NORTHWESTERN UNIVERSITY

Consolidation and Consequence in Web Delivery Infrastructure

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Computer Science

By

Rashna Kumar

EVANSTON, ILLINOIS

February 2026

Abstract

Modern web services rely heavily on third-party providers for foundational functions such as content hosting, authoritative DNS, and certificate authorities. As a result, a small number of organizations now play an increasingly important role in how websites are delivered and reached. This dissertation investigates consolidation in core web infrastructure at Internet scale, asking not only how concentrated these service layers are, but also when similar concentration patterns reflect different underlying conditions and consequences for resilience, exposure, and performance.

The dissertation proceeds in two parts. The first measures consolidation across commercial and public-sector web infrastructure around the world. It shows that concentration is substantial, including in government web services, but also that concentration by itself does not tell the full story. Similar levels of concentration can emerge for very different reasons. To explain this variation, the dissertation introduces a comparative framework that distinguishes structural consolidation, which reflects broader ecosystem constraints, from strategic consolidation, which reflects choices made by particular actors. By comparing government websites with popularity-stratified commercial websites in the same national context, the dissertation shows how concentration can be interpreted relative to the surrounding provider environment rather than viewed in isolation.

The second part studies what consolidation means for users. It shows that concentration can expand shared failure domains and reduce practical redundancy, increasing fragility when widely used providers fail. It also examines exposure through both foreign service dependencies and the on-path intermediaries involved in reaching government websites, showing that consolidation can deepen cross-border and network-level dependencies with implications for jurisdiction and security. Finally, it analyzes performance, showing that consolidation shapes latency both through where infrastructure is deployed and through how users are directed to content.

Taken together, this dissertation argues that Internet infrastructure consolidation cannot be understood from concentration metrics alone. The same observed concentration may reflect unavoidable ecosystem constraints or deliberate deployment choices, and these different conditions produce measurably different consequences for resilience, exposure, and performance. By providing a comparative framework for distinguishing these regimes, this work offers a foundation for more rigorous and accountable interpretation of consolidation across sectors and countries.

Thesis Committee

Fabián E. Bustamante	Committee Chair
Northwestern University	
Yan Chen	Committee Member
Northwestern University	
Peter Dinda	Committee Member
Northwestern University	
Aleksandar Kuzmanovic	Committee Member
Northwestern University	
Marcel Flores	External Committee Member
Netflix	
Phillipa Gill	External Committee Member
Google	

Table of Contents

CHAPTER 1

Introduction

Over the past decade, the Internet has increasingly shifted from on-premises infrastructure toward third-party services, driven by scalability, operational efficiency, and improved security. This transition has also concentrated critical service layers onto a small number of providers. However, concentration alone is an ambiguous signal: the same measured level of consolidation can reflect unavoidable constraints of a local ecosystem (structural consolidation) or deliberate deployment and procurement choices within that ecosystem (strategic consolidation). These regimes can produce measurably different consequences for resilience, exposure, and performance even when scalar concentration metrics appear similar.

This dissertation examines how the Internet's growing reliance on third-party infrastructure reshapes dependency and concentrates control across foundational service layers, including content hosting, authoritative DNS, and certificate authorities. It first measures consolidation at Internet scale across countries and across sectors (commercial and public-sector web services), showing that comparable concentration levels can arise either from ecosystem constraints or from deployment choices, making concentration alone an ambiguous indicator of control, exposure, and resilience. It then resolves the resulting ambiguity by introducing a comparative baseline: within each country, it treats the domestic commercial ecosystem as a counterfactual reference for what levels of concentration are structurally induced, and identifies when public-sector choices meaningfully diverge from those constraints. Finally, it uses this structural/strategic distinction to interpret the user-facing implications of consolidation for resilience, exposure, and performance.

1.1. Part I: Measuring Consolidation Across Service Layers and Sectors

The first part of this dissertation measures consolidation in infrastructure services. Using worldwide vantage points collected over a two-year measurement campaign, we analyze how popular websites rely on third-party providers for content delivery networks, authoritative DNS, and certificate authorities. We find substantial concentration within each service category, with a small number of providers serving a dominant share of global demand. We then extend the same measurement lens to public-facing government websites across diverse countries and regions. These results establish a motivating puzzle: similarly high concentration can appear in countries with very different provider ecosystems and very different exposure and control outcomes. Chapters 4 and 5 document this phenomenon; Chapter 6 introduces the comparative framework that explains it.

These measurements establish that consolidation is pervasive, but they also reveal why measurement alone is insufficient. Identical concentration values can arise in countries with radically different provider ecosystems and governance constraints, making it unclear whether consolidation reflects ecosystem limits or discretionary choices. This ambiguity motivates the comparative approach in Chapter 6, which reframes consolidation as a two-dimensional phenomenon: how concentrated and why concentrated.

1.1.1. Explaining Consolidation: Structural and Strategic Consolidation

Observing that government websites are often similarly consolidated (as we quantify in Chapter 5), we ask what drives that consolidation. In some countries, high concentration is structural, reflecting a limited domestic provider ecosystem that also constrains the commercial web. In others, concentration is strategic, reflecting public-sector deployment and procurement choices that centralize on a narrower set of providers than the surrounding market would require. Without distinguishing these regimes, scalar consolidation measurements are ambiguous with respect to control, failure domains, and user-facing impact. Chapter 6 resolves this ambiguity using a within-country, stratified observational design that treats the domestic commercial ecosystem as a counterfactual baseline. Across 61 countries, we compare government websites to popularity-stratified commercial sites operating in the same national network environment and measure consolidation across hosting, authoritative DNS, and certificate authorities.

1.2. Part II: User-Facing Implications of Consolidation

The second part of this dissertation examines user-facing implications of consolidation through the structural/strategic lens, with a focus on three outcomes: resilience, exposure, and performance. Conditioning on consolidation type clarifies when concentration reflects shared ecosystem constraints versus discretionary centralization, and therefore when observed resilience, exposure, and performance outcomes should be interpreted as structural limitations or as the consequences of deployment choices.

1.2.1. Resilience Implications

First, consolidation can heighten systemic fragility when a small number of providers become shared failure domains for large fractions of the web, such that even routine faults can cascade into widely visible incidents. Recent disruptions in 2025 illustrate this amplification effect: Cloudflare experienced a global outage on November 18, 2025 after an internal change triggered failures in core traffic handling, disrupting access to many sites and services. AWS saw a major US-EAST-1 disruption around October 20, 2025, with knock-on effects across many dependent services. Microsoft similarly faced a global Azure outage on October 29, 2025, which also impacted Microsoft 365 and other downstream services. Crucially, these incidents show that even sophisticated and well-resourced providers are not immune to failures, and when so much of the Internet depends on them, a single disruption can ripple outward at scale.

At the same time, an open question is whether the type of consolidation changes the redundancy that users can realistically rely on: if governments centralize strategically, do they obtain meaningfully better fallback than what structurally constrained ecosystems permit? Chapter 7 shows that, across service layers, practical fallback is limited regardless of consolidation regime. In particular, authoritative DNS exhibits near-zero organizational redundancy (most domains depend on a single independent DNS provider), and hosting redundancy is limited and often reflects multi-origin complexity rather than true failover capacity.

1.2.2. Exposure Implications

Second, consolidation can increase exposure by increasing dependence on infrastructure outside local control. One form is jurisdictional exposure: reliance on foreign providers and hosting locations introduces cross-border dependencies that complicate accountability, legal jurisdiction, and operational control. Conditioning on consolidation regime clarifies why similar concentration levels can imply different exposure outcomes: structurally consolidated countries exhibit more uniform exposure, while strategically consolidated countries show substantially greater dispersion, indicating that when governments diverge from market constraints, exposure becomes more a consequence of policy choice than a fixed feature of the ecosystem (Chapter 8).

Exposure, however, is not only a property of where services are hosted, but also of which networks citizens traverse to reach them. This dissertation therefore complements hosting-level analysis with path-level measurement: Chapter 8.3 analyzes on-path infrastructure dependencies to government websites, including transit networks and exchange points. We find that access to government services often traverses a small set of shared foreign intermediaries, revealing an additional and often overlooked layer of concentration that hosting-only analysis cannot detect. This pattern also has security implications: countries with greater foreign on-path exposure tend to exhibit weaker HTTPS adoption, increasing the risk that sensitive interactions are not protected by encryption and authentication, particularly when paths traverse third countries outside both the government's country and the hosting country.

1.2.3. Performance Implications

Third, consolidation can affect the performance experienced by end users, and conditioning on consolidation type clarifies the mechanisms. Chapter 9 shows a direct pathway through infrastructure placement: across countries, structurally consolidated regimes tend to impose higher domestic latency to hosting, DNS, and CA endpoints with tighter government–commercial overlap, consistent with shared ecosystem constraints, whereas strategically consolidated regimes exhibit lower median latency but greater dispersion, consistent with uneven localization and deployment choices (Section 9.1).

Consolidation can also shape performance indirectly through content delivery steering. As CDNs consolidate, a growing fraction of the web is served through a small number of delivery platforms, centralizing not just where content is hosted, but also how users are steered to replicas (Sections 9.2–9.3). Replica selection is the CDN control plane that maps users to specific points of presence, shaping the latency they experience.

CDNs typically select replicas using one of three approaches. With anycast-based routing, Internet routing delivers a user to a nearby replica, but catchments can be difficult to predict and may shift over time. With regional anycast, DNS first maps the user to a region, and anycast then selects a nearby replica within that region, limiting extreme catchment stretch while retaining anycast within-region dynamics. With DNS-based replica selection, the CDN explicitly chooses a replica based on the apparent location of the user's DNS resolver, using DNS as the primary steering signal.

To study how this centralized steering is implemented in practice, we introduce a methodology that infers a CDN's predominant replica selection approach from the outside, at scale. We use it to characterize which steering mechanisms are most commonly used to serve popular content. We find that DNS-based replica selection is the dominant approach. This result matters because it means that content delivery performance is often mediated by DNS behavior rather than only by Internet routing.

We find that DNS-based replica selection is the dominant approach. This result matters because it implies that content delivery performance is often mediated by DNS behavior rather than only by Internet routing. The resulting chain is: CDN consolidation increases reliance on centralized steering platforms; DNS-based steering makes resolver location the dominant steering signal; resolver consolidation weakens that signal when resolver deployment is decoupled from end users; and user-facing latency increases when users are mapped to suboptimal replicas even when closer replicas exist. In combination, CDN consolidation, reliance on DNS-based steering, and consolidation in DNS resolvers create an indirect pathway through which concentration at one layer can translate into user-facing performance degradation.

1.3. Summary

Taken together, this dissertation shows that observed consolidation is ambiguous without a comparative baseline: similar concentration levels can reflect structural constraints or strategic centralization, and these regimes produce measurably different outcomes for resilience, exposure, and performance. Through large-scale measurements across service layers and sectors, and a within-country comparative design that separates how concentrated from why concentrated, this work provides a unified account of Internet consolidation and demonstrates why explanation is necessary for interpreting its technical and societal implications.

