysis is server logs of one of the largest global CDNs. In 2026, this CDN serves end-users globally by operating over 350,000 servers in over 1,600 ASes in over 130 countries. Each time a client fetches a Web object from a CDN edge server, a log entry is created, which is then aggregated through a distributed data collection framework. We have access to the exact number of requests issued per hour from individual IPv4 addresses and for IPv6 /64s. We note that in IPv6, an entire /64 is a commonly used assignment size [10]. We observe hourly request counts, or "hits", for IPv4 addresses and IPv6 /64s. We additionally use the term "traffic" to refer to all requests that the CDN receives from an IP or /64. Request counts include cached hits and all content types, including Web, video, and API calls. We refer to an IP as "active" if the IPv4 address or IPv6 /64 has at least one successful request during the analysis period. For the bulk of our analysis, we focus on the week of July 13th, 2025. Later, we show long-term trends from January 2024 to October 2025.

3.2 Contextual Datasets

Identifying Cellular Clients. We contextualize our dataset using the CDN's Real User Monitoring (RUM) system to measure whether the client is using a cellular or WiFi connection. The RUM system leverages the Network Information API [69] implemented in Android's browser, among others, to report this data. We note that the primary data described in Section 3.1 is not limited to certain devices and covers all the requests the CDN receives. We follow prior work [64] to measure and classify prefixes (/24s in IPv4, /48s in IPv6) as cellular or non-cellular, including their treatment of mixed-access prefixes and the use of a 0.5 cellular-to-WiFi measurement ratio threshold for cellular classification. Prior work validated that this threshold was shown to provide a conservative estimate of cellular prefixes [64]. Furthermore, to ensure representative data per prefix, we only use prefixes where we have at least 10 measurements. In total, we use 112M measurements to label some 2.5M IPv4 /24s and 1.4M IPv6 /48s address blocks. We do not observe traffic for the remaining IP address space, indicating that other prefixes are inactive or unobservable from our vantage. As we will show, this method provides coverage for over >90% of massively shared IPv4 IPs that we identify.

Measuring IPv6 Connectivity for IPv4 Networks. Similar to prior work [53], for dual-stack clients, we record their prefixes in each IP version (/24s in IPv4 and /48s in IPv6). When a client



(a) US broadband ISP A.

(b) US broadband ISP B.

Fig. 1. Traffic over 3 weekdays (72 hours) for different IP aggregate sizes within US ISPs A and B, known to deploy single-subscriber IPv4. In each plot, we sum and normalize the number of hits from all IPs.

makes a request to a dual-stack domain hosted by the CDN, the server records the current client prefix (IPv6 /48 or IPv4 /24, but typically IPv6 because of happy eyeballs [65]), then launches an additional Web request towards a domain that only supports IPv4 and records the /24. We use this to tag some 1.7M IPv4 /24s as also having IPv6 connectivity to contextualize deployments.

User-Agent data. We additionally log the User-Agent (UA) string present in a random sample of 0.1% of client requests. We use this data to contextualize client traffic patterns in Section 8.

Geolocating Data. We geolocate all client IP addresses using the proprietary CDN geolocation database to country-level granularity. We point out that the Service Level Agreement (SLA) of the geolocation database covers 99% country-level accuracy and point out that country-level geolocation for clients has been found to be consistent across different methods and datasets [27, 57].

Labeling AS Types. Throughout our analysis, we label ASN types primarily using data from IPInfo [30] to classify ASNs into 5 categories: ISP, hosting, education, business, and government. To verify our labels, we manually review the labels of the top 10 providers in each of the top 10 countries per continent. If necessary, we change the labels by consulting with other data sources [7, 55] and looking at the websites of the company owning the ASN. Furthermore, we add labels for unlabeled ASNs until we cover over 99% of traffic per continent.

4 Method

In this section, we present our method to detect *diurnal* and *smooth* patterns in client requests. We leverage the Fast Fourier Transform (FFT) to decompose a time series of requests per hour into frequency-domain components. We first show a motivating example, review related work on detecting diurnal network characteristics, and then introduce our method and parameter choices. We include a statement on ethics in Appendix A.

4.1 Motivating Example: Synthesizing Massively Shared Traffic

We first provide a motivating example to highlight the desired attributes our method detects. We exemplify what shared traffic looks like by *synthesizing* what traffic from many subscribers looks like. To do this, we inspect traffic patterns of address blocks that connect end-users to the Internet and are known to assign a unique IPv4 IP per subscriber [4, 8]. That is, these IPs may be shared locally, i.e., within a household via NAT; however, they are not *massively* shared, e.g., as opposed to a CGN IP. We then show how the shape of requests changes when aggregating the traffic from

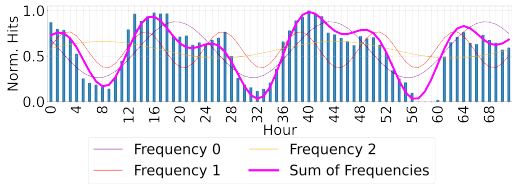


Fig. 2. To illustrate how we use the FFT output to reconstruct the traffic pattern, we plot the top 3 strongest frequencies (centered at the mean of the hit volume to visually highlight how the curve fits the data) and their sum for the traffic from an IP in a proxy service. Each sinusoidal function captures a periodic behavior in the traffic, and when added together, they fit the curve.

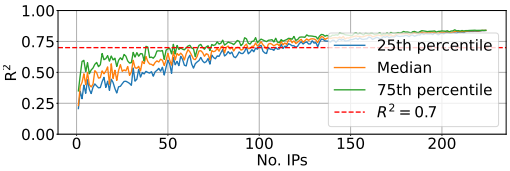


Fig. 3. We experiment with traffic aggregates in a /24 from US ISP A. Starting with 1, we randomly choose N IPs 25 times, reconstruct the traffic with the top 3 frequencies from the FFT, and compute the R^2 value between the reconstruction and the data. We plot the R^2 values of each quartile for each experiment with N IPs. Broadly, the R^2 value (smoothness) increases as we add more IPs (aggregate more subscribers).

individual IP addresses together. To illustrate why this is a strong indication of sharedness, we present motivating examples from two large US broadband providers. We collect these address blocks by examining probe tags on RIPE Atlas [63] and by manual collection from colleagues who use these providers. Figure 1 shows examples of traffic patterns from IPs in two IPv4 /24 networks. In traffic patterns of individual IPv4 addresses, while some diurnal patterns emerge (i.e., no traffic late at night), we see high variability. Aggregating traffic from 10 IPs within the same /24 starts to provide more diurnal visual clues, with peaks and dips in traffic at logical times. Finally, we see that predictable patterns emerge when we look at traffic from the entire /24.

Our method is centered around systematically detecting these resultant *diurnal* and *smooth* traffic patterns across traffic from individual IP addresses to detect massive sharing. That is, on top of strong daily patterns, we seek traffic that is also predictable (i.e., less variable). In the next sections, we will formalize our method, which allows us to detect these patterns at scale.

4.2 Related Work: Detecting Diurnality

Prior work has used FFT analysis to determine diurnality, specifically by analyzing whether 1/24 is the strongest frequency and its energy [5, 59]. However, this definition was used in an analysis of the number of online IPs in a network in a given time period, aligning with common patterns in IP churn (i.e., 24 hours). Because our data is based on traffic from IPs and therefore subject to different patterns and trends, we build upon this definition in Section 4. Other work has detected diurnal patterns in IPv6 traffic over entire ISPs [66], or identified diurnal IPv6 patterns over all traffic generated by individual households [67]. However, we highlight that the authors measured 5 residences and noted high variation in traffic patterns. Others have used diurnal patterns in traffic as an indication that traffic is not spoofed or part of an attack [37], or more broadly to forecast backbone traffic [54] and anomaly detection [6].

4.3 Detection Method and Criteria

Fourier Analysis to Identify Diurnal Traffic. Following our observation of diurnal traffic shapes in Section 4.1 and prior work [5, 59], we use Fourier Analysis to quantify diurnality in traffic patterns. We perform a Fast Fourier Transform (FFT) on a time series of the number of "hits" per hour to define and reconstruct the periodic signal. Similarly to prior work, we consider traffic diurnal if 1/24 (a 24-hour period) is the frequency with the most energy.

Modeling Smoothness. Based on our observations, we introduce one more criterion: smoothness. We want to identify traffic that follows a relatively smooth sinusoidal signal without much randomness. An FFT decomposes a time series into sinusoidal curves with different strengths and frequencies. When all these are added together, the signal is reconstructed and, for example, would match the traffic pattern of the single IP in Figure 1. In general, when fewer frequencies are used, the match to the original signal degrades. Our idea is that the notion of *smoothness* is when the match to the original signal remains good even when just a few of the strongest frequencies are used. That is, if only a few sinusoidal signals can reconstruct the shape of the traffic, it must be smooth, as it is not possible to reconstruct a chaotic shape with so few frequencies.

After experimentation, we choose to use just the 3 strongest frequencies (plus the DC offset, the y-axis offset added to the reconstruction) to reconstruct the signal. We show the three top frequencies and their addition to reconstruct traffic from a shared IP using a proxy service as an example in Figure 2 as a visual example of how the FFT deconstructs the shape of the curve. Choosing this parameter is a trade-off: we potentially have more false negatives but we have higher confidence that the IPs we do identify are massively shared.

