A Global Inference and Assessment of Large Shared IP Addresses

Vasileios Giotsas
Cloudflare, Inc.

Loqman Salamatian
Columbia University

Antoine Cordelle
Cloudflare, Inc.

Nick Wood
Cloudflare, Inc.

Marwan Fayed
Cloudflare, Inc.

Abstract

The number of clients and users behind an IP address can differ by orders of magnitude, owing to large shared IPs such as VPNs, proxies, and Carrier-Grade NAT (CGN) gateways. However, limitations on visibility, the absence of Internet-wide data, and the dynamic nature of IP allocations make it difficult to disambiguate multi-user IPs (M-IPs) and their impact on service provision. In this paper we devise an inference technique to detect M-IPs. We train a classifier by annotating data from public sources with features extracted from a global CDN request log. To demonstrate reproducibility, we build a parallel model using public M-Lab data and achieve comparable accuracy with different features. An unanticipated result was the dominance of /24 feature importance over individual IPs. Using the CDN logs, we then evaluate implications of CGNs along multiple dimensions including user impact, relationship with IPv6 networks, and regional distributions. Our results reinforce various intuition, and potentially challenge some assumptions.

CCS Concepts

• **Networks** → **Network measurement**; *Network performance analysis*; *Middle boxes / network appliances*.

Keywords

CGNAT, Network Measurement, Shared IP Addresses

ACM Reference Format:
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 ACM SIGCOMM 2026 Conference (SIGCOMM '26).* ACM, New York, NY, USA, 19 pages. https://doi.org/3789240.3829192

1 Introduction

IP addresses are the gateways that link users and services to the Internet. IP addresses neither have, nor ascribe, any meaning to those users and devices—they are equivalent by design [58, 61]. In contrast to their design, they are often used to make higher-layer characterizations about services, connections, and the connected clients (§2.1)—giving *identifier* semantics to IP addresses that otherwise have none. IP addresses are increasingly poor surrogates for the diversity of devices, services, and users, that connect to the Internet.

Characterizations of client-side IPs are particularly problematic. *Large-scale* multi-user IP addresses (M-IPs) belonging to Carrier-Grade NATs (CGNs), VPN exit nodes, and enterprise proxies may serve thousands of clients (§2.3). Furthermore, a map view of Internet penetration figures normalized by the IPv4 space in a region suggests implications may vary along socio-political boundaries (see Figure 1, §2.2).

The sensitivities are well known to content and security services that detect and protect against unwanted or malicious activity. These systems use heuristics that may be influenced by connection rates or request volumes per IP [19, 34, 77]. Without additional knowledge, heuristics implicitly regard source IPs as approximately equivalent. The assumed equivalence is correctly broken by new or unrecognizable attacks, but also falsely broken by high volumes of legitimate users behind large-scale multi-user IPs (M-IPs). In extreme cases operators may block or rate-limit all traffic from an IP address [26, 52], equally affecting all users behind it (§2.1). Because IP-level actions apply indiscriminately, their consequences on unrelated users are unavoidable.

In this paper we design, evaluate, and deploy, a methodology to infer large M-IPs from high-volume but high-noise data. Our approach is shaped by the needs of CDNs and three operational constraints at scale. First, recognize the legitimate clients behind shared IPs that would otherwise be blocked or rate-limited, since the cost of false positives outweighs missed detections, so being "not wrong" is preferable to being "perfectly right". Second, analysis must succeed on aggressively sub-sampled and noisy data; because CDNs observe traffic at massive scale, only a small fraction of HTTP requests (on the order of 1% offline down to 0.01% in production) are logged. Alongside, lower sampling rates should only reduce coverage and not degrade the inference reliability or accuracy. Third, ground-truth labels are unavailable in practice because networks do not disclose their internal address-sharing configurations; inference must therefore rely on labels constructed by aggregating multiple public datasets, drawing on a mix of long-standing measurement techniques and newer approaches.

Overview: We begin by constructing a 'reference' dataset for training and validation (§3). Source IPs are seeded from public sources including active scanning projects, traceroutes,

SIGCOMM '26, Denver, CO, USA

ACM ISBN 979-8-4007-2467-1/26/08
https://doi.org/3789240.3829192

and reverse DNS PTR records. Each of these data gives some indication or signal that identifies different types of M-IPs. We evaluate and extract connection- and application-level features of the reference M-IPs, as well as the IPs in the covering /24 prefix, using the CDN HTTP request logs (§4). We train a multi-class classification model using the aggregate features of each reference IP and its /24 prefix (§5). Analysis shows that IPs selected by the M-IP classifier have profiles consistently different from those of other IPs. For example, inferred M-IPs generate proportionally 16× more requests than their non-M-IP counterparts.

The importance of /24 features emerged as an unanticipated insight from our analysis: Feature diversity across the /24 turns out to be a significantly stronger discriminator than volume, AS characteristics, or the same features of individual IPs (§5). Our finding is consistent with observations that CGNs are often deployed as contiguous address blocks: Since every individual connection behind a CGN can be mapped to any public IP of the CGN, all IPs are needed to capture the user population's diversity. In contrast, VPNs and proxies are given isolated addresses assigned by a hosting network, and all of a user's connections are bound to that single IP.

To encourage reproducibility, we also develop a parallel model using only public M-Lab data (§6). The public model uses features unique to the M-Lab datasets and achieves high classification accuracy (ROC AUC > 0.97) comparable to the CDN model where users behind CGNs have run speed tests to M-Lab's servers, so the IPs appear in the dataset. We share the software used to develop the public model, including retrieval of public data to build a training set [14].

Our evaluations, insights, and discussion (§7, §8) focus on CGNs because of their unique attribute among M-IPs: Unlike VPNs and proxies, users do not choose CGNs and have no agency over their use. Users cannot reasonably anticipate, and have no visibility nor recourse, if adversely affected when behind a CGN. We use our inferred results and the CDN logs to assess their impact on users, and existence in IPv6 networks. We also evaluate various per-country CGN distributions to find stark contrast in some cases, nuance in others. We also discuss their impact on security and operations, while projecting into a future where botnets test the limits of defenses suggesting opportunities for further study. The strength of reported results have motivated the CDN to operationalize our design in the production environment.

1.1 Ethics, Data Sharing & Privacy

Our study follows guidelines established by the Menlo Report [25], and underwent a detailed review to ensure compliance with the CDN's policies on data collection, processing, and privacy, which meet or exceed the privacy and security requirements of all major data protection regulations, including GDPR[18]. The proprietary CDN data was available only to certain CDN employees with strict access controls, retention and expiration policies, and processed on the CDN premises. CDN data was gathered strictly in aggregates, i.e., raw numbers of distinct fields (e.g., source ports, SNIs, user agents), with no attempts to track users or reconstruct individual browsing histories. M-Lab data was used strictly to validate our models, and in accordance with M-Lab's policies [48].

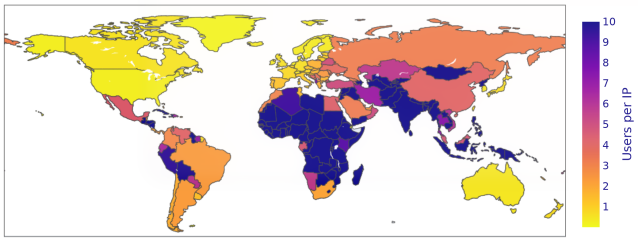


Figure 1: An Internet-user penetration map normalized by the number of IPs registered in the country (§2.2). The disproportionate distribution of users-per-IPs suggests that an understanding, inference, or action according to IP addresses leads to outcomes along socio-economic and geopolitical boundaries.

2 Background and Motivation

2.1 Actions on IP addresses at layers-4+

M-IP user-populations impact multiple research domains. Numbers of subscribers per ISP help characterize persistent last-mile congestion [33], measure the proximity of cloud services to end users [6], analyze topological vulnerabilities in overlays [53], or evaluate CDN performance [11, 71]. Knowledge of user populations also informs infrastructure planning and policy [29–31, 75].

IP-oriented security protections necessarily incur a fate-sharing quality. Providers may block or rate-limit IP addresses [7, 8] because of non-human activity [74], without regard to the effect on legitimate clients behind the same IP [59]. At Wikimedia Commons, 3–5% of editors may be affected by IP-level filtering, as some communities block edits originating from multi-user IPs that prevent attribution to individual users [21]. Anecdotally, engineers at our target CDN recounted an incident that was declared after customers blocked known VPNs, which prevented the host operating system from retrieving critical security patches.

2.2 Global Internet Penetration at layer-3

User penetration rate, defined as the estimated number of Internet users divided by the population of a country, is a common indicator of global development and varies widely across regions [64, 78].

The network-layer sees only IP addresses, and is opaque to user penetration. We visualize this gap by computing country-level ratios of users estimated by the World Bank [78], to the number of in-region IP addresses according to Regional Internet Registry records [5]. The colour-coded map that emerges in Figure 1 inverts our understanding of Internet populations: the usage of each IP address is dominated by regions that are otherwise under-represented in conventional user penetration maps. For example, Australia, Canada, Europe, and the USA have lowest user-to-IP ratios, while large portions of Africa and South Asia present strongly.

The scarcity of IPv4 address space means that regional differences can only worsen as penetration rates increase. A natural consequence of increased demand in developing regions is more CGN deployments, causing human users to unknowingly share their interface to the Internet—potentially with unintentional effects on their Internet experience (§2.1).

2.3 Shared IPs and User Choice

Commercial VPN and proxies intentionally multiplex traffic from many users behind shared IP addresses. Commercial VPN services route subscriber traffic through provider-managed gateways for privacy or security; users may connect from diverse access networks and locations [32, 47]. Enterprise proxies and firewalls similarly aggregate traffic from multiple devices by either terminating connections on behalf of the client directly or forwarding them transparently. More recently, oblivious proxies have emerged to explicitly decouple client identity from network-layer identifiers [38, 70, 76]. These designs trade IP-level attribution for privacy and are typically adopted explicitly by users.

Carrier-grade NATs (CGNs) perform network address translation at large scales [37]. CGNs tend to be affiliated with last-mile access networks to which users subscribe for service, and have been the subject of multiple studies across disciplines and domains [1, 39, 62, 75]. A detailed example is shown in Figure 2. Users do not choose to use CGNs, making their detection more important to content service operators. Unlike other large shared IP types, users cannot reasonably anticipate or understand the implications of their use. Other multi-user IPs can be disabled via a user-action or are expected in the workplace. When behind a CGN, users have no recourse or escape from onward actions triggered by voluminous or malicious activity behind the same IP.

2.4 A Comment on IPv6

IPv6 is not a reason to ignore large-scale address sharing in IPv4. Although CGNs were largely motivated by IPv4 address exhaustion, address sharing also serves other purposes: VPNs and oblivious proxies show continued demand for shared egress as a privacy abstraction. In addition, even IPv6-capable ISPs may continue to operate CGNs so subscribers can reach IPv4-only services; we return to this relationship between IPv6 deployment and CGN usage in §7.2.

Our methods may also apply beyond IPv4 because the underlying problem is aggregation, not only address scarcity. Large services often reason about IPv6 at prefix granularity because individual addresses are too numerous to treat as stable units. Operational guidance recommends delegations no smaller than /56 [81], and addresses within a delegated block can rotate freely [55, 65]. As a result, the operational unit of analysis in IPv6 is often closer to a /56 or /64 than to an individual address, echoing the role of /24s in our IPv4 measurements.

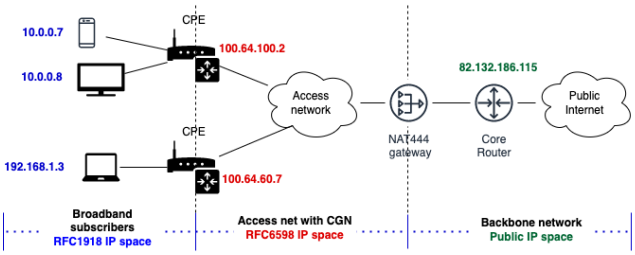


Figure 2: Traceroute CGN discovery. IP addresses behind a CGN, 'NAT444' above, use shared or private IP space. Intuitively, between user-devices and the public Internet there must be at least two hops in the non-public IP space.