1.4. Dissertation Organization

The remainder of this dissertation is structured as follows. Chapter 2 presents the thesis statement. Chapter 3 presents background and context for the dissertation. Chapter 4 discusses a large-scale longitudinal study of third-party dependency and consolidation trends in the service infrastructure supporting popular commercial websites, with a focus on foundational services including content hosting, authoritative DNS, and certificate authorities. Chapter 5 focuses on web hosting strategies and the extent of consolidation in the infrastructure behind public-facing government websites across countries. Chapter 6 is the core contribution: it examines what drives observed consolidation patterns and introduces a comparative framework to distinguish structural consolidation from strategic centralization using within-country comparisons between government websites and popularity-stratified commercial websites. Chapters 7–9 then use this distinction to interpret implications that would otherwise be conflated under scalar concentration metrics. Chapter 7 analyzes resilience implications, including shared failure domains and limited practical redundancy across critical services. Chapter 8 examines exposure implications, including foreign service dependencies and on-path infrastructure dependencies involved in access to government websites, and their implications for jurisdiction and security. Chapter 9 analyzes performance implications, connecting latency outcomes under consolidation regimes with content-delivery steering, and introduces a methodology to infer CDN replica selection approaches at scale to study how resolver consolidation interacts with DNS-based steering to affect end-user latency. Chapter 10 presents related work and puts this dissertation in perspective. Finally, Chapter 11 concludes the dissertation by synthesizing the main findings, discussing their technical and governance implications, and outlining directions for future work.

CHAPTER 2

Thesis

Observed consolidation in Internet infrastructure is ambiguous without a comparative baseline: the same concentration level can reflect unavoidable ecosystem constraints or deliberate deployment choices, and these two regimes produce measurably different outcomes for resilience, exposure, and performance.

CHAPTER 3

Background

Over the past decade, the Internet has increasingly shifted from on-premise infrastructure toward third-party services, driven by clear operational advantages such as elastic resource allocation, global deployment footprints, high availability, and lower capital and operational costs [**19, 90, 104**]. At the same time, this shift has changed who operates the infrastructure underlying web access, trust, and delivery, raising concerns about consolidation and centralization across critical Internet services [**193, 231, 6, 147, 115, 82, 143, 119, 98**]. A growing body of work quantifies this consolidation, but concentration metrics alone do not explain why consolidation arises or when similar concentration levels imply different forms of control and risk.

Modern websites depend on multiple external services that determine reachability, authentication, and performance, including content hosting and delivery infrastructure, authoritative DNS, and certificate authorities [**115, 27, 204, 114, 62, 221**]. While the web appears decentralized at the level of domain names, operational control over these foundational services is frequently concentrated in a small number of providers, creating shared failure domains and centralized control points [**15, 99, 101, 131, 193, 17**]. Real-world incidents, including CDN and cloud outages, certificate revocation events, and coercive routing disruptions, illustrate how failures at dominant providers can propagate across otherwise independent sites and produce Internet-scale disruption [**173, 124, 176, 175, 174**].

Two observations motivate studying these dependencies comparatively at Internet scale. First, web demand is not globally uniform. Beyond a small set of globally dominant sites, popularity is often region specific, so the services and providers that matter to users vary across countries and markets [**184**]. Second, although third-party infrastructures such as CDNs and cloud platforms have expanded significantly, they are not omnipresent, and their deployment, performance, and relative advantages vary across locales [**82, 28, 229, 191**]. As a result, websites and content providers often make different service choices across regions based on connectivity, availability, and cost [**191**]. These observations imply that consolidation should be studied across countries and across different parts of the web, rather than only through a single global ranking or a single vantage point [**115, 215**].

These issues are particularly salient for government websites, which increasingly serve as a primary interface between states and citizens for services such as taxation, healthcare, licensing, and immigration [**214, 225, 149**]. As governments move these services onto the public web, they must make infrastructure choices between in-house operation and third-party providers such as commercial hosting and content delivery networks. These choices carry consequences beyond performance and cost: reliance on third-party infrastructure can expand shared failure risk, introduce external dependencies, and shift aspects of control over essential civic services outside direct public oversight. At the same time, governments operate within domestic market constraints, making it necessary to interpret public-sector concentration relative to what the surrounding commercial ecosystem makes feasible.

Interpreting consolidation, however, requires more than concentration metrics alone. Comparable concentration levels may arise either from unavoidable ecosystem constraints or from deliberate deployment choices, and these cases can imply different risk profiles. In some countries, dependence reflects structural limits in the domestic provider ecosystem; in others, it reflects strategic deployment or procurement choices by governments within that ecosystem. Distinguishing these regimes requires a comparative baseline that situates government deployments within the surrounding commercial environment, holding national network conditions fixed. This motivates treating consolidation not only as a descriptive measurement problem, but also as a comparative inference problem.

This comparative perspective is essential for analyzing the implications of consolidation, because resilience, exposure, and performance depend not only on how concentrated services are, but also on why those dependencies arise and how they are embedded in the broader infrastructure. Conditioning on consolidation regime clarifies when observed outcomes should be interpreted as structural limitations versus the consequences of deployment choices (as examined in Chapters 7–9). The consequences of consolidation extend beyond service ownership and market share to the mechanisms that govern how users reach services and how traffic is steered across the Internet.

Consolidation also shapes end-user performance through content-delivery steering. At the heart of CDN operation is *replica selection*, the process that determines which server handles a user request. Replica selection is not only a performance optimization mechanism, but also a control point that determines how traffic is distributed and which infrastructure mediates user experience. CDNs commonly rely on DNS-based redirection, anycast, or regional anycast, each of which allocates steering authority differently between the CDN, DNS resolvers, and interdomain routing [**154**, **7**, **41**, **142**, **232**]. The EDNS Client Subnet (ECS) extension further highlights the tradeoff between steering accuracy and privacy by exposing client prefix information to authoritative DNS infrastructure [**55**, **?**, **?**, **52**]. Although prior work has extensively studied CDN performance and mapping efficiency [**234**, **75**, **159**, **97**, **113**, **83**], steering choices remain difficult to observe at scale and are often opaque to users and policymakers.

This performance story is tightly linked to DNS resolver consolidation. DNS directly affects user experience through lookup latency, and indirectly affects performance because many CDNs still use resolver location as a proxy for user location in DNS-based steering [**33**, **30**, **31**, **95**, **39**]. In practice, modern DNS resolution often involves complex chains of ingress, forwarding, hidden, and egress resolvers, sometimes operated by different organizations and located in different networks or countries [**186**, **4**]. As resolver infrastructure consolidates onto a small set of third-party providers, resolver location may become a weaker proxy for end-user location, creating hidden dependencies and increasing the risk of suboptimal steering [**2**, **182**, **99**, **171**, **147**, **220**, **146**].

Studying these issues raises substantial measurement challenges. Many dependencies are not directly observable and must be inferred from partial views of service infrastructure, routing, and control-plane behavior. Service-level studies require provider attribution across heterogeneous hosting, DNS, and CA deployments. Path-level studies must contend with traceroute opacity, incomplete hop visibility, MPLS tunneling, ICMP filtering, and route variation over time. Exposure analyses require jurisdiction-aware attribution of ASes, IXPs, and service operators despite multinational operation and incomplete metadata. Performance analyses require geographically diverse vantage points in access networks, not only in well-provisioned backbone locations. These constraints motivate measurement methodologies that prioritize broad coverage, conservative inference under uncertainty, and transparent attribution across service, network, and control-plane layers.

Taken together, this background motivates the central approach of this dissertation: to measure consolidation across foundational Internet services (Chapters 4–5), distinguish structural from strategic forms of concentration through within-country comparative analysis (Chapter 6), and connect these patterns to user-facing implications for resilience, exposure, and performance (Chapters 7–9).

CHAPTER 4

Third-Party Dependency and Consolidation in the Commercial Sector

The shift from on-premise infrastructure to third-party providers has become a global trend, fundamentally reshaping the Internet over the past decade. This shift is driven by the benefits offered by major providers, including access to computing resources in data centers across multiple regions, flexible resource allocation, high service availability, and relatively lower capital and operational costs [**19**, **90**, **104**]. Yet the same shift has a second, less well characterized consequence: it makes "consolidation" observable at scale, while simultaneously making its interpretation ambiguous. Similar concentration levels in third-party services can arise because local ecosystems constrain feasible alternatives or because providers and sites deliberately converge on a small set of platforms. This chapter establishes the measurement puzzle that motivates the dissertation's structural versus strategic distinction developed in Chapter 6.

This shift, however, has also led to increasing concerns about Internet consolidation and centralization, the concentration of traffic, infrastructure, services, and users on a handful of providers. Importantly, concentration alone does not explain whether consolidation reflects unavoidable ecosystem constraints or discretionary deployment and procurement choices. Throughout Part I, we therefore treat consolidation measurements as evidence of a phenomenon that is real and consequential, but not self interpreting without an explicit comparative baseline

The 2019 Global Internet Report [**193**] provides an early overview of this trend in every aspect of the Internet economy, from access provision to service infrastructure and applications. It argues that while consolidation is often seen as an expected result of maturing markets and industries, the combination of society's increased dependency on the Internet, business agility, and the almost total lack of regulation is leading to a handful of platforms in control of much of the Internet's functionality and interoperability. Subsequent measurement efforts have documented consolidation across layers and regions, but largely stop at quantifying concentration rather than distinguishing its origins or identifying when similar measured levels correspond to fundamentally different underlying regimes. Since that report, several efforts have explored this trend [**231, 6, 147, 115, 82, 143, 119, 98**] and its economic, political and reliability implications [**15, 99, 101, 131, 231, 193, 17**].

This chapter begins Part I by measuring third-party dependency and concentration in commercial web infrastructure at global scale, and by surfacing the core interpretability challenge that motivates the dissertation: similar concentration levels can mask very different underlying drivers.

The Web provides a relatively accessible environment to characterize these centralization trends in public-facing content and, incidentally, to understand their concerning implications. Accessing a website depends on several services provided by third parties, including DNS, CDNs, and CAs. To visit a site, a user must interact with at least one DNS authoritative nameserver to retrieve the IP address of the web server(s) hosting the content. These servers may be operated by one or more CDNs to enhance reliability and performance. If the servers use HTTPS, the client may also need to consult one or more CAs to verify the validity of the servers' SSL certificates. In fact, a popular website may rely entirely on third-party providers for these critical services. This layered dependency structure is precisely what makes consolidation measurable using public signals, and also what makes its implications non-obvious without separating what is constrained by ecosystem structure from what is chosen strategically.

Several recent studies have leveraged this observation to assess third-party dependencies in popular websites, though typically from only a single vantage point [**115**, **27**, **68**, **204**, **114**] or only a few vantage points [**215**].

We build on these prior work ***to characterize how third-party dependency and concentration vary across countries and regions of the world, and to show why cross-country variation alone is insufficient to interpret consolidation without a baseline for comparison***.

Our work is motivated by two simple observations. *First*, while websites could potentially be accessed anywhere, not all websites are popular everywhere; beyond a few top-ranked sites, popularity is region specific. *Second*, while many third-party services such as DNS and CDNs have been building global infrastructures, on and off-networks [**82**], their deployment is not (yet) omnipresent, and their relative performance compared to competitors varies across markets. Together, these observations imply that a single global consolidation statistic can conceal distinct country-level realities. In Part I, we therefore use country-level measurement to establish where concentration looks similar across contexts and where it does not, setting up the need for the within-country, stratified comparisons introduced in Chapter 6.

Motivated by these observations and concerns, we explore third-party and centralization trends in the wider Internet for public web content. We present a methodology that builds on prior work [**115**, **215**] to carry out a large-scale, longitudinal analysis of third-party dependency and centralization around the world. Using this methodology, we analyze the dependencies of top regional websites in 50 countries, covering approximately 78% of the global Internet population (a total of 16,774 unique websites). We provide results from two consecutive years, offering insights into how these trends have evolved over time.