Evaluating Curve Fit. Then, we measure how well the reconstructed signal fits the time series by calculating the R^2 (also known as the coefficient of determination) as a measure of the variance captured and goodness of fit for our curve [12, 70]. The R^2 value is between 0 and 1, with higher values indicating a better fit, i.e., better *smoothness*.

To evaluate how the curve fits vary, we repeatedly aggregate a random set of N IPs from a /24 in US ISP A and calculate the R^2 value from the reconstruction. We perform this from N=1 to the number of active IPs in the /24 (235 total) and present our findings in Figure 3. As we aggregate traffic from more IPs, R^2 increases and shows less variability, indicating a smoother fit and more uniform behavior for larger aggregates of subscribers. In this work, we choose our threshold for R^2 at 0.7, which we find to be consistently exceeded once 75 subscribers/IPs are aggregated. We also note that 0.7 is a commonly accepted value for a good curve fit [25]. We study the sensitivity of our choice in Appendix B.

Criteria for Massively Shared IP addresses. In this work, we label an IPv4 address, or an IPv6 /64, as massively shared if its respective request pattern meets two criteria: *(i)* the strongest frequency is 1/24, and *(ii)* when using the 3 strongest frequencies to reconstruct the signal using an FFT, the R^2 value between the data and the reconstruction is greater than 0.7. This definition results in a binary classification of IPs as either massively shared or non-massively shared. Practically speaking, we label an IP address as massively shared if the hourly request pattern reaches a strongly diurnal pattern with a degree of smoothness that matches our synthesized traces.

We evaluate how our findings change alongside different R^2 parameter values longitudinally, finding that results do change slightly with different values, but traffic quantities are consistent over time. We discuss the R^2 parameter sensitivity in more detail in Appendix B.

Implementation: Time Period. We apply our method to 3 days of data per IP on Tuesday, Wednesday, and Thursday to avoid differences in diurnal traffic patterns due to weekend behavior (e.g., for business networks) and differences in days of the week considered weekends globally. Furthermore, we find that including weekend traffic is more likely to create a traffic shape with a strongest frequency of 1/7, not 1/24, due to daily patterns changing on weekend days. Thus, this would reduce our ability to detect consistent daily patterns rather than patterns across an entire week of data. We elaborate on this choice in Appendix C. We acknowledge that identifying differences in traffic on weekends is interesting in itself, and discuss it along with other future directions in Section 10. Finally, we also acknowledge that IP churn in our analysis period could result in a poor fit. We discuss this and other limitations in Section 5.4.

5 Method Evaluation

In this section, we initially contextualize the IPs our method identifies. We estimate the number of subscribers behind the IPs, compare our identified massively shared IPs with relative traffic volumes, and compare them to the IPs of several specific known cases. We clarify that we contextualize our broader characterization at a global scale in Sections 6, 7, 8, and 9.

5.1 Estimating Usage Behind Massively Shared IPs

To estimate how many individual subscribers are behind a massively shared IP, we aggregate traffic from individual IPs and analyze how many it takes to reach our massively shared criteria, as presented in Figure 3. While we cannot claim to precisely quantify the number of users behind a massively shared IP, we use hits aggregated from individual subscribers as a heuristic to estimate the user scale at which traffic begins to satisfy our massively shared criteria. As shown in Figure 3, in a US ISP, we find that our criteria for massively shared are consistently reached around 75 single-subscriber IPs. We repeat this experiment with other /24s in this US ISP and other ISPs with single-subscriber IPv4 IPs in Germany, France, Japan, and Iran. We find that we reach the massively shared criteria after aggregating traffic from 45-160 IPs. We discuss these analyses in more detail in Appendix D. These values vary with CDN presence, browsing activity, and other network infrastructure, but indicate that the IPs we label are generally shared by *at least* 50-100 subscribers. We emphasize that this number is a lower bound, as after this number is reached, synthesized traffic consistently meets our criteria for massively shared. In Section 5.2, we further characterize the IPs we are able to detect relative to the country's total traffic volume. In these experiments, we use IPv4 /24 subnets from ISPs that publicly document not using CGN or use the subnets of RIPE Atlas probes with "native IPv4" tags.

5.2 Massively Shared Addresses and Traffic Volume

We next compare massively shared IPs with their traffic concentration per country. Specifically, in Figure 4, we analyze the traffic from the top 10% of IPs (ranked by volume of traffic sent). In 85% of countries, the top 10% of IPs account for over 90% of traffic from massively shared IPs, showing that these IPs are typically high-volume. However, high volume alone is not sufficient for massively shared IP detection: in only 17% of countries do massively shared IPs account for over 90% of traffic from the top 10% of IPs. That is, high-volume IPs can also correspond to infrastructure (e.g., API calls), crawlers, and unusually high-usage individuals. We emphasize the importance of this distinction; the top 10% of IPs account for over 60% of all traffic in 77% of countries, making policies that target high-volume IPs broadly impactful.

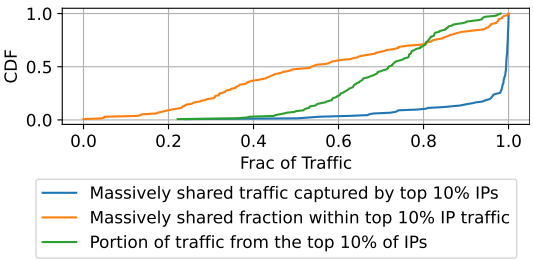


Fig. 4. The traffic concentration per country and its intersection with the massively shared IPs.

5.3 Validating with Known Cases

In this section, we showcase examples of different cases that we expect to be massively shared or not massively shared. While we do have expectations for these technologies broadly, we have no ground truth for massively shared IPs, especially which addresses will be massively shared at a

given point in time. Thus, we present these examples to characterize their deployment and our method. We discuss additional manual validation of individual technologies in Appendix E.

Cellular Networks in Mobile-Reliant Countries. We analyze three African countries where we observe high rates of mobile reliance and IPv4 scarcity, implying massive sharing [41, 49]: Mali, Algeria, and Ethiopia [1]. Both of these characteristics lead us to expect that cellular IPs would be massively shared due to high reliance and low numbers of IPs. Using the cellular identification methodology described in Section 3, we cross-check cellular IPs with massively shared IPs. We find that in Mali, 100% of cellular traffic is from massively shared IPs. In Algeria and Ethiopia, deployment varies slightly; 87% and 76% of cellular traffic is from massively shared IPs, respectively. Thus, our findings align with our expectations that high-traffic cellular IPs in these settings are shared by many users. We explore the role of cellular connectivity in more detail in Section 7.1.

Anonymization Proxy Services. We use publicly available geofeeds to study the global networks of two proxy services: Proxy 1 and Proxy 2. Proxy 1 is a secure service for businesses. Proxy 2 can be used on mobile devices or computers but only works on a few applications and requires a paid subscription. In each case, we analyze these deployments in countries where these proxies appear to have high usage, making it likely that they are massively shared. The geofeeds for these proxy services contain thousands of prefixes in both IPv4 and IPv6. Thus, for this exercise, we count the proxy as highly used in a country if over 100 proxy IPs are in the top 10% of IPs for that country, indicating they are indeed heavily used relative to the country's overall traffic.

For Proxy 1, we observe that traffic is highly concentrated in massively shared IPs. In 79% of countries with high usage, over 99% of proxy traffic comes from massively shared IPs, and in all countries with high usage, over 90% of proxy traffic comes from massively shared IPs. We hypothesize that most IPs from this service are massively shared because entire business networks use this service for all or nearly all of their traffic, resulting in high numbers of users. Proxy 2 aligns with massively shared IPs, although it varies more per country. In Proxy 2, in 75% of countries with high usage, over 70% of proxy traffic comes from massively shared IPs. Thus, although this service sends a high relative traffic volume per country, they are not necessarily always used by a high volume of users. Because this service requires a subscription and only works on certain applications, we hypothesize that actual usage and adoption vary. Thus, in both proxies, our expectation that their networks broadly encompass massively shared IPs is confirmed.

IPv4 Deployments That Are Not Massively Shared. To study false positives "in the wild", we evaluate against two /16s (with over 40,000 active IPs each) from single-subscriber IPv4 deployments in ISP A. Less than 1-2% of IPs are labeled as massively shared. We additionally sample from 100 RIPE Atlas [63] probes across 30 countries tagged as native IPv4 and active. Our method does not detect any of these IPs as massively shared. Expanding our analysis to each IP in the /24 of the probes (over 11,000 IPs total), 2% are massively shared. We emphasize that, although we do not have ground truth for these networks (some IPs may indeed have a multitude of users, e.g., in a business network or public WiFi, or improperly labeled probes), this result lends itself to the conservative nature of our method. That is, we favor false negatives over false positives.

5.4 Limitations

We are foremost limited to the CDN's vantage and usage regionally. While we tested our method empirically in multiple countries, our data may be biased towards certain types of traffic. We may hence underestimate massively shared traffic in regions with relatively fewer customers or users,

[1]We infer that these countries are reliant on mobile connectivity due to reports of high cellular usage and, in some cases, low reported fixed broadband subscriptions [16, 31–33].

resulting in traffic shapes that do not aggregate smoothly. In such cases, higher degrees of address sharing may be required to meet our criteria, reducing the reliability of our method.