3 Reference dataset: Multi-user IPs

In this section we construct a reference dataset of large-scale[1] IPs for training, feature extraction (§4), and validation (§5).

We deliberately refer to these IPs as a *reference dataset* rather than ground truth, because every label is itself derived from public signals and carries some residual noise. We do not treat the reference set as infallible. Instead, we (i) enumerate the expected noise of each source and the mitigation we apply in Table 3, and (ii) treat the labels as a starting point that must be corroborated by independent behavioral evidence.

3.1 Data and Feature Collection

Our reference multi-user IP addresses are assembled from various public sources and active scans. Features are drawn from a global CDN HTTP request log, as described below.

VPNs and proxies from scanning services Port scanning tools such as Zmap [28] or Masscan [35] can be used to scan the entire public IPv4 range on well-known ports, and inspect the corresponding banners that are returned. Services such as Shodan [45], Censys [27], and Rapid7 Project Sonar [60] periodically enumerate network services using similar strategies. They conduct a two-state IP scan, first for open well-known ports, followed by a request tailored to the service identified on that port. Responses are recorded as a service banner of application-layer details that can include protocol headers, log-in messages, HTML landing pages, and software versions. The metadata identifies different types of M-IP addresses, including proxy and VPN servers.

In our dataset we drew VPN IPs from the `ipinfo.io` service[2], which actively subscribes to VPN services to identify exit nodes. We also drew VPN and proxy IPs from Censys and Shodan, which detect VPN endpoints by sending protocol-specific probes (IKE for IPsec, client hellos for OpenVPN, TLS handshakes for SSL VPNs) and fingerprinting responses using vendor IDs, certificates, and HTTP artifacts.

Carrier-grade NATs via active traceroutes We execute a traceroute campaign using RIPE Atlas [63], which gives more than 14K vantage points in 182 countries. We apply the NATRevelio technique [43] to find CGN IPs since ISPs often

[1] Publicly addressed home broadband routers and other on-premise CPE devices are small-scale and out of scope.
[2] API data using the VPN tag, `https://ipinfo.io/tags/vpn`

do not follow best practice [1], which recommends use of dedicated shared IP address space [23].

Given a traceroute, NATRevelio flags a CGN when the path exhibits the following three properties: (i) the path segment between the probe and its globally routable address (GRA) contains at least one hop in the shared address space[3]; (ii) the first hop IP address is in the private address space, corresponding to customer premises equipment (CPE); and (iii) the RTTs between consecutive hops along the path up to the GRA exhibit a tenfold increase. This abrupt RTT increase is a strong indication that the corresponding link lies within the ISP access network [43].

We ran our traceroute campaign from 7–10 April 2025 using 14,104 RIPE Atlas probes (all the responsive Atlas probes at the time). From those, 8,937 received a response from their GRA, with 1,400 having a direct IP path to their GRA to give a set of true negatives (non M-IPs). Among the remaining 7,537 probes, there are 815 probes found behind a CGN gateway, located in 329 ASes. From those, 325 traverse shared IP space; 428 traverse the public IP space of the GRA's AS; and 62 probes cross both. Where two or more GRAs appear in the same `/24`, we treat the entire `/24` as a CGN, consistent with existing research findings [39, 67]. If a /24 has only one probe behind it, we check if there are multiple GRAs in the same /24 by getting all the probe source IPs for a period of 10 days after the measurement. If the probe is assigned a new source IP within the same `/24` prefix, we repeat the measurement to confirm that the probe is still behind CGN. We similarly test from probes that do not appear to be behind a CGN, to extend the characterization of non-CGN for all the IPs in the same `/24` prefix.

Carrier-grade NATs via DNS IPv4 PTR responses Many operators encode metadata about their IPs in the corresponding 'reverse' DNS pointer (PTR) record that can signal administrative attributes [41] and geographic information [40]. We first query the DNS for PTR records for the full IPv4 space, and then filter for a set of known keywords from the responses that indicate a CGN deployment [39, 67]. We conducted a zDNS scan [36] between 2025-04-04 and 2025-04-18 targeting the entire set of source IP addresses that appear in the CDN logs at the same period. We found 123,134 IPs in 383 ASes whose PTR records indicated the presence of CGNs. From those, 42 of the ASes in the PTR dataset also appeared in the RIPE Atlas dataset, meaning that the reference data for CGNs cover 670 ASes. For the 50 ASes with the most PTR records we manually verified the fact that they deploy CGN through official documentation or user reports.

Layer-4+ features from CDN HTTP request logs To validate assembled IPs we use data from a CDN HTTP requests log, with the expectation that various characteristics are measurably distinct between M-IPs and non-M-IPs. The CDN receives 50/75M average/peak requests per second, respectively, from more than 13K interconnections with other networks at points of presence in over 300 cities in more than 100 countries. The log is a random 1% sample of HTTP requests that we use for feature extraction and evaluation: For each reference IP found in the log (≈ 400 million IPs/day), we tabulate aggregate statistics and standard deviations over a set of application- and connection-level attributes: timestamps, source and destination IP addresses, port numbers, HTTP version, HTTP User Agent, hostname, TCP Round-Trip Time (RTT), TLS handshake completion time.

We also tabulate the same characteristics on the logged IP addresses in the covering /24: Prior work suggests that CGN deployments typically occupy contiguous IP blocks [39, 67] that are heavily and uniformly used, whereas VPN and proxy services tend to use only a small number of IPs within a given prefix. Therefore, we posit that aggregate behavior at the /24 level helps distinguish CGNs from other multi-user IP types. We test and confirm the hypothesis in §4.

Table 1: Composition of our reference dataset.

Reference IP Dataset	Addrs Found	No. of ASes	Countries
VPNs & Proxies	948,283	7,122	187
CGNs (traceroutes)	194,317	329	73
"Non M-IP" (traceroutes; non-CGNs)	355,600	813	110
CGNs from DNS PTR Records	123,134	383	64

Assembled training dataset sizes The number of IP addresses identified in our reference dataset from each source is listed in Table 1, along with the number of ASes in which they appear. The "non M-IP" label corresponds to client IPs observed in traceroutes that are not identified as CGNs and appear as a direct hop from their public IP address (i.e., not behind multiple layers of NAT). The reference dataset covers a wide set of countries allowing us to avoid regional biases.

Not all reference IPs are observed in the CDN HTTP request logs. At a 1% sampling rate, the logs contain 48,613 CGN IPs from traceroutes (25%), 283,618 VPN/Proxy IPs (29.9%), and 252,459 non M-IPs (71%).

The comparatively small fraction of observed VPN/Proxy IPs was somewhat surprising but the existence of a proxy or VPN endpoint does not imply sustained use. Many such IPs are lightly used, intermittent or not serving user traffic (but can be scanned). Under an 1% request sampling, an IP that generates at least 300 requests has ≤5% chance of being missed entirely, implying that VPN/Proxy IPs absent from the logs overwhelmingly correspond to endpoints that send no more than a few hundred requests to this CDN during the sampling period. In contrast, heavily used shared IPs—particularly CGNs—remain visible and align with our goal to capture IPs even under aggressive subsampling.

3.2 Limitations, Bias, and Reproducibility

Limitations Our methodology targets *passive* detection of *large-scale* shared IPs in context of large operational environments with low sampling rates. We prioritize orders of magnitude differences; individual networks are out of scope.

[3]We conservatively exclude the private address space despite having first-hand knowledge of ISPs that use the non-canonical 10./8 range, since private addresses may also indicate an enterprise network.

Our focus is where collections of unrelated users co-exist in their own unrelated networks behind single IP addresses.

We additionally prioritize carrier-grade NATs since users do not choose to use them. Inferred CGN IPs may be either fixed line or wireless. Their user and location distributions behind CGNs may differ, but the detection methods are similar. Separately, in our evaluations we combine VPNs with proxies because their metrics are similar (§4), and their use is known to their users. Table 3 in Appendix H summarizes the noise we expect each reference source to introduce, and the mitigation we apply before a label is used for training or evaluation.

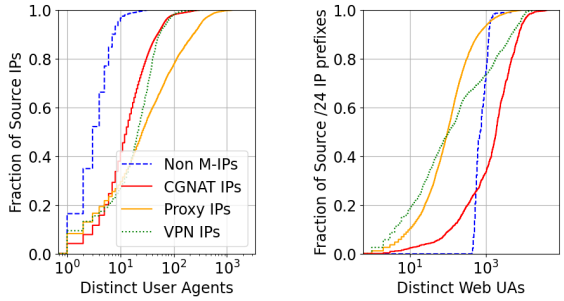
Sources of bias—public sources The reference dataset may reflect biases inherent in the public data sources. RIPE Atlas probes are volunteer-deployed and unevenly distributed across regions and networks, potentially underrepresenting types of access networks. Similarly, PTR-based CGN identification depends on operator-maintained naming conventions and keyword matching, which may vary across regions and languages or be intentionally obfuscated. Active scanning and third-party datasets likewise preferentially capture visible, server-facing infrastructure. These factors primarily affect coverage rather than feature validity: The public sources influence which IPs enter the reference set, but the behavioral signatures learned by the model remain consistent across the networks and deployments we observe (§5). We posit this is an artifact of our methodology's focus on connection- and request-level distributions rather than user activity.

Sources of bias—request logs Our inference is shaped by the diversity of CDN customers and their users represented in the request log, and their locations. While we are unable to report exact numbers, the CDN serves content for millions of domains. We also have confidence of reasonable user representation from the distribution of observed HTTP user-agents (UAs) across the world, which follows the distribution of the global Internet population. We analyzed the full request logs and found that UAs have a strong monotonic relationship with the number of Internet users in a country, having a Spearman's rank correlation coefficient of 0.941, and a linear association according to a Pearson correlation coefficient of 0.631. Similarly strong indicators present when comparing UAs to APNIC estimates of users per ASN [4, 66]. The analysis indicates that the CDN has a relatively balanced view of the Internet-connected population; less clear is how well it represents the user population. Full details of the analysis are available in Appendix A.

Reproducibility Reference IPs are derived from public data. Also, our extracted features are usable for owners of similar logs. We demonstrate that our approach generalizes beyond privileged data sources by developing a parallel, fully public pipeline using M-Lab's `tcpinfo` dataset (§6).

4 M-IP features selection

In this section, we extract and analyze a 1% sample of CDN HTTP request logs for features of IPs in our reference set, as well as for their covering /24 prefixes. We examine how these features differ across IP types (VPNs, proxies, CGNs) introduced in §3 and summarized in Table 1. Our goal is to identify measurable behavioral differences in CDN traffic that distinguish large shared IP addresses from other sources.



(a) UAs per IP. (b) UAs per covering /24.

Figure 3: Cumulative distribution of distinct web browser user-agents (UAs) for different IP types.

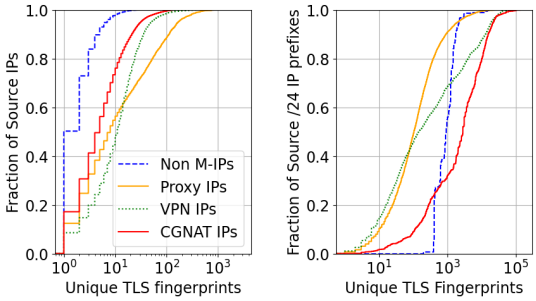
Our working hypothesis is that characteristics in the aggregate for large M-IPs is distinguishable from activity generated by other source IP addresses. Our feature extraction is an evaluation of that hypothesis: We analyze various characteristics of traffic from our reference IPs, against those that are not. Since networks do not disclose CGN and other uses of IPs, the quality of our inference is strictly dependent on our confidence in the training data.

We claim the key discriminator is diversity, not (just) volume. For example, VM-hosted scanners may generate high request volume, but with few TLS signatures and low RTT variability. Similarly, publicly routable CPEs may have as many TLS signatures as software clients, but with smaller volumes (that are less likely to be sampled).