Our findings reveal that third-party dependencies and critical dependencies vary significantly across regions. We report that between 19% and as much as 76% of websites, across all countries, depend on a DNS, CDN, or CA third-party provider. Critical dependencies, where a service necessary to ensure access to a website depends on a single provider, while lower are equally spread ranging from 5% (CDN in Costa Rica) to 60% (DNS in China). At the same time, and despite this heterogeneity in dependence, the market of third-party providers is highly concentrated: the top three providers across all countries serve an average of 92% of sites and Google, *by itself*, serves an average of 70% of all websites. We further observe that these values increase over time (for example, by ≈14% on average for CDNs in the second year of our data). This combination of wide variation in dependence alongside persistently high concentration highlights the core puzzle that motivates the dissertation's reframing: concentration is visible, but its meaning is not. The remainder of Part I (Chapters 5 and 6) generalizes these measurements and then introduces the baseline needed to distinguish structural from strategic consolidation.

Our findings reveal that third-party dependencies and critical dependencies vary significantly across regions. We report that between 19% and as much as 76% of websites, across all countries, depend on a DNS, CDN, or CA third-party provider. Critical dependencies, where a service necessary to ensure access to a website depends on a single provider, while lower are equally spread ranging from 5% (CDN in Costa Rica) to 60% (DNS in China). Interestingly, despite this high variability, the market of third-party providers seems highly concentrated: the top-three third-party providers across all countries serve an average of 92% of all sites and Google, *by itself*, serves an average of 70% of all websites. Perhaps more problematically, we find these values have increased, a year later, by ≈14% on average, for CDNs.

While the remainder of this chapter focuses on measuring dependency and concentration, a key point to keep in mind is that these quantities are not self-interpreting. A country can appear highly consolidated because the ecosystem offers few viable alternatives, or because adoption converges on a dominant platform despite available options. We return to this distinction in Chapter 6, where we formalize a within-country comparative baseline to separate structural constraints from strategic choices.

4.1. Methodology and Dataset

In this section, we describe a measurement methodology that builds on Kashaf et al. [**115**]'s to characterize service dependencies and centralization around the world. Our

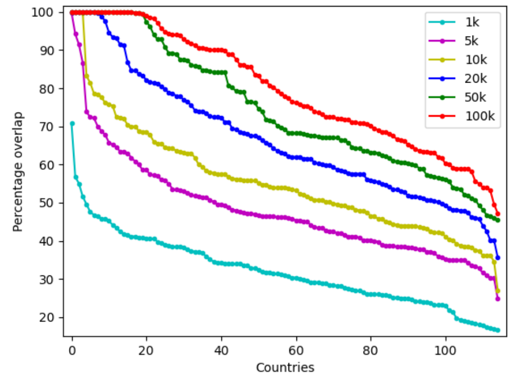


Figure 4.1. Overlap between global and top regional sites. Countries are sorted based on their overlap and plotted left to right.

Overlap Class	Country Codes
High	AE, AR, AU, BR, CA, CN, DE, ES, FR, GB, GR, HK, ID, IN, IT, JP, KR, MX, MY, SG, TR, TW, US, VN
Medium	BE, CH, CL, CR, IL, UA, PL, NL, NO, RO, SE, TH, ZA
Low	AL, BA, BG, CZ, DK, EE, GE, HU, LV, MD, NZ, PT, RS

Table 4.1. Countries grouped by degrees of overlap between top-regional sites and the global ranking.

extensions to Kashaf et al.'s are meant for carrying out a country-specific analysis, including the selection of countries based on the ranking overlap and the use (and validation) of VPN vantage points for measurement; we also analyze the number of websites using OCSP must-staple and quantify the resources that do not have CNAMES and can be mapped to a CDN. We close with a description of the dataset we collect using it.

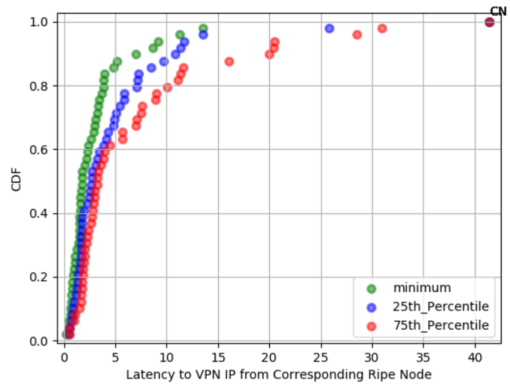


Figure 4.2. Latency (ms) from RIPE Atlas Nodes to corresponding VPN Nodes.

4.1.1. Country Selection

As Fig. 4.1 shows, the set of top-ranked websites is clearly determined by a handful of countries, while half of all countries have less than ≈77% overlap, even with the full top-100k list of websites. Thus to understand the degree of third-party dependency around the world, we should use a set of countries that together capture the range of possible overlaps.

To this end, we first group all 115 countries with regional rankings in Alexa based on the degree of overlap between their regional popular sites and the list of top global sites. We divide the countries based on this overlap into three groups: *high-*, *medium-* and *low-overlap* for the top, medium and bottom third of countries with the highest overlap. We then select 50 countries from across these three groups ensuring that (*i*) the sample of countries in the analysis captures a sufficiently large fraction of the Internet user population, and (*ii*) there are VPN vantage points in those countries with claimed locations that can be verified. We select countries that have a range of overlap to study the third-party dependency of the popular sites of countries that are not well represented in the global rankings and are therefore not investigated in the broader global analysis.

Figure 4.1 lists the selected countries (using their two-letter codes) which includes 38.4%, 25.9% and 35.7% from the high, medium and low-overlap sets based on regional websites.

In total, our dataset covers all inhabited regions of the world and captures 78.1% of the world's Internet user population [**226**]. We group countries in five different regions: the Americas, Europe, Asia Pacific, Africa, and the Middle East and calculate the percentage of the world's Internet population covered by the corresponding countries in those regions.

In order to make observations about each country, we consider vantage points available via VPN providers. Of the 50 countries, all but China have vantage points available through the Nord VPN [**152**]. To include China we use the Hotspot Shield VPN, since Nord and many other VPNs do not provide service in China [**116**].

To gain confidence in the claimed location of the vantage points, we obtain their public IPs (resolving a domain whose authoritative server we control) and use a set of five RIPE Atlas nodes within the same country to send a sequence of three ICMP pings to the vantage point. We would expect most nodes geographically close to the vantage point's claimed location to have minimum latencies in the 10-20ms range and below 50ms. As a reference, China, the largest country in our set, is 5,250km East-West or ≈40.25ms considering a 2.3x median inflation over *c*-latency [**192**].

Figure 4.2 shows the minimum, 25th and 75th percentile of ping latency from the set of RIPE Atlas nodes to the corresponding vantage point. The minimum latency to all vantage points, with the exception of the one in China, is within 15ms and 86% of minimum latencies are below 5ms, suggesting that these nodes are within the claimed country. We further geolocate the vantage points' IPs using two popular geolocation databases: MaxMind GeoLite2 [**141**] and IP2Location BD11.Lite [**109**]. Past work has shown the geolocation databases to be reliable at the country-level [**168**]. Both databases place the IPs of all 49 Nord VPN nodes in the claimed countries. China proved to be more challenging. The two geolocation databases we use disagree with Maxmind geolocating the node in China, while the IP2Location database placing it in Japan. The ping latency to the VPN node in China, however, are consistent with our estimations and thus we consider the claimed location to be correct.

4.1.2. Data Collection Methodology

Using VPN vantage points in the selected countries, we launch measurements to their country's set of top-500 regional websites, and use a range of heuristics for labeling three major services – DNS, CA and CDN.

4.1.2.1. DNS Measurements. We use a number of heuristics to label all nameservers used by a website as *private* or *third-party*. For instance, the site belgocontrol.be uses two distinct nameservers, i.e. *skynet.be which is a third-party nameserver and *belgocontrol.be which is a private nameserver. For each website, we find all the nameservers used by the website, that belong to different logical entities, by issuing NS queries to the domain name from the selected country's VPN vantage point. Note that we do not perform resolution at this stage so our results are not affected by caching, we just find the unique set of DNS providers used by the website. We start by labeling each nameserver used by a website as of an *unknown* type. We then compare the second level domain (2LD)[1] of the website and the nameserver with a match suggesting this is a *private* nameserver [**118**].

While the *2LD-matching* heuristic works well in most cases, it may result in misclassifications of some nameservers. For instance the nameserver of youtube.com is *google.com and though both belong to the same logical entity, the *2LD-matching* heuristic will classify the nameserver as third-party. To resolve this we make use of an additional heuristic based on Subject Alternative Names (*SANs List*) [**34**]. If the website uses HTTPS, we find the site's SANs list via the SSL certificate of the website. For each unclassified nameserver, we then look for the presence of the second level domain of the nameserver in the SAN list, whereby the presence indicates a *private* DNS nameserver. This heuristic correctly identifies cases like youtube.com using a private DNS.

We use a third heuristic based on Start of Authority records (SOA) – *SOA-record-matching* – to label the unclassified nameserver [**115**]. In this case, we compare the entity pointed to by the SOA records of the website and the DNS provider pointed to by the SOA records of the nameserver; a mismatch here indicates a *third-party* DNS nameserver. For instance, the SOA record for the website imdb.com is *amazon.com and its nameservers

[1]By 2LD we refer to 2LD + TLD in this work

are Dynect and UltraDNS. Since the SOA records of these nameservers do not match the SOA record of imdb.com, we label imdb.com as using two third-party DNS providers.

For the remaining unknown servers, if the concentration of the nameserver (i.e. the number of websites dependent on a given provider) is large, we label it as *third-party*. We set the value of the threshold as >50 (i.e. if an unlabeled nameserver serves greater than 50 websites, we label it as a third-party).[2] For sanity check, we manually investigated the servers that were labeled by this heuristic and they were all, in fact, popular third-party DNS providers, such as Amazon, Akamai, NsOne, Cloudflare, DnsPod and Alibaba DNS. We also performed a sensitivity analysis on this threshold and observed that reducing the value of the threshold resulted in some nameservers that we could not manually determine as third-party with full confidence such as *gandi.net in France and *.hyp.net in Norway.

The second condition of Algorithm 1 summarizes the heuristic when the service type is instantiated as *DNS*. This basic three-step classification logic, involving 2LD-matching, SANs-List and SOA-record-matching, is described in Algorithm 2 where the *service.url* is provided to the algorithm is the DNS nameserver.

DNS Redundancy. We also measure the percentage of websites that are served by a single DNS provider (i.e., critically dependent on this provider) or served by multiple third-party DNS providers, and the percentage of websites that are served by private and third-party DNS providers.

[2]Following Kashaf et al. [**115**].

4.1.2.2. CDN Measurements. To find the set of CDNs hosting the targeted website and determine whether the CDNs used are private or a third-party service, we find the CNAME of the internal resources of the website. We start by using the webdriver capabilities of the Selenium library in python to generate a HAR file for each website which gives us all the resources of a website. We filter internal resources from the set of *all* resources by matching the 2LD of the website to that of the resource, checking the presence of the 2LD of the resource in the SAN List of the website, and comparing the SOA records of the website and resource, a match in any of the three cases indicates an internal resource. We additionally use public suffix lists [**129**, **209**] to identify any remaining internal resources.

To find third-party dependence, we find the CNAMEs of all the internal resources of a website by issuing dig CNAME queries on all the internal resources of the webpage. We then obtain the set of CDNs used by the internal resources from our self-populated CNAME-CDN map [**44**, **227**]. An alternative way to identify the CDN hosting an internal resource without a CNAME redirect would be to compare the autonomous system number (ASN) of the resource with those of popular CDNs [**227**, **139**]. We find an additional 17% of all resources, on average across countries, can be mapped to their CDNs using this approach. However, since the classification algorithm depends on CNAME for the labelling, we can not leverage the AS mapping approach here. Our results, therefore, show a lower bound on the third-party dependency trends for the CDN service. The process is summarized in the third condition of Algorithm 1.