Moreover, we acknowledge the lack of validation with ground-truth data on shared IPs and massively shared IPs. This does not allow us to provide evaluation metrics such as precision and recall for our method and tune our parameters (though we provide parameter sensitivity analysis in Appendix B). However, we highlight that our method is conservative by design. As discussed in Section 4.3, our method has trade-offs, but we purposely favor false negatives over false positives in order to report lower bounds on the global use of massively shared IPs and their traffic. In other scenarios, the parameters of our method may be tuned in the opposite direction as appropriate. In this work, we may not detect all massively shared IPs; for example, in Figure 3, we show that there is a chance that we do not detect traffic shared across high numbers of individual subscriber IPs. Another case that may impede our detection is IP churn in the middle of our analysis period, as mentioned in Section 4.3, which could lead to a poor fit and mislabeled IPs. However, this is dependent on specific network properties and may happen within *any* time period we choose. Thus, we instead optimize our time period for broad factors, such as weekends and workdays. In the opposite direction, we still may mislabel non-massively shared IPs if they exhibit the traffic characteristics we observe, but Section 5.3 supports that such cases are generally rare.

Finally, we cannot quantify users behind massively shared IPs precisely; we can only estimate subscriber counts with our heuristics to synthesize traffic, as patterns from individual users vary and are dependent on deployment characteristics. Thus, we have limited vantage as to what is *actually* behind massively shared IPs, but have strong reason to infer that this is many subscribers. Additionally, we acknowledge that we cannot generally associate massively shared IPs with specific types of services at scale, such as CGN, proxies, and VPNs. Therefore, we cannot evaluate the deployment of specific IP sharing technologies.

Defining "Massively Shared" IP Addresses:
We introduce the term massively shared IPs to describe IP addresses for which the hourly request counts we observe exhibit smooth, diurnal patterns. Intuitively, this pattern can only be observed when requests from many users are aggregated behind the same IP, since the traffic of a single user or household is shaped by individual browsing behavior (Section 4.1). We detect smooth and diurnal patterns using Fourier analysis and curve fitting metrics to formally detect 24-hour traffic patterns with low variability (Section 4.3). We estimate that this pattern is formed by traffic equivalent to at least 50-100 single-subscriber IPs (Section 5.1).

6 A Global View of Massively Shared IPs

In this section, we analyze the results of deploying our method over the entirety of the IPv4 and IPv6 traffic from the CDN's access logs.

Overall Prevalence. Globally, we find that 41.1% of IPv4 traffic (CDN client requests) comes from massively shared IPs, meaning that *close to half of today's IPv4 traffic served by a CDN flows via massively shared address infrastructure.* In terms of IPv4 addresses, we find that only 1.6% of active IPv4 addresses are massively shared, highlighting a heavily skewed distribution of traffic over the IPv4 space. In IPv6, we find that some 5.6% of IPv6 client traffic comes from massively shared /64s, and that these account for 0.35% of active /64s during our measurement window. We are surprised to find substantial (though much lower than in IPv4) address sharing in IPv6, since IPv6 massive sharing cannot be driven by address scarcity.

Regional Prevalence. Next, we present the fraction of massively shared traffic per continent in Figure 5. In IPv4, we observe that more than half of client traffic seen in Asia, Africa, and Oceania

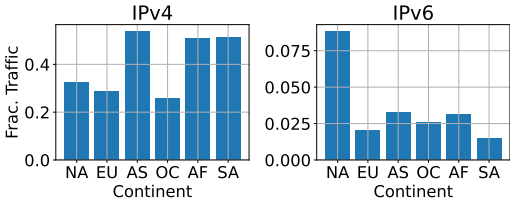


Fig. 5. The fraction of traffic from massively shared IPs per continent. Across all continents, IPv4 massive sharing is pervasive and particularly pronounced in Africa, Asia, and South America. IPv6 massive sharing, while present, is less widely deployed.

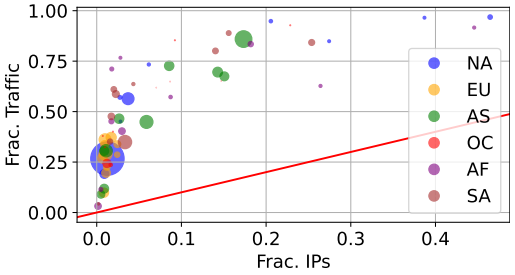


Fig. 6. Traffic-to-IP fraction ratio for massively shared IPs out of all active IPs and traffic per country. We show the top 10 countries with the most traffic per continent. The marker size is relative to the number of requests the CDN receives from the country.

comes from massively shared IP addresses, with massive sharing in Europe and North America slightly less prevalent at 29% and 32%, respectively. We point out that North America, especially the USA, has high volumes of IPv4 addresses, and Europe also has a large fraction of IPv4 addresses relative to population size [29], alleviating the need to share addresses widely. In IPv6, we observe that this is not the case, and IPv6 massive sharing represents a much smaller fraction of traffic in each case. Notably, we find that North America dominates massively shared IPv6 traffic (at 9%). Once again, we emphasize that while simply sharing IPv4 IPs is somewhat expected, our results highlight the scale at which a high number of users depend on massively shared addressing.

Country View of Traffic Concentration. Next, we analyze massively shared traffic at the country level. Although we already know from Section 5.2 that traffic per country is typically concentrated in a small fraction of active IPs, we now observe the magnitude of this concentration in massively shared IPs. Given that address scarcity, technological deployment, and usage differ across regions, we complement our metric of massively shared traffic fraction with the proportion of massively shared IP addresses among all active IPs within each country. Figure 6 shows this ratio for the top-10 traffic-ranked countries on each continent. In all cases, the fraction of massively shared traffic is significantly higher than the fraction of IPs. While all selected European countries, as well as the USA, exhibit a small fraction of massively shared IPs and shared traffic between 20% and 40%, many other nations show traffic ratios exceeding 50%, and in some cases even 80%. Specifically, in North America, traffic from Guatemala and Nicaragua is mostly from massively shared IPs that encompass high fractions of all active IPs. This ratio is 97% of traffic from 44% of IPs and 97% of traffic from 39% of IPs, respectively. Similarly, in Africa, 97% of traffic from Somalia originates from massively shared IPs that encompass 62% of all active IPs. With a high proportion of all active IPs being massively shared and sending nearly all

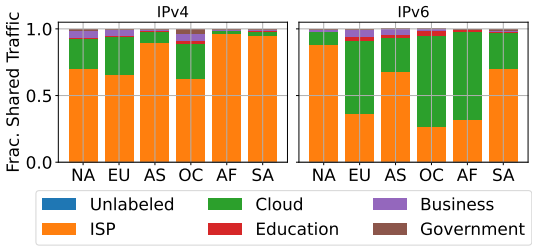


Fig. 7. The distribution of massively shared traffic among different AS types, per continent. Massively shared traffic is dominated by ISPs and cloud providers across all continents. Recall that the massively shared IPv6 traffic shown is a small fraction of total IPv6 traffic.

of the traffic for each country, these findings align with prior observations that limited Internet infrastructure and later access to IPv4 resources result in heavy reliance on CGN [62].

In IPv6, massively shared/64s produce far less traffic. We find that 88% of the countries we analyze have less than 3% of their traffic from massively shared IPv6 /64s. However, we highlight notable outliers. The United States has one major broadband and cellular provider that sends 47% of its traffic from massively shared IPs. Hong Kong has 43% of its traffic from massively shared IPs. Upon closer inspection, we observe that most of this traffic comes from cloud providers. We elaborate on this and other more granular observations on IPv6 sharing in Sections 7.3 and 8.

Traffic By AS Type. Next, we categorize ASNs by type and observe the amount of massively shared traffic the CDN receives from each. In Figure 7, we find that the bulk of massively shared traffic comes from End-User ISPs (82% of all massively shared traffic in IPv4, 80% in IPv6) and cloud providers (12% in IPv4, 16% in IPv6). While this is intuitive for End-User ISPs, which must connect many users over a limited address space, we emphasize the scale that our findings of massively shared IPs show. Moreover, we are surprised to observe a substantial portion of massively shared traffic coming from cloud networks, a finding examined further in Section 8.

Takeaway: Massively shared traffic compromises 41% of the CDN's IPv4 client requests, but less than 2% of active IPv4 addresses. Although present in all countries, massive sharing is particularly pronounced in areas with later Internet deployment. On the other hand, we find just over 5% of IPv6 traffic coming from massively shared /64s.

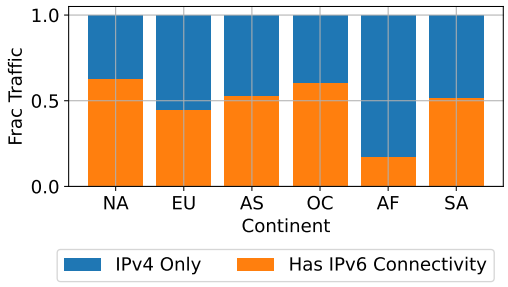
7 Massively Shared Addresses in End-User ISPs

As discussed in Section 6, most massively shared traffic originates from End-User ISPs. Next, we focus on ASes, their use of massive sharing in IPv4, the presence of IPv6 capabilities, and unexpected cases of massive sharing in IPv6.