4.1 User-agent Diversity

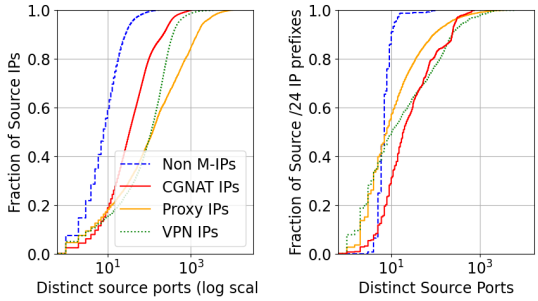
A user-agent (UA) identifies the client's software stack and architecture to the remote HTTP server so that it may respond accordingly [22]. The UA HTTP header is a characteristic string consisting of the client's application, operating system, vendor, and version and/or device where the application runs. UA strings in our dataset are reasonably diverse and correlate with population distribution (§3.2).

Figure 3(a) plots the number of unique agents across all HTTP requests per IP. Individual multi-user IPs have an inflated number of unique UAs compared to the rest of the IPs. Almost all of the non-M-IPs have fewer than 10 distinct UAs, compared to 40% of CGN and 50% of VPN and proxy IPs in our reference dataset. Figure 3(b) plots the aggregate number of unique UAs for the /24 IP prefixes that cover each reference IP. Interestingly, prefixes covering CGN IPs exhibit the highest total number of unique UAs, but prefixes hosting VPNs and proxies have, on average, a much smaller number of unique UAs compared to non M-IPs. This gap in the total number of unique UAs between IPs and their covering /24 prefixes matches observations that CGNs are deployed using contiguous prefixes (§3.1), while proxies and VPNs tend to operate on a small number of isolated IPs within a prefix.



(a) TLS signatures per IP. (b) TLS signatures per /24.

Figure 4: Cumulative distribution of distinct TLS signatures per IP type, and covering /24 prefix. Relative differences and ordering among individual IPs are near opposite of their /24s.



(a) Source ports per IP. (b) Source ports per /24.

Figure 5: Cumulative distribution of unique source ports for different type of IPs and corresponding /24 prefix. Source port usage over /24 appears to converge, which may be unsurprising given that OSes select ports over a fixed range.

Explaining the /24 difference. We believe the shifts in individual distributions in Figure 3 reflect deployment-level distinctions, which are also mirrored in TLS signatures (§4.2), source ports (§4.3), and number of visited domains (§4.6).

To illustrate, consider a neighborhood where users connect either through a CGN or directly via publicly routable IPs. Behind a CGN, connections from all users may be mapped to any address in the shared block, so observing a single IP reveals only a fraction of the population. Aggregation at the /24 level is therefore required to capture full diversity. In contrast, publicly routable IPs typically correspond to individual households, making each IP's UA diversity both smaller and fully observable; the /24 distribution is simply the sum of these independent users.

Proxy UA distributions are largely unchanged, suggesting that proxies are deployed as isolated IPs in an address block (for example in the enterprise). VPN utilization appears to vary from one to many addresses in a block, and approach CGN levels of block utilization only in a few cases.

4.2 Diversity of TLS Signatures

A TLS signature captures client-specific behavior during the TLS handshake, using fields in the ClientHello such as protocol versions, cipher suites, extensions, supported groups, and EC point formats. Because client implementations tend to populate these fields consistently, their combination can distinguish device or application stacks [3], and TLS fingerprinting is widely used for security monitoring and bot mitigation [13]. We characterize each request using a canonicalized ClientHello fingerprint recorded in the CDN logs, with GREASE values removed [9] and extension order canonicalized to reduce per-connection randomization. The fingerprint is similar in spirit to JA3 [68] and JA4 [2], but preserves the underlying cryptographic parameters rather than relying only on hashed or truncated representations.

We use this signal only in aggregate: the feature is the number of distinct TLS signatures observed for an IP address or covering /24 over a fixed time window, not the identity of any specific signature. This count is a noisy measure of client-stack diversity, but unusually high signature diversity after canonicalization is consistent with many distinct client stacks sharing the same address, as expected under CGN.

Figure 4 shows the distribution of distinct TLS signatures per IP and per covering /24. As shown in Figure 4(a) , almost all non M-IPs in the reference set have fewer than 10 distinct TLS signatures. In the /24s we observe a much narrower distribution of distinct signatures for non M-IPs, around 1,000 per prefix. CGN/24s distinguish themselves with nearly every IP in the prefix showing higher count of distinct TLS signatures than prefixes containing VPN and proxy IPs.

4.3 Source ports per IP

In HTTP2+, any single client can re-use a single connection to make multiple requests, that could be logged separately. To mitigate protocol bias in our feature extraction, we look to source ports since every new connection binds to a port; two different connections from a source to a destination must use two distinct source ports.

We plot the cumulative distribution of unique source ports per IP for each IP type in Figure 5. Each of the shared IP types uses a greater number of distinct source ports compared to non-shared IPs. For example, about 50% of the proxy and VPN IPs have more than 100 source ports per IP, and 15% of the proxy IPs have more than 1,000 unique ports. CGN IPs have a greater number of source ports than non M-IPs, but they generally have smaller number of source ports compared to proxy and VPN IPs. The aggregate source port numbers per /24 prefix are plotted in Figure 5(b) , where CGN prefixes tend to use higher numbers of source ports than the rest of shared IP types, which is in contrast to distributions per-IP. Interestingly, 40% of non-M-IPs have a higher number of source ports than /24 prefixes that host VPNs and proxies.

4.4 RTT Variability

Another distinctive characteristic of M-IPs in the dataset compared to other IPs is the TCP RTTs' variability between source IPs and the CDN servers. Figure 6 shows the distribution of the standard deviation of TCP RTTs per IP in our dataset. CGN IPs show the highest TCP RTT variability

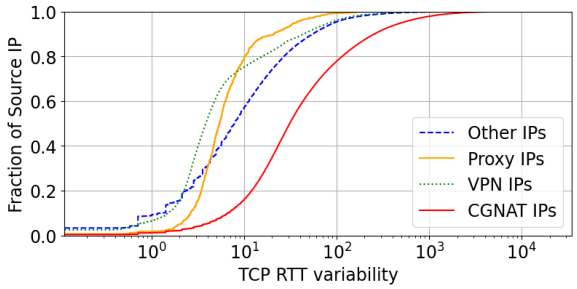


Figure 6: Cumulative distribution of TCP RTT standard deviation for different types of IPs. CGN IPs exhibit higher RTT variability compared to all other types of IPs.

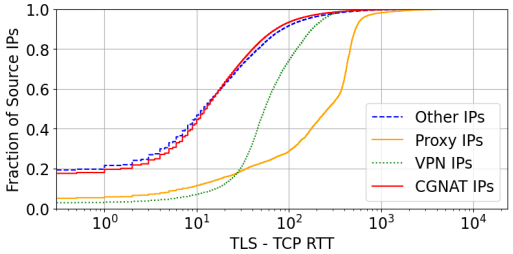


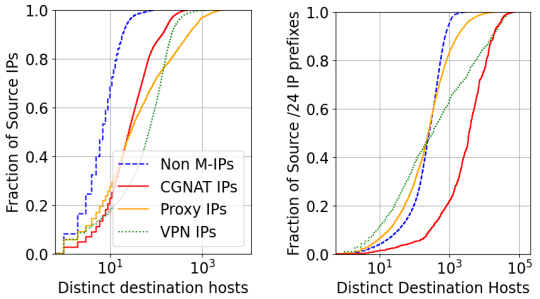
Figure 7: Cumulative distribution of the difference between the TLS handshake and TCP RTT for IPs in our dataset.

among the connections coming from the same IP, with 80% of CGN IPs exhibiting a standard deviation of over 30 ms. In contrast, only 20% of proxy IPs and IPs that are not shared (non M-IPs) have a standard deviation of more than 30 ms. The difference in RTT variability among different IP types is explained by their deployments. For example, CGN clients are typically located in wide geographical areas, such as in mobile networks, or have different last-mile characteristics to the CPE and its private network. In contrast, we expect clients behind typical non M-IPs to be on the same device, or behind a single CPE device; therefore, the RTT variability is expected to be lower.

VPN and proxy IPs exhibit the lowest RTT variability. This may seem counter-intuitive because users of VPNs and proxy can be anywhere in the world. However, VPN and proxy servers often terminate and initiate new TCP connections on behalf of the client. The services are also often hosted in well-connected cloud providers (e.g., AWS, OVH, Azure) with small RTTs to the CDN. Also, the VPN or proxy service is likely shared with other processes that initiate connections from the server itself, which we confirm via the HTTP request log for a small sample of VPN and proxy IPs.

4.5 TLS/TCP RTT Difference

Chiapponi et al. [16, 17] explored a VPN and proxy detection scheme based on the differences between TCP connection RTTs and their TLS handshake times. The intuition is based on the observation that various proxies initiate new TCP connections to servers, while the TLS connection is end-to-end between the true client and the target server. In a direct connection, we expect similar TCP RTT and TLS RTT because they originate from the same node. We therefore compute the TLS/TCP RTT difference as a potential feature that can help us distinguish proxy IPs from CGN IPs or IPs that are not shared. Figure 7 shows the difference between TCP RTT and TLS handshake is consistently higher among VPN and especially proxy IPs in the reference dataset, with more than 100ms of difference for 30% of proxy IPs and up to 1000ms in a few cases—results that are consistent with the findings in [16, 17].



(a) Destination hostnames per source IP. **(b) Destination hostnames per source /24 prefix.**

Figure 8: Cumulative distribution of unique destination hosts per IP for different types of IPs.

4.6 Diversity of Destination Hosts

Past research has demonstrated that the browsing histories of individual users are highly distinctive to the degree that individual users can be identified only based on the web pages they request [10, 54]. Based on this observation, we count the number of unique destination hosts visited from each IP type, and its covering /24, in Figure 8. In our data, the M-IP types visit a higher number of destination hosts. Comparing /24s, the CGNs distinguish themselves, while the median number of hosts is comparable for VPNs, proxies and non M-IPs.

5 Inference of Shared IP Addresses

We train a supervised classifier to infer whether an IP address is multi-user and, if so, whether it corresponds to a CGN or a VPN/Proxy. Due to their similar behaviors, VPN and proxy IPs are merged into a single class, while distinguishing CGNs is operationally more relevant. The classifier uses features derived from the HTTP request logs (§4), computed at both the IP level and the covering /24 prefix level. Features are computed over 24-hour windows to increase sample density while limiting noise from IP re-allocation. We split the reference dataset into training (70%) and testing (30%) sets, ensuring no /24 prefix appears in both sets. The classifier is implemented using XGBoost with class-weighting to account for label imbalance. A complete description of the feature set, preprocessing, and implementation is in Appendix B.

5.1 Model Validation

On the held-out test set, the classifier achieves an accuracy of 0.98, a weighted F1-score of 0.97, and a log loss of 0.04.

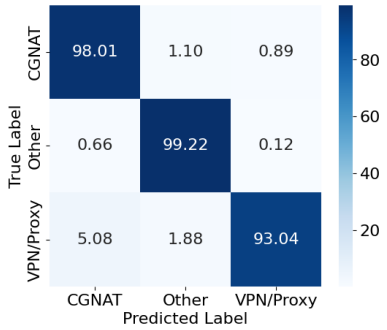


Figure 9: Confusion matrix of IP type inferences.

Figure 9 shows the confusion matrix. Most errors correspond to VPN/Proxy IPs misclassified as CGNs, which we attribute to mixed-use /24 prefixes that blur RTT-based signals.

We further evaluate robustness using 10-fold cross-validation with disjoint /24 prefixes, achieving a macro-average AUC of 0.99 (std. 0.006) and a log loss of 0.04 (std. 0.01). To assess external validity, we test the classifier on an independent dataset of 368 live Socks4/Socks5 proxy IPs, of which 96% are correctly classified. We also verify that several ISPs publicly known to deploy CGN are inferred as such by our model. For example, most of the Starlink prefixes are consistently classified as CGNs, matching public observations of their CGN deployment [73]. We also obtain independent confirmations from several ISPs for prefixes classified as CGNs.