Next, we determine whether each CDN that hosts the internal resources of a given site is a private service or a third-party-provided one. For each *(website, CDN)* pair, we extract the CNAMEs of the internal resources of the website which uses that CDN. Then for each of these CNAMEs, if the 2LD of the CNAME is the same as the 2LD of the website, we classify the CDN as private. If the website uses HTTPS and the 2LD of the CNAME is present within the SAN list obtained from the SSL certificate of the website, the CDN is again classified as private. For example, the website twitch.tv has resources fetched from the CDN Fastly and contains CNAMEs such as *.fastly.net. The 2LD of the CNAME and website do not match but the presence of the 2LD of the CNAME in the SAN list of the website indicates a private CDN in this case. We finally label the remaining websites by comparing the DNS SOA records of both the website and the CNAME; a mismatch here indicates a third-party CDN. For instance, the website reddit.com also has resources fetched from the CDN Fastly. Since the 2LD matching and the SAN list check do not indicate a private CDN, we finally look at the SOA information. In this case, the SOA of the CNAME of the CDN is *.fastly.net, and the SOA of the website is *awsdns.net; the mismatch indicates a third-party CDN. For CDNs that have multiple CNAMEs, we iterate over all CNAMEs and if any of the CNAME is identified as private, we label the CDN as private. For instance, the website facebook.com uses the CDN Facebook (CNAMEs: ”*.fbcdn.net”, “*.facebook.com” and “*.cdninstagram.com”). Our heuristic classifies the first two CNAMEs as private so we label the CDN in this case as private. Then, for an unclassified CDN, if any of the CNAME is identified as third-party, we label the CDN as third-party. We manually sampled websites and verified the cases where websites have multiple CNAMEs and find that this heuristic correctly labels all the CDNs. So to classify whether each CDN used by the internal resources of the website is a private or third-party service, we follow the same three-step heuristics of Algorithm 2 using the CNAMEs of the internal resources as the *service.url*.

CDN Redundancy. As with DNS, we measure websites that are (*i*) redundantly provisioned by CDNs (host content from more than one private and/or third-party CDNs), (*ii*) critically dependent on a third-party CDN (host content from that one CDN), (*iii*) use multiple third-party CDNs and (*iv*) use both private and third-party CDNs. We measure CDN redundancy by finding the unique set of CDNs that the CNAMEs of a website map to using our self-populated CNAME-CDN map. For instance, the website zoom.us is redundantly provisioned by CDNs as it uses the CDN Cloudfront (CNAME: *.cloudfront.net), Google (CNAME: googlehosted.com) and Cloudflare (CNAME: *.cloudflare.com).

4.1.2.3. CA Measurements. For each website that supports HTTPS, we want to find its CA and also identify whether the CA is a third-party (e.g. DigiCert used by yahoo.com) or private CA (e.g. Google Trust Services used by google domains or Microsoft Corporation used by microsoft domains). In addition, we want to know if the website has enabled Online Certificate Status Protocol (OCSP) stapling. This means that before accessing a site, clients do not need to explicitly contact the CA, which manages the Certificate Revocation List (CRL) distribution Points (CDP) and OCSP servers, to verify the validity of the certificate. With OCSP enabled, the certificate's revocation status comes included with the TLS/SSL handshake. This reduces the criticality of the third-party dependency on the CA which means an outage of OCSP responders and CDPs does not translate into the website becoming inaccessible. To this end, we first make a request using OpenSSL to find a website's listed CA. If the request, which is to port 443, fails, then we assume the website is HTTP-only. At this stage, if the request succeeds, we also check if it has enabled OCSP stapling through information in the request response. The second condition of Algorithm 1 summarizes our heuristic when the service type is CA.

Next, we find the CA's url from the name of the CA. To classify the CA's url as third-party or not, we make use of the same three step heuristics described in Algorithm 2 in order to prevent misclassification by using a single approach. If the 2LD of the website and the CA's url match, then we classify the CA as private. If there is a mismatch, but if the 2LD of CA's url is in the SAN list for the website, then we also classify the CA as private. Finally, if neither of the previous two conditions are met, we check if the DNS SOA record for the CA and the website match. If they do not match, then we classify the CA as third-party. If a website does not fit the previous conditions, then we classify the CA as unknown.

4.1.2.4. Third-party Service Centralization. We are particularly interested in the degree of service centralization in markets around the world. The hypothesis is that third-party dependencies and centralization are positively correlated (i.e., high degrees of centralizations in markets with a high level of third-party dependencies) as consolidation of third-party services leads to centralization. However, different markets could be centralized around different or the same set of key service providers.

To measure the degree of centralization across each service, we find the number of third-party websites served by the top-1, top-3 and top-5 providers of each service across the countries and the websites critically dependent on these top providers. We analyze market trends across infrastructures and countries in Sec. 9.3.3.

Pseudocode 1 ThirdPartyDependence(w)

```
service ::= DNS | CDN | CA
if service = DNS then
    NS ← digNameservers(w)
    for ns ∈ NS do
        nstype ← FindserviceType(w, ns)
        if nstype = unknown ∧ concentration(ns) > 50 then
            nstype ← third
        end if
    end for
end if
if service = CA then
    CA ← findCertificate(w)
    CAURL ← findCAURL(w, CA)
    catype ← FindserviceType(w, CAURL)
end if
if service = CDN then
    IR ← findInternalResources(w)
    cnamesIR ← digCnames(IR)
    CDNs ← findCDN(cnamesIR)
    for cdn ∈ CDNs do
        cnames ← findCnames(w, cdn)
        for cname ∈ cnames do
            cnametype ← FindserviceType(w, cname)
        end for
    end for
end if
```

Pseudocode 2 FindserviceType(w, service.url)

```
service ::= DNS | CDN | CA
service.type ← unknown
if 2ld(w) = 2ld(service.url) then
    service.type ← private
else if isHTTPS(w) ∧ 2ld(service.url) ∈ SANList then
    service.type ← private
else if SOAProvider(w) ≠ SOAProvider(service.url) then
    service.type ← third
end if return service.type
```

4.1.3. Dataset

We collected the final set of regional websites in April 2021 for the 50 countries. The set includes a total of 25,000 websites with 15,774 unique sites. The average number of unique sites across these countries is 280 and China has the most unique set of sites (448 out of the top-500) and Singapore has the least (160 out of top-500). We run the study in April 2021 and again in 2022. In the 2022 snapshot, we find that from 15,774 total unique websites, 11,138 use CDNs and 9,766 use HTTPS. For our CDN analysis, we find a total of 1,339,871 unique resources and use 877,337 unique internal resources. Across countries, we find 68 unique third-party CAs, 60 unique third-party CDNs and 740 unique third-party DNS Providers. Note that, each year for every country, we select three probes and for each probe, we run the measurements thrice. However, in a given year the set of nameservers, CAs and CDNs identified and their classification for each country was the same in all three runs.

4.2. Dependency and Centralization

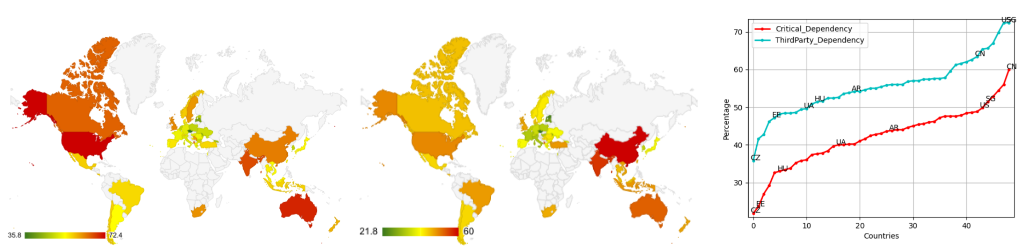
In the following paragraphs, we use the dataset collected in 2022 to study the degree of third-party dependencies, critical dependency, and market centralization around the world. Our analysis looks to answer the following questions:

- How common is the third-party dependency of websites around the world?
- How much of this dependency is critical, dependent on a single third-party DNS or CDN provider?
- How concentrated is the market of third-party service providers within a country, region, and globally?

We first look at DNS third-party dependencies, including critical dependencies, around countries and regions.

4.2.1. DNS Findings

Figure 4.3(a) and 4.3(b) plot the map of DNS third-party and critical dependency in each country, with red-colored countries having the highest dependence, yellow-colored countries having moderate dependence and green-colored countries having least dependence. Figure 4.3(c) plots a line graph to show the variation in the degree of DNS third-party dependency and critical dependency across countries. We find an average third-party dependency of 55.4% and most noticeable a wide range of dependency, from as low as 35.8% for the Czech Republic, to as high as 72.4% for Singapore. While critical dependency is lower than third-party dependency, with an average of 42.0%, the spread is similar ranging from the Czech Republic's lowest critical dependency of ≈21.8% to the critical dependency of China close to 60.0%. While the US and Singapore have the highest third-party dependency, China has the highest critical dependency. Generally, countries that have higher third-party dependency also tend to have a higher critical dependency on a single third-party DNS provider.



(a) Third-Party Dependency (b) Critical Dependency (c) Third-Party and Critical Dependency

Figure 4.3. DNS third-party dependency and critical dependency of each country's top-500 sites.

We also characterize the fraction of websites that use multiple DNS providers (redundant), that use multiple third-party DNS providers, and that use both third-party and private DNS providers. Figure 4.4 plots a map indicating the degree of DNS redundancy in each country, with red-colored countries having the least redundancy and green-colored countries having the highest redundancy. Figure 4.5 plots a line graph to further show the variation in overall redundancy, multiple third-party providers and third-party and private DNS providers across the different countries. We see that, on average, 14% of regional sites have redundant DNS, with Estonia having a maximum redundancy of 24% and China having minimum redundancy of 4.0%. We find that 3.8% of sites, on average, have multiple third-party DNS providers with the US having a maximum of 13.8% and China having a minimum of 0.8%. On average, 3.2% sites use third-party and private DNS services (maximum of 7.2% for France and minimum of 0.4% for China).

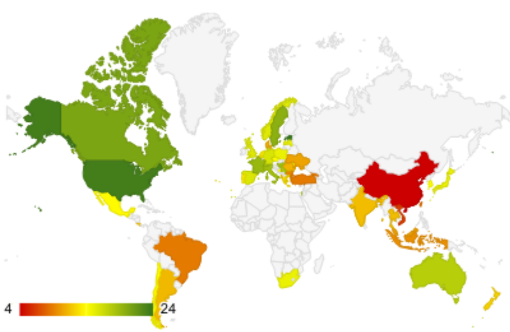


Figure 4.4. Plot of DNS redundancy

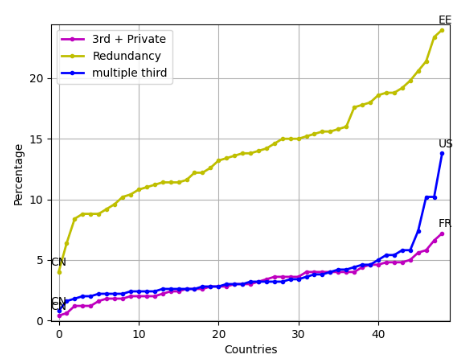


Figure 4.5. Line plot of redundant, multiple third-party, and both third-party and private DNS.

We then identify the level of dependency of websites on the five most popular DNS providers across countries. Table 4.2 shows the average number of websites relying on each DNS provider. We see that Cloudflare and Amazon DNS alone are used by 73% of the websites that use a third-party DNS, on average, across the countries.