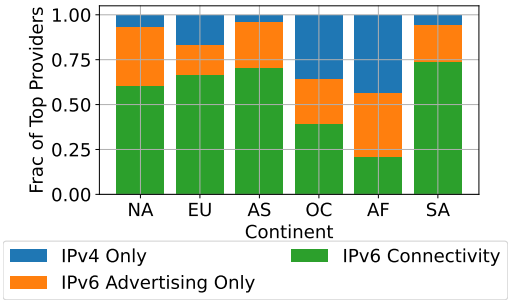
7.1 Massively Shared IPv4 in ISPs

End-User ISPs present a natural deployment setting for massively shared IPv4 addressing, given the need to connect large numbers of end users ("eyeballs"). As shown earlier in Figure 5, we find that Asia, Africa, and South America show particularly high levels of massively shared addressing deployments in IPv4. This trend directly extends to ISPs, where 54%, 51%, and 50% of all traffic originates from massively shared IPv4 addresses in Asia, Africa, and South America, respectively, compared to 29%, 23%, and 20% in North America, Europe, and Oceania.

Cellular Subnets and Massive Sharing. Prior work has shown that IPv4 address sharing is ubiquitous in cellular networks [62]. Next, we ask what fraction of massively shared IPv4 traffic in End-User ISPs is driven by cellular deployments, using context from the RUM measurements described in Section 3 to identify /24s that are cellular. We note that this dataset contextualizes over 90% of massively shared IPv4 IPs as either WiFi or cellular, providing wide coverage over our dataset. We find that the fraction of massively shared ISP traffic from cellular connections varies by continent, with 26% in North America, 52% in Europe, 36% in Asia, 10% in Oceania, 66% in Africa, and 20% in South America. Thus, cellular plays a significant role in massive IP sharing, but not all. Outside of Africa and Oceania, half or more of massively shared IPv4 traffic from ISPs is from non-cellular deployments, likely with CGN. However, in Africa and Oceania, 7/10 and 6/10 of the countries, respectively, with the most massively shared traffic have over 60% of their massively shared traffic from cellular IPs. In the most extreme cases, Ethiopia's IPv4 traffic is 92% massively shared, and 70% of that traffic is cellular; in Nigeria this is 83% and 92%, in Papua New Guinea this is 93% and 91%, and in Fiji this is 85% and 84%.



(a) The fraction of massively shared IPv4 ISP traffic from clients with IPv6 connectivity. Across continents, many clients lack IPv6 capabilities. Africa has exceptionally low deployment, where many who share IPv4 have no IPv6 connectivity.



(b) The IPv6 capabilities of the top-10 providers per country on each continent with the most massively shared IPv4 traffic.

Fig. 8. An overview of IPv6 capabilities of the networks with associated massively shared IPv4 traffic. We first provide an overview of IPv6 capabilities by amount of traffic, then by provider, including how many advertise any IPv6 BGP prefixes. We highlight that Africa has the highest fraction of IPv4-only providers, further characterizing this region as IPv4-dependent, as it also has the highest rate of massively shared IPv4.

Takeaway: Massively shared IPv4 is deployed widely by broadband and cellular providers. Although cellular is responsible for most massively shared IPv4 traffic in Africa, implying a broader dependence on this type of connectivity, in other continents, the majority of massively shared IPv4 traffic originates from non-cellular deployments.

7.2 Massively Shared IPv4 and IPv6 Connectivity

Next, we ask whether massively shared IPv4 deployments also have IPv6 connectivity (i.e., dual stack), indicating they have a path away from dependence on IPv4. Alternatively, standalone massively shared IPv4 deployments may prolong IPv4 use, increasing reliance on exhausted address space while leaving networks misaligned with growing global IPv6 adoption [23].

Measuring IPv6 Connectivity. For massively shared IPv4 addresses, we are interested in whether the respective clients also have IPv6 connectivity. For this, we use the IPv4-IPv6 association data described in Section 3. We label a massively shared IPv4 address as IPv6-capable if the respective /24 has an IPv6 association, labeling over 96% of massively shared IPv4 addresses.

Figure 8a shows the fraction of massively shared IPv4 ISP traffic from IP addresses that have IPv6 connectivity. [2] At a high level, we find that about half of the massively shared IPv4 traffic comes from IPs with IPv6 connectivity. Africa is the exception, with less than 20% of the massively shared traffic coming from dual-stack clients. Recall from Figure 5 that Africa also has the highest fraction of massively shared traffic overall. Our findings indicate that while massively shared IPv4 networks are commonly dual-stack [17], it is also often deployed as a standalone solution.

A Multi-Dimensional View on IPv6 Deployment. Next, we analyze the degree of IPv6 deployment we observe in massively shared IPv4 networks at the AS level. In Figure 8b, we categorize the top 10 ASes that send the most traffic in each country (in the top 10 countries per continent) into three categories: *(i)* IPv4 only, where we see massive sharing in IPv4 and no indication of IPv6; *(ii)*

[2]Note that several factors can affect whether objects get fetched via IPv4 or IPv6 from the CDN, including CDN mapping considerations, customer and client-specific configurations, and network conditions [65, 67].

IPv6 Connectivity: ASes where we see massive sharing in IPv4 and direct IPv6 associations; *(iii) IPv6 Advertising only*: ASes where we do not measure an IPv6 association, but we do see that the ASes announce *any* IPv6 prefix in BGP. We use the CAIDA Pfx2AS dataset, built from Routeviews BGP data, to identify ASes that originate IPv6 BGP prefixes [9].

AS-Level Deployment Strategies. As per Figure 8b, more than half of ISPs, in most regions of the world (Africa being the exception), do deploy massively shared IPv4 and IPv6 that is not massively shared. Specifically, 52% of all providers we analyze use this deployment when classified by traffic. Although we do not explicitly detect single-subscriber or native IP deployments, we highlight that a common dual-stack deployment when IPvX to IPvY transition mechanisms are in place [17, 44, 45] uses shared IPv4 and native IPv6. Our findings imply that these technologies *may* be deployed widely across ASes with massive sharing. We elaborate on massively shared IPv6 in Section 7.3.

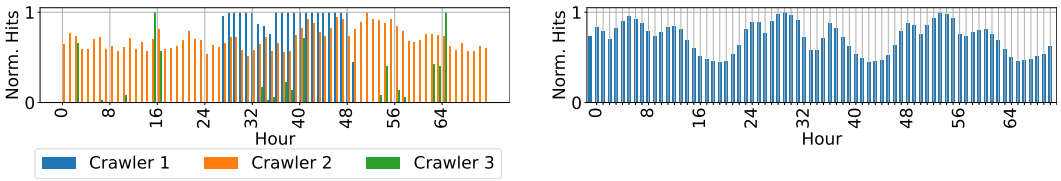
We highlight the importance of the differences between analyzing traffic and capabilities by provider. Although our results in Figure 8a largely align with other IPv6 deployment reports [2, 23], our per-provider results in Figure 8b paint a more complex picture. Looking at the top 10 providers per country for the top 10 countries per continent (100 total), IPv6 connectivity is commonly available at some point in the AS, but we commonly see in Figure 8a that a large quantity of massively shared IPv4 traffic has no IPv6 associations. This finding in large providers is impactful as they serve a significant part of the country's population, suggesting that IPv6 deployment metrics can overstate the extent to which end users behind massively shared IPv4 infrastructure have IPv6 connectivity. Although we focus on large providers here, where we expect more IPv6 deployment, we see 79% and 61% of top ISPs in Africa and Oceania use IPv4 only (by connectivity). Importantly, we also find a sizable number of ISPs globally that advertise at least one IPv6 BGP prefix, but no measurable IPv6 connectivity from massively shared IPv4 clients. That is, a provider may deploy IPv6 unevenly within the network, be making early attempts at configuring and setting up IPv6, advertise IPv6 as a part of an agreement with transit providers, or deployment may simply be a work in progress.

Takeaway: In 52% of providers and in 50% of traffic, dual-stack massively shared IPv4 clients have non-massively shared IPv6. We also find significant IPv4-only sharing deployments, particularly concentrated in Africa, where up to 79% of top providers have no IPv6 clients.

7.3 Massively Shared IPv6 in ISPs

While IPv4 shows widespread massively shared addressing, likely driven by scarcity, we unexpectedly observe that some ISPs also employ massively shared address mechanisms in IPv6. Looking at massive sharing in IPv6 ISPs, we find that North American (8% of total IPv6 traffic) and Asian ISPs (2% of total IPv6 traffic) are slightly higher than those on other continents.

Massively Shared IPv6 Cellular Traffic. Toward understanding IPv6 massive sharing, we analyze the fraction of cellular traffic among massively shared IPv6 traffic. In Europe, Africa, and South America, cellular traffic represents less than 3% of massively shared IPv6 traffic, and in Oceania, it is just 10%. However, in North America and Asia, cellular traffic represents the majority of massively shared IPv6 traffic (92% and 69%, respectively). In North America, we note again that this phenomenon is due to one large provider (over two sibling ASes) that appears to share /64 prefixes across multiple subscribers. Within each AS, over 98% of massively shared traffic comes from cellular subnets. We also highlight that this aligns with findings from prior work that analyzed end-user IPv6 addressing, which found similar behavior in the same ISP, finding *over 10,000 users in a single /112* [36]. In Asia, we also find that this is due to one large telecom provider in Indonesia, which is responsible for almost 70% of all massively shared IPv6 traffic in Asia.



(a) Traffic from bots on IPs that are not shared. These bots self-identify as web crawlers for search engines. We observe different traffic patterns, with some constantly crawling web services hosted by the CDN, and others only presenting sporadic traffic in bursts.