Note that checks draw on evidence disjoint from the reference labels used for training (live proxy captures, public operator disclosures, and direct ISP confirmations), so agreement here indicates that the classifier recovers generalizable address-sharing behavior rather than memorizing noise particular to any single reference source in Table 3(Appendix H).

5.2 Feature Importance

Using permutation-based feature importance, which measures the drop in performance when a feature is randomly permuted, we find that /24 prefix-level features are more informative than IP-level features (Figure 10). The most important feature at the /24 level is RTT variability. While this result may appear counter-intuitive given that predictions are made per IP, it reflects the fact that CGNs typically apply uniformly across contiguous IPs in a prefix [67], allowing the model to distinguish CGN from VPN/Proxy IPs that may share similar IP-level behavior. Consistent with this interpretation, a model trained with only IP-level features reduces the macro F1-score to 0.57 and increases log loss to 0.46, underscoring the importance of prefix aggregation.

5.3 Results

We apply our classification model to CDN request logs on 29 April 2025. We identify over 40 million CGN IPs spanning 14,319 ASNs, and over 2 million VPN/Proxy IPs across 22,696 ASNs. The CGN IPs count is approximately 20× larger, but VPN/Proxy IPs are spread across a broader AS set.

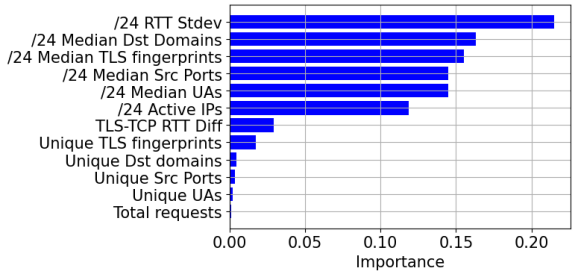


Figure 10: Feature importance value. Prefix-level features are significantly more informative than IP-level features for predicting IP address type.

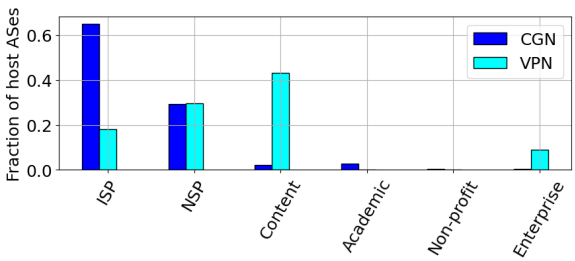


Figure 11: AS types that host M-IP addresses by the different M-IP types. CGN gateways are predominantly hosted in ISPs; VPN and proxy servers in Content ASes.

To assess temporal stability, we repeat the analysis on a snapshot collected one month earlier (26 March 2025) and compute the symmetric difference between the two snapshots. Despite modest differences in raw IP counts (1.25% fewer CGN IPs and 11% fewer VPN/Proxy IPs), we observe symmetric differences of 10.2% for CGN IPs and 13.6% for VPN/Proxy IPs. At the ASN level, the differences are 3.1% for CGN deployments and 7.5% for ASes hosting VPN/Proxy IPs. These results indicate that classifications evolve non-trivially over time, highlighting the need for regular updates in operational settings. Daily inference stability over a full month is examined in Appendix G.

Figure 11 shows the distribution of inferred CGN and VPN/Proxy IPs by AS type, based on self-reported PeeringDB records [56]. CGN IPs are predominantly associated with ISPs and NSPs, while VPN/Proxy IPs are mainly hosted by Content ASes (e.g., cloud providers and CDNs), as well as NSPs and, to a lesser extent, ISPs (e.g., residential proxies). We note that PeeringDB reports AS types for only 21,330 ASes at the time of writing; ASes without PeeringDB records are excluded from this analysis.

6 A Public Reproducible Model

We next ask whether the CGN signal learned from proprietary CDN logs can be reproduced using only public measurements. We build a parallel, narrower model using M-Lab NDT data [49] and transport-layer features from NDT `tcp-info` [50]. Unlike the CDN model, this public model performs only CGN/non-CGN inference over IP space visible to M-Lab; it is not designed to identify VPN or proxy exits. This

difference reflects the available signals: CDN logs expose application-level diversity, whereas NDT exposes transport-layer state from speed-test flows. We use the public model to test whether shared-address aggregation leaves a reproducible transport-layer signature consistent with the CDN-based inference.[4]

6.1 Feature Design and Inference Results

The features from M-Lab's `tcpinfo` dataset capture three dimensions of behavior that emerge when many clients share a single public IP: (i) temporal activity spread, where a prefix active across many time windows is more likely to represent continuously shared infrastructure than an intermittent single-user address; (ii) port usage, where broad active ephemeral-port use reflects multiple connections from many hosts behind the same public address; (iii) PMTU/MSS and receive-MSS variability, which exposes heterogeneity in device stacks and access-network paths aggregated behind a shared address; (iv) destination diversity, where many distinct test endpoints or destination addresses are consistent with traffic from multiple users rather than one household or device. Detailed definitions are provided in Appendix C.

Classifier and Validation. We use XGBoost [15], a gradient-boosted decision tree classifier, with feature vectors derived from all observed /24 prefixes in the M-Lab dataset. The resulting public classifier is binary: it separates CGN from non-CGN prefixes, rather than attempting the CDN model's CGN/VPN-proxy/non-M-IP taxonomy. We evaluate the model using standard cross-validation and by comparing its predictions against CGN prefixes inferred independently from CDN request logs, assessing whether a public, shareable model recovers signals consistent with the CDN-based model.

To evaluate agreement with the CDN-based inference, we restrict comparison to the overlapping address space: only 64.6% of CDN-inferred prefixes are observed in M-Lab, reflecting known biases in speed-test coverage (§D.1). Within this subset, the public model achieves an AUC of 0.97 and an overall accuracy of 94%, indicating strong agreement between two independently trained inference pipelines operating on disjoint data sources.

Interpretability. To understand how the classifier reaches its decisions, we quantify the contribution of each feature to individual predictions. In contrast to the CDN-based model that relies on application-layer diversity and prefix-level aggregation to reveal address sharing, the public model recovers analogous signals purely from transport-layer behavior.

The most informative signal captures how persistently an IP prefix remains active throughout the day—a strong proxy for address sharing, since CGNs tend to be continuously in use. Other influential signals reflect the breadth of ephemeral ports allocated, indicating multiplexed connections from multiple hosts, and variability in transport parameters. These features serve as proxies for device and access-network heterogeneity behind a single public IP. More details about the information derived from each feature can be found in Appendix D.

[4]The CDN request logs contain client-to-edge TCP RTT but not `tcp-info`-like statistics such as MSS, PMTU, or congestion-control counters. Although the CDN collects TCP/IP telemetry in sampled-packet datasets and flow logs, those samples are independent of the HTTP request sample and cannot be aligned at the request or connection level, so they cannot be incorporated into our proprietary model.

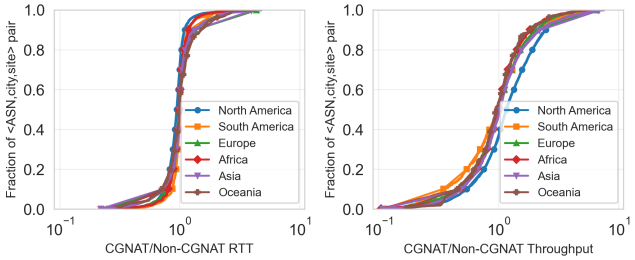


Figure 12: CDF of RTT and throughput ratio between inferred CGN and non-CGN/24 prefixes, stratified by continent. Positive throughput values indicate better throughput under CGN, while positive RTT values indicate worse latency.

7 Internet-wide Insights

In this section we connect inferred CGN IPs to their operational environment, impact, and global distribution.

7.1 CGNs and User Performance

A natural assumption is that CGNs negatively impact performance for users behind them. Our observations suggest a more nuanced reality. In the M-Lab data, differences in connection characteristics between CGN and non-CGN are notably restrained: CGN and non-CGN traffic are similar, and in some cases slightly better.

Figure 12 compares CGN versus non-CGN within the same network context: For each inferred CGN/24 and a non-CGN/24 from the same ISP and city, we compute the difference in median RTT and median throughput observed in M-Lab measurements across one week in December 2025. Each difference is normalized by the non-CGN baseline (so values represent a percent change, with 0 meaning no difference). We stratify these normalized differences by continent to separate broad geographic effects from CGN effects. Across regions, the median RTT difference is close to zero and the median throughput difference is modest, indicating that in the typical case CGN does not materially change user-visible performance. In some regions (e.g., North America), CGN prefixes show slightly better average performance. One possible explanation may be that networks optimize routing more aggressively where most of their users are located, which could make performance for those destinations appear better. We leave a careful study of this hypothesis to future work.

7.2 CGNs and IPv6 Deployment.

CGN is often framed as a transitional response to IPv4 scarcity, prompting us to examine how inferred CGN deployment relates to IPv6 deployment at the organization level. We group organizations into deciles by the number of IPv6 prefixes they announce; each decile shows the distribution of

inferred CGN/24 prefixes (fig. 21 in Appendix F). Most organizations in every decile have no CGN deployment, and typically small CGN numbers when present. However, the largest CGN deployments appear almost exclusively among organizations with the largest IPv6 footprints. This suggests that IPv6 and CGN are not substitutes: IPv6 can accommodate network growth but IPv4 cannot, which concentrates traffic behind NAT devices to ensure reachability with IPv4-only networks. Rather than being optional, extensive investment in IPv6 networks may force organizations to also operate sizeable CGN pools. We validate this interpretation in Appendix F by showing that networks with larger IPv6 address footprints also carry a higher fraction of IPv6 traffic.

7.3 Feature Granularity: CGN Prefix Sizes

Inference in our system is performed using /24 prefix–level features. In earlier models, we augmented these features with AS-level attributes, expecting that organizational context would improve accuracy. Surprisingly, this approach did not outperform a model based only on /24 features.

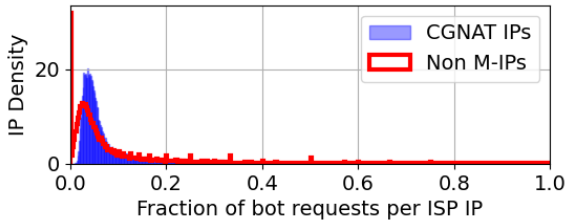
The intuition is that CGN deployment aligns more closely with address-space boundaries than with AS boundaries. When a CGN is deployed, connections are dynamically mapped to any of a shared pool of public addresses, producing similar aggregate behavior across the addresses in the pool. As a result, the behavior of the population behind a CGN is coherent at the prefix level, not necessarily at the individual address. Aggregating features at the AS level would therefore pool together address blocks that serve different roles, while address-level inference would be too sparse and unstable.

A limitation of our /24 granularity is that CGN deployments may occupy only part of a /24, or may span multiple adjacent /24s. In principle, our approach could be made prefix-adaptive: starting from a /24, the classifier could recursively split into smaller subprefixes, such as /25s or /26s, and stop when the resulting subprefixes exhibit statistically consistent CGN behavior. This approach would trade finer localization for substantially higher cost: features would need to be recomputed over many candidate subprefixes, and each split would reduce the number of observations available for inference. We leave such adaptive-granularity inference to future work. Under our current sampling regime, however, we expect large CGN pools to remain visible at /24 granularity, since high-volume shared addresses are more likely to appear as heavy hitters in the CDN's observations.

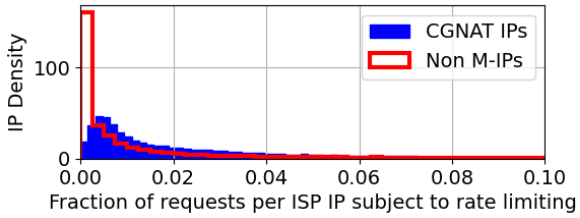
7.4 Implications for Global South?

One of the early motivations of this work was to understand if or by how much our (lack of) knowledge about large shared IP addresses might hide a bias along socio-economic boundaries – and whether an action on those IP addresses may disproportionately affect global populations. A necessary first step is to locate them.

