Who Holds the Steering Wheel? Opacity and Consolidation in CDN Replica Selection

Rashna Kumar
Northwestern University, USA

Fabián E. Bustamante
Northwestern University, USA

Marcel Flores
Edgio, USA

Abstract

Replica selection, the process by which CDNs decide which server delivers content, has become a hidden lever of power and fragility in today's Internet. Most users, operators, and policymakers remain blind to how these decisions are made, yet they shape latency, resilience, and sovereignty at global scale. DNS resolver centralization further distorts this function, concentrating influence in the hands of a few global actors.

We present the first methodology to systematically infer CDN replica selection strategies at global scale, enabling third-party visibility into opaque steering mechanisms. Using RIPE Atlas probes and a geographically distributed set of DNS resolvers, we construct latency fingerprints that distinguish DNS-based, anycast, and regional anycast deployments. We validate our approach on well-documented global providers before applying it to a diverse set of 17 global and regional CDNs serving the top 1,000 websites across 19 countries, covering 66% of Internet users. We also examine ECS support and its interaction with DNS-based redirection.

Our findings show that DNS-based steering remains the dominant approach, used by over 70% of CDNs and responsible for most delivered bytes, yet regional variation and mixed strategies complicate the picture. These results highlight replica selection not only as a technical optimization, but as a sociotechnical risk: opaque steering decisions, particularly among regional CDNs, amplify the effects of resolver consolidation and shape the Internet's future resilience and control.

2012 ACM Subject Classification Networks → Location based services; Networks → Network measurement

Keywords and phrases CDN, replica selection, Anycast, Regional Anycast, DNS, consolidation, centralization

Digital Object Identifier 10.4230/OASIcs.NINeS.2026.23

Acknowledgements This work was supported by National Science Foundation grant CNS-2246475. The opinions, findings, conclusions, and recommen-dations expressed are those of the authors and do not necessarily reflect the views of the funding sources. The authors would like to thank the reviewers for their valuable feedback and our shepherd for their comments on early versions of this manuscript.

1 Introduction

Replica selection is a core function of content delivery networks (CDNs), determining which server ultimately delivers content in response to a user request [14, 13, 41]. Although replica selection directly shapes latency, availability, and fault tolerance, its operation is largely opaque to external observers. Requests are steered according to internal policies and resolver interactions that are not visible to end users, researchers, or policymakers, yet these decisions collectively influence traffic localization, cross-border flows, and failure behavior at global scale.

1st New Ideas in Networked Systems (NINeS 2026).
Editors: Katerina J. Argyraki and Aurojit Panda; Article No. 23; pp. 23:1–23:24
OpenAccess Series in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

This lack of visibility is increasingly consequential as DNS resolution infrastructure consolidates. Many CDNs rely on DNS-based redirection to map users to replicas [47, 61], making their steering behavior dependent on a small number of large, globally deployed DNS resolvers [36, 38]. As a result, replica selection decisions may reflect resolver placement and policy rather than client location, particularly when regional infrastructure is sparse or unavailable. These effects are difficult to observe directly, yet they can influence where traffic is served from and how delivery systems respond to failures.

At a high level, CDNs implement replica selection using a small set of mechanisms. The most common approaches include DNS-based redirection [14], IP anycast [13], and variants such as EDNS0 Client Subnet (ECS) [21] and regional anycast [65]. These mechanisms differ in how much control they delegate to resolvers, how they balance locality and scalability, and how steering decisions vary across geographic scope.

A substantial body of prior work has examined these mechanisms, primarily through the lens of performance optimization [42, 47, 25, 12, 7, 32, 34, 57, 60, 52, 53, 1, 26, 61]. These studies characterize latency tradeoffs, cache efficiency, and mapping accuracy, and often focus on individual CDNs or controlled deployments. What remains less well understood is which replica selection strategies CDNs actually deploy in practice across regions and across the long tail of providers, particularly from an unprivileged, external vantage point.

In this paper, we develop a lightweight methodology to infer CDN replica selection strategies as they are experienced by clients. Our key observation is that different steering mechanisms produce distinct latency patterns when the same content is resolved via DNS resolvers operating at different geographic scopes. We operationalize this observation using latency measurements from globally distributed vantage points (RIPE Atlas probes) resolving CDN-hosted resources through carefully selected DNS resolvers. The methodology is unprivileged and scalable, requiring no cooperation from CDNs and no access to client traffic or internal configuration.

We validate the approach on three well-characterized CDNs – Akamai, Cloudflare, and Edgio[1] – each employing a different steering strategy. We then apply it at global scale, analyzing the landing pages of the top 1,000 websites in 19 countries, covering approximately 66% of the global Internet user population. This allows us to infer the replica selection strategies used by 17 CDNs across regions and customer bases.

Our results show that DNS-based steering dominates both in prevalence and in delivered bytes, particularly in North America and Oceania, while anycast-based approaches are more common in Europe. We further evaluate the role of ECS in fine-grained redirection and discuss how reliance on resolver-based steering interacts with consolidation in the DNS ecosystem.

1.0.0.1 Contributions.

Taken together, our work makes the following contributions:

- We reframe replica selection not only as a performance optimization, but as a sociotechnical risk shaped by opacity and consolidation.
- We present a lightweight, unprivileged methodology that infers whether a CDN uses DNS-based, global anycast, or regional anycast steering, leveraging latency fingerprints from diverse resolvers.

[1] As of 2025, Edgio is no longer operating commercially but provides a useful reference deployment for this study.

- We apply this approach at global scale, analyzing 17 CDNs that serve the top 1,000 websites across 19 countries and cover ≈66% of Internet users.
- We show that DNS-based steering dominates, and that its dependence on resolvers makes it acutely sensitive to consolidation – with implications for resilience, sovereignty, and control.

Together, these contributions provide new visibility into CDN steering and reveal how resolver consolidation reshapes who controls replica selection and, ultimately, the resilience of the Internet's delivery infrastructure.

2 Background

At the heart of CDN operation lies *replica selection*: the decision of which server should serve a user's request. This choice is not just a matter of performance tuning, but also a lever of control. Each steering mechanism encodes implicit assumptions about who gets to decide where traffic goes, and how much visibility others have into that process. We focus here on the three dominant approaches: DNS-based redirection, anycast, and regional anycast, as well as the EDNS Client Subnet (ECS) extension that exposes tradeoffs between privacy and control.

DNS-based redirection. The most common approach is to use the client's recursive resolver as a proxy for its location [46]. The CDN's authoritative DNS server maps the resolver's IP to a nearby replica, returning that IP to the client. This method is simple and flexible, but it gives resolvers outsized influence. When clients rely on remote or centralized resolvers (e.g., Google's 8.8.8.8, Cloudflare's 1.1.1.1), steering decisions are effectively delegated to a third party, and locality may be distorted [54]. Thus, DNS-based redirection turns resolvers into hidden intermediaries in traffic routing.

ECS as a visibility–control tradeoff. To mitigate locality mismatches, CDNs may leverage the EDNS0 Client Subnet (ECS) extension [21], which allows resolvers to embed a truncated portion of the client's IP prefix in DNS queries. ECS sharpens geolocation, but at the cost of privacy (exposing client networks) and operational complexity (larger caches, inconsistent deployments). Some providers (Akamai, Google [3, 29]) embrace it, while others, like Cloudflare [19], reject it on principle. ECS thus illustrates the broader tension: gaining finer control over steering often reduces transparency and weakens privacy guarantees.

Anycast. In anycast [7, 13], a single IP prefix is announced from multiple geographic sites, and BGP routing policies determine which site receives the traffic. Anycast simplifies operation and provides natural failover, but it also shifts steering authority away from CDNs to intermediate networks. CDNs cannot directly dictate which replica serves which client; instead, outcomes depend on opaque inter-domain routing policies. While this decentralization can aid resilience, it leaves both operators and users blind to how routing decisions are actually made.

Regional anycast. Regional anycast [44, 65] blends the two paradigms. A CDN advertises multiple anycast prefixes, each scoped to a region, and relies on DNS to assign clients to the right prefix. Within the region, BGP decides the specific replica. This approach provides more fine-grained control than global anycast, while still leveraging routing simplicity inside each region. But here too, authority is shared: CDNs decide the region, networks decide the path, and neither is fully transparent.

In practice, CDNs mix these approaches – serving static assets via anycast, video streams via DNS, and specialized services via regional deployments. Selection strategies vary across domains, customers, and regions, and their opacity complicates efforts to assess resilience or diagnose failures. While prior work has measured mapping efficiency and geolocation accuracy [66, 25, 7, 13], most has treated replica selection purely as a technical optimization problem. What is missing, and what motivates this work, is a systematic understanding of *who holds the steering wheel* across CDNs, and how their opaque choices interact with the consolidation of DNS resolvers to shape global resilience and control.

3 Methodology

Replica selection is deliberately opaque: CDNs rarely disclose how requests are steered, and the policies that determine where traffic goes are hidden in DNS responses and BGP routing. Our methodology is designed to pierce this opacity. By treating DNS resolvers as experimental levers—resolving the same CDN-hosted resource through resolvers at progressively broader geographic scope – we induce conditions under which hidden selection mechanisms leave observable traces in client-perceived latency.