(b) Traffic from a popular AI search engine. Searches are initiated *on demand* when users query the AI. This traffic meets our massively shared criteria, where, although the end users are not (directly) behind the IP, the traffic is triggered by their actions.

Fig. 9. Examples of traffic patterns from cloud IPs, identified by their UA strings. The traffic from a user-initiated crawling bot meets our massively shared criteria, whereas the others present constant or bursty traffic patterns, likely driven by different crawling targets at the CDN.

Although uncommon, IPv6 massive sharing by large ISPs and cellular providers can generate substantial traffic. We hypothesize that providers may deploy it to maintain symmetry with IPv4 [26]. If such practices become widespread, IPv6 may inherit shared IPv4's challenges.

Takeaway: While massively shared IPv6 is not common in ISPs, we find surprising exceptions among large mobile providers who appear to share IPv6 /64s among multiple subscribers. If IPv6 sharing practices spread, known issues from the IPv4 sharing space may surface in IPv6 as well.

8 Massively Shared Addresses in the Cloud

Although we find that majority of massively shared traffic originates from End-User ISPs, we also find 14-16% of massively shared traffic (in IPv4 and IPv6, respectively) from cloud providers. We ask the question: What is this type of traffic, and are end users behind these IP addresses?

Showcasing Cloud Traffic. Before analyzing massively shared traffic from cloud ASes systematically, we discuss common traffic patterns that we see from cloud IP addresses. As shown in Figure 12 in Section 4 and discussed in Section 5.3, secure proxy services are often massively shared and thus exhibit expected smooth and diurnal patterns similar to those in ISPs. In Figure 9 we present two other examples: *(a)* Web crawlers of major search engines and *(b)* an interactive AI bot that accesses the Web on demand when users interact with the respective AI chatbot. We see that the Web crawlers, Fig. 9a, exhibit a variety of traffic patterns, none of which meet our criteria for massively shared IPs. However, we find instances of bots that do, such as the on-demand AI crawler shown in Figure 9b. We will discuss these cases in more depth later in Section 8.2.

8.1 Toward Classifying Massively Shared Cloud IPs

As both proxy-like services and some instances of bots meet our criterion for massively shared IPs, we dissect this set of cloud IPs with the goal of separating both cases.

User-Agent Strings to Identify Users and Bots. We use the HTTP User-Agent (UA) strings in requests from massively shared cloud IPs as an auxiliary data source to explore *why* this behavior manifests in cloud traffic. Recall that the User-Agent field identifies the respective client application (e.g., a Web browser version, an app, a crawler or bot [18, 24, 46, 51]).

Homogeneous vs. Diverse IPs. Prior work found that IPs presenting few (or just a single) UA strings are more likely to be bots [61], and a larger number of unique UA strings indicates end users

and the various applications they use to access the Web [34, 39, 40]. We build on this insight and tag IPs that issue requests with a single UA string as *homogeneous*, i.e., they come from bots, and IPs with multiple as *diverse*, i.e., they come from end-user proxies. We acknowledge that our separation is oversimplified, as bots may leverage multiple User-Agents, and cloud end-user gateways may be confined to single applications, leading proxy gateways to show up with a single UA. However, we present this as a first step toward teasing proxy-like IPs and bots apart. [3]

Among massively shared cloud IPv4 traffic, 81% of massively shared cloud traffic comes from *diverse* IPs and 19% comes from *homogeneous* IPs. In IPv6, the contrast is even starker, with 96% of massively shared cloud traffic coming from diverse /64s and only 4% from homogeneous /64s. We find similar proportions of IP addresses themselves, with 77% of massively shared IPv4 addresses and 84% of massively shared IPv6 /64s tagged as diverse.

8.2 Of Proxies and Bots in the Cloud

Diverse IPs – Likely End-Users. First, analyze the cloud IPs that meet our massively shared criteria and present multiple UA strings. Studying the ASes of massively shared cloud IPs, we find that more than 70% of all shared IPv4 traffic originates from 10 cloud providers, including hyperscalers, multi-tenants, CDNs, and cloud security companies. The top cloud AS (a hyperscaler) accounts for 17% of the *diverse* massively shared IPv4 traffic from cloud providers. The second, a cloud security company, originates 15% of this traffic and advertises secure web gateway services.

In IPv6, we see even higher levels of concentration, with 94% of all massively shared traffic coming from the top 10 cloud ASes. The top two providers with the most activity are large cloud providers that offer a multitude of services, including proxy services, Web hosting, and cybersecurity products. The top provider (a large CDN and cloud provider) is responsible for 37% of all massively shared cloud IPv6 traffic. Upon closer inspection, we find that all of the IPs from this provider are from a popular anonymization proxy service that publicly identifies its IPs through a geofeed.

While not exhaustive, our analysis shows that a large amount of end-user massively shared cloud traffic stems from proxy infrastructures. Importantly, these massively shared deployments (in both IPv4 and IPv6) are likely also driven by anonymity, operational convenience, and business models centered on centralized traffic mediation and control.

Homogeneous IPs – Likely Bots. Next, we analyze the smaller fraction of massively shared IPs from the cloud where UA strings are homogeneous, accounting for 19% and 4% of massively shared IPv4 and IPv6 traffic, respectively. These cases are distinct from others of massively shared IPs, as these IPs meet the traffic shape criteria but are unlikely to directly connect end-users. Instead, we hypothesize that much of this activity reflects *user-triggered* bot actions. That is, bots are triggered by link previews[22, 47, 48], ad bidding [58], user-triggered AI crawlers [51, 56, 68], or an application that initiates API calls [1]. While these types of IPs do not encompass end users as directly as proxies, they do capture a large amount of end-user browsing behavior that culminates in the same traffic shapes. Moreover, several companies that run these crawlers consider these to be user requests, thus not following the robots.txt protocol [35, 51, 56].

Indeed, in IPv4, the top cloud AS originating homogeneous, massively shared traffic (28% of traffic), the most common UA is from a popular business application suite that interacts with web services for the user; the second is associated with a large e-commerce platform that fetches Web resources. In the second most popular AS (14% of traffic), the only UA string presented is for an advertising platform that is likely collecting analytics for their customers or carries out ad bidding.

[3]We refer the reader to Figure 15 in Appendix F for the full distribution of unique UA strings per IPv4 address and IPv6 /64s.

Takeaway: A substantial amount of cloud traffic from massively shared IP addresses is end-user proxies, including consumer VPN/exit services and enterprise secure gateway deployments. A smaller amount of this traffic comes from bots, many of which appear to be user-triggered.

9 Massively Shared IPv4 Over Time

Finally, we study how massive address sharing has evolved in IPv4. For each month, we choose a week that does not include a major holiday globally and repeat our methodology. We observe that, *over the course of two years, the fraction of overall IPv4 traffic from massively shared IPs increases by nearly 10%*, ranging from 36% in January 2024 to 45% in October 2025.

Specifically in ISPs, massive sharing in IPv4 increases in all regions, with Asia, Africa, and South America starting at higher rates of massively shared traffic out of all traffic (45%, 43%, 45%, respectively) and increasing in traffic from massively shared IPv4 addresses (in Asia 11%, in Africa 8.8%, and in South America 7.7%). We also observe massive sharing in North America, Oceania, and Europe, starting at 24%, 16%, and 26% of all traffic and increasing by 10%, 4.5%, and 0.9%, respectively. Given that Internet penetration is increasing [28], we can expect this trend to continue.

Takeaway: The fraction of massively shared IPv4 traffic out of all IPv4 traffic has increased by 10% between early 2024 and late 2025. While pronounced in regions with scarce IPv4 resources, we see increasing massive sharing globally.

10 Concluding Discussion

In this work, we characterized large-scale IP sharing, which we coined massively shared IPs. We analyzed the prevalence of this deployment over both IPv4 and IPv6 traffic as seen globally by a major CDN. Our method for massive sharing detection relies solely on traffic shape, detectable by Fourier analysis and curve-fitting evaluation. We find that over 40% of IPv4 requests to a major CDN come from massively shared IP addresses, deployed across IPv4 networks globally, and increasingly over time. In this section, we discuss some of the key takeaways from our work.

IPv4 Sharing and IPv6 Deployment. Massively shared IPv4 is widely deployed in End-User ISPs (Section 6), and about half of this traffic originates from addresses that also have IPv6 connectivity (Section 7.2). This suggests these providers are prepared to reduce reliance on massively shared IPv4. However, IPv4-only deployments and partial IPv6 deployments are prevalent, particularly in Africa, introducing another single point of failure between clients and servers. Unless IPv6 deployment accelerates, the Internet will continue to broadly rely on IPv4 sharing, exacerbating reliance on a relatively small quantity of Internet resources.

Massive Sharing Is Not Always Driven By Scarcity. Contrary to initial intuition, we find many cases of massively shared addressing that are *not* driven by IPv4 scarcity. In fact, we observe massively shared IPv6 /64s in major ISPs (Section 7.3), as well as many instances of massively shared IPv4 (and to some degree IPv6) in the cloud (Section 8), underpinning proxy, enterprise gateway, and anonymization services. From this, we point to broader shifts in Internet infrastructure, illustrating that large quantities of end-user traffic originate outside of ISPs.

Massive IP Sharing Is Here To Stay. From both of these observations, one thing becomes clear: IPv4 exhaustion and IPv6 deployment aside, massive sharing is widespread, increasing (Section 9), and if observed trends continue, we may see even higher proportions of traffic carried over massively shared IP addresses in the future. Massively shared addressing complicates security, geolocation, host reputation, and blocklisting, underscoring the need for robust sharing-detection techniques

and for security systems to move beyond IP-based attribution. To combat these, detection methods and characterization must be continuously deployed and studied.