We additionally find that only three DNS providers are used by an average of 77.5% of the websites that use third-party DNS across countries. Taiwan has the highest degree of DNS Centralization with the top three DNS providers being used by 88.7% of the country's websites that use third-party DNS providers and the Czech Republic has the least DNS centralization (60.9%). Table 4.3 shows the top-3 popular DNS providers across all regions of our vantage points and the average number of websites dependent on them. We see that in Europe, the Asia Pacific and Africa and the Middle East, Cloudflare is the most popular third-party DNS provider, whereas, in the Americas, Amazon Route 53 is the most popular provider.

Summary. We observe that more than half of the countries have more than 55% of their regional sites dependent on a third-party DNS provider and more than 43% of their regional sites critically dependent on a third-party DNS operator. Across the countries, most websites have lower redundancy in their use of different DNS providers. When comparing third-party DNS dependency across regions, we learn that Eastern Europe has the lowest third-party DNS dependence, whereas North America and some parts of Asia Pacific have a high third-party DNS dependence. The top-3 providers across all regions are the same (highly centralized) except in Eastern and Southern Europe where NsOne is among the top-3 instead of Akamai as in other regions. Additionally, the top-3 providers

DNS Provider	Avg	Std Dev
CloudFlare	**43.5**	**14.2**
Amazon Route 53	**29.3**	**9.6**
NsOne	8.3	3.0
Akamai	7.6	5.0
UltraDNS	4.3	2.1

CDN	Avg	Std Dev
Google	**70.0**	**26.3**
Akamai	26.9	9.7
Fastly	18.7	7.7
Cloudflare	16.6	5.9
Amazon Cloudfront	15.7	8.7

CA Provider	Avg	Std Dev
Digi Cert	**36.3**	**7.3**
Comodo CA Limited	15.2	4.6
IdenTrust Inc.	14.8	7.3
GlobalSign	14.0	4.4
Starfield Technologies, Inc.	6.4	3.3

Table 4.2. Top-5 DNS, CDN and CA providers across countries (average percentage of websites).

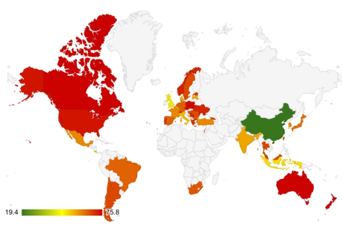
are responsible for 70% or more websites in each region. We also see that Cloudflare and Amazon DNS alone are used by 73% of the websites, on average, across the countries.

Region	DNS Providers	Avg	Std Dev	CA Providers	Avg	Std Dev	CDN Providers	Avg	Std Dev
The Americas	Amazon Cloudflare Akamai	79.1	4.2	DigiCert GlobalSign Comodo	69.1	2.7	Google, Akamai, Amazon Cloudfront	90.8	9.0
Europe	Cloudflare Amazon Akamai	74.6	6.5	DigiCert IdenTrust Inc. Comodo	70.6	5.3	Google, Akamai, Fastly	92.2	6.7
Asia Pacific	Cloudflare Amazon Akamai	81.6	4.5	DigiCert GlobalSign Comodo	72.8	7.4	Google, Akamai, Fastly	91.7	8.2
Africa and Middle East	Cloudflare Amazon Akamai	80.4	5.5	DigiCert GlobalSign Comodo	70.8	0.8	Google, Akamai, Fastly	86.4	14.9

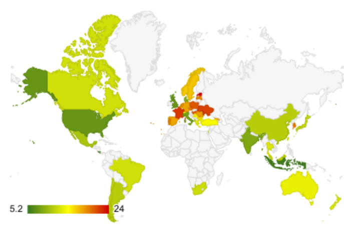
Table 4.3. Top three DNS, CA and CDN providers with their corresponding market share per region (average percentage of websites).

4.2.2. CDN Findings

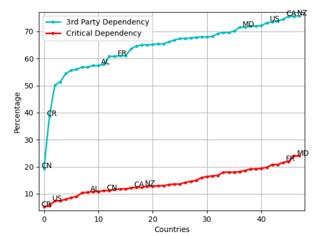
Figure 4.6(a) and 4.6(b) plot the map of CDN third-party and critical dependency in each country, again with red colored countries having the highest dependence and green colored countries having least dependence. Figure 4.6(c) plots a line graph to show the variation in the degree of CDN third-party dependency and critical dependency across countries. We find an average third-party CDN dependency across all countries of 64.1%. The country with the lowest dependency of 19.4% is China while the country with the highest dependency of 75.8% is New Zealand. Next, we aim to see the criticality of these CDNs. Critical dependency means when a website is solely hosted on one CDN. Figure 4.6 shows that the country with a maximum critical dependency on a third-party CDN is Moldova with a value of 24.0% and the country with a minimum value of critical dependency is Costa Rica with a value of 5.2%.



(a) Third-Party Dependency



(b) Critical Dependency



(c) Third-Party and Critical Dependency

Figure 4.6. CDN Third-party dependency and critical dependency of each country's top-500 sites.

We also show the similarity between the CDN usage trends - the percentage of websites using more than one CDN, the percentage of websites using only third-party CDNs, and the percentage of websites using both private and third-party CDNs. Figure 4.7 plots a map indicating the degree of CDN redundancy in each country, with red-colored countries having the least redundancy and green-colored countries having the highest redundancy. Figure 4.8 plots a line graph to show the variation in overall redundancy, on multiple third-party providers and on third-party and private CDN providers across the different countries. On average 51.2% of Alexa regional sites were redundantly provisioned with CDNs, with the US having the maximum redundancy of 67.2% and China having minimum redundancy of 8.2%. 39.7% sites on average use multiple third-party CDN providers with Canada having a maximum of 59.6% and China having a minimum of 5.0%. 6.6% sites on average use third-party and private DNS with Israel having a maximum value of 43.0% and China having a minimum value of 0.0%.

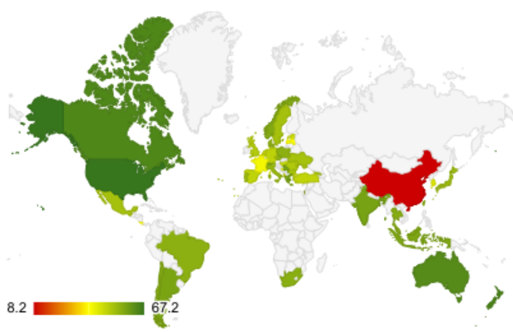


Figure 4.7. Plot of CDN redundancy

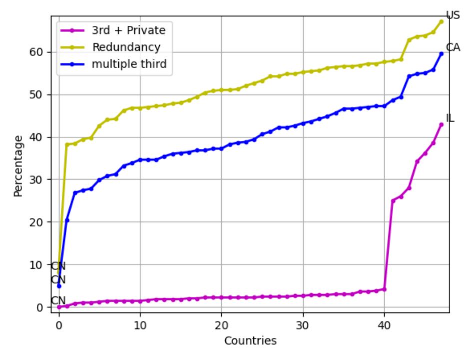


Figure 4.8. Line plot of redundant, multiple third-party, and both third-party and private CDN.

We observe that more than half of the countries have a third-party CDN dependency greater than 68% and a critical CDN dependency higher than 14%. Interestingly, for some countries such as the US, Canada, and New Zealand we see high third-party dependency but low to moderate critical dependencies. Whereas, countries like France show a higher critical dependency but a lower third-party dependency. Overall, we notice lower critical dependency on the CDN infrastructure and higher redundancy since many sites today use content from multiple CDNs.

Table 4.2 shows the top-5 most popular CDNs across the countries we used for measurement, the average percentage of websites hosted on them along with the standard deviation. This shows that of the top-500 Alexa websites that use third-party CDN, 70% sites use Google as their CDN, 27% use Akamai and 19% use Fastly.

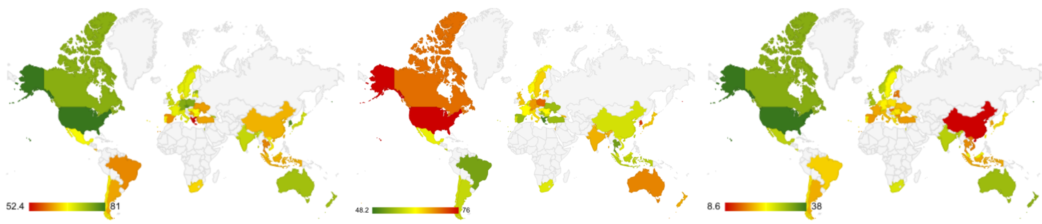
We find that only three CDNs are used to host an average of 91.5% of the websites that use third-party CDN across countries. Albania has the most CDN centralization with the top-3 CDN providers hosting 98.3% of the country's websites that are served by a third-party CDN and Denmark has the least centralization with 68.6% of the country's websites served by top-3 third-party CDNs. These results demonstrate a high degree of centralization of the CDN service. Table 4.3 shows the top-3 CDN providers that are popular across different regions of our vantage points and the average percentage of websites using them. The results show that Google and Akamai are the top-2 CDN providers across all regions. On average, more than 86% of websites that use a third-party CDN use the top-3 CDN providers across all regions showing a high degree of centralization.

Summary. We observe that more than half of the countries have more than 68% of their regional sites dependent on a third-party CDN provider and more than 14% critically dependent on a third-party CDN. We see more redundancy (and therefore lower critical dependency) on CDN providers compared to DNS providers as more than half of the countries have greater than 53% sites using multiple CDN providers. We see a higher third-party dependency in the Americas and most of Europe and the Pacific regions and the lowest third-party CDN dependency in China. Across all regions, more than 86% of websites are dependent on top-3 CDNs. Google and Akamai are among the top-3 CDNs across all the regions with Google significantly dominating the market (average of 70% of the websites).

4.2.3. CA Findings

In the case of CA, Fig. 4.9 plots the map of HTTPS Support, CA third-party dependency and OCSP Stapling support in each country. For the CA third-party dependency, red-colored countries have the highest value and green-colored countries have the lowest value and vice versa for HTTPS Support and OCSP Stapling.

We find that the average percentage of sites using HTTPS across countries was 67.4%. This average is dragged down by countries in Latin America, and a few countries in Europe and Asia with Greece having the lowest number of websites using HTTPS (52.4%). The US has the highest rate of HTTPS adoption, with 81.0% of the top-500 sites using HTTPS. In terms of average third-party CA dependency across all countries, 61.6% percent of sites within our dataset are using a third-party CA. The results ranged from Albania at the bottom with 48.2% of its top-500 sites, and the US at the top with 76.0% of its top-500 sites using a third-party CA. OCSP stapling is much less popular. On average, 22.3% of countries' top-500 sites use OCSP stapling with China having the lowest usage at only 8.6% and the US having the highest usage at 38.0%. The low popularity of OCSP stapling is perhaps because of the lower OCSP support across web servers and browsers. For instance, the browser with the highest market share, Google Chrome, does not support OCSP stapling [**49**, **198**]. Additionally, since OCSP servers are unreliable so practically all clients implement OCSP in soft fail mode. OCSP must-staple addresses this, however, it is yet to gain widespread adoption [49]. None of the websites in our set support OCSP must-staple.



(a) HTTPS Support (b) Third-Party Dependency (c) OCSP Stapling

Figure 4.9. Percentage of each country's sites using HTTPS, third-party CA and OCSP stapling.