In this section, we describe our methodology to experimentally identify a CDN's replica selection mechanism. The approach uses RIPE Atlas probes [9] as clients and a set of strategically selected DNS resolvers. Each client resolves CDN-hosted resources through multiple DNS resolvers and measures the latency to the assigned replicas. By analyzing how latency distributions vary with the resolver's geographic scope, we infer the CDN's underlying selection approach. The following paragraphs describe the approach in detail and illustrate its use.

3.1 Finding CDN Server Assignments

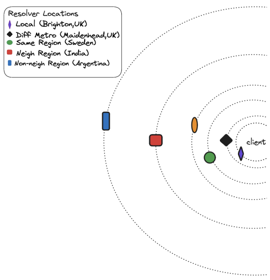


Figure 1 DNS resolvers at different distances from the client.

As a first step, Atlas clients resolve CDN-hosted domains using five DNS resolvers chosen to span distinct geographic *scopes* relative to the client: (*i*) the same metro area, (*ii*) a different metro in the same country, (*iii*) a different country within the same region, (*iv*) a neighboring region, and (*v*) a non-neighboring region (Fig. 1).

To illustrate, a client in Brighton, UK could use resolvers located in Brighton (same metro), Maidenhead (same country), Sweden (same region), India (neighboring region), and Argentina (non-neighboring region).[2]

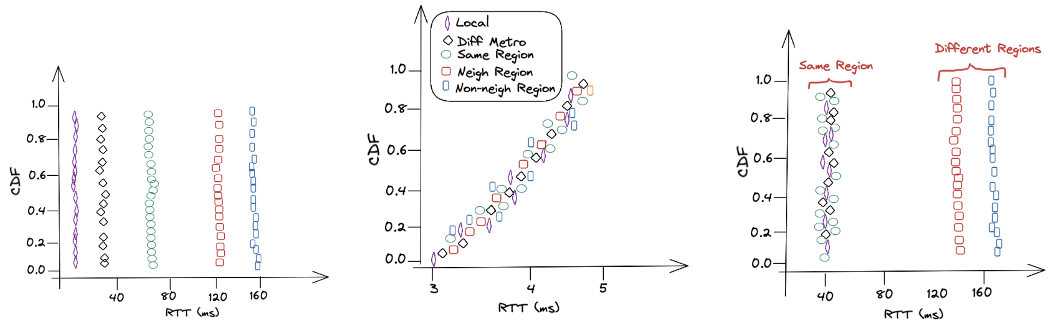
[2] While India and Argentina may appear equally distant from the UK geographically, our classification reflects observed network behavior, not physical proximity.

This process is repeated for all CDN-hosted resources of interest. The resulting latency distributions (one per resolver scope) form the basis of our inference methodology.

3.2 Canonical Examples for Illustration

We begin by illustrating the expected behavior of latency distributions for the three main replica selection strategies.

If a CDN uses DNS-based replica selection, latency distributions will shift based on the geographic scope of the resolver, i.e., resolvers farther from the client will yield replicas farther away (Fig.2a). If a CDN uses global anycast, all resolvers will result in similar assignments, regardless of their location (Fig.2b). Finally, if a CDN uses regional anycast, resolvers within the same region will yield similar (and low) latencies, while those from other regions will yield higher latencies (Fig. 2c).



(a) DNS-based **(b)** Anycast **(c)** Regional-Anycast

Figure 2 Expected CDF sets for DNS-based ($CRV[i] \approx 1$ for all i), Anycast ($CRV[i] \approx 0$) and Regional-Anycast replica selection ($CRV[1] \approx 0$, $CRV[3] \approx 1$).

3.3 Replica Selection Approaches

To systematically characterize replica selection behavior, we define the *Coefficient of Regionalization Vector (CRV)*. This is a 4-element vector measuring the distance between the latency distribution for the client's local resolver and those for the four progressively more distant resolver scopes.

Formally, $CRV[i]$ is computed as the Kolmogorov–Smirnov (KS) distance between the latency CDF from the local resolver and that from the $i+1$th resolver scope. KS distance measures the maximum vertical difference between two CDFs, ranges from 0 (identical) to 1 (completely different), and makes no assumptions about underlying distribution shape.

We focus primarily on $CRV[1]$ and $CRV[3]$, which capture inter-country and inter-regional variation. These provide the strongest signals for distinguishing between replica selection strategies. Other entries in the CRV may vary depending on CDN deployment granularity but are less decisive. We validated this in Sec. 5 using known CDN behaviors.

We use the CRV pattern to classify each CDN:

- DNS-based: High variation across all scopes → $CRV[i] \approx 1$
- Anycast: No variation → $CRV[i] \approx 0$
- Regional anycast: Low CRV within-region (e.g., $CRV[1] \approx 0$), high CRV across regions ($CRV[3] \approx 1$)

Regional scope mismatches (e.g., CDNs defining South America as one region) are mitigated by our multiple-scope design, which allows for fuzzy regional partitioning.

We note that earlier approaches have relied on IP prefix similarity to cluster CDN responses [55, 13], but these require threshold tuning and often struggle with overlapping or sparse deployments. Our latency-based technique avoids these issues by directly observing client-visible variation. As we show in Appendix A.4, latency yields a more robust and precise signal of selection behavior than IP prefix clustering.

3.4 Checking for ECS Support

To evaluate ECS support, we issue A-record queries for CDN domains and their CNAME chains using Google Public DNS, a resolver that supports EDNS0 Client Subnet.

Each query includes a spoofed /24 subnet via the edns-client-subnet option – one US-based (e.g., 74.125.0.0/24) and one Asia-based (e.g., 203.0.113.0/24). We then compare the IPs returned. If the responses differ and match the expected geolocation of the provided subnet, we infer that the CDN supports ECS and uses it for replica selection. We also check for non-zero EDNS scope values in the response to confirm ECS metadata is returned.

4 Detecting CDN Replica Selection

We now apply our methodology to experimentally identify the replica selection approaches used by CDNs to deliver Web content worldwide.

The design of this step is deliberate: by placing vantage points in representative countries and issuing controlled queries, we cut through the current opacity of replica selection and expose, at scale, the steering decisions that shape Internet performance and control.

Following the United Nations geo scheme [62] and published statistics on Internet penetration [56], we select vantage points to ensure that our analysis (*i*) captures a large fraction of the global Internet user population, and (*ii*) relies on VPN locations we can independently verify. Our chosen set spans 19 countries across all inhabited continents and covers 66% of the world's Internet users (Table 1). For each continent, the selected countries represent at least 50% (and up to 89%) of the region's Internet population.

Region	% of Internet Population
North America	89.3
Oceania	75.0
Asia	70.1
South America	60.0
Europe	60.0
Africa	50.1
Worldwide	66.0

Table 1 Share of Internet population by region.

For each of these countries, we collect the top 1,000 sites based on the Google CrUX dataset [27]. These sites serve as a proxy of the most commonly accessed content, and in aggregate across countries and continents serve as a good starting point for a global study of commonly used replica selection approaches.

We use popular commercial VPN providers [45, 31, 59] in each country to gather the resources of these top sites and the CDN(s) hosting them. To collect all resources of each website, we generate and utilizate the HTTP Archive (HAR) file for each website. To find the set of CDNs used by the resources of top sites in each country, we find the Canonical Name (CNAME) records for all website resources and map them to their corresponding CDNs using our self-populated CNAME-CDN map (which inherently includes both on-net and off-net resources), following the methodology of previous works [18, 64, 36, 39]. Additionally, to identify the CDNs hosting resources without a CNAME redirect, we compare the autonomous system number (ASN) of the unlabeled resource with those of popular CDNs [43, 64]. We validate the claimed locations of our VPN providers, by geolocating the VPN vantage points' IPs using IPInfo [33], a widely-used open geolocation database. Darwich et al. [22] report that 89% of the geolocation targets in IPInfo have an error of less than 40 km (i.e., within a city).

We use a single RIPE Atlas probe per country as a representative client vantage point. According to CDN operator guidance, replica selection is typically consistent within a country, making one probe sufficient to capture country-level behavior. We empirically validate this assumption by selecting stable, responsive resolvers in different geographic scopes and observing consistent selection outcomes.

For local resolution, the client probe uses its default resolver. For remote scopes, we select well-known Regional DNS resolvers [23] and verify that they are stable, responsive, and non-anycasted. We validate resolver locations using the same methodology applied to VPN servers.[3] The clients then issue DNS resolution queries for the resources to the selected set of DNS resolvers. We collect RTTs from the RIPE Atlas probes to the different CDN replicas assigned for each resolver scope by issuing a sequence of three ICMP pings. We record the minimum RTT from the repeated runs to ignore any transient spikes in the latency. Finally, we use the collected latency distributions to compute the CRV of each CDN for different regions.

5 Known CDNs for Validation

Before presenting our global analysis, we validate our methodology against three major CDNs – Akamai, Cloudflare, and Edgio – whose replica selection strategies are well documented in public sources [50]. These serve as ground-truth illustrations of three steering archetypes: DNS-based, global anycast, and regional anycast. For each CDN, we apply our methodology and compare the resulting latency patterns and CRV values to the expected behaviors illustrated in Fig. 2. Our goal here is not to re-characterize these well-studied providers, but to show that our approach faithfully recovers their known behaviors, thereby supporting its accuracy and generalizability.