In Figure 15(a) we plot the number of inferred CGNs on a country-level map according to the CDN's geolocation service (Appendix A). The top-3 countries by raw CGN counts are



(a) Bot scores are comparable at the median; non-CGNs have a longer tail.



(b) Fractions of requests per CGN IPs are three times more likely to be rate-limited by customers of the CDN.

Figure 13: Bot scores and rate-limiting on connections from CGN and non-CGN IPs at various ISPs.

perhaps unsurprising: India's large user base relies heavily on mobile cellular networks; the USA has a large IP address count overall with $1,000\times$ more addresses than human users; and Brazil networks are known to rely heavily on CGN.

In contrast, a clearer picture emerges in Figure 15(b) , which shows the fraction of a country's inferred CGN IPs over all IPs observed in the country. Taking this perspective highlights stark geodiversity of CGNs in terms of importance. For example, the USA and Western Europe are diminishing in importance. Conversely, this proportional view indicates that much of Africa, Central and South-East Asia, and parts of South America, rely so much on CGNs as to be otherwise invisible in raw counts. These are also high-population regions with potentially fastest-growing user-penetration rates.

Any imbalance in CGN locations and user-bases can lead to biases and disproportional outcomes from IP-level actions for security and abuse mitigations.

8 Security and Abuse Mitigation

In this section, we share insights and experience from the CDN to connect CGNs with their operational implications. The core issue is fate-sharing for users who have no agency over their use of CGNs: When many unrelated users share one source IP, any IP-level action affects the whole population behind that address. Operational deployment details are provided in Appendix E.

8.1 IP-level Classification and Actions

A strong motivation for the CDN to identify CGNs is to admit benign clients that would otherwise be filtered or rate-limited by IP-level characterizations. We first analyze the CGN and non-CGN IPs' distributions of bot 'scores', as assigned by the CDN's internal classification system to each request [12],

The bot score label is ingested by the CDN's application-layer firewall; it is also available to the CDN's customers who can use it to create or configure filtering policies.

Figure 13(a) shows the fraction of bot requests from inferred CGN and non-CGN IPs at ISPs[5]. The median bot request rate is virtually equal (4.8% for CGN IPs and 4.7% for non CGN IPs), yet mean bot requests from CGN IPs is nearly half of the non-CGNs (7% versus 13.1%), suggesting different distributions despite similar statistics. The observation is supported by a much wider spread of bot request rates from non-CGN IPs (up to 100% for some), with about 10% of the non-CGN IPs having no requests tagged as bots. In contrast, the bot request rates for CGN IPs are concentrated almost entirely below 15%. We surmise this is because any non-CGN IP is likely to be either more bot or more human, while CGNs mix both but are dominated by human requests.

Customer-configured actions in Figure 13(b) show a slightly different picture when they ingest the bot score before initiating firewall actions, rate-limits, and CAPTCHA-like challenges. Where we previously observed that CGN *requests* are less likely to be labelled as bots, customers appear to use the CGN IPs to rate-limit 3× more than non-CGNs. It therefore appears that CGN users are subjected to more aggressive filtering policies by site administrators because multiple clients are sharing the same public IP address. Based on these observations, the CDN is fine-tuning its data representation and guidance to educate its customers.

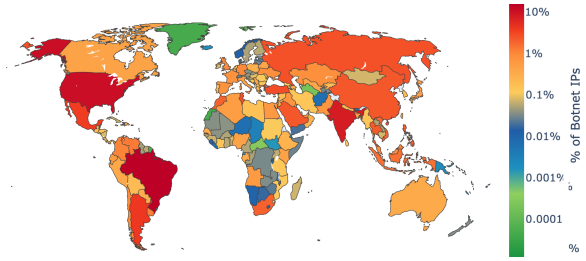
8.2 Internal Incident Response

During efforts to operationalize our design, inferred CGN IPs helped diagnose an incident involving the CDN's zero-trust (ZT) service. ZT proxies enterprise customers' connections through egress addresses in the CDN's own IP space. Those egress addresses are used regardless of whether the customer is visiting an external site or a service that is itself served through the CDN's regular request-processing pipeline. This means that when ZT users accessed CDN-protected services, the regular CDN protection stack observed requests sourced from ZT egress IPs.

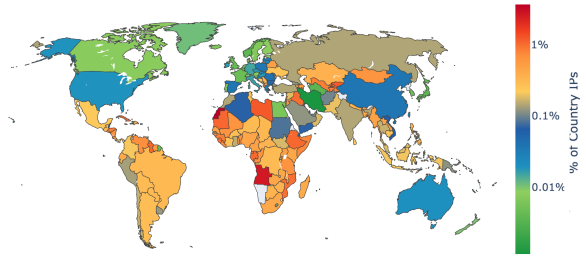
After a DDoS protection software release, requests from ZT egress IPs began experiencing elevated blocking and rate limiting in this regular CDN pipeline. The release had passed internal tests and appeared to behave as intended, so there were two plausible explanations: either ZT traffic had some unusual or previously hidden behavior, or the new protections were reacting to the high volume of benign traffic concentrated behind shared egress addresses.

Inferred CGN IPs provided a comparison class. Like ZT egress IPs, CGN IPs aggregate traffic from many users behind source addresses that users do not choose and cannot easily avoid. Operators found that requests from inferred CGN IPs were affected in a similar way, suggesting that the mitigation was responding to traffic concentration at shared egress addresses rather than to behavior unique to ZT customers.

[5]AS types taken from PeeringDB [56]. We focus on ISPs to compare networks of similar type (cloud/hosting providers do not typically operate CGNs)

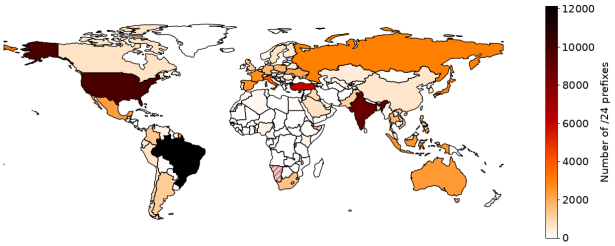


(a) Per-country portion of all CGNs emanating botnet traffic.

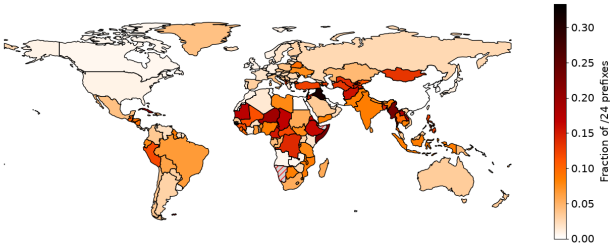


(b) CGNs with botnet traffic as ratio of country's active IPs.

Figure 14: Country distributions of inferred CGN IPs that emanate botnet traffic (§8.3). (a) CGNs in BR, IN, PH, US, and MX, account for ≈48% of all CGNs emanating botnet traffic globally. (b) CGNs with botnet traffic in a country as a fraction of each country's observed IP space highlights risks to most of Africa and others: IP-level responses triggered by botnet sources in five countries could disproportionally affect countries, even continents, with far less IPv4 space.



(a) Number CGN IPs per country.



(b) Fraction of each country's observed IPs that are CGNs.

Figure 15: Distributions of inferred CGN IPs per country (§7.4).

8.3 Botnets as a thought experiment

We describe a worst-case thought experiment prompted by a large 31.4 Tbps botnet attack on the CDN's infrastructure [80], and informed by CGN data that suggests geographic

differences are inevitable for IP-induced actions. Our analysis assumes that the growth of botnets eventually may force even the largest service providers to block source IPs. We confirmed that the CDN's protections do not rely on source IP (§7.1), but also that the CDN does keep a record of source IPs emanating botnet traffic—which suggests the CDN could use source IPs for worst-case possibilities.

Using the inferred CGN data, we ask what role CGN IPs play in botnets, and potential effects on users if blocked. We were given access to a list of IP addresses that participated in various DoS attacks during January 2026, each labeled with the time and duration of the attack. Consistently, ≈40% of all source IPs emanating botnet traffic were among our inferred CGN IPs. In Figure 14(a) , we plot each country's proportion of those CGN IPs emanating botnet traffic. Among the CGN IPs comprising ≈40% of attack sources, almost half (≈48%) are in BR, IN, PH, US, and MX. On appearance, the global distribution of CGNs emanating botnet traffic roughly follows raw CGN counts (Fig.15(a)).

A very different picture emerges when plotting each country's proportion of CGN IPs in botnets and the country's observable IP space, as shown in Figure 14(b) . This view represents the proportion of a country's blocked IP space if CGNs were blocked. From this perspective, the most affected regions are in Africa, parts of Central and South-East Asia, and Central and South America. In these regions, at least one order of magnitude greater proportions of IP space are affected by heavy-handed IP-level actions.

Our findings highlight a tension that is likely to intensify: The Internet increasingly relies on shared addressing for scalability, while its defensive mechanisms remain fundamentally constrained by address-level abstractions as 'nuclear' solutions. Making large-scale IP sharing visible and measurable does not resolve the tension but does make it explicit and possible to reason about—and potentially protect against, using both technical and non-technical tools. IPv6 is not immune since IP address volumes and rotations mean that prefixes up to /64 are the de facto unit of record.

9 Related Research

Prior work has shown that estimating the number of users behind an IP address can substantially improve system performance and security. HTTP cookies were used to infer application users per IP to improve ad click-fraud at Google [51]. Follow-up work [72] demonstrated similar gains in identifying machine-generated traffic from ad impression logs. Those systems use application identifiers available in a single service to estimate user counts for abuse detection. Our goal and signals differ: there are no cookies or account identifiers and we do not estimate exact user numbers behind each address. Instead, we infer broad shared-address classes across the Internet using aggregate IP- and prefix-level characteristics.

Several studies have focused on detecting CGN deployment. Lutu et al. [43] introduced NATRevelio, which uses active measurements from edge clients to identify hosts behind NAT and CGN gateways; subsequent work refined these heuristics and validated them using RIPE Atlas and FCC MBA measurements [44]. Richter et al. [62] combined Netalyzr measurements with BitTorrent DHT crawling to identify CGN usage, reporting 421 deploying ASes. Salamatian et al. [67] used dense traceroute measurements to identify CGN deployments that rely on squatted address space, distinguishing CGN hosts based on topological properties of a large corpus of traceroutes. This approach requires extensive traceroute measurements originating from end-users and is not adequate for our setup. Most closely related to our work, Livadariu et al. [39] presented a large-scale passive methodology to detect CGN-used /24 IPv4 blocks by combining repeated IP observations across BitTorrent identifiers, network telescope traffic, and M-Lab NDT logs, ultimately identifying roughly 4K CGN-deploying ASes. We also leverage NDT data, but with no need to repeat IP observations over long time windows. Instead, we extract transport-layer features that capture large-scale address sharing, enabling inference at short time scales and under heavy sampling.

10 Conclusion

In this paper we develop and validate techniques for detecting large multi-user IP addresses by combining public labels with behavior features extracted from logs of a global CDN. An unexpected dominance of /24 feature importance over individual IPs highlights the subtleties of accurately characterizing shared IPs. The findings are significant to understand user activity, improve performance metrics estimation, and enhance threat detection or mitigation strategies. To make these results reproducible beyond proprietary logs, we build a parallel, fully public pipeline using M-Lab NDT `tcp-info` that we share with the community upon acceptance. Despite relying on different features and a different measurement context, the public model recovers the CDN-based inferences on the address space seen by both, showing that the underlying technique is generalizable.

We use the inferred IPs to revisit common assumptions about CGN at Internet scale. We demonstrate that IPv6 does not eliminate CGN and that the largest IPv6 footprints often coincide with the largest CGN deployments, suggesting that IPv6 adoption can increase reliance on CGN pools for IPv4 reachability. We also find that average user-visible performance differences between CGN and non-CGN prefixes are modest, but that major operational risk arises in exceptional cases, like fate-sharing risks when botnets also share the CGNs. These risks are more acute in regions where address sharing and risks of device infection are most present. Results suggest the impact most affects users in the Global South, underscoring the need for policies and operational practices that avoid disproportionate collateral damage.