Additionally, we find DigiCert is the most popular CA in all countries' top-500 Alexa sites except for Estonia, Latvia, and Moldova. On average, 36.3% of websites in our dataset that used a third-party CA used DigiCert (including Baltimore CyberTrust certificates, which were purchased by DigiCert). The other popular CAs are Comodo CA Limited with 15.2% popularity, IdenTrust Inc. with 14.8% popularity, and GlobalSign with 14.0% popularity. Table 4.2 shows the average percentage of websites relying on each CA provider across the different countries.

Table 4.3 shows the top-3 CA providers across all regions of our vantage points and the average percentage of websites using them. We see that DigiCert and Comodo are in the top-3 CA providers across all regions. On the other hand, GlobalSign is one of the top-3 providers in all regions but Europe, whereas IdenTrust Inc. is one of the top-3 in Europe but not in other regions. Countries, such as the Czech Republic, China, and Serbia tend to be the most centralized around popular CAs, with more than 80% of websites using third-party CA choosing one from the top-3 CAs in their country. Other countries like

Taiwan and Switzerland show less centralization: for these countries, less than 60% of websites use third-party CA from the top-3.

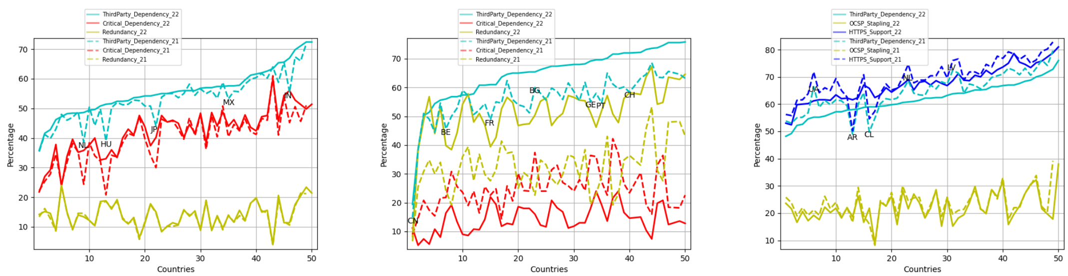
Finally, in total, we identified 68 unique CAs, 15% higher than the 59 CAs reported in Kashaf et al. for the top-100K sites. We hypothesize this is because we have a better representation of websites from different countries and thus a better representation of country-specific CAs. Some examples of country-specific CAs that we find include TWCA, a Taiwanese CA, and Microsec Ltc., a Hungarian CA.

Summary. We observe that more than half of the countries have more than 62% of their regional sites dependent on a third-party CA provider. We see a higher third-party CA dependency in North America and the least dependency in South America, Eastern Europe and most of Asia. Across all regions, more than 65% of websites are dependent on top-3 CAs. DigiCert, GlobalSign, Comodo, and IdenTrust Inc. are among the top CAs across the regions with DigiCert being the most dominant CA, used by 36.3% websites on average, across all the countries except Estonia, Latvia, and Moldova.

4.2.4. Third-party dependency across services

Overall, we observed that some countries have higher third-party dependency across all of the three DNS, CDN and CA infrastructures. These countries, ranked in the order of their third-party dependency, include the United States, Australia, Canada, Singapore, New Zealand, Sweden, Norway, India and Japan(range:58%-76%). Whereas, in Europe, South America and many Asian countries (except for China) only the CDN infrastructure is responsible for higher third-party dependency. We note that China has considerably lower third-party CDN dependency(19.4%) and this may be caused by top CDN providers in our study having almost no deployments in China [**51, 85, 8, 3**].

4.3. Snapshot Measurements



(a) DNS (b) CDN (c) CA

Figure 4.10. Trends across services. Each vertical line corresponds to a country, ordered by 2022 values.

To gain an initial understanding of longitudinal trends on third-party dependency and centralization around the world, we carry out our measurement campaign in two consecutive years, during April 2021 and again in April 2022. Figure 4.10 shows these trends for the different services in our analysis – DNS, CDN, and CA. For each service plot, the order of countries corresponds to the countries ordered by the third-party dependency of that service in 2022.

DNS. Figure 4.10(a) plots the percentage of websites, for the different countries, and each country's corresponding DNS third-party dependency, critical dependency, and redundancy. We find that on average across all countries, DNS third-party dependency increased by 4% and the critical dependency on a third-party DNS provider increased by 5%, in just one year. For both years, Cloudflare and Amazon remained the most dominant

DNS third-party providers across the countries, with on average 73% websites that use a third-party provider dependent on them.

CDN. Figure 4.10(b) plots the percentage of websites, for the different countries, and each country's corresponding CDN third-party dependency, critical dependency, and redundancy in 2021 and 2022. We observe that on average across all countries, the third-party dependency on CDNs increased by 15% and the redundancy increased by 60% in one year. Average critical dependency on a third-party CDN provider, across all countries, however, decreased by 44% in one year. Google significantly dominated the CDN market across both years (an average of 70% of the websites). Our findings show that more websites are increasingly using multiple CDNs. Though it should be noted that we do not look at criticality in terms of the content served by the CDN.

CA. Figure 4.10(c) similarly shows, for each country, the corresponding change in the country's websites having third-party CA dependency, HTTPS Support, and OCSP stapling over the two-year period. We see that across all countries, on average third-party CA dependency decreased by 4.8% and HTTPS support decreased by 2.9%. The percentage of websites supporting OCSP stapling stayed the same for most countries. The plot shows that the drop in third-party dependency corresponds with the drop in HTTPS support. Upon manually investigating sites that no longer support HTTPS, we found that all these sites return an SSL failure perhaps due to an expired certificate. However, DigiCert remained the most popular CA across the countries (serving an average of 36% websites).

4.4. Discussion

To the best of our knowledge, this is the first large-scale study of global trends towards third-party dependency and centralization across services. Our study, however, is subject to a number of limitations. First, Scheitel et al. [**184**] raises concern about the ranking opacity and stability of top lists, including the Alexa regional ranking list we rely on for our analysis. While the Tranco [**166**] list addresses some of these issues, Alexa's regional ranking offered the largest regional ranking available to the community, and includes in its listing top regional domain aliases (e.g., yahoo.com.jp and yahoo.com). Tranco allows filter-based selection of regional rankings, however, that list contains an intersection of Tranco's global ranking and domains that also appear in the country-specific Chrome User Experience Report list. This introduces biases in Tranco's regional rankings as Chrome's user Experience list represents only a fraction of users who use Chrome as their browser, have opted in to sync their browsing history, have not set up a Sync passphrase, and have usage statistics enabled [**84**]. Between 1% to 4.5% of websites appearing in Alexa regional rankings have aliases in different countries (e.g., google.com and yahoo.com use google.co.rs and yahoo.co.rs, respectively, in Serbia). These domain aliases can have different dependencies even if they belong to the same entity; for instance, rakuten.com uses Akamai as its DNS provider and rakuten.jp uses a private DNS. Alexa's regional rankings include these domain aliases, while Tranco's regional ranking is derived from its global rankings that may not include these less globally popular, regional versions. We follow recommended best practices [**184**], as Alexa's regional ranking is regarded as the best match for our human-centered study of global third-party dependencies, and we make available the downloaded list (including download date) to enable basic replicability. We are currently exploring alternative sources of regional rankings.

Second, we characterize centralization and third-party dependencies using the list of top-500 regional websites. While we acknowledge this is a relatively limited view of the most popular sites in a region or country, it is however the largest available list.

Third, the study focuses on server-side centralization only thus the implications of centralization and third-party dependence on the end-user side are not the focus of the study. This focus allows us to rely on the use of VPN nodes as vantage points as a proxy to study the extent of third-party dependency the users of different countries are exposed to when visiting the popular sites of the country. This same method, however, precludes the analysis of services assignments (e.g., CDN replicas) that may depend on users' DNS resolvers.

Fourth, we do not measure physical and network infrastructure dependencies; e.g., physical hosting (content providers, IT operators, common landing point in the submarine network, etc.), routing, or third-party dependencies on the web content.

Last, while we limit our analysis of dependency and criticality to 50 out of close to 200 world countries, the set of countries was selected to ensure it covers the range in terms of overlap between the top-500 regional sites and the global-ranked lists, has vantage points highly likely located within the claimed country, and, together, captures over three-fourths of the Internet user population. We are exploring alternative approaches to expand our analysis, particularly of under-sampled locations such as Africa and the Middle East.

4.5. Conclusion

We presented the first large-scale study of third-party dependency and concentration around the world for commercial websites. Using vantage points in 50 countries across all inhabited continents, and focusing on the most popular regional websites, we find a wide range of third-party service dependencies across countries. Despite this heterogeneity in dependence, the provider market remains highly concentrated, and in some cases concentration increases even over a single year. Taken together, these results establish two key premises for the dissertation's central claim. First, consolidation is empirically real, widespread, and measurable at global scale using public signals. Second, measurement alone does not explain consolidation: the same observed concentration can reflect different underlying constraints and choices. Chapter 6 builds on the country-level evidence developed here (and extended to the public sector in Chapter 5) to introduce the comparative baseline that distinguishes structural from strategic consolidation, enabling the implication analyses in Chapters 7–9.

CHAPTER 5

A Comparative Analysis of Government Hosting

For governments, the growing reliance on third-party infrastructure presents a particularly challenging dilemma. While third-party providers offer specialized content delivery solutions with clear operational benefits, they also introduce risks, including reduced control over data placement [**47**], multi-tenancy [**5**, **57**, **112**], and increased concentration of critical services [**15**, **99**, **101**, **131**, **231**, **193**, **17**]. These risks are especially salient for governments because consolidation is not only an economic outcome but also a sovereignty signal: it can amplify foreign dependence and create exposure to foreign jurisdictions, platforms, and infrastructure [**72**, **169**, **197**, **63**, **70**]. At the same time, the presence of concentration alone does not reveal whether it is an unavoidable consequence of the local ecosystem or a discretionary strategy. This chapter provides the empirical foundation needed to pose that interpretability question.

Although early discussions on cyber sovereignty – from data sovereignty to digital privacy and security and Internet governance – followed the 2013 Snowden revelations of widespread surveillance [**135**], various initiatives around the world have since focused on this issue in the context of geopolitical and economic tensions and the growing recognition of the Internet as critical infrastructure [**32**, **80**, **156**, **185**, **188**]. These concerns have motivated the development of legal frameworks, including the European Union's General Data Protection Regulation (GDPR) [**54**], California's Consumer Privacy Act

(CCPA) [**155**], and Brazil's General Data Protection Law (LGPD) [**125**]. Together, these frameworks reflect a concerted effort to protect and manage data within the respective jurisdictions, highlighting the increasing importance of data sovereignty.

At the same time some cloud providers have began offering solutions tailored to specific governments. For example, Amazon Web Services [**78**] and Microsoft Azure [**79**] have developed solutions tailored to meet the requirements of the U.S. government. Nevertheless, for the majority of countries, third-party services are foreign-based, forcing them to strike a balance between external expertise and maintaining sovereign control over their digital assets.

Our work aims ***to empirically characterize the various ways in which governments navigate and resolve this emerging dilemma***, and to provide the measurement foundation needed for the structural versus strategic consolidation baseline introduced in Chapter 6.

Understanding this is crucial because digital transformation has fundamentally altered how governments communicate, creating new channels for disseminating policies and information while giving citizens direct access to essential services [**23, 212**]. The importance of digital government is evident in cases like federal websites in the US, which attract nearly two billion visits monthly and result in approximately 80 million hours of public interaction [**96**], and in the Asia-Pacific region where 77% of citizens primarily use a digital platform to access government services [**69**]. This transformation underscores the need for understanding the infrastructure behind public-facing government websites.