Future Work. As our method is purely centered on traffic shape, it can be used by operators and Internet researchers who might have access to volumetric telemetry but limited other details. Thus, future work can apply it to other vantage points, including transit networks and sampled telemetry such as IPFIX, NetFlow, and sFlow. Future work may fine-tune parameters for specific traffic types and vantage points, and exploit additional FFT properties such as frequency magnitudes and amplitudes or incorporate weekday–weekend variations. Finally, it would be valuable for future work to extend this work to study shared IPs in addition to massively shared IPs for further insight into Internet deployment.

11 Acknowledgements

The authors are grateful for the assistance and support of the Custom Analytics group at Akamai, in particular from KC Ng, and for the feedback provided by the anonymous reviewers. This material is based upon work supported by the National Science Foundation (NSF) Graduate Research Fellowship Program under Grant No. 2545526 and NSF CNS award 2319315. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] Akamai. 2019. Akamai State of the Internet Security Report: Retailers Most Common Credential Stuffing Attack Victim; Points to Dramatic Rise in API Traffic as Key Trend. https://www.ir.akamai.com/news-releases/news-release-details/akamai-state-internet-security-report-retailers-most-common.

[2] APNIC Labs. 2026. IPv6 Capable Rate by country (%). https://stats.labs.apnic.net/ipv6/.

[3] Apple Support. 2026. iCloud Private Relay. https://support.apple.com/en-us/102602.

[4] AT&T Internet. 2026. About Static IP addresses. https://www.att.com/support/article/u-verse-high-speed-internet/KM1002300/.

[5] Guillermo Baltra, Xiao Song, and John Heidemann. 2024. Ebb and Flow: Implications of ISP Address Dynamics. In *Passive and Active Measurement: 25th International Conference, PAM 2024, Virtual Event, March 11–13, 2024, Proceedings, Part II*. Springer-Verlag, Berlin, Heidelberg, 132–149. https://doi.org/10.1007/978-3-031-56252-5_7

[6] Paul Barford, Jeffery Kline, David Plonka, and Amos Ron. 2002. A signal analysis of network traffic anomalies. In *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet Measurment* (Marseille, France) *(IMW '02)*. Association for Computing Machinery, New York, NY, USA, 71–82. https://doi.org/10.1145/637201.637210

[7] bgp.tools. 2026. bgp.tools. https://bgp.tools/.

[8] John Brzozowski. 2011. Technical Details for Our IPv6 Deployment. https://corporate.comcast.com/comcast-voices/ipv6-deployment-technology.

[9] CAIDA. 2025. RouteViews Prefix to AS mappings. https://doi.org/10.21986/CAIDA.DATA.ROUTEVIEWS-PFX2AS-MAPPING.

[10] B. Carpenter, T. Chown, F. Gont, S. Jiang, A. Petrescu, and A. Yourtchenko. 2015. Analysis of the 64-bit Boundary in IPv6 Addressing. RFC 7421. https://doi.org/10.17487/RFC7421

[11] Martin Casado and Michael J. Freedman. 2007. Peering Through the Shroud: The Effect of Edge Opacity on IP-Based Client Identification. In *4th USENIX Symposium on Networked Systems Design & Implementation (NSDI 07)*. USENIX Association, Cambridge, MA. https://www.usenix.org/conference/nsdi-07/peering-through-shroud-effect-edge-opacity-ip-based-client-identification

[12] Samprit Chatterjee and Ali S. Hadi. 2012. *Regression Analysis by Example* (5th ed.). John Wiley & Sons, Hoboken, NJ.

[13] Alex Chen and Nate Sales. 2021. Multi-User IP Address Detection. https://blog.cloudflare.com/multi-user-ip-address-detection/

[14] Cloudflare. 2026. 1.1.1.1. https://one.one.one.one/.

[15] Cloudflare. 2026. Split Tunnels. https://developers.cloudflare.com/cloudflare-one/team-and-resources/devices/cloudflare-one-client/configure/route-traffic/split-tunnels/.

[16] Destatis. 2025. Fixed broadband Internet subscribers. https://www.destatis.de/EN/Themes/Countries-Regions/International-Statistics/Data-Topic/Tables/BasicData_BroadbandInternet.html.

[17] Alain Durand, Ralph Droms, Yiu Lee, and james woodyatt. 2011. Dual-Stack Lite Broadband Deployments Following IPv4 Exhaustion. RFC 6333. https://doi.org/10.17487/RFC6333
[18] Roy T. Fielding and Julian Reschke. 2014. Hypertext Transfer Protocol (HTTP/1.1): Semantics and Content. RFC 7231. https://doi.org/10.17487/RFC7231
[19] O. Gasser, R. Bush, M. Candela, and R. Housley. 2026. Publishing End-Site Prefix Lengths. RFC 9977. https://doi.org/10.17487/RFC9977
[20] Vasilis Giotsas and Marwan Fayed. 2025. One IP address, many users: detecting CGNAT to reduce collateral effects. https://blog.cloudflare.com/detecting-cgn-to-reduce-collateral-damage/
[21] Vasileios Giotsas, Loqman Salamatian, Antoine Cordelle, Nick Wood, and Marwan Fayed. 2026. A Global Inference and Assessment of Large Shared IP Addresses. In *Proceedings of the ACM SIGCOMM 2026 Conference*. Association for Computing Machinery, New York, NY, USA, 990–1008. https://doi.org/10.1145/3789240.3829192
[22] Google. 2025. List of Google user-triggered fetchers. https://developers.google.com/search/docs/crawling-indexing/google-user-triggered-fetchers.
[23] Google. 2026. Google IPv6 Statistics. https://www.google.com/intl/en/ipv6/statistics.html.
[24] Google Search Central. 2026. Google Crawler (User Agent) Overview. https://developers.google.com/crawling/docs/crawlers-fetchers/overview-google-crawlers.
[25] Avi Gupta, Thor Stead, and Latha Ganti. 2024. Determining a Meaningful R-squared Value in Clinical Medicine. *Academic Medicine & Surgery* (11 2024). https://doi.org/10.62186/001c.125154
[26] Scott Hogg. 2021. You Thought There Was No NAT for IPv6, But NAT Still Exists. https://blogs.infoblox.com/ipv6-coe/you-thought-there-was-no-nat-for-ipv6-but-nat-still-exists/.
[27] B Huffaker, M Fomenkov, and k claffy. 2011. *Geocompare: a comparison of public and commercial geolocation databases*. Technical Report. Cooperative Association for Internet Data Analysis (CAIDA). https://catalog.caida.org/paper/2011_geocompare_tr
[28] International Telecommunication Union. 2025. ICT Statistics. https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx.
[29] IP2Location. 2025. IPv4 Address 2025 Infographics. https://blog.ip2location.com/knowledge-base/ipv4-address-2025-infographics/.
[30] IPInfo. 2026. IPInfo. https://ipinfo.io/.
[31] Simon Kemp. 2025. Digital 2025: Algeria. https://datareportal.com/reports/digital-2025-algeria.
[32] Simon Kemp. 2025. Digital 2025: Ethiopia. https://datareportal.com/reports/digital-2025-ethiopia.
[33] Simon Kemp. 2025. Digital 2025: Mali. https://datareportal.com/reports/digital-2025-mali.
[34] Jeff Kline, Paul Barford, Aaron Cahn, and Joel Sommers. 2017. On the structure and characteristics of user agent string. In *Proceedings of the 2017 Internet Measurement Conference* (London, United Kingdom) *(IMC '17)*. Association for Computing Machinery, New York, NY, USA, 184–190. https://doi.org/10.1145/3131365.3131406
[35] Martijn Koster, Gary Illyes, Henner Zeller, and Lizzi Sassman. 2022. Robots Exclusion Protocol. RFC 9309. https://doi.org/10.17487/RFC9309
[36] Frank Li and David Freeman. 2020. Towards A User-Level Understanding of IPv6 Behavior. In *Proceedings of the ACM Internet Measurement Conference* (Virtual Event, USA) *(IMC '20)*. Association for Computing Machinery, New York, NY, USA, 428–442. https://doi.org/10.1145/3419394.3423618
[37] Franziska Lichtblau, Florian Streibelt, Thorben Krüger, Philipp Richter, and Anja Feldmann. 2017. Detection, classification, and analysis of inter-domain traffic with spoofed source IP addresses. In *Proceedings of the 2017 Internet Measurement Conference* (London, United Kingdom) *(IMC '17)*. Association for Computing Machinery, New York, NY, USA, 86–99. https://doi.org/10.1145/3131365.3131367
[38] Ioana Livadariu, Karyn Benson, Ahmed Elmokashfi, Amogh Dhamdhere, and Alberto Dainotti. 2018. Inferring Carrier-Grade NAT Deployment in the Wild. In *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications* (Honolulu, HI, USA). IEEE Press, 2249–2257. https://doi.org/10.1109/INFOCOM.2018.8486223
[39] Gregor Maier, Fabian Schneider, and Anja Feldmann. 2010. A First Look at Mobile Hand-Held Device Traffic. In *Passive and Active Measurement*, Arvind Krishnamurthy and Bernhard Plattner (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 161–170. https://doi.org/10.1007/978-3-642-12334-4_17
[40] Gregor Maier, Fabian Schneider, and Anja Feldmann. 2011. NAT Usage in Residential Broadband Networks. In *Passive and Active Measurement*, Neil Spring and George F. Riley (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 32–41. https://doi.org/10.1007/978-3-642-19260-9_4
[41] Patrick Maigron. 2026. RIR Delegations and RIPE NCC Allocations. https://www-public.telecom-sudparis.eu/~maigron/rir-stats/rir-delegations/world/world-ipv4-by-number.html.
[42] Anna Maria Mandalari, Miguel Angel Diaz Bautista, Francisco Valera, and Marcelo Bagnulo. 2017. NATwatcher: Profiling NATs in the Wild. *IEEE Communications Magazine* 55, 3 (2017), 178–185. https://doi.org/10.1109/MCOM.2017.1600776CM