5.1 Replica Selection by Akamai

We begin by applying our methodology to Akamai, a CDN well known for its use of DNS-based replica selection [14, 46, 48]. Using all resources hosted on Akamai, we compute CRV vectors across all countries and analyze the resulting latency distributions.

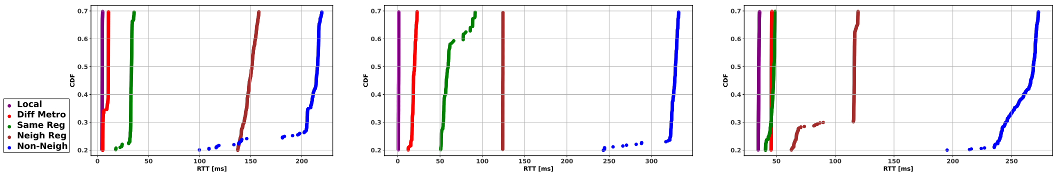
Figure 3 shows latency CDFs for Akamai-assigned replicas from clients in the UK, India, and the US, for each resolver scope. As expected for DNS-based selection (see Fig. 2a), the distributions are cleanly separated across scopes. Corresponding CRV values (Fig. 6a) show $CRV[1]$ and $CRV[3]$ between 0.9 and 1.0 – consistent with strong regional sensitivity in replica assignment.

Our analysis clearly confirms Akamai's use of DNS-based replica selection across all measured countries. In addition, Akamai supports ECS for replica assignment (see § 6.4).

[3] Sensitivity checks with alternate resolvers per scope showed no change in the identified mechanisms.

Vantage	Local	Diff Metro	Same Reg	Neigh Reg	Non-Neigh
UK	Brighton, UK	Maidenhead, UK	SE	IN	AR
IN	Mumbai, IN	Bengaluru, IN	PK	DE	AR
US	Indianapolis, US	San Francisco, US	CA	BR	IN
BR	Anápolis, BR	Cotia, BR	AR	CA	IN
AU	New Castle, AU	Sydney, AU	NZ	IN	AR
ZA	Piet Retief, ZA	Prestondale, ZA	ZW	IN	AR
TR	Antalya, TR	Istanbul, TR	IT	IN	AR
RU	Ryazan', RU	Polyany, RU	PL	IN	AR

Table 2 Resolver placements by scope for each vantage country.



(a) Akamai-United Kingdom **(b)** Akamai-India **(c)** Akamai-US

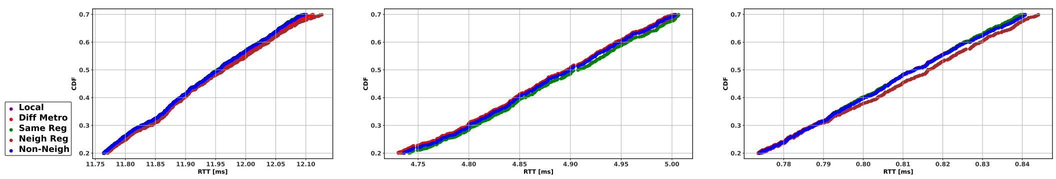
Figure 3 Akamai's latency fingerprints shift with resolver scope, confirming DNS-based steering across regions.

The US-based measurements also highlight a useful edge case: a vantage point in Indianapolis observes similar latencies to replicas assigned via a Canadian resolver and one in San Francisco, despite the >2,000 mi geographic separation. This underscores how CDN infrastructure topology, not just geography, shapes replica selection—and the value of comparative CRV analysis over naive reliance on distance or prefix overlap. It also illustrates a broader point: steering traffic across national borders, even when not strictly necessary for performance, raises questions of jurisdiction and data sovereignty that remain hidden under today's opaque redirection practices.

5.2 Replica Selection by Cloudflare

Cloudflare is a prominent CDN widely recognized for its use of anycast for replica selection [20, 49]. We apply our methodology to Cloudflare-hosted resources across Brazil, the UK, and India.

Figure 4 shows the latency CDFs for each resolver scope. In all three countries, the latency distributions largely overlap – closely matching the prototypical anycast pattern in Fig. 2b. The corresponding CRV values (Fig. 6b) show $CRV[1]$ and $CRV[3] \approx 0$, confirming Cloudflare's reliance on global anycast for replica selection.



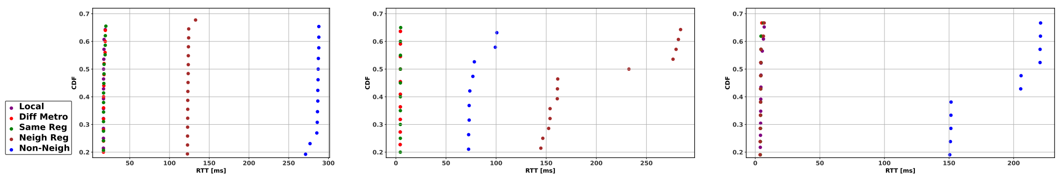
(a) Cloudflare-Brazil **(b)** Cloudflare-United Kingdom **(c)** Cloudflare-India

Figure 4 Cloudflare shows overlapping latency across resolvers, consistent with global Anycast.

5.3 Replica Selection by Edgio

We next analyze Edgio, one of the few CDNs reported to use regional anycast [44, 51].[4]

Figure 5 shows the latency CDFs for clients in Brazil, the UK, and Australia. Across all three countries, latency distributions are consistent across resolver scopes within the same region, but diverge sharply for resolvers in non-neighboring regions. This behavior matches the expected pattern of regional anycast (Fig. 2c).

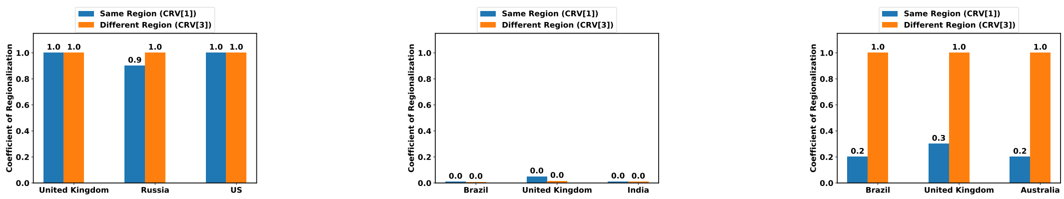


(a) Edgio-Brazil **(b)** Edgio-United Kingdom **(c)** Edgio-Australia

Figure 5 Edgio exhibits stable latencies within regions but divergence across them, matching regional Anycast.

The Australia case (Fig. 5c) highlights an interesting edge case: resolvers in Sydney and India show overlapping latency distributions. Further analysis revealed that 81% of the measured Edgio objects returned the same IP for both, suggesting they belong to the same anycast region. Personal communication with Edgio operators confirmed this interpretation.

Figure 6c shows the corresponding $CRV[1]$ and $CRV[3]$ values. As expected, $CRV[1]$ remains low (0.2–0.3) while $CRV[3] \approx 1$, confirming Edgio's use of regional anycast for replica selection.



(a) Akamai **(b)** Cloudflare **(c)** Edgio

Figure 6 CRV values cleanly separate the three archetypes: Akamai (DNS), Cloudflare (Anycast), Edgio (Regional Anycast).

Together, these results demonstrate that our methodology correctly distinguishes the canonical steering approaches. More importantly, they show that we can now apply it to the broader and less transparent ecosystem of global and regional CDNs, cutting through opacity to reveal the steering logics that shape today's Internet.

Takeaway: Our methodology reliably recovers the known steering strategies of Akamai (DNS), Cloudflare (Anycast), and Edgio (Regional Anycast), confirming its accuracy for broader use.

[4] EdgeCast (AS15133) and Limelight (AS22822) merged to form Edgio in 2022. We focus here on measurements to the EdgeCast network.

6 Inferring CDN Replica Selection

Having validated our methodology against three well-known CDNs, we now apply it at scale to the broader set of CDN providers identified across the top 1,000 websites in each of 19 countries. These include both large international CDNs (e.g., CloudFront, Fastly) and regional providers (e.g., Azion, NGENIX). For each, we classify the predominant replica selection approach used in practice based on CDF analysis and CRV inference.

While our study focuses on 17 CDN providers, this set covers the providers responsible for the vast majority of content delivered across the 19 countries we measure. Our methodology scales to additional CDNs, and expanding coverage to further regional or niche providers is part of future work. We prioritized CDNs with sufficient measurement visibility across multiple vantage points and resource domains.

In addition to identifying replica selection mechanisms, we also examine ECS (EDNS Client Subnet) support as an orthogonal yet complementary mechanism that further refines client-to-replica mapping. While not a standalone selection strategy, ECS plays an important role in mitigating location mismatches introduced by public DNS resolvers.

6.1 CDNs Replica Selection

We extend our methodology to the remaining CDN providers identified across the top-ranked websites in our dataset. This includes major global CDNs such as Amazon CloudFront and Fastly, as well as regional providers like Azion (Brazil), NGENIX (Russia), and Taobao (China). Google presents a unique challenge due to its hybrid architecture and is discussed separately. Given space constraints, we focus here on representative examples that illustrate the diversity of observed behaviors.