We present the first comprehensive study of hosting models employed by public-facing government digital services. Together with the commercial-sector measurements in Chapter 4, this chapter highlights why observed concentration is not self-interpreting: similar levels of consolidation can reflect different underlying constraints and choices, motivating the comparative baseline introduced in Chapter 6.

Our analysis draws on data from 61 countries, covering every continent and representing over 82% of the world's Internet population. We identify government-related sites within these countries, collect resources from the landing pages of government websites and recursively crawl internal pages up to seven levels deep [**190**]. Our dataset comprises over 1 million unique resources, providing a broad and detailed snapshot of government digital service hosting. Building on this dataset, we conduct an extensive measurement study to analyze government hosting strategies, cross-border dependencies, and the level of centralization in government web services.

Our contributions are threefold. First, we describe a methodology to characterize government approaches to domain hosting by identifying their service infrastructure and geographic location. Second, we apply this methodology to build a comprehensive dataset of government URLs and annotated networks spanning 61 countries. Finally, we present the first extensive measurement study that investigates government hosting strategies, and the degree of centralization in government web services.

Our analysis reveals several key findings regarding governments' reliance on third-party infrastructure for data delivery. Governments predominantly use third-party providers to deliver 62% of URLs and 53% of bytes, though the adoption of these providers varies significantly across and within regions. For example, in North America, 68% of government bytes are delivered via third-party providers, while in South Asia, this reliance drops to just 5%. Neighboring countries also show contrasting patterns: Argentina relies on third-party providers for 90% of its government data, whereas Uruguay's reliance is only 2%.

Furthermore, consolidation on third-party providers in government services appears more pronounced than in other sectors [**115**, **120**], with Cloudflare serving 49 governments—nearly double the number of the next two providers, Microsoft and Amazon. Diversification also seems correlated with reliance on third-party providers: 63% of countries that primarily use government infrastructure serve the majority of their content from a single network, compared to just 32% of countries that rely mainly on global providers.

5.1. Methodology

To characterize governments' approaches to domain hosting, we (*i*) collect government sites, and (*ii*) identify the resources they rely on, excluding those of external contractors. We then determine (*iii*) the serving infrastructure of those resources and (*iv*) their location. The following paragraphs describe this process in detail.

5.1.1. Gathering Government Websites

The first step in our methodology is to compile a comprehensive list of government sites. In this study, we focus specifically on federal-level (or equivalent) resources, including various segments of the federal administration (e.g., the presidency, ministries, and secretaries), federal agencies, often referred to as decentralized agencies (e.g., the US National Science Foundation and the US Internal Revenue Service) and state-owned enterprises. To consider State-Owned Enterprises (SOEs), we follow the International Monetary Fund (IMF) guidelines and only include companies where the federal government holds more than 50% of the shares [**81**].

This step requires searching through a country's government sources that may provide insights into the organizational structure, identifying digital directories and authoritative resources that provide details on these structures and links to corresponding government sites. As this information is typically in the country's official languages, we rely on translation tools for this part of the process.

5.1.2. Scraping Government Websites

We scrape the collected government websites to identify the resources they rely on. For this, we use Selenium [**187**], a web automation tool, to capture the URL of each resource that constitutes the queried websites, which are then consolidated into an HTTP Archive (HAR) file. We move beyond the landing pages using the collected HAR files to recursively navigate internal pages up to seven levels deep, a threshold informed by previous work [**190**].

The geographic location of our vantage points can impact website rendering, replica selection, or determine resource accessibility, with some sites restricting access to non-domestic devices.[1] To avoid these and other potential problems, we rely on different VPN services including NordVPN [**152**], Surfshark [**203**], Hotspot Shield [**94**], to access these sites from within the target country.

5.1.3. Internal Government URLs

As we scrape seven levels deep into a government domain, we run the risk of leaving the government domain (e.g, into an external contractor's site). After completing data collection we identify internal government URLs and filter out non-government ones following the steps summarized in Table 5.1.

We first label as government resources those with domains under government top-level domains (TLDs). We adopt the pattern-matching rules defined by Singanamalla et al. [**190**], which account for the different government TLDs that vary based on each country's definitions and official languages. This includes TLDs such as `.gov`, `.gouv`, `.gob`, and `.go`, among others, as listed in Table 5.1.

We then identify government resources that do not fall under government top-level domains (TLDs), either because the country does not utilize government TLDs or chooses not to use them for some agencies or state-owned enterprises (§5.1.1). If the hostname of an internal page matches the hostname of any of the sites comprised in our list of government websites, we classify it as a government hostname.

[1]For instance, Mexico's Taxpayers Defense Attorney (in Spanish Procuraduría de la Defensa del Contribuyente, `www.prodecon.gob.mx`).

Approach	Method
Government TLDs	All domains including `.gov`, `.govern`, `.government`, `.govt`, `.mil`, `.fed`, `.admin`, `.gouv`, `.gob`, `.go`, `.gub`, `.guv`
Domain Matching	If the hostname of the internal page aligns with those listed in the government websites section (§5.1.1).
SAN	If the hostname is included under domains specified as Subject Alternative Names (SANs) in the TLS certificates of landing pages

Table 5.1. Steps of the methodology to identify government domains.

Finally, we identify government resources included under domains specified as Subject Alternative Names (SANs) in the TLS certificates of landing pages [35]. When the hostname of an internal page appears in the SANs list of landing pages, we manually verify that the hostname corresponds to a government resource. This last step allows us to select additional government-affiliated resources that may not be directly evident through their domain names or top-level domains (e.g., `orniss.ro`, `energia-argentina.com.ar`). At this stage, any hostnames that cannot be verified as government hostnames are discarded from our analysis.

5.1.4. Identifying the Serving Infrastructure

We identify the serving infrastructure utilized by government hostnames. This involves determining the IP address, Autonomous System (AS) number, organization, the registered location and the geolocation of the serving infrastructure. Table 5.2 shows an example of the information we collect for a government hostname in Uruguay.

Field	Value
URL	`www.gub.uy`
IP address	179.27.169.201
ASN	6057
Organization	Administracion Nac. de Telecom.
Registration	Uruguay
Geolocation	Uruguay

Table 5.2. An example of the information of serving infrastructure that is collected for each government resource.

To obtain registration and topological data on government website infrastructure, we connect to a VPN within the country and resolve all government hostnames to their IP addresses. Once we have the IP addresses, we determine the corresponding AS number, organization and country of registration using public WHOIS services managed by organizations responsible for IP address registration.

We then determine whether content is served from on-premise infrastructure within government-operated networks. While a recent study has made progress in identifying state-owned Internet providers [42], there is no dataset with annotations of government networks. We thus manually examine the entity behind all identified ASes to determine government ownership. It is important to differentiate between state-owned Internet providers – government-controlled companies participating in the Internet market – and government networks used exclusively by government institutions.

We combine various data sources to identify government ownership of networks. We examine PeeringDB records, searching for any indicator of government ownership, which may be revealed in the network's name, associated organization, or note, as in the entry of AS26810 indicating the organization as "U.S. Dept. of Health and Human Services". We also leverage the website reported on PeeringDB records and investigate whether the associated website reveals any information that could indicate a connection with the government. Given the limitations of PeeringDB's coverage, we use WHOIS records to complement our classification. This involves querying WHOIS databases to check if the organization's name refers to the government (e.g., ministry) or the domain of the contact person's email is linked to a government domain (e.g., ".gov"). Finally, for cases where we are unable to find direct matches, we resort to Google searches. We utilize domain information extracted from WHOIS records to search for these companies' websites. This process also allows us to identify domains of state-owned enterprises that may not always be identifiable as government domains (e.g., AS27655 - Yacimientos Petrolíferos Fiscales).

5.1.5. Server Geolocation

The last step of our process consists of determining the geographic location of the infrastructure serving government websites. Given the limitations of existing geolocation heuristics and databases, we outline our specific methodology to address these challenges.

Step #1: Geolocation databases. We first query IPInfo [**110**], a widely-used open geolocation database, with the addresses of all the collected government hostnames. Darwich et al. [**60**] report that 89% of the geolocation targets in IPInfo have an error of less than 40 km (i.e., within a city).

Step #2: Identifying Anycast addresses. IP Anycast challenges latency-based geolocation. To determine if a server address is anycast, we rely on a recent data snapshot from MAnycast2, generated based on the idea of using anycast IPs as VPs to launch active measurements to candidate anycast destinations [**195**].

Step #3: Verifying country-level geolocation. To enhance the accuracy of our geolocation data, we deploy active-probing measurements to validate the reported geolocations.

For anycast addresses, we select five RIPE Atlas probes situated in the vantage country to send three pings to anycast addresses and calculate the minimum latency to each address. Our methodology integrates active measurements with the country's road infrastructure data to derive a threshold that determines whether a server address is located within a country. When the latency to a specific server address is less than the threshold, we conclude that the anycast address has servers within the country. Anycast addresses with latencies higher than this threshold are excluded from the analysis.

For unicast addresses, we also use five RIPE Atlas probes in the country assigned by IPInfo to send pings to each reported address in that country. To confirm the server's location reported by IPInfo, we calculate the minimum latency to each address and, following [**11**], check if the latency to a specific server address is less than this threshold calculated using the country's road infrastructure data. Discrepancies trigger additional verifications for unicast addresses, explained in Step #4.

Given the different shapes and sizes of countries, rather than settling for a single global threshold, we determine a per-country threshold based on the intercity road distance between the two furthest cities in that country and convert this distance into latency values.

Step #4: Geolocating Unicast Addresses. To verify the location of remaining unicast addresses we use CAIDA's HOIHO methodology [**133**], which leverages geolocation hints found in PTR DNS records, with additional regular expressions (e.g., for NTT). We also consult the cached results from RIPE's IPmap [**179**] and, if not available, we resort to active probing following a single-radius approach for geolocation.

5.2. Government Hosting Dataset

To capture a global view of trends in government hosting, we select a sample of 61 countries across all world regions, and apply our methodology for identifying government approaches to domain hosting. We first describe our criteria for including a country in our sample before providing general statistics on the collected dataset.

5.2.1. A Sample of Countries

We create a representative dataset encompassing countries from all regions worldwide. Regional divisions allow us to identify global and regional trends for governments' digital approaches. We set criteria for sampling countries across these regions, balancing our scope with technical and logistic limitations (such as the absence of verifiable VPN servers[2] or insufficient information on e-governments).

World's Regional Slicing. To explore regional patterns in government digital strategies, we rely on the World Bank's regional division [**22**]. This division groups countries into seven regions: North America (NA), Latin America and the Caribbean (LAC), Europe and Central Asia (ECA), North Africa and the Middle East (MENA), Sub-Saharan Africa (SSA), South Asia (SA), and East Asia and Pacific (EAP).

Country Selection Criteria. Covering each region, we select countries that, combined, capture a wide range of key development indices, specifically: (1) the E-Government Development Index (EGDI) [**150**], (2) the Human Development Index (HDI) [**170**], and (3) the International Telecommunication Union/World Bank Internet Penetration rates [**200**]. This combination of indices allows us to capture a broad spectrum of countries in various stages of development and digital advancement. We integrate these indices at a regional level and select countries from five different quintiles.

[2]We gained confidence in the claimed VPN locations of the countries in our set, by validating the VPN vantage points' IPs using the geolocation approach described in (§5.1.5)

While aiming for uniform coverage across these quintiles, we encounter some limitations. Specifically, the challenge is set by the lack of commercial VPN services in countries from the lower quintile, particularly in regions like Sub-Saharan Africa and Latin America and the Caribbean.