[43] Anna Maria Mandalari, Andra Lutu, Amogh Dhamdhere, Marcelo Bagnulo, and KC Claffy. 2017. Tracking the Big NAT across Europe and the U.S. arXiv:1704.01296 [cs.NI] https://arxiv.org/abs/1704.01296

[44] Philip Matthews, Iljitsch van Beijnum, and Marcelo Bagnulo. 2011. Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers. RFC 6146. https://doi.org/10.17487/RFC6146

[45] Masataka Mawatari, Masanobu Kawashima, and Cameron Byrne. 2013. 464XLAT: Combination of Stateful and Stateless Translation. RFC 6877. https://doi.org/10.17487/RFC6877

[46] MDN Web Docs. 2026. Browser Detection Using the User Agent String (UA Sniffing). https://developer.mozilla.org/en-US/docs/Web/HTTP/Guides/Browser_detection_using_the_user_agent.

[47] Meta. 2025. Meta Web Crawlers. https://developers.facebook.com/documentation/sharing/webmasters/web-crawlers.

[48] Microsoft. 2025. Overview of Bing crawlers (user agents). https://www.bing.com/webmasters/help/which-crawlers-does-bing-use-8c184ec0.

[49] Towela Nyirenda-Jere and Tesfaye Biru. 2015. *Internet Development and Internet Governance in Africa.* Technical Report. Internet Society. https://www.internetsociety.org/wp-content/uploads/2017/08/InternetInAfrica-2015070820Final.pdf

[50] Lai Yi Ohlsen and Sabina Zejnilovic. 2026. Iran's Internet is partially restored, Cloudflare Radar data shows. https://blog.cloudflare.com/iran-internet-partially-restored-may-2026/.

[51] OpenAI. 2026. Overview of OpenAI Crawlers. https://developers.openai.com/api/docs/bots.

[52] Opera. 2026. Browser with Free VPN. https://www.opera.com/features/free-vpn.

[53] Ramakrishna Padmanabhan, John P. Rula, Philipp Richter, Stephen D. Strowes, and Alberto Dainotti. 2020. DynamIPs: analyzing address assignment practices in IPv4 and IPv6. In *Proceedings of the 16th International Conference on Emerging Networking EXperiments and Technologies* (Barcelona, Spain) *(CoNEXT '20)*. Association for Computing Machinery, New York, NY, USA, 55–70. https://doi.org/10.1145/3386367.3431314

[54] Konstantina Papagiannaki, Nina Taft, Zhi-Li Zhang, and Christophe Diot. 2003. Long-Term Forecasting of Internet Backbone Traffic: Observations and Initial Models. *Proceedings - IEEE INFOCOM* 2 (03 2003). https://doi.org/10.1109/INFCOM.2003.1208954

[55] PeeringDB. 2026. PeeringDB. https://www.peeringdb.com/.

[56] Perplexity. 2026. Perplexity Crawlers. https://docs.perplexity.ai/guides/bots.

[57] Ingmar Poese, Steve Uhlig, Mohamed Ali Kaafar, Benoit Donnet, and Bamba Gueye. 2011. IP geolocation databases: unreliable? *SIGCOMM Comput. Commun. Rev.* 41, 2 (April 2011), 53–56. https://doi.org/10.1145/1971162.1971171

[58] Enric Pujol, Philipp Richter, Balakrishnan Chandrasekaran, Georgios Smaragdakis, Anja Feldmann, Bruce MacDowell Maggs, and Keung-Chi Ng. 2014. Back-Office Web Traffic on The Internet. In *Proceedings of the 2014 Conference on Internet Measurement Conference* (Vancouver, BC, Canada) *(IMC '14)*. Association for Computing Machinery, New York, NY, USA, 257–270. https://doi.org/10.1145/2663716.2663756

[59] Lin Quan, John Heidemann, and Yuri Pradkin. 2014. When the internet sleeps: correlating diurnal networks with external factors. In *Proceedings of the 2014 Conference on Internet Measurement Conference* (Vancouver, BC, Canada) *(IMC '14)*. Association for Computing Machinery, New York, NY, USA, 87–100. https://doi.org/10.1145/2663716.2663721

[60] 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

[61] 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

[62] 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 *Proceedings of the 2016 Internet Measurement Conference* (Santa Monica, California, USA) *(IMC '16)*. Association for Computing Machinery, New York, NY, USA, 215–229. https://doi.org/10.1145/2987443.2987474

[63] RIPE NCC. 2026. RIPE Atlas. https://atlas.ripe.net/.

[64] John P. Rula, Fabián E. Bustamante, and Moritz Steiner. 2017. Cell spotting: studying the role of cellular networks in the internet. In *Proceedings of the 2017 Internet Measurement Conference* (London, United Kingdom) *(IMC '17)*. Association for Computing Machinery, New York, NY, USA, 191–204. https://doi.org/10.1145/3131365.3131402

[65] Patrick Sattler, Matthias Kirstein, Lars Wüstrich, Johannes Zirngibl, and Georg Carle. 2025. Lazy Eye Inspection: Capturing the State of Happy Eyeballs Implementations. In *Proceedings of the 2025 ACM Internet Measurement Conference* (USA) *(IMC '25)*. Association for Computing Machinery, New York, NY, USA, 213–221. https://doi.org/10.1145/3730567.3732925

[66] Stephen D. Strowes. 2016. Diurnal and Weekly Cycles in IPv6 Traffic. In *Proceedings of the 2016 Applied Networking Research Workshop* (Berlin, Germany) *(ANRW '16)*. Association for Computing Machinery, New York, NY, USA, 65–67.

https://doi.org/10.1145/2959424.2959438
[67] Sulyab Thottungal Valapu and John Heidemann. 2025. Towards a Non-Binary View of IPv6 Adoption. In *Proceedings of the 2025 ACM Internet Measurement Conference* (USA) *(IMC '25)*. Association for Computing Machinery, New York, NY, USA, 727–745. https://doi.org/10.1145/3730567.3764467
[68] Kha Van. 2025. Getting Started with Amazon Nova Act: A Complete Sample Demos Collection. https://builder.aws.com/content/31K6XtmBdnZqOB5hsPGIsSOpJh5/getting-started-with-amazon-nova-act-a-complete-sample-demos-collection.
[69] W3C Community Group. 2026. The Network Information API. https://wicg.github.io/netinfo/.
[70] Sanford Weisberg. 2014. *Applied Linear Regression* (4th ed.). John Wiley & Sons, Hoboken, NJ.

A Ethics

Measurements in this study were conducted using confidential operational data provided by the participating CDN. Data processing took place exclusively on secure machines maintained and controlled by the CDN. Access to the data was restricted to authorized personnel with valid credentials, and all activities adhered to the CDN's internal policies, security controls, and applicable regulatory requirements. The reported results are aggregated such that no personally identifiable information of any end-user could be inferred, nor any proprietary information of the CDN, such as individual customers or network partners.

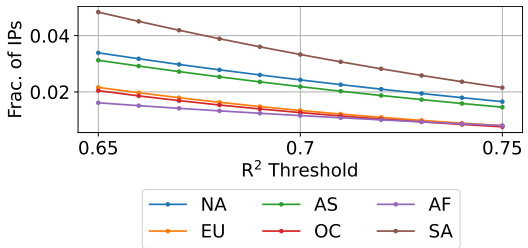


Fig. 10. The fraction of active IPs that are massively shared with different R^2 thresholds per continent. We experiment with values from 0.65 to 0.75 in increments of 0.01.

B Parameter Sensitivity Analysis

***R^2* Sensitivity.** Next, we experiment with how the R^2 threshold affects the fraction of massively shared IPs and traffic we detect. In Figure 10, we present how the fraction of massively shared IPs changes per continent with different thresholds, seeing that this fraction does change. This presents a trade-off between the strictness of a definition of smoothness (i.e., a higher R^2 value is stricter) and the number of IPs we capture. We note that this pattern also applies to the fraction of traffic from these IPs to a higher degree. When varying the value from 0.65 to 0.75, the total massively shared traffic volume changes by 7-12% (depending on continent). We proceed with using 0.7 as our threshold based on our experimentation, but emphasize that in future work this can be adjusted based on the nature of the traffic.

Number of Frequencies Used for Signal Reconstruction. To further validate our choice to use the top 3 frequencies to reconstruct the curve, we show how using too many frequencies allows non-shared IPs to meet our fitting criteria. In Figure 11, we run the FFT on the individual IPs presented in Figure 1 and reconstruct the signal using different numbers of frequencies (the strongest frequencies, i.e., 5 refers to the 5 strongest frequencies). We observe that using more frequencies allows these IPs to meet our R^2 threshold of 0.7. To ensure we filter out these cases, we use 3 frequencies to reconstruct the traffic shape.