6.1.1 Global CDNs

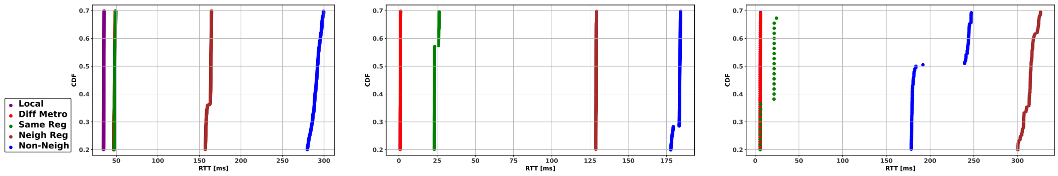
Amazon CloudFront, operated by AWS, is a global CDN with over 450 points of presence in more than 49 countries. Figure 7 shows the latency CDFs to replicas assigned through different resolver scopes from vantage points in the US, India, and South Africa. While the deployment scale is smaller than Akamai's, the CDFs are generally non-overlapping across resolver scopes, indicating DNS-based replica selection. Overlap in narrower scopes (e.g., metro and regional) in the US reflects CloudFront's relatively smaller infrastructure footprint.

In South Africa, the CDF from the same-region resolver (Zimbabwe) splits into two vertical clusters. About 40% of replicas align with those seen in other South African metros, suggesting variation tied to specific domains or content categories.

To validate this classification, we compute the $CRV[1]$ and $CRV[3]$ values across all three countries (Fig. 9a). The consistently high CRVs (0.8–1.0) confirm that CloudFront's replica selection is influenced by resolver geography, a hallmark of DNS-based mechanisms.

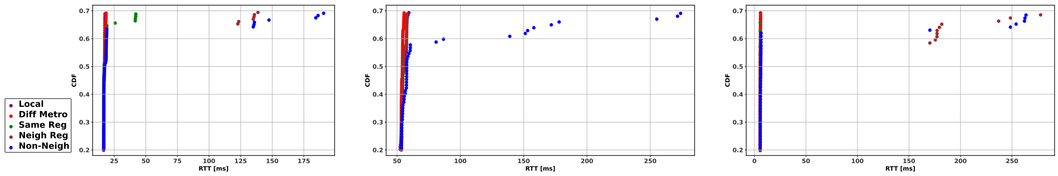
We also probed for IP prefixes associated with other Amazon services [8] to explore whether CloudFront's selection strategy generalizes across Amazon-hosted applications. However, we found limited overlap in our dataset, leaving that question open for future investigation.

Fastly is a global CDN with approximately 80 PoPs across six continents. In addition to content delivery, it provides services such as real-time streaming and private CDN configurations. Figure 8 shows the latency CDFs for resolver scopes from vantage points in Brazil, Turkey, and South Africa. Across all three countries, the CDFs exhibit substantial



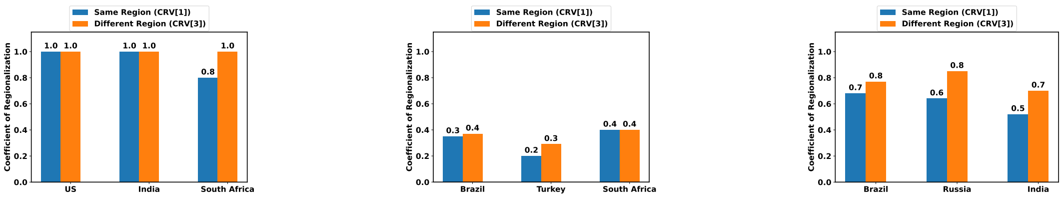
(a) Amazon-US **(b)** Amazon-India **(c)** Amazon-South Africa

Figure 7 Cloudfront matches DNS-based replica selection mechanism, as shown in the US, India and South Africa.



(a) Fastly-Brazil **(b)** Fastly-Turkey **(c)** Fastly-South Africa

Figure 8 Fastly predominantly matches Anycast replica selection mechanism, as shown in Brazil, South Africa and Turkey.



(a) CloudFront **(b)** Fastly **(c)** Google

Figure 9 Coefficient of regionalization for same ($CRV[1]$) and different regions ($CRV[3]$) for other CDNs.

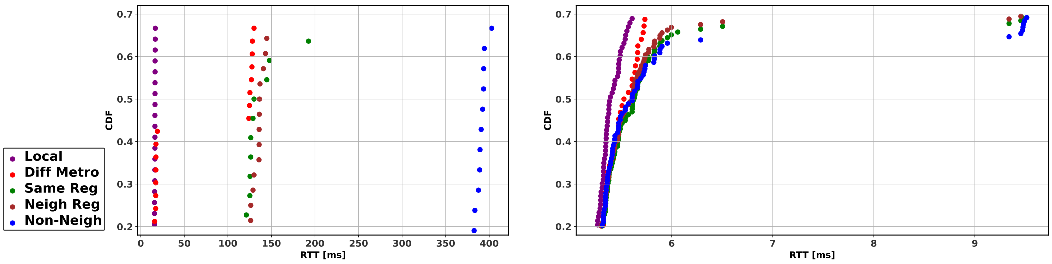
overlap regardless of resolver location, consistent with an Anycast-based replica selection strategy.

While the dominant pattern aligns with Anycast, we observe occasional deviations resembling DNS-based behavior, possibly due to customer-specific configurations or specialized services. Compared to Akamai, Fastly's smaller deployment may expose greater variability across content. The corresponding $CRV[1]$ and $CRV[3]$ values range from 0.2 to 0.4 (Fig. 9b), reinforcing this classification. These values are lower than those in DNS-based CDNs like Akamai, but not as close to zero as Cloudflare, suggesting a predominantly Anycast model with pockets of variation.

6.1.2 Regional CDNs

Regional CDNs, while smaller in scale, are critical to content delivery in local markets—particularly for government and public service websites [40]. Our methodology is designed to work across deployment scales and is thus well suited to characterizing these providers when sufficient measurement visibility exists.

Figure 10 illustrates two such cases: **Azion**, based in Brazil, and **NGENIX**, based in Russia. For Azion, the CDFs exhibit clear divergence across resolver scopes, with latencies



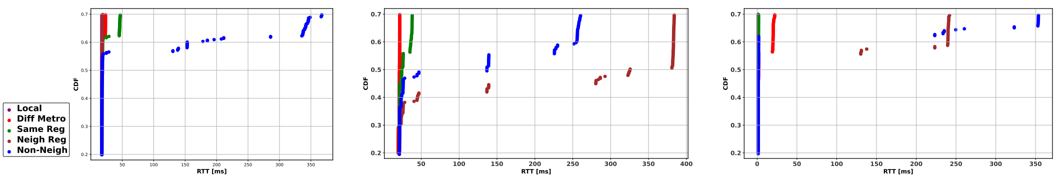
(a) Azion-Brazil **(b)** NGENIX-Russia

Figure 10 Replica selection mechanisms used by regional CDNs.

increasing as resolver distance grows. Corresponding CRV values ($CRV[1] \approx 1$, $CRV[3] \approx 1$) confirm a DNS-based selection strategy. In contrast, NGENIX shows strongly overlapping CDFs and low CRV values (≈0.3), indicating an Anycast-based approach.

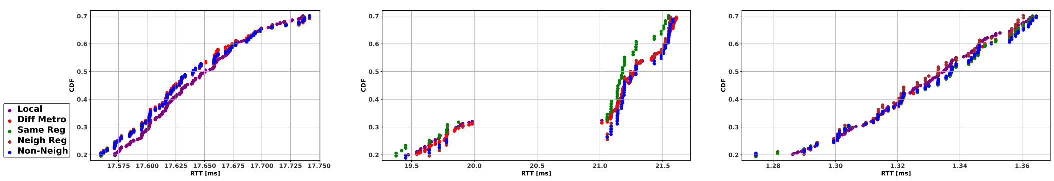
These findings highlight that regional CDNs are not uniformly configured—they adopt different selection strategies depending on infrastructure constraints, client base, and operational goals. Our methodology provides a scalable, unified framework to study both global and regional CDN behavior in a comparable way.

6.2 Google: Mixed Replica Selection



(a) Google-Brazil **(b)** Google-Russia **(c)** Google-India

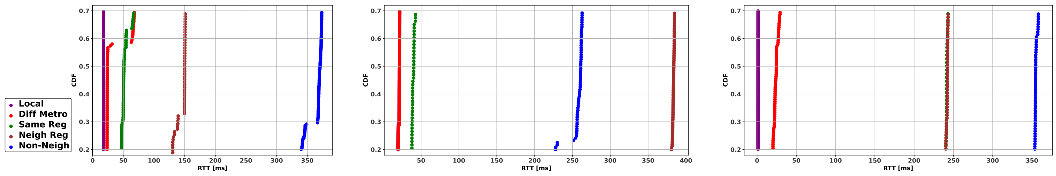
Figure 11 Google uses a combination of replica selection mechanisms that matches the DNS-based for some content and Anycast for other, as shown in Brazil, Russia, and India.