References

[1] Brian Aitken. 2013. *Report on the implications of carrier grade network address translators.* Technical Report. Technical report, InterConnect Communications.
[2] John Althouse. 2023. JA4+ Network Fingerprinting. FoxIO Blog. `https://blog.foxio.io/ja4+-network-fingerprinting`.

[3] Blake Anderson and David McGrew. 2019. TLS Beyond the Browser: Combining End Host and Network Data to Understand Application Behavior. In *Proceedings of the Internet Measurement Conference* (Amsterdam, Netherlands) *(IMC '19)*. Association for Computing Machinery, New York, NY, USA, 379–392. doi:10.1145/3355369.3355601

[4] APNIC. 2022. Visible ASNs: Customer Populations (Est.). https://stats.labs.apnic.net/aspop/.

[5] APNIC. 2025. Resource ranges by RIR. https://www.apnic.net/manage-ip/manage-resources/address-status/by-rir/.

[6] Todd Arnold, Jia He, Weifan Jiang, Matt Calder, Italo Cunha, Vasileios Giotsas, and Ethan Katz-Bassett. 2020. Cloud provider connectivity in the flat internet. In *Proceedings of the ACM Internet Measurement Conference*. Association for Computing Machinery, New York, NY, USA, 230–246.

[7] Auth0. 2022. Attack Protection. https://auth0.com/docs/secure/attack-protection. Accessed 01/2022.

[8] Barracuda Networks. 2022. Web Application Firewall. https://campus.barracuda.com/product/webapplicationfirewall/doc/4259869/enabling-brute-force-protection/. Accessed 01/2022.

[9] David Benjamin. 2020. Applying Generate Random Extensions And Sustain Extensibility (GREASE) to TLS Extensibility. RFC 8701. doi:10.17487/RFC8701

[10] Sarah Bird, Ilana Segall, and Martin Lopatka. 2020. Replication: Why we still can't browse in peace: On the uniqueness and reidentifiability of web browsing histories. In *Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020)*. USENIX Association, Berkeley, CA, USA, 489–503.

[11] Jeremias Blendin, Fabrice Bendfeldt, Ingmar Poese, Boris Koldehofe, and Oliver Hohlfeld. 2018. Dissecting Apple's Meta-CDN during an iOS Update. In *Proceedings of the Internet Measurement Conference 2018*. Association for Computing Machinery, New York, NY, USA, 408–414.

[12] Alex Bocharov. 2020. Cloudflare Bot Management: machine learning and more. https://blog.cloudflare.com/cloudflare-bot-management-machine-learning-and-more/.

[13] Tomasz Bujlow, Valentín Carela-Español, Josep Solé-Pareta, and Pere Barlet-Ros. 2017. A Survey on Web Tracking: Mechanisms, Implications, and Defenses. *Proc. IEEE* 105, 8 (2017), 1476–1510.

[14] CGNAT 2026. CGNAT Detection and Analysis Repository. https://github.com/Burdantes/cgnat-detection.

[15] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. Association for Computing Machinery, New York, NY, USA, 785–794.

[16] Elisa Chiapponi, Marc Dacier, and Olivier Thonnard. 2023. Inside residential ip proxies: Lessons learned from large measurement campaigns. In *2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, IEEE, Piscataway, NJ, USA, 501–512.

[17] Elisa Chiapponi, Marc Dacier, Olivier Thonnard, Mohamed Fangar, and Vincent Rigal. 2022. Badpass: Bots taking advantage of proxy as a service. In *International Conference on Information Security Practice and Experience*. Springer, Springer, Cham, Switzerland, 327–344.

[18] Cloudflare. [n. d.]. Privacy and Data Protection. https://www.cloudflare.com/trust-hub/privacy-and-data-protection/.

[19] Cloudflare. 2023. Rate limiting best practices. https://developers.cloudflare.com/waf/rate-limiting-rules/best-practices.

[20] Cloudflare. 2026. Cloudflare Radar. https://radar.cloudflare.com/. Accessed: 2026-02-06.

[21] WikiMedia Commons. 2022. Apple iCloud Private Relay. https://meta.wikimedia.org/wiki/Apple_iCloud_Private_Relay.

[22] World Wide Web Consortium. 2011. Definition of User Agent. https://www.w3.org/WAI/UA/work/wiki/Definition_of_User_Agent.

[23] Michelle Cotton, Leo Vegoda, Ron Bonica, and Brian Haberman. 2013. Special-Purpose IP Address Registries. RFC 6890. https://rfc-editor.org/rfc/rfc6890.txt

[24] Omar Darwich, Hugo Rimlinger, Milo Dreyfus, Matthieu Gouel, and Kevin Vermeulen. 2023. Replication: Towards a Publicly Available Internet scale IP Geolocation Dataset. In *Proceedings of the 2023 ACM on Internet Measurement Conference*. Association for Computing Machinery, New York, NY, USA, 1–15.

[25] David Dittrich and Erin Kenneally. 2012. *The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research*. Technical Report. U.S. Department of Homeland Security.

[26] Alexandre Dulaunoy. 2023. HTTP/2 Rapid Reset DDoS Attack. GitHub Gist. https://gist.github.com/adulau/7c2bfb8e9cdbe4b35a5e131c66a0c088

[27] Zakir Durumeric, David Adrian, Ariana Mirian, Michael Bailey, and J. Alex Halderman. 2015. A search engine backed by Internet-wide scanning. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*. Association for Computing Machinery, New York, NY, USA, 542–553.

[28] Zakir Durumeric, Eric Wustrow, and J. Alex Halderman. 2013. ZMap: Fast Internet-wide Scanning and Its Security Applications.. In *USENIX Security Symposium*, Vol. 8. USENIX Association, Berkeley, CA, USA, 47–53.

[29] European Commission. 2021. Broadband Connectivity in the Digital Economy and Society Index. https://digital-strategy.ec.europa.eu/en/policies/desi-connectivity.

[30] Federal Communications Commission. 2021. Broadband Data Collection. https://www.fcc.gov/BroadbandData.

[31] Federal Communications Commission. 2021. Measuring Broadband America. https://www.fcc.gov/general/measuring-broadband-america.

[32] Roy T. Fielding and Julian Reschke. 2014. Hypertext Transfer Protocol (HTTP/1.1): Semantics and Content. RFC 7231. doi:10.17487/RFC7231

[33] Romain Fontugne, Anant Shah, and Kenjiro Cho. 2020. Persistent Last-mile Congestion: Not so Uncommon. In *Proceedings of the ACM Internet Measurement Conference*. Association for Computing Machinery, New York, NY, USA, 420–427.

[34] GitLab. 2026. Rate limits. https://docs.gitlab.com/ee/security/rate_limits.html.

[35] Robert Graham. 2023. MASSCAN: Mass IP port scanner. https://github.com/robertdavidgraham/masscan.

[36] 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*. Association for Computing Machinery, New York, NY, USA, 33–43.

[37] Sheng Jiang, Brian E. Carpenter, and Dayong Guo. 2011. An Incremental Carrier-Grade NAT (CGN) for IPv6 Transition. RFC 6264. doi:10.17487/RFC6264

[38] Eric Kinnear, Patrick McManus, Tommy Pauly, Tanya Verma, and Christopher A. Wood. 2022. *Oblivious DNS Over HTTPS*. Internet-Draft draft-pauly-dprive-oblivious-doh-11. Internet Engineering Task Force. https://datatracker.ietf.org/doc/html/draft-pauly-dprive-oblivious-doh-11 Work in Progress.

[39] 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*. IEEE, IEEE, Piscataway, NJ, USA, 2249–2257.

[40] Matthew Luckie, Bradley Huffaker, Alexander Marder, Zachary Bischof, Marianne Fletcher, and KC Claffy. 2021. Learning to extract geographic information from internet router hostnames. In *Proceedings of the 17th International Conference on emerging Networking EXperiments and Technologies*. Association for Computing Machinery, New York, NY, USA, 440–453.

[41] Matthew Luckie, Alexander Marder, Marianne Fletcher, Bradley Huffaker, and KC Claffy. 2020. Learning to Extract and Use ASNs in Hostnames. In *Proceedings of the ACM Internet Measurement Conference*. Association for Computing Machinery, New York, NY, USA, 386–392.

[42] Scott M. Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Long Beach, California, USA) *(NIPS'17)*. Curran Associates Inc., Red Hook, NY, USA, 4768–4777.

[43] Andra Lutu, Marcelo Bagnulo, Amogh Dhamdhere, and Kimberly C Claffy. 2016. NAT Revelio: Detecting NAT444 in the ISP. In *International Conference on Passive and Active Network Measurement*. Springer, Springer, Cham, Switzerland, 149–161.

[44] Anna Maria Mandalari, Andra Lutu, Amogh Dhamdhere, Marcelo Bagnulo, and KC Claffy. 2017. Tracking the Big NAT across Europe and the U.S.

[45] John Matherly. 2015. Complete Guide to Shodan. Shodan, LLC.