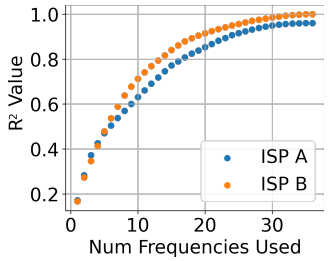


Fig. 11. We show the R^2 value when using different numbers of frequencies to reconstruct the signal on the single IPs from 1.

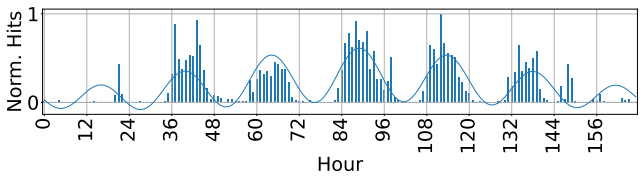


Fig. 12. 7 days of traffic from an IP in an enterprise proxy service. We see little activity over the weekend, motivating our choice to select Tuesday-Thursday as the 72-hour period we analyze weekly.

Stability over Time. The bulk of our analysis is performed on data from a 3-day period in a single non-holiday week in July 2025. While we analyze a relatively small time period, we observe that our results are consistent when repeated over the course of 1 month (throughout 4 weeks in May 2025). When we experiment per continent, the fraction of massively shared IPs changes by at most 0.2% over 1 month, and the fraction of massively shared traffic changes by at most 3%. Therefore, our chosen time period does not substantially affect our results.

C Implementation: Time Period

Figure 12 shows a case of this disruption in an enterprise gateway. The full 168 hours of traffic does not meet our criteria due to poor fit. However, isolating to Tuesday-Thursday (72 hours) does. We find several such examples in our exploration and therefore apply our method to 72 hours of consecutive traffic from Tuesday to Thursday.

D Examples From Other Countries

We also perform the multi-user simulation experiment in Section 4.1 in /24s from ISPs in other countries and find similar results: aggregating individual IPs from each subnet eventually reaches our massively shared criteria. We identify these IPs through research of the networks and through networks identified through RIPE Atlas [63]. We identify probes that are tagged with "Native IPv4." Although we do not have ground truth that these IPs in fact do have native IPv4 or that these ISPs use single-subscriber IPv4 addressing, we highlight that this aligns with our claims of our estimations being lower bounds. That is, if more subscribers are behind individual IPs, this means that more user traffic was necessary to reach our criteria.

We show the results in Figure 13. We observe that this experiment is successful in multiple distinct regions globally. We specifically highlight the subnet in Iran, where we see significantly fewer IPs total active in the /24 than in those of other countries. We hypothesize that one reason for this lack of activity from IPs is due to both the CDN's presence in the region and potential

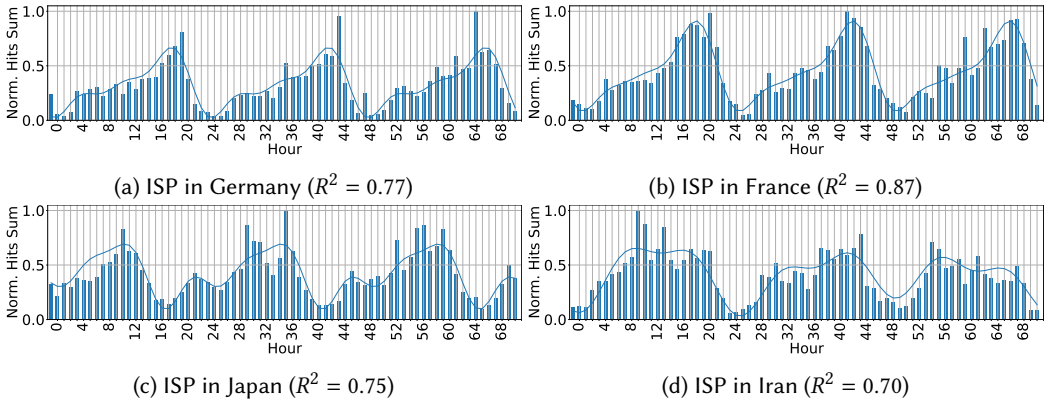


Fig. 13. We present traffic aggregated from /24s in 4 different countries with varying levels of CDN presence. We find that all traffic aggregated together meets our massively shared criteria.

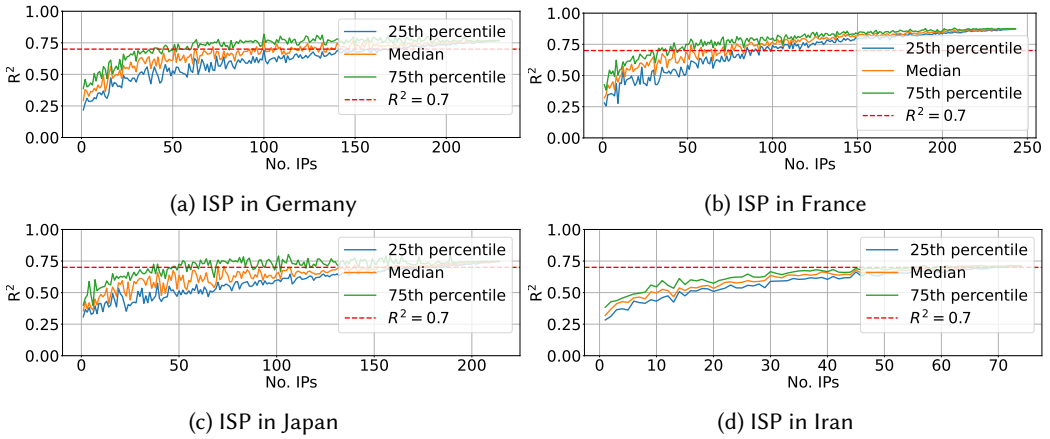
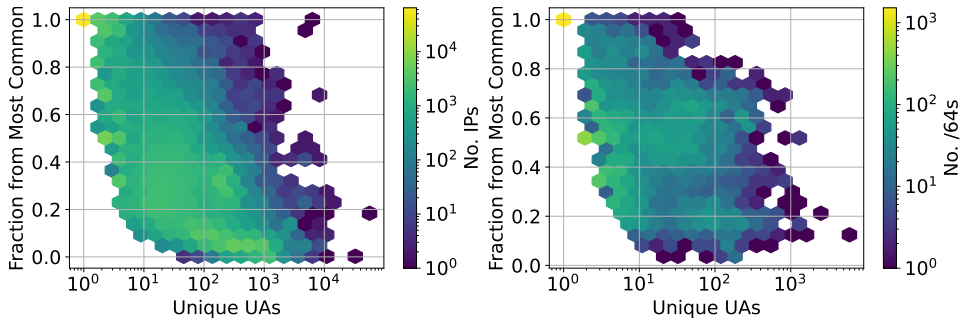


Fig. 14. We repeat the aggregation experiment presented in Figure 3 in /24s in 4 different countries. We find that all traffic aggregated together meets our massively shared criteria between 65 and 75 individual IPs.

Internet disruptions [50]. Despite this, we are still able to aggregate traffic that meets our diurnal and smooth criteria. This shows that when the traffic of many users is aggregated, these signals can appear even with sparse data. We also use this to build confidence in other geographic areas that appear similarly in the CDN's access logs.

Estimating Single Subscriber Count Behind Massively Shared IPs. To further quantify how many single subscribers are captured in massively shared IPs, we repeat the analysis presented in Section 5.1. We show the results of these experiments in Figure 14. In these examples specifically, we find that the median number of single subscriber IPs we need to aggregate to meet our massively shared criterion is 65-75 in all cases. We remind the reader that this is the lower bound for the number of subscribers needed to meet this criterion; adding more single-subscriber IPs past this amount consistently meets our criteria.



(a) IPv4 (b) IPv6

Fig. 15. UA string diversity from massively shared traffic in cloud providers. On the y-axis, we plot the fraction of UA strings that the most common UA string represents. For example, if we observe 10 hits from an IP, and 9 of them have the same UA string, the fraction will be 0.9. On the x-axis, we compare this to the total unique UA strings we observe for this IP (note the log scale). We observe a concentration of IPs that present a single UA string, which prior work has identified as a bot characteristic [61].

Of course, this method has caveats, as we have no ground truth for these particular subnets, only context from operator reports on RIPE Atlas and ISP documentation. However, as stated in Section 5.1, we provide this as a heuristic to estimate subscriber count behind massively shared IPs.

E Manual Validation of Known Technologies

In addition to the widespread validation across large proxy geofeeds or entire countries described in Section 5.3, we additionally validate individual IPs by manually collecting them. We collect cellular IPs from 4 different providers across the US and Europe, and IPs from 2 different proxy services from our colleagues. In all cases, we find that these cases meet our massively shared criteria. Although small in scale, these cases provide ground truth validation.

F UA String diversity

Here, we present a more detailed view of UA string diversity across massively shared IPs in cloud providers. In Figure 15, we present the UA strings presented by these IPs and /64s. We observe that a large fraction of IPs (19.2% in IPv4, 4.2% in IPv6) present a single UA string, which we refer to in Section 8 as *homogeneous* IPs.

Received 5 June 2026; accepted 10 September 2026