(a) Google-Brazil **(b)** Google-Russia **(c)** Google-India

Figure 12 Google content using Anycast replica selection mechanism, as shown in Brazil, Russia, and India.

Google presents a complex case that highlights the importance, and limitations, of replica selection inference at scale. Figure 11 shows CDFs of latencies to Google-assigned replicas across resolver scopes for clients in Brazil, Russia, and India. The observed patterns are clearly bimodal: about half of the responses (e.g., ≈50% in Brazil) exhibit characteristics



(a) Google-Brazil **(b)** Google-Russia **(c)** Google-India

Figure 13 Google content using DNS-based replica selection mechanism, as shown in Brazil, Russia, and India.

of DNS-based selection (divergent latencies), while the rest follow an Anycast-like pattern (overlapping latencies).

This heterogeneity is reflected in Google's coefficient of regionalization values: for Brazil, Russia, and India, $CRV[1]$ and $CRV[3]$ range between 0.52 and 0.85 (Fig. 9c), diverging from the patterns typical of purely DNS-based or anycast approaches. While these mid-range values may initially seem ambiguous, they in fact indicate a genuine hybrid strategy: Google concurrently uses both DNS-based and anycast selection, depending on the resource or customer.

To verify this, we isolate resources matching each pattern. When separated, the resulting CDFs align cleanly with their expected forms: Anycast-based resources exhibit fully overlapping latency distributions and CRVs near zero (Fig. 12), while DNS-based resources show divergent latencies and CRVs near one (Fig. 13). This supports the use of CRV analysis even in the presence of complex, mixed deployments.

We further examined which domains correspond to each selection model. Key Google services (`google.com`, `youtube.com`, `googleadservices.com`, and `googlesyndication.com`) consistently rely on DNS-based replica selection. In contrast, externally hosted content, e.g., `snapchat.com`, `chess.com`, is more often served via Anycast. Table 3 quantifies the split across countries, showing that Google's replica selection is almost evenly divided between DNS-based and Anycast.

Country	DNS(%)	Anycast(%)
France	43.8	56.1
Germany	42.1	57.9
Russia	63.7	36.3
Spain	63.9	36.1
Turkey	58.1	41.9
United Kingdom	44.1	55.9
US	45.4	54.6
Argentina	47.4	52.6
Brazil	53.3	46.7
China	46.4	53.6
India	45.3	54.7
Indonesia	62.9	37.1
United Arab Emirates	60.2	39.8
Australia	42.4	57.6
Algeria	52.0	48.0
Egypt	26.7	73.3
Ghana	44.6	55.4
Nigeria	55.1	44.9
South Africa	56.5	43.5

Table 3 % of Google resources that match DNS and Anycast approaches.

Lastly, we analyzed Google Cloud by isolating responses from IP prefixes Google designates for its cloud services [28]. All of these responses show Anycast behavior, with CRV values ≈0. This distinction further underscores the value of our method: not only can it detect mixed strategies across a CDN, but it can also differentiate between infrastructure components or service classes within a provider.

6.3 Geographic Trends in Replica Selection

Location	Users(%)	DNS-based(%)	Anycast(%)	Reg. Anycast(%)	Mixed(%)
Europe	60.0	29.5	39.2	2.0	29.2
France	1.1	33.7	38.8	4.4	23.2
Germany	1.5	34.5	38.1	1.0	26.3
Russia	2.3	10.6	58.0	0.5	30.9
Spain	0.8	41.8	30.1	1.6	26.4
Turkey	1.3	22.7	41.5	0.6	35.2
United Kingdom	1.2	33.0	33.4	4.3	29.4
North America	89.3	36.8	23.8	6.5	32.8
→US	*5.5*	*36.8*	*23.8*	*6.5*	*32.8*
South America	60.0	27.7	35.2	1.3	35.8
Argentina	0.8	24.1	35.1	0.8	40.0
Brazil	3.3	29.4	35.9	1.8	33.0
Asia	70.1	25.0	44.6	1.4	29.0
China	18.8	38.6	39.5	2.8	19.2
India	15.5	32.8	40.4	1.0	25.9
Indonesia	3.9	16.2	48.2	0.8	34.8
United Arab Emirates	0.2	21.1	43.3	1.6	34.0
Oceania	75.0	34.7	28.3	2.4	34.7
Australia	0.4	34.7	28.3	2.4	34.7
Africa	50.1	20.1	49.2	1.5	29.2
Algeria	0.7	21.2	53.7	0.75	24.3
Egypt	1.0	18.0	56.0	0.7	25.3
Ghana	0.3	22.4	43.1	1.9	32.6
Nigeria	2.9	20.8	42.8	3.0	33.4
South Africa	0.6	15.5	55.9	0.6	28.1
World Total	66.0	26.7	40.8	2.0	30.6

Table 4 By resource count, Anycast dominates in Africa/Asia, but DNS remains substantial in Europe and the Americas.

We now examine the global landscape of replica selection strategies across the top 1,000 websites in 19 countries, capturing ≈66% of the world's Internet population. Table 4 reports, per country and region, the percentage of web resources delivered via each approach or a combination. These breakdowns allow us to compare both the technical strategies CDNs employ and the resulting user experience across geographies.

For example, in the US (highlighted), 36.8% of resources rely on DNS-based selection, 23.8% on Anycast, 32.8% on a mixed model, and only 6.5% on Regional Anycast. In contrast, countries like India, Turkey, and those in Africa show a stronger reliance on Anycast. Europe and China exhibit a more balanced mix, often reflecting dominant CDNs like Google that employ hybrid strategies.

To account for the fact that not all resources are equally weighted, we also analyze content volume. Table 5 shows the proportion of total bytes delivered via each mechanism. This byte-weighted view reveals an important trend: across most continents, including North and South America, Africa, and Oceania, DNS-based selection still dominates in terms of content volume, even when Anycast is prevalent by resource count. This suggests that heavier or more critical content (e.g., video, rich media) may still favor DNS-based strategies, possibly due to finer-grained performance tuning.

These trends reflect differences in CDN scale, customer base, and infrastructure. Larger DNS-based CDNs often serve heavier content; Anycast-based CDNs host lighter or static

Location	Users(%)	DNS-based(%)	Anycast(%)	Reg. Anycast(%)	Mixed(%)
Europe	60.0	27.6	41.4	2.6	28.4
France	1.1	27.5	38.5	3.9	30.2
Germany	1.5	30.8	42.2	1.6	25.4
Russia	2.3	9.9	63.1	0.6	26.4
Spain	0.8	44.5	27.4	3.1	25.0
Turkey	1.3	26.7	46.0	3.3	24.0
United Kingdom	1.2	28.0	33.8	3.0	35.3
North America	89.3	34.7	32.3	3.8	29.1
→ US	*5.5*	*34.7*	*32.3*	*3.8*	*29.1*
South America	60.0	35.5	33.1	1.4	30.0
Argentina	0.8	34.2	33.7	1.1	31.0
Brazil	3.3	35.7	33.1	1.8	29.5
Asia	70.1	26.3	43.8	1.8	28.1
China	18.8	48.1	30.0	1.3	20.6
India	15.5	28.1	40.1	0.8	31.1
Indonesia	3.9	13.6	53.5	2.7	30.2
United Arab Emirates	0.2	20.5	40.0	1.8	37.7
Oceania	75.0	40.1	27.1	1.8	30.6
Australia	0.4	40.1	27.1	1.8	30.6
Africa	50.1	56.7	26.1	0.9	16.4
Algeria	0.7	22.1	51.0	1.7	25.1
Egypt	1.0	16.9	55.4	0.3	27.4
Ghana	0.3	82.0	9.6	0.4	8.0
Nigeria	2.9	36.5	37.3	2.5	23.7
South Africa	0.6	14.3	52.9	0.9	31.9
World Total	66.0	40.9	33.6	1.7	23.9

Table 5 By delivered bytes, DNS steering dominates globally — showing heavier traffic (e.g., video) still prefers DNS-based selection.

sites.

While our study is scoped to public, unlogged landing pages, the broad adoption of DNS-based selection – especially for large content – suggests that trends are likely to extend to internal or authenticated traffic as well. Future work will explore this hypothesis in depth.

6.4 ECS Support Across CDNs

While ECS support is orthogonal to the underlying replica selection mechanism, it remains a critical capability for improving selection granularity—particularly in DNS-based systems. ECS allows CDNs to tailor responses based on a more accurate client location, mitigating issues like client–LDNS mismatch. Understanding ECS support is thus essential for assessing a CDN's ability to optimize per-user performance.

To detect ECS usage, we issued DNS queries with embedded client subnets to public resolvers and compared the responses from CDN-authoritative nameservers. We then validated whether the selected server locations matched the geolocation of the specified subnet.

Out of the 17 CDNs in our dataset, 10 respond to ECS queries with a scope greater than zero (Table 6), indicating support for the extension. Among these, most also return different server assignments for different subnets, confirming that ECS is used in their replica selection process. In the majority of cases, the geolocation of the assigned server closely matches that of the provided subnet, reinforcing ECS's role in fine-grained, location-aware request routing.