[46] Maxmind, Inc. 2024. GeoIP and GeoIPLite. https://dev.maxmind.com/geoip.
[47] Patrick McManus. 2018. Bootstrapping WebSockets with HTTP/2. RFC 8441. doi:10.17487/RFC8441
[48] Measurement Lab. 2025. M-Lab Acceptable Use Policy. https://www.measurementlab.net/aup/.
[49] Measurement Lab. 2026. Measurement Lab. https://www.measurementlab.net/.
[50] Measurement Lab (M-Lab). 2026. TCP INFO. https://www.measurementlab.net/tests/tcp-info/. Accessed: 2026-01-19.
[51] Ahmed Metwally and Matt Paduano. 2011. Estimating the number of users behind ip addresses for combating abusive traffic. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. Association for Computing Machinery, New York, NY, USA, 249–257.
[52] National Institute of Standards and Technology. 2023. CVE-2023-44487. NIST National Vulnerability Database. https://nvd.nist.gov/vuln/detail/CVE-2023-44487
[53] Rishab Nithyanand, Rachee Singh, Shinyoung Cho, and Phillipa Gill. 2016. Holding all the ASes: Identifying and Circumventing the Pitfalls of AS-aware Tor Client Design. arXiv:1605.03596. arXiv:1605.03596 https://arxiv.org/abs/1605.03596
[54] Lukasz Olejnik, Claude Castelluccia, and Artur Janc. 2014. On the uniqueness of web browsing history patterns. *annals of telecommunications-annales des télécommunications* 69 (2014), 63–74.
[55] 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)*. Association for Computing Machinery, New York, NY, USA, 55–70. doi:10.1145/3386367.3431314
[56] PeeringDB. 2022. https://peeringdb.com.
[57] Ingmar Poese, Steve Uhlig, Mohamed Ali Kaafar, Benoit Donnet, and Bamba Gueye. 2011. IP geolocation databases: Unreliable? *ACM SIGCOMM Computer Communication Review* 41, 2 (2011), 53–56.
[58] Jon Postel. 1981. Internet Protocol. RFC 791, https://rfc-editor.org/rfc/rfc791.txt. doi:10.17487/RFC0791
[59] 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*. Association for Computing Machinery, New York, NY, USA, 360–369.
[60] Rapid7. 2013. Scanning All The Things. https://www.rapid7.com/blog/post/2013/09/26/internet-wide-probing-rapid7-sonar/.
[61] Yakov Rekhter, Jon Crowcroft, and Brian E. Carpenter. 1997. IPv4 Address Behaviour Today. RFC 2101. https://www.rfc-editor.org/info/rfc2101
[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*. Association for Computing Machinery, New York, NY, USA, 15 pages.
[63] RIPE NCC. 2026. RIPE Atlas. https://atlas.ripe.net/.
[64] Hannah Ritchie. 2019. How many internet users does each country have? https://ourworldindata.org/how-many-internet-users-does-each-country-have.
[65] Erik Rye, Robert Beverly, and KC Claffy. 2021. Follow the Scent: Defeating IPv6 Prefix Rotation Privacy. In *Proceedings of the 21st ACM Internet Measurement Conference*. Association for Computing Machinery, New York, NY, USA, 739–752.
[66] Loqman Salamatian, Calvin Ardi, Vasileios Giotsas, Matt Calder, Ethan Katz-Bassett, and Todd Arnold. 2024. What's in the Dataset? Unboxing the APNIC per AS User Population Dataset. In *Proceedings of the 2024 ACM on Internet Measurement Conference* (Madrid, Spain) *(IMC '24)*. Association for Computing Machinery, New York, NY, USA, 165–182. doi:10.1145/3646547.3688411
[67] Loqman Salamatian, Todd Arnold, Italo Cunha, Jiangchen Zhu, Yunfan Zhang, Ethan Katz Bassett, and Matt Calder. 2023. Who squats IPv4 Addresses? SIGCOMM Comput. Commun. Rev.
[68] Salesforce Engineering. 2019. TLS Fingerprinting with JA3 and JA3S. Salesforce Blog. https://engineering.salesforce.com/tls-fingerprinting-with-ja3-and-ja3s-247362855967/.
[69] scikit-learn. 2024. scikit-learn: Machine Learning in Python. https://scikit-learn.org/stable/.
[70] Sudheesh Singanamalla, Suphanat Chunhapanya, Jonathan Hoyland, Marek Vavruša, Tanya Verma, Peter Wu, Marwan Fayed, Kurtis Heimerl, Nick Sullivan, and Christopher Wood. 2021. Oblivious DNS over HTTPS (ODoH): A Practical Privacy Enhancement to DNS. *Proceedings on Privacy Enhancing Technologies* 2021, 4 (2021), 575–592.
[71] Rachee Singh, Arun Dunna, and Phillipa Gill. 2018. Characterizing the deployment and performance of multi-cdns. In *Proceedings of the Internet Measurement Conference 2018*. Association for Computing Machinery, New York, NY, USA, 168–174.
[72] Fabio Soldo and Ahmed Metwally. 2012. Traffic anomaly detection based on the IP size distribution. In *2012 Proceedings IEEE INFOCOM*. IEEE, IEEE, Piscataway, NJ, USA, 2005–2013.
[73] Starlink Help Center. 2026. *What IP address does Starlink provide?* Starlink (SpaceX). https://starlink.com/support/article/1192f3ef-2a17-31d9-261a-a59d215629f4 Accessed: 2026-02-02.
[74] Brett Stone-Gross, Marco Cova, Lorenzo Cavallaro, Bob Gilbert, Martin Szydlowski, Richard Kemmerer, Christopher Kruegel, and Giovanni Vigna. 2009. Your botnet is my botnet: analysis of a botnet takeover. In *Proceedings of the 16th ACM conference on Computer and communications security*. Association for Computing Machinery, New York, NY, USA, 635–647.
[75] Emily Taylor. 2013. *MC/159 Report on the Implications of Carrier Grade Network Address Translators Final Report*. Technical Report. Ofcom, United Kingdom.
[76] Martin Thomson and Christopher A. Wood. 2022. *Oblivious HTTP*. Internet-Draft draft-ietf-ohai-ohttp-01. Internet Engineering Task Force. https://datatracker.ietf.org/doc/html/draft-ietf-ohai-ohttp-01 Work in Progress.
[77] Antoine Vastel. 2022. How to Use Machine Learning to Detect Residential Proxies. https://datadome.co/bot-management-protection/how-to-use-machine-learning-to-detect-residential-proxies/.
[78] World Bank and International Telecommunication Union. 2020. Individuals using the Internet. https://data.worldbank.org/indicator/IT.NET.USER.ZS/.
[79] XGBoost. 2024. XGBoost Python Package. https://xgboost.readthedocs.io/en/stable/python/index.html.
[80] Omer Yoachimik, Jorge Pacheco, and Cloudforce One. 2026. 2025 Q4 DDoS threat report: A record-setting 31.4 Tbps attack caps a year of massive DDoS assaults. https://blog.cloudflare.com/ddos-threat-report-2025-q4/.
[81] Jan Žorž, Sander Steffann, Primož Dražumerič, Mark Townsley, Andrew Alston, Gert Doering, Jordi Palet, Jen Linkova, Luis Balbinot, Kevin Meynell, and Lee Howard. 2017. Best Current Operational Practice for Operators: IPv6 prefix assignment for end-users - persistent vs non-persistent, and what size to choose. https://www.ripe.net/publications/docs/ripe-690.

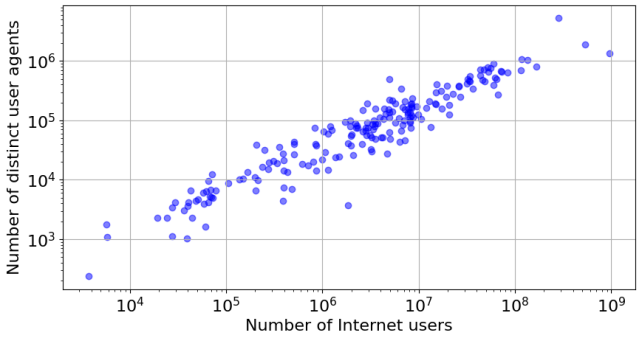
Appendix

Appendices are supporting material that has not been peer-reviewed.

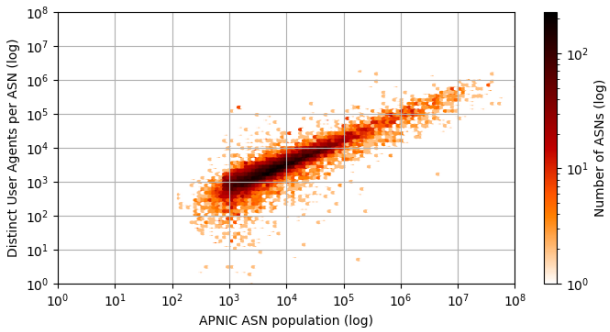
A User-agents as Proxies for Population

We first geolocate IPs to countries with the *single radius* (also called *shortest ping*) technique [24]. We take the minimum recorded latencies among all connections to client IPs seen by each PoP. An IP is assigned to the same country as the PoP for RTTs of up to 2ms. For IPs that cannot be geolocated using the TCP RTTs, we use the free Maxmind geolocation database [46]. Past research has shown that geolocation databases, including Maxmind, are reliable for country-level geolocation of end hosts [57].

We then use the geolocation to count the unique UA strings per country over a full day, and we compare them to the estimated number of Internet users per country according to the statistics published by the World Bank and the ITU



(a) The number of unique User-Agent HTTP header strings from each country compared to World Bank estimated number of Internet users per country (log-log scale).



(b) The number of unique User-Agent strings per ASN compared to the estimated number of users per ASN (log-log scale). The number of ASNs is indicated by colour density.

Figure 16: User-agent and population counts.

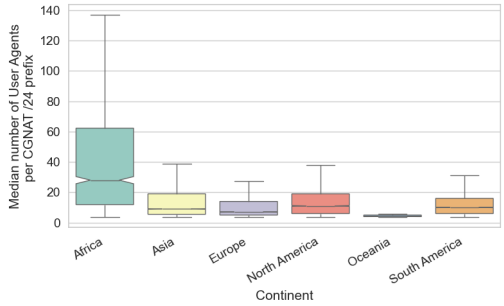


Figure 17: Distribution of median number of distinct User Agents per IP of each /24 prefix inferred to be CGN, per continent.

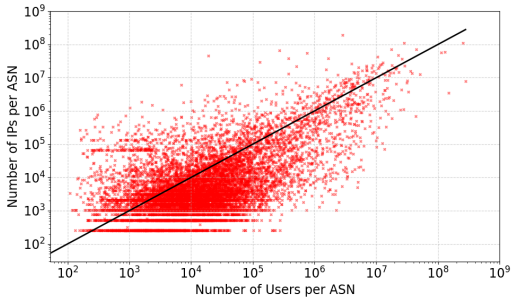


Figure 18: Number of users per ASN (APNIC estimates) compared to number of available address space for each ASN (as observed in BGP RouteViews), for ASNs that have been inferred to deploy CGN in at least one of their prefixes.

(International Telecommunication Union) [78]. As shown in Figure 16(a) , UAs have a strong monotonic relationship with the number of Internet users in a country, with a Spearman's rank correlation coefficient of 0.941. In contrast, we observe a linear association evaluated by a Pearson correlation coefficient of 0.631. The observation that UAs provide a reasonable approximation of user population is an indication that the CDN has a relatively balanced view of the Internet-connected population; but it should not be considered an exact measurement of user populations since there is non-trivial variation around the line of best fit.

We also evaluate UA strings as potential approximations for client populations at the autonomous system (AS) level using APNIC user estimates per AS [4] in Figure 16. Here, too, we observe a strong positive linear correlation, with a Pearson correlation coefficient of 0.895, as well as a Spearman's rank correlation coefficient of 0.892. The linear association appears to be stronger at the ASN-level compared to country-level, providing evidence that the view we obtained from the CDN does not suffer from significant biases with regards to how the users' AS networks are connected to the CDN.

It should be noted that the APNIC dataset contains only 23,510 (approximately half of ASNs observed); these are presumably the ASNs in which clients have interacted with the APNIC measurement platform.

Figure 17 shows the distribution of distinct user agents (UAs) per IP across /24 prefixes inferred to be part of a CGN deployment, separated by continent. Notably, Africa has a much higher number of UAs than other regions, suggesting more end-users share the same IP in African ASNs. This observation aligns with the map in Figure 1, even though our classifier does not use the number of IPs per ASN or APNIC user estimates as features. Hence, it serves as a solid sanity check for the consistency of our inferences.

Figure 18 shows that while most CGN-deploying ASNs have fewer available IP addresses than their user base (according to APNIC), many still deploy CGN despite having more addresses than users. Presumably, these ASNs provide additional services beyond broadband, preventing them from dedicating their entire address pool to subscribers. Thus, while the ratio of users to IPs can indicate potential CGN usage, it is not sufficient by itself to confirm deployment.

B Model training details

Based on the analysis of the characteristics of different types of shared IPs in the previous section, we compiled the following features per IP and then per covering /24 prefix:

- Median number of requests per IP, and the median number of requests per IP in the same prefix.
- Number of unique *User Agents* per IP, and the median number of unique User Agents per IP in the prefix.

- Number of unique *Source Ports*, and the median number of unique Source Ports per IP in the prefix
- Number of unique *TLS Fingerprints* per IP, and the median number of unique TLS fingerprints per IP in the prefix.
- Number of unique requested *domain names* per IP, and the median number of domain names per IP in the prefix.
- The median *TCP - TLS RTT difference* for HTTPs requests per IP, and the median TCP RTT standard deviation per IP in the prefix.
- The total number of requests.

The above features from the HTTP requests log are computed for a 24 hour time interval to derive a high number of data points, while also shielding from possible IP address re-allocation (e.g. due to DHCP) noise in our statistics. After computing the above features we split our reference dataset into training (70%) and testing (30%) sets. In the split, we ensure that the training and the testing datasets have no overlapping /24 prefixes, so that the address space used for training and testing is disjoint.

We train a multi-class classification model using XGBoost, which enhances the gradient boosting technique by using optimized decision trees as base learners [15]. It achieves high performance using advanced regularization features to prevent over-fitting, and a gradient descent optimization method that considers both first- and second-order derivatives to minimize the loss function effectively.

We implement the classifier using the open-source Python XGBoost package [79] and the scikit-learn library [69]. We use the `multi:softprob` objective which uses the softmax function but also assigns a probability for each class of IPs, to evaluate the loss of the classifier. Since the IP types in the reference dataset are imbalanced, we use the XGBoost's `sample_weight` parameter to assign weights to class instances based on the observed class frequencies. For each prediction, we assign an IP to the class with the highest predicted probability.

C More details about the M-Lab model

We extract the following set of transport-layer features from M-Lab's `tcpinfo` dataset to capture behavioral signals indicative of CGN deployments. These features fall into three broad categories: (1) ephemeral port allocation, (2) PMTU and MSS variability, and (3) activity scope and temporal spread. Below, we describe each category and the rationale behind its inclusion.

Ephemeral Port Allocation Behavior. CGN devices multiplex traffic from many users through a small number of external IP addresses, leading to distinctive ephemeral port usage patterns:

(1) **Port range**: Defined as the difference between the maximum and minimum ephemeral ports observed for a given IP or /24 prefix. A large range suggests that the address is handling many concurrent flows, typical of CGN nodes.