6.4.0.1 ECS and Replica Selection Interaction

We find that ECS support complements DNS-based replica selection, helping address the client-LDNS mismatch problem. Among CDNs using DNS-based approaches (e.g., Akamai,

CDN	DNS-based	Anycast	Regional Anycast	ECS
Akamai	✓			✓
Azion	✓			✗
BunnyCDN	✓			✓
CDN77	✓			✓
Cloudflare		✓		✗
Cloudfront	✓			✓
Edgio			✓	✓
Facebook	✓			✗
Fastly	✓	✓		✗
Google	✓	✓		✓
Level3	✓			✓
Medianova	✓			✓
NGENIX		✓		✓
StackPath		✓		✗
Taobao	✓			✗
Tencent	✓			✗
Yahoo			✓	✓

Table 6 Most CDNs (12/17) rely on DNS steering; a smaller set on Anycast or Regional Anycast. Ten support ECS, mainly those using DNS.

CloudFront, Google), most exhibit strong ECS support and return location-sensitive responses based on client subnets. Conversely, CDNs that rely primarily on Anycast (e.g., Cloudflare, StackPath) do not support ECS, likely because ECS plays no meaningful role in routing under anycast-based models. For CDNs with mixed strategies (e.g., Google, Fastly), ECS support appears more variable: DNS-mapped resources tend to support ECS, while anycast-mapped domains do not.

6.5 Summary: CDN Techniques and ECS

Table 6 summarizes the predominant replica selection strategy used by each CDN in our dataset. DNS-based selection remains the most common (12 of 17 CDNs), followed by anycast (Cloudflare, StackPath, Fastly, NGENIX), and regional anycast (Edgio, Yahoo). Google and Fastly show mixed behavior, illustrating how a single CDN may adopt different strategies across services or regions.

Though we focus on 17 CDNs, they serve most content across the 19 countries measured. Our methodology generalizes to other providers with sufficient visibility.

We also evaluate ECS support, a complementary mechanism to improve client mapping in DNS-based systems. Ten CDNs respond to ECS queries with non-zero scope and vary responses based on client subnet. In most cases, the assigned replica aligns with the subnet's geolocation, confirming ECS-informed assignment.

Overall, while selection strategies vary across CDNs and regions, ECS is a widely adopted enhancement among DNS-based systems. The rise of hybrid models and ECS support reflects a shift toward more precise and flexible user steering.

Takeaway: DNS-based steering dominates global traffic volume, but regional variation and hybrid deployments (e.g., Google, Fastly) show that replica selection remains diverse and opaque.

7 Discussion

Our results should be interpreted with care. Inferring replica selection from client-side latency is necessarily approximate, and several factors complicate interpretation. We outline key methodological limits and our mitigations, then return to the broader implications: while our method cannot speak to operator intent, it makes opaque steering behaviors measurable and raises questions about resilience and governance.

7.1 Methodological Limits

Deployment density. In sparsely covered regions, resolvers may return the same replicas simply because alternatives do not exist, regardless of the CDN's steering strategy. For instance, in India, CloudFront often maps different metros to the same servers (Fig. 7b), likely reflecting limited infrastructure [63]. To reduce sensitivity, we focus on the most informative CRV entries – $CRV[1]$ (local vs. same-region) and $CRV[3]$ (local vs. distant region).

Regional boundaries. To define resolver scopes, we adopt the UN geoscheme, but CDNs often define regions idiosyncratically (e.g., treating all of South America as a single region). Such mismatches can blur the distinction between intra- and inter-region steering. Repeated measurements and consistency in CRV[1] and CRV[3] mitigate this effect by focusing on coarse-grained shifts rather than exact regional labels.

Transient conditions. Network RTTs vary over time due to congestion, routing changes, and load. To reduce the impact of this variability, we repeat measurements, analyze complete latency distributions, and rely on CrUX landing pages as stable reference points. Our dataset primarily reflects web traffic and underrepresents video and software delivery, which we leave to future work.

Prefix-based methods. An alternative to our latency-based inference is to reason about replica selection using prefix-level control-plane signals, such as IP clustering or anycast prefix enumeration. However, these approaches face well-known challenges, including variable prefix lengths, overlapping announcements, and prefix reuse across sites and services. Prior clustering-based techniques [55, 13] can therefore be brittle when applied at scale, while tools such as MAnycast2 identify anycast IP addresses but provide limited insight into how replica selection varies across geographies or client populations. In contrast, our approach infers steering behavior directly from client-visible latency outcomes, complementing prefix-based methods with a view of how replica selection is experienced in practice.

7.2 Broader Implications

These limits underscore a central point: **replica selection remains deliberately opaque**. Operators disclose little about how resolvers or clients are mapped to replicas, yet these hidden choices shape traffic flow. By making steering behavior empirically visible, our methodology enables debates that extend beyond performance:

- **Resilience:** Should global rerouting power rest with a few resolvers and opaque CDN policies? How does this concentration of control affect fault tolerance in times of crisis?
- **Sovereignty:** When government or regional traffic is steered abroad, what are the implications for jurisdiction and autonomy?
- **Control:** Does global anycast, while operationally simple, cede too much control to BGP? Do hybrids mixing DNS and anycast compound opacity?

Our measurements cannot answer these questions or reveal operator intent. But by turning replica selection into something measurable, they provide a basis for both scientific classification of CDN strategies and normative debates about accountability in Internet infrastructure.

8 Related Work

Prior work on replica selection has predominantly framed the problem as one of *performance optimization*. A substantial body of research has proposed and evaluated mechanisms such as DNS-based redirection, anycast, and end-user mapping, characterizing their tradeoffs in terms of latency, cache efficiency, and client affinity [1, 13, 30, 26, 42, 12, 14, 7, 32, 34, 35, 37, 60, 52, 53, 66, 44, 4, 6, 65]. Other studies have taken a provider-centric perspective, examining the infrastructure and steering behavior of individual CDNs such as Akamai, Google, and Yahoo [5, 46, 57, 25, 11]. Complementary efforts have developed techniques for detecting and enumerating anycast deployments from external vantage points [10, 16, 17, 24, 15].

Beyond performance-focused analyses, broader Internet measurement has investigated the structural context in which replica selection operates, including where popular Web content is hosted and how hosting infrastructure is distributed across providers and regions [2]. This line of work highlights patterns of concentration and shared dependency that shape the environment in which CDN steering mechanisms are deployed. Related research has further demonstrated that these architectural choices have security and abuse implications, for example by identifying CDNs that remain susceptible to domain fronting through DNS-based traffic analysis [58].

Collectively, this literature provides a detailed understanding of the performance characteristics and deployment properties of different steering mechanisms. However, it offers limited visibility into the *replica selection strategies CDNs employ in practice at global scale*, particularly in the presence of DNS resolver consolidation. Moreover, existing studies stop short of connecting steering mechanisms to their broader structural implications for resilience, jurisdictional exposure, and consolidation. In this work, we address these gaps by introducing an experimental methodology that infers whether a CDN relies on DNS-based steering, global anycast, or regional anycast, and applying it systematically across providers and regions.

9 Conclusions

Replica selection is the hidden steering wheel of the Internet: every request is mapped to a server, shaping latency, resilience, and sovereignty in ways few can observe. We showed how DNS resolver consolidation magnifies the risks of this opacity, concentrating power over content delivery into the hands of a few actors.

Our contribution is twofold. First, we frame replica selection not as performance optimization, but as a sociotechnical risk linking infrastructure design to control and resilience. Second, we introduce a lightweight methodology that infers whether a CDN relies on DNS, global anycast, or regional anycast by exploiting distinct latency fingerprints across resolvers. Applied to CDNs serving two-thirds of global users, we find that **DNS-based steering dominates and is acutely sensitive to resolver consolidation**.

These findings raise broader questions: Should the mechanisms steering the Internet remain hidden inside DNS resolvers? Does anycast offer resilience at the expense of control, or partly counter centralization? How should policymakers, operators, and researchers grapple with a future where global traffic is steered by opaque intermediaries?

We do not claim to offer final answers. Instead, we open a conversation: providing a tool for visibility and a call to recognize replica selection as a structural vulnerability that demands both technical innovation and new thinking about governance and accountability.

References

1 J. Abley and K. Lindqvist. Operation of anycast services. *www.rfc-editor.org*, 2006. URL: `https://www.rfc-editor.org/rfc/rfc4786`.

2 Bernhard Ager, Wolfgang Muhlbauer, Georgios Smaragdakis, and Steve Uhlig. Web content cartography. In *Proc. of ACM SIGCOMM*, 2011. URL: `https://doi.org/10.1145/2068816.2068870`.

3 Akamai Technologies. What is ecs (edns0 client subnet)?, 2022. URL: `https://community.akamai.com/customers/s/article/What-is-ECS-EDNS0-Client-Subnet?language=en_US`.

4 Rami Al-Dalky, Michael Rabinovich, and Kyle Schomp. A look at the ecs behavior of dns resolvers. In *Proc. of IMC*, Oct 2019.

5 Mohammad Al-Fares, Khaled Elmeleegy, Benjamin Reed, and Igor Gashinsky. Overclocking the yahoo! cdn for faster web page loads. In *Proc. of IMC*, 2011.

6 Hussein A. Alzoubi, Seungjoon Lee, Michael Rabinovich ad Oliver Spatscheck, and Jacobus Van Der Merwe. A practical architecture for an anycast cdn. *ACM Trasactions on the Web*, 5(4), 2011.

7 Hussein A. Alzoubi, Seungjoon Lee, Michael Rabinovich, Oliver Spatscheck, and Jacobus Van der Merwe. Anycast cdns revisited. In *Proc. of the WWW*, 2008.

8 Amazon. Amazon aws, 2023. URL: `https://ip-ranges.amazonaws.com/ip-ranges.json`.

9 RIPE ATLAS. Atlas console, 2023. URL: `https://atlas.ripe.net`.

10 Rui Bian, Shuai Hao, Haining Wang, Amogh Dhamdere, Alberto Dainotti, and Chase Chuang:evaluatingcdnsotton. Towards passive analysis of anycast in global routing: Unintended impact of remote peering. *ACM SIGCOMM CCR*, 49(3), 2019.

11 Matt Calder, Xun Fan, Zi Hu, Ethan Katz-Bassett, John Heidemann, and Ramesh Govindan. Mapping the expansion of google's serving infrastructure. In *Proc. of IMC*, 2013.

12 Matt Calder, Xun Fan, and Liang Zhu. A cloud provider's view of edns client-subnet adoption. In *Proc. of TMA Conference*, 2019.

13 Matt Calder, Ashley Flavel, Ethan Katz-Bassett, Ratul Mahajan, and Jitendra Padhye. Analyzing the performance of an anycast cdn. In *Proc. of IMC*, 2015.

14 Fangfei Chen, Ramesh K. Sitaraman, and Marcelo Torres. End-user mapping: Next generation request routing for content delivery. In *Proc. of ACM SIGCOMM*, 2015.

15 Danilo Cicalese, Jordan Auge, Diana Joumblatt, Timur Friedman, and Dario Rossi. Characterizing ipv4 anycast adoption and deployment. In *Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies*, 2015. URL: `https://doi.org/10.1145/2716281.2836101`.

16 Danilo Cicalese, Diana Joumblatt, Dario Rossi, Marc-Olivier Buob, Jordan Auge, and Timur Friedman. A fistful of pings: Accurate and lightweight anycast enumeration and geolocation. In *Proc. of IEEE INFOCOM*, 2015.

17 Danilo Cicalese and Dario Rossi. A longitudinal study of ip anycast. *ACM SIGCOMM CCR*, 48(1), 2018.

18 Cisagov. findcdn, 2023. URL: `https://github.com/cisagov/findcdn`.

19 Cloudflare, 2023. URL: `https://news.ycombinator.com/item?id=19828702`.

20 Cloudflare. Cdn · cloudflare reference architecture docs, 2023. URL: `https://developers.cloudflare.com/reference-architecture/cdn-reference-architecture/`.

21 C. Contavalli, W. van der Gaast, D. Lawrence, and W. Kumari. Client subnet in dns queries. RFC 7871, IETF, 2016.

22 Omar Darwich, Hugo Rimlinger, Milo Dreyfus, Matthieu Gouel, and Kevin Vermeulen. Replication: Towards a publicly available internet scale ip geolocation dataset. In *Proc. of IMC*, 2023.

23 Public DNS. Public dns server list, 2023. URL: `https://public-dns.info/#countries`.

24 Xun Fan, John Heidemann, and Ramesh Govindan. Evaluating anycast in the domain name system. In *Proc. of IEEE INFOCOM*, 2013.

25 Xun Fan, Ethan Katz-Bassett, and John Heidemann. Assessing affinity between users and cdn sites. In *Proc. of TMA Conference*, 2015.

26 Utkarsh Goel, Mike P Wittie, and Moritz Steiner. Faster web through client-assisted cdn server selection. In *Proc. of ICCN*, Oct 2015.

27 Google. About crux, Jun 2022. URL: https://developer.chrome.com/docs/crux/about/.

28 Google. Obtain google ip address ranges, 2023. URL: https://www.gstatic.com/ipranges/cloud.json.

29 Google Developers. Edns client subnet (ecs) guidelines, 2024. URL: https://developers.google.com/speed/public-dns/docs/ecs.

30 Shuai Hao, Yubao Zhang, Haining Wang, and Angelos Stavrou. End-users get maneuvered: Empirical analysis of redirection hijacking in content delivery networks. In *Proc. of USENIX Security*, 2018.

31 hotspotshield. Hotspotshield vpn, 2023. URL: https://www.hotspotshield.com.

32 Cheng Huang, Angela Wang, Jin Li, and Keith Ross. Measuring and evaluating large-scale cdns. In *Proc. of IMC*, 2008.

33 ipInfo. ipinfo, 2023. URL: https://ipinfo.io.

34 K. L Johnson, J. F Carr, M. S Day, and M. F Kaashoek. The measured performance of content distribution networks. *Computer Communications*, 24(2), 2001.

35 J Kangasharju, K.W Ross, and J.W Roberts. Performance evaluation of redirection schemes in content distribution networks. *Computer Communications*, 2001.

36 Aqsa Kashaf, Vyas Sekar, and Yuvraj Agarwal. Analyzing third party service dependencies in modern web services: Have we learned from the mirai-dyn incident? In *Proc. of IMC*, 2020.

37 Balachander Krishnamurthy, Craig Wills, and Yin Zhang. On the use and performance of content distribution networks. In *Proc. ACM IMW*, 2001.

38 Rashna Kumar, Sana Asif, Elise Lee, and Fabian E. Bustamante. Each at its own pace: Third-party dependency and centralization around the world. In *Proc. of ACM SIGMETRICS*, 2023.

39 Rashna Kumar, Sana Asif, Elise Lee, and Fabian E. Bustamante. Each at its own pace: Third-party dependency and centralization around the world. *Proc. ACM Meas. Anal. Comput. Syst.*, 2023.

40 Rashna Kumar, Esteban Carisimo, Lukas De Angelis Riva, Mauricio Buzzone, Fabián E. Bustamante, Ihsan Ayyub Qazi, and Mariano G. Beiró. Of choices and control - a comparative analysis of government hosting. In *Proceedings of the 2024 ACM on Internet Measurement Conference*, 2024.

41 Zhihao Li, Dave Levin, Neil Spring, and Bobby Bhattacharjee. Internet anycast: Performance, problems, & potential. In *Proc. of ACM SIGCOMM*, 2018.

42 Zhuoqing Morley Mao, Charles D. Cranor, Fred Douglis, Michael Rabinovich, Oliver Spatscheck, and Jia Wang. A precise and efficient evaluation of the proximity between web clients and their local dns servers. In *Proc. of USENIX ATC*, 2002.

43 Srdjan Matic, Gareth Tyson, and Gianluca Stringhini. Pythia: A framework for the automated analysis of web hosting environments. In *The World Wide Web Conference*, 2019.

44 Stephen Mcquistin, Sree Uppu, and Marcel Flores. Taming anycast in the wild internet. In *Proc. of IMC*, 2019.

45 nordvpn. Nord vpn, 2023. URL: https://nordvpn.com.

46 Erik Nygren, Ramesh Sitaraman, and Jennifer Sun. The akamai network: A platform for high-performance internet applications. *ACM SIGOPS Operating Systems Review*, 44(3), July 2010.

47 John Otto and Mario Sánchez John Rula Fabián E. Bustamante. Content delivery and the natural evolution of DNS: remote DNS trends, performance issues and alternative solutions. In *Proc. of IMC*, 2012.