Table 2: Average cross-validation performance of the public XGBoost CGN classifier.

Metric	Value
AUC	0.972
Accuracy	0.940
Precision	0.943
Recall	0.781
F-1 Score	0.854

(2) **Port density**: The ratio of the number of unique source ports used to the size of the port range. High density implies a densely packed allocation pattern, consistent with NAT state tables under pressure.
(3) **Port reuse ratio**: The number of total connections divided by the number of unique ports used. High reuse indicates repeated port allocation, often seen under CGN-induced port exhaustion.

PMTU and MSS Variability. Transport-layer parameters such as PMTU (Path Maximum Transmission Unit) and MSS (Maximum Segment Size) reflect the characteristics of the local client network and negotiated end-to-end path. When multiple heterogeneous devices share a public IP, variability across these metrics increases:

(1) **PMTU/MSS standard deviation**: High standard deviation signals that connections from the same public IP exhibit heterogeneous MTU-related behavior, often due to device or access-network diversity.
(2) **Average and maximum MSS**: These contextualize the spread and highlight significant outliers compared to typical residential values.

Residential users typically exhibit consistent PMTU and MSS values, while CGN addresses aggregate traffic from devices with diverse link-layer and OS constraints.

Table 2 reports cross-validation results.

Temporal and Spatial Activity. CGN IPs often exhibit broader usage over time and across destination addresses:

(1) **Time buckets**: The number of hourly intervals in which the IP or /24 appears. Persistent activity across many hours suggests IP sharing.
(2) **Destination diversity**: The number of distinct destination IPs contacted from a given source prefix. Higher diversity is expected when many users operate behind the same NAT address.

D Interpretability via SHAP.

To understand how the classifier reaches its decisions, we use SHAP (SHapley Additive exPlanations) to quantify the contribution of each feature to individual predictions [42]. The top predictive signals (`agg_time_buckets` (Time buckets), `agg_port_range` (Port range), and `max_rcvmss` (Maximum MSS)) capture sustained activity, wide ephemeral port allocation, and device diversity, respectively. As shown in Figure 20,

Figure 19: Geographic distribution of CGN prefix recovery using M-Lab's public data. Darker shades indicate higher recovery rates; white regions correspond to areas with insufficient data for estimation.

high values along these dimensions strongly increase the likelihood of a CGN classification, while the remaining features contribute finer-grained distinctions.

D.1 Geographic Coverage of CGN Recovery.

Figure 19 illustrates the fraction of CGN prefixes that can be recovered using public M-Lab data, aggregated at the country level. While recovery is high across regions with many users running speed-tests, such as North America, Western Europe, and Australia, coverage drops sharply in parts of Africa, South and Southeast Asia, and Latin America. In many of these regions, public measurement data are either sparse or entirely absent, leaving substantial blind spots. Ironically, these are also the regions where CGN usage is expected to be most prevalent, due to limited IPv4 availability and slower IPv6 adoption. As a result, the very networks most dependent on CGN are also those least visible to public measurement platforms.

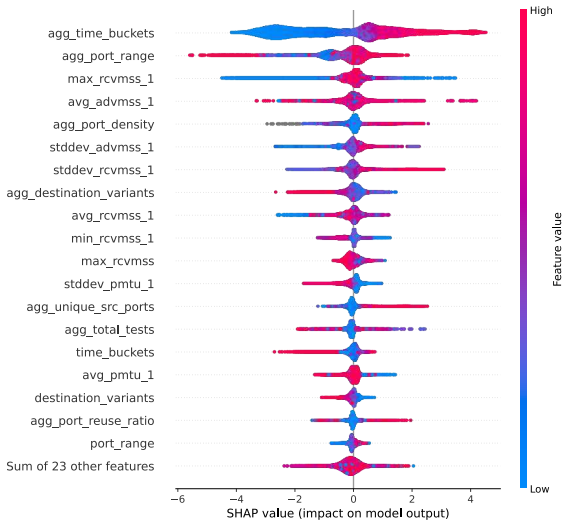


Figure 20: SHAP beeswarm plot showing the impact of each feature on the model's CGN predictions. High `agg_time_buckets`, `agg_port_range`, and `max_rcvmss` values are the most consistent indicators of CGN behavior.

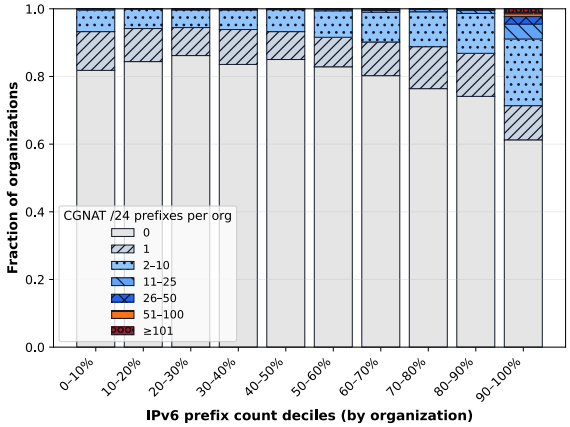


Figure 21: CGN deployment versus IPv6 deployment at the organization level. Large CGN deployments occur primarily among organizations with large IPv6 footprints.

E Operational feedback

All reporting with CDN data in this work derives from a standalone proof-of-concept (PoC). Queries and compute were out of the critical path and production services. However, the CDN is also transitioning our inference technique to production with notable changes for testing.

Query limits. The logs we used are also used for various production services and customer data analytics. In that environment, queries are subject to strict limits on execution time and volume. To stay within limits in the PoC we elected to initiate 0.1% sample queries, which produces a substantial subset of inferred results from the 1.0% sample.

Compute and execution. As a frame of reference, PoC runs on a 4 vCPUs (AMD EPYC 7642, 2.3 GHz), 32 GB memory machine. Under these constraints, a complete inference cycle over 24 hours of data (e.g., querying, local aggregation, and IP classification) takes five to seven hours, depending on database load and the number of IPs observed that day.

F Do large IPv6 footprints correspond to IPv6-carrying traffic?

A potential concern in §7.2 is that an organization's large *advertised* IPv6 footprint (e.g., many IPv6 prefixes in BGP) might not reflect substantial *IPv6 usage* in practice. To sanity-check that our "IPv6 footprint" axis is a reasonable proxy for IPv6-carrying traffic, we compare prefix counts against independent estimates of IPv6 traffic share.

For each organization, we attach an estimate of IPv6 usage from Cloudflare Radar [20] by mapping the organization's ASNs to Radar's per-AS IPv6 traffic metric and taking the maximum across ASNs within the organization.[6] Figure 22 plots the resulting distribution of IPv6 traffic share across the

[6]Taking the maximum is conservative for our purpose: it tests whether organizations with large IPv6 footprints include ASNs that carry substantial IPv6 traffic. Results are qualitatively unchanged using mean or traffic-weighted aggregates where available.

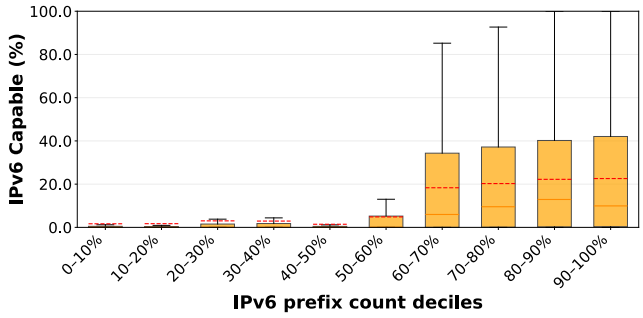


Figure 22: IPv6 adoption by organizations according to Cloudflare Radar, grouped by IPv6 prefix footprint. While organizations with larger IPv6 footprints exhibit higher median IPv6 adoption, even the top deciles retain substantial IPv4 traffic, indicating continued reliance on IPv4 connectivity despite extensive IPv6 deployment.

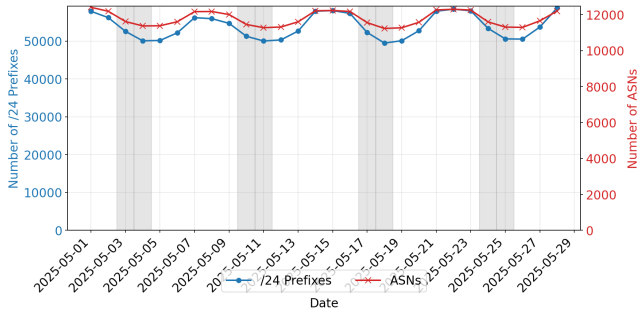


Figure 23: Number of inferred /24 prefixes inferred to be part of CGN deployments, and number ASNs with inferred CGN deployments over a period of a month.

IPv6-prefix deciles[7]. Overall, organizations in the upper IPv6-prefix deciles exhibit consistently higher IPv6 traffic shares than those in the lower deciles, indicating that large IPv6 footprints typically correspond to meaningful IPv6-carrying traffic rather than dormant advertisements. This supports interpreting "large IPv6 footprint" as a proxy for real IPv6 deployment in §7.2.

G Longitudinal CGN Inferences

Figure 23 shows the number of daily inferences produced between 2025-05-01 and 2025-05-29 by our service that is deployed in the CDN to detect CGN IPs. We observe that the number of inferences is relative stable, with small variations especially during weekends (marked with gray shadow). Note that the times in the plot are UTC, so weekend has a small drift for networks based in North America which hosts the majority of CGN clients. This behaviour is largely expected given that during holidays the CDN generally receives lower volume of traffic, and therefore some IP blocks either become inactive or they produce much lower volume of requests that is not captured by the 0.1% sampling rate used by our production service. Note that even with the 0.1% the daily number of detected ASNs deploying CGN is 2x higher than what previously reported for longer measurement periods [39, 67].

H Robustness of the Reference Dataset

Every label in our reference dataset is derived from public data that may be noisy. Scans for VPN and proxy endpoints reach server-facing infrastructure, and miss endpoints that are dormant. PTR keyword matches depend on operator naming conventions that vary across regions and languages, and may be stale or misleading. Traceroute heuristics can be inaccurate when transient congestion inflates the RTTs that NATRevelio treats as an access-network signal, or when the CGN deployment is unconventional. Therefore, we treat each label as a starting point that behavioural evidence must corroborate, and we sanity check each reference data source as summarized in Table 3.

[7]We suppress outliers using an IQR filter for readability

Table 3: Reference-label robustness. For each source in our reference dataset, the expected source of labeling noise and the mitigation we apply.

Reference source	Expected source of noise	Mitigation
VPN/Proxy endpoints from ipinfo.io, Censys, Shodan	Server-facing scan bias. Dormant or lightly used endpoints that don't serve user traffic are missing.	VPN/Proxy IPs that are dormant are outside of the scope of our work, since for the purposes of our work we focus on IPs that are indeed multi-user. We merge VPN and proxy into one class to absorb inter-source tag disagreement.
CGNAT from RIPE Atlas traceroutes	Non-canonical CGNAT deployments may use private, squatted, or publicly routed IPs instead of shared IP space. Transient RTT inflation due to congestion or rate-limited ICMP may mimic the access-network RTT jump.	We require the NATRevelio signals [43] to persist over a period of 10 days. We require the hops between the probe and its GRA to be either in the shared IP space, or the IP space of the probe's AS.
CGNAT IPs derived by DNS PTR keyword scan	Operator naming conventions vary by region/language, PTRs may be stale or aspirational, or may be used to deliberately obfuscate the purpose of the IP.	We adopted our keyword set from validated prior work [39, 67]. We manually verified the 50 highest-volume ASes (spanning 17 countries); all were confirmed via ISP-website documentation or user-forum posts. We also cross-referenced PTR data against traceroute-derived CGNAT IPs and Shodan scans to ensure agreement across reference sources.
Non-M-IP (true negatives) from direct-path traceroutes	An IP on a direct path may be re-assigned to a CGNAT deployment.	Require a direct hop from the probe's public IP to persist over a period of 10 days.