48 CDN Planet. Akamai cdn, 2023. URL: https://www.cdnplanet.com/cdns/akamai/.

49 CDN Planet. Cloudflare cdn, 2023. URL: https://www.cdnplanet.com/cdns/cloudflare/.

50 CDN Planet. Content delivery networks, 2023. URL: `https://www.cdnplanet.com`.
51 CDN Planet. Edgio cdn, 2023. URL: `https://www.cdnplanet.com/cdns/edgio/`.
52 Ingmar Poese, Benjamin Frank, Bernhard Ager, Georgios Smaragdakis, Steve Uhlig, and Anja Feldmann. Improving content delivery with padis. *IEEE Internet Computing*, 2012. `doi:10.1109/mic.2011.105`.
53 Stefan Saroiu, Krishna P. Gummadi, Richard J. Dunn, Steven D. Gribble, and Henry M. Levy. An analysis of internet content delivery systems. In *Proc. of USENIX OSDI*, 2002. `doi:10.1145/844128.844158`.
54 Anees Shaikh, Renu Tewari, and Mukesh Agrawal. On the effectiveness of dns-based server selection. In *Proc. of IEEE INFOCOM*, 2001.
55 Raffaele Sommese, Leandro Bertholdo, Gautam Akiwate, Mattijs Jonker, Roland van Rijswijk-Deij, Alberto Dainotti, KC Claffy, and Anna Sperotto. Manycast2: Using anycast to measure anycast. In *Proceedings of the ACM Internet Measurement Conference*, 2020.
56 Internet World Stats. World internet users statistics and 2019 world population stats, 2022. URL: `https://www.internetworldstats.com/stats.htm`.
57 Ao-Jan Su, David R. Choffnes, Aleksandar Kuzmanovic, and Fabián E. Bustamante. Drafting behind akamai (travelocity-based detouring). In *Proc. of ACM SIGCOMM*, 2006.
58 Karthika Subramani, Roberto Perdisci, Pierros-Christos Skafidas, and Manos Antonakakis. Discovering and measuring cdns prone to domain fronting. *Proceedings of the ACM Web Conference 2024*, 2024. URL: `https://doi.org/10.1145/3589334.3645656`.
59 surfshark. Surfshark vpn, 2023. URL: `https://surfshark.com`.
60 Sipat Triukose, Zhihua Wen, and Michael Rabinovich. Measuring a commercial content delivery network. In *Proc. of the WWW*, 2011.
61 Zheng Wang, Jun Huang, and Scott Rose. Evolution and challenges of dns-based cdns. *Digital Communications and Networks*, 4(4):235–243, Nov 2018. `doi:10.1016/j.dcan.2017.07.005`.
62 Wikipedia. United nations geoscheme, Jul 2022. URL: `https://en.wikipedia.org/wiki/United_Nations_geoscheme`.
63 Wikipedia. Amazon cloudfront, May 2023. URL: `https://en.wikipedia.org/wiki/Amazon_CloudFront#:~:text=CloudFront%20has%20servers%20located%20in`.
64 Jing'an Xue, David Choffnes, and Jilong Wang. Cdns meet cn an empirical study of cdn deployments in china. *IEEE Access*, 5:5292–5305, 2017. `doi:10.1109/ACCESS.2017.2682190`.
65 Minyuan Zhou, Xiao Zhang, Shuai Hao, Xiaowei Yang, Jiaqi Zheng, Guihai Chen, and Wanchun Dou. Regional IP Anycast: deployments, performance, and potentials. In *Proc. of ACM SIGCOMM*, 2023.
66 Jiangchen Zhu, Kevin Vermeulen, Italo Cunha, Ethan Katz-Bassett, and Matt Calder. The best of both worlds. In *Proc. of IMC*, 2022.

A Appendix

A.1 Ethics

This work does not raise any ethical issues.

A.2 DNS Resolvers Used in the Study

We selected public DNS resolvers to span five geographic scopes relative to each RIPE Atlas probe: same metro, same country (different metro), same region (different country), neighboring region, and non-neighboring region. Resolvers were chosen based on responsiveness, stability, and accessibility from RIPE Atlas probes.

To ensure accurate geographic attribution and to avoid bias from anycasted infrastructure, we validated resolver locations using IP geolocation (via IPInfo), AS registry data, and anycast status as reported by IPInfo. Resolvers exhibiting routing behavior indicative of anycast were excluded.

A.3 CRV Vector

The CRV is a four-tuple vector where each entry, $CRV[i]$, is the distance between the distribution associated with the DNS local resolver and that of the $(i+1)$th resolver scope. Table 7 describes the $(i+1)$th resolver scope corresponding to each entry in the CRV vector.

CRV Index	$(i+1)$**th resolver scope**
$CRV[0]$	Resolver in a different metro area within the same country
$CRV[1]$	Resolver in a different country within the same region
$CRV[2]$	Resolver in a neighboring region
$CRV[3]$	Resolver in a non-neighboring region

Table 7 Table describing each entry of the CRV vector, where $CRV[i]$ is the distance between the distribution associated with the DNS local resolver and that of the $(i+1)$th resolver scope.

A.4 Detecing Replica Selection Using IP Prefixes

An alternative method for identifying a CDN's replica selection approach involves comparing the sets of replica prefixes assigned to clients across different resolver scopes. This method requires matching each replica IP to its corresponding BGP prefix announcement. While this might initially seem simpler, it quickly becomes complex due to various factors. These include CDNs using varying prefix lengths in their BGP announcements, intricate configuration policies such as announcing the same prefixes from all sites while managing per-site selection internally, and large anycast networks involving multiple prefixes. In contrast, our latency-based approach effectively addresses these challenges as we show in our analysis.

In Table 8, we present a comparison of identification of replica selection mechanism used by CDNs across different countries using two types of vectors: the Coefficient of Regionalization Vector (CRV) and the Prefix Divergence Vector (PDV). The CRV captures the distance

[4] From personal communication we understand that Yahoo uses other approaches in other locales.

CDN	Country	CRV[1]	CRV[3]	PDV[1]	PDV[3]
Akamai	UK	1.00	1.00	0.67	0.70
Akamai	RU	0.90	1.00	0.69	0.75
Akamai	US	1.00	1.00	0.58	0.79
Cloudflare	BR	0.01	0.01	0.07	0.07
Cloudflare	UK	0.05	0.01	0.16	0.13
Cloudflare	IN	0.01	0.01	0.18	0.13
EdgeCast	BR	0.20	1.00	0.18	0.93
EdgeCast	UK	0.30	1.00	0.33	0.97
EdgeCast	AU	0.20	1.00	0.32	0.88
Cloudfront	US	1.00	1.00	0.78	0.84
Cloudfront	IN	1.00	1.00	0.67	0.81
Cloudfront	ZA	0.80	1.00	0.83	0.84
Fastly	BR	0.35	0.37	0.28	0.41
Fastly	TR	0.20	0.29	0.59	0.65
Fastly	ZA	0.40	0.40	0.52	0.54
Azion	BR	1.00	1.00	0.43	0.38
NGENIX	RU	0.31	0.39	0.39	0.39
Google	BR	0.68	0.77	0.24	0.52
Google	RU	0.64	0.85	0.47	0.53
Google	IN	0.52	0.70	0.37	0.49

Table 8 CRV and PDV values for various CDNs and countries.

between the latency distributions associated with the DNS local resolver and those of resolvers at increasing geographic scopes. Specifically, each entry, $CRV[i]$, represents the distance between the local resolver's latency distribution and that of the $(i+1)$th resolver scope. Complementing the CRV, the PDV (Prefix Divergence Vector) is introduced to capture the diversity of replica IP prefix sets associated with various resolver scopes. Each entry in the PDV, $PDV[i]$, represents the Jaccard distance between the replica IP prefix set of the local resolver and that of the $(i+1)$th resolver scope. This metric quantifies the degree of similarity or dissimilarity between the sets, where a value closer to 0 indicates high overlap (similar sets), and a value closer to 1 signifies greater divergence (less overlap). This table includes the CDNs and countries for which CRV values are shown in the main paper, and here, we provide the corresponding PDV values for the same CDN-country pairs.

For anycast-based replica selection, we expect that both $CRV[1]$ and $CRV[3]$, as well as $PDV[1]$ and $PDV[3]{\approx}0$. For DNS-based replica selection, we expect $CRV[1]$ and $CRV[3]$ along with $PDV[1]$ and $PDV[3]{\approx}1$. For regional-anycast replica selection, we expect $CRV[1]$ and $PDV[1]{\approx}0$ and $CRV[3]$ and $PDV[3]{\approx}1$.

However, this table shows that $PDV[1]$ can be as low as 0.58 even for Akamai, while $CRV[1]$ and $CRV[3]$ remain consistently high across all Akamai locations. Similarly, for Azion, a regional CDN in Brazil, $CRV[1]$ and $CRV[3]{\approx}1$ but $PDV[3]$ is as low as 0.38. These findings highlight that, while the overlap in replica prefixes across resolver scopes can vary significantly for different replica selection approaches, latency measurements reliably indicate whether the DNS resolver's location was used as a proxy for the client's location. Specifically, anycast-based replica selection consistently shows low latency to nearby replicas across resolver scopes, whereas DNS-based selection exhibits latency variations based on the resolver's location. This underscores latency as a clearer and more reliable metric for identifying a CDN's replica selection mechanism.

A.5 Detecing Anycast-based Replica Selection Using MAnycast2

While the objective of our methodology is the identification of the main replica selection approach used by any given CDN, and not the determination of a IP address as anycast or unicast, we use the MAnycast2 [55] technique to assess whether the IP prefixes associated with each CDN's replicas are anycast addresses or not.

Within the subset of CDNs categorized as DNS-based, our analysis revealed that, with the exception of Akamai and Cloudfront, none of the IPs associated with their replicas were identified as anycast. In the case of Akamai, out of 4070 Akamai prefixes, 4 prefixes were identified as anycast and 17 out of the 924 Cloudfront prefixes were identified as anycast. Conversely, among CDNs identified as anycast, all prefixes of Cloudflare and StackPath were recognized as anycast by MAnycast2. Notably, none of NGENIX's prefixes were identified as anycast, likely due to potential coverage limitations. In instances where CDNs are identified to use a combination of DNS-based and anycast approaches, MAnycast2 identified 583 out of 1092 Google prefixes and 50 out of 213 Fastly prefixes as anycast. For Regional Anycast, MAnycast2 identified all prefixes of Edgio as anycast and none of Yahoo's prefixes as anycast.

As noted by Sommese et al. [55], the variability in the accuracy of MAnycast2 across regions, can be influenced by differences in connectivity density relative to various vantage points in their testbed. This variability may hinder the detection of regional anycast services (e.g., Yahoo) or smaller deployments (e.g., NGENIX).

In addition, differentiating between anycast and regional anycast solely based on IP assignment patterns is not straightforward due to the absence of consistent patterns in the use of allocated IP blocks per region.