Our final selection of 61 countries from across the globe includes 2 countries from North America, 8 from Latin America and the Caribbean, 29 from Europe and Central Asia, 5 from North Africa and the Middle East, 2 from Sub-Saharan Africa, 3 from South Asia, and 12 from East Asia and Pacific (EAP). These countries combined represent 82.70% of the global Internet population. To access government URLs across these countries, we use 3 VPN services: NordVPN (49), Surfshark VPN (10), and Hotspot Shield VPN (2).

5.2.2. Dataset Characteristics

We apply our methodology to the set of countries in our sample. Table 5.3 offers a high-level overview of the extent and scope of our data collection.

Category	Element	Value
Government Websites	Landing URLs	15,878
	Internal URLs	1,017,865
	Total Unique URLs	1,033,743
	Total Unique Hostnames	13,483
Serving Infrastructure	ASes	950
	Govt ASes	347
	Unique IP addresses	4,286
	Anycast addresses	433
	Countries with servers located	68

Table 5.3. Landing URLs, unique hostnames and unique URLs in our dataset.

Government Websites. The dataset includes 15,878 unique landing pages from governments of 61 countries, and 1,017,865 internal government URLs obtained through scraping across seven levels. In total, the dataset comprises 13,483[3] unique hostnames and 1,033,743 distinct URLs. The vast majority of URLs, 84%, were collected directly from the landing pages, with 95% obtained from one additional level below the landing page.

Internal Government URLs. We apply a set of heuristics (Table 5.1) to identify government URLs and filter out non-government ones from the set of URLs obtained. This step identified 285,767 (27.6%) internal government URLs using the government TLDs, 745,358 (72.1%) using the domain-matching approach and 2,618 (0.3%) using SANs.

Serving Infrastructure. We identified 950 ASes connecting to 4,286 server addresses associated with 13,483 hostnames. We discovered 347 (36.5%) of these ASes are operated by government entities.

We localize the serving infrastructure of the 4,286 addresses. MAnycast2 identified 433 (10.10%) of them as anycast addresses. Active-probing confirmed that 361 anycast addresses (83.37% of all anycast addresses identified) are within the country's borders. We excluded the remaining 72 anycast addresses from the analysis due to insufficient confidence in their location.

From the 3,853 unicast addresses, IPInfo identifies 3,349 addresses (86.92%) in the same country as the government they are serving and 504 unicast addresses (13.08%) outside the country borders. To increase our confidence, we tried to confirm IPInfo geolocation.

[3]Note that the number of unique hostnames is less than the number of unique landing pages. This is because landing pages can include URLs like https://www.gov.br/secretariageral/pt-br, https://www.gov.br/abin/pt-br, representing different pages with the same hostname.

Type of Address	AP	MG	UR
Unicast Addresses	0.41	0.57	0.02
Anycast Addresses	0.83	0.00	0.17

Table 5.4. Fraction of unicast and anycast addresses validated by Active Probing (AP) and Multistage Geolocation (MG), or Unresolved (UR).

Through active-probing, we confirmed the location of 40.77% (1,571) of these addresses. Through a multistage geolocation approach (§5.1.5) we confirmed the geolocation of an additional 2,198 addresses. In total, we confirmed 3,769 (97.8%) of all unicast addresses. We exclude 84 instances where the geolocation obtained at this stage conflicts with IP-Info. Table 5.4 summarizes the output of this validation process for unicast and anycast addresses spanning 68 countries.

5.3. Trends in Government Hosting

Building on the collected dataset, in this section, we explore global and regional trends in government domain hosting and compare them with trends among popular websites. We close the section examining the similarities in governments serving strategies across the countries in our study.

5.3.1. Global Trends

We first take a global perspective, exploring governments' preferences in choosing the serving infrastructure powering their websites. *Do governments prefer on-premises or third-party hosting?* For governments opting for third-party providers, we further explore their preferences towards global, regional, or local providers.

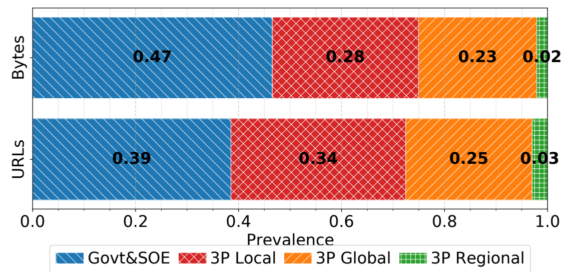


Figure 5.1. Global fraction of URLs and Bytes served by each provider category.

We examine the adoption of on-premises solutions, labeled *Government and State-Owned Enterprises* (Govt&SOE), versus third-party providers. We categorize third-party (3P) services into three groups: (1) Local (3P Local), (2) Regional (3P Regional), and (3) Global providers (3P Global), with 3P Global defined as networks that serve governments across multiple continents, and 3P Local as those registered in the same country as the government they serve. The remaining category, 3P Regional, includes networks registered outside the country they serve, but that do not span beyond one continent.

Using this classification, Figure 5.1 illustrates the global prevalence of each server URL category and quantifies the content by aggregating the total bytes of government URLs to account for variations in URL sizes.

Overall, governments show a preference for 3P infrastructure for data delivery, using them to deliver 62% of URLs and 53% of bytes, compared to only 39% of URLs and 47% of bytes hosted by Govt&SOE. When focusing on the categories of 3P, the figures show a more balanced reliance on Govt&SOE, 3P Global, and 3P Local although with a preference for Govt&SOE for bytes.

Interestingly, the analysis reveals that governments rarely consider 3P Regional, preferring to depend on their own infrastructure, collaborate with global partners, or engage with local providers. Utilizing their own infrastructure provides the maximum degree of control, but involves capital and operational expenditures. Global partners, on the other hand, offer the benefit of mature, large-scale infrastructure, while local providers may combine the benefits of third-party expertise and specialization under government jurisdiction.

Governments vs. Topsites. To compare the hosting strategies of governments and popular websites, we select a subset of 14 countries (described in Table 5.5), including two from each region from different digital development strata and compare the adoption of third-party providers between those countries' governments and regional popular sites. We use Google's Chrome User Experience Report (CrUX) to compile a list of popular websites in these countries. To mirror our methodology, we employ VPNs and limit our scraping to resources one level beyond the landing pages. This depth limit is due to the intensive nature of deeper scraping of commercial sites (i.e., particularly broad trees) and the observation that a significant majority (95%) of government URLs are found just one level down. By leveraging the methodology described in (§5.1.4) and (§5.1.5), we then determine the serving infrastructure and geolocation of the organizations responsible for the infrastructure serving these top sites in each selected country. We also identify the fraction of non-government topsites that use either on-premise or third-party solutions to deliver content. This mirrors our government site analysis and redefines categories as (1) self-hosting, (2) global, (3) local, and (4) foreign providers. To identify self-hosted solutions, we use a heuristic from previous research [115, 119].

Region	**Country Code**
North America (NA)	Canada United States
Latin America and the Caribbean (LAC)	Mexico Brazil
Europe and Central Asia (ECA)	France Bosnia
North Africa and the Middle East (MENA)	UAE Israel
Sub-Saharan Africa (SSA)	South Africa Egypt
South Asia (SA)	India Pakistan
East Asia and Pacific (EAP)	Japan New Zealand

Table 5.5. Two countries per region were selected to compare content delivery strategies between government websites and top sites. Our selection criteria focus on capturing countries with varying levels of digital development within each region.

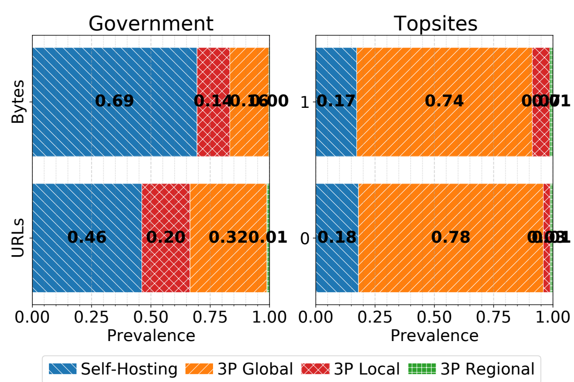


Figure 5.2. Comparison of self-hosting (on-premises) and third-party hosting between government websites and top sites within our selected subset of countries.

This comparison (Fig. 5.2) shows that top sites predominantly rely on 3P Global ●, using them to deliver 78% of URLs and 74% of bytes, more than twice as commonly as government sites with 32% for URLs and 16% for bytes. In contrast, on-premise infrastructure is much more prevalent across governments, with an average of 46% of URLs and 69% of bytes, compared to only 18% and 17% for top sites. The difference suggests the relative weight of considerations, beyond market forces, behind government hosting decisions.

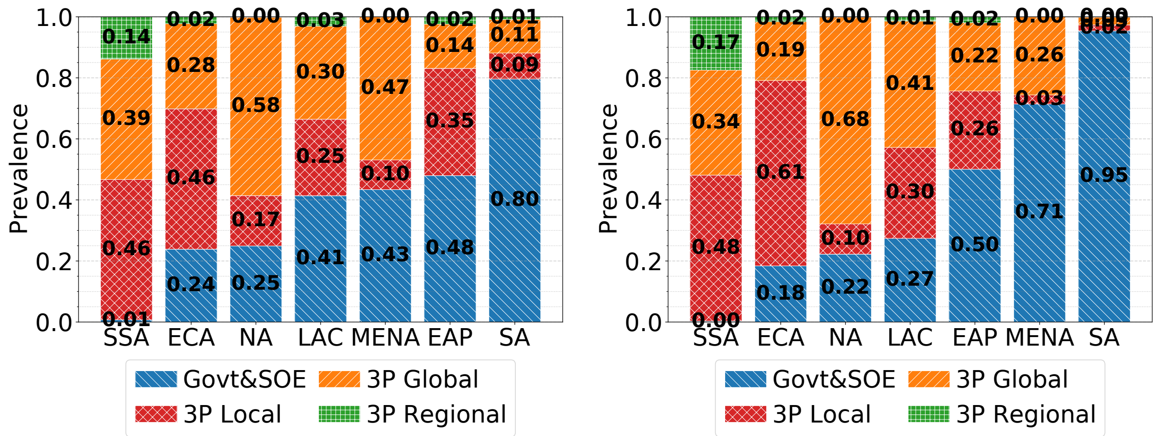
5.3.2. Regional Trends

In this section, we replicate our previous analysis now using the World Bank's regional division (§5.2.1) to investigate unique patterns or singularities that might exist in different regions. This regional-focused approach provides valuable insights into how factors like shared geography[4] and common cultural backgrounds may influence government decisions regarding digital infrastructure.

We assess both on-premises and third-party providers using the same four categories (Govt&SOE ●, 3P Global ●, 3P Local ● and 3P Regional ●) from a regional perspective. Figure 5.8 illustrates the regional prevalence of each category, represented separately for URLs (Fig. 5.8(a)) and bytes (Fig. 5.8(b)).

Both perspectives consistently reveal a significant variation in adopting Govt&SOE ● or 3P infrastructures across different regions. For instance, in regions like South Asia (SA) and North Africa and the Middle East (MENA) most bytes originate from government

[4]Geographical considerations affect choices to host content with providers whose serving infrastructure is distant from a particular country.



(a) URLs

(b) Bytes

Figure 5.3. Fraction of URLs and Bytes served by each provider category per region.

infrastructures (95% and 71%, respectively). In the case of North America, most bytes and URLs originate from 3P Global (68% and 58%, respectively). Sub-Saharan Africa (SSA), on the other hand, delivers most of their URLs and bytes through a combination of 3P Global and 3P Local infrastructures (85% and 82%), highlighting the complexity and variability in regional hosting strategies.

5.3.3. Countries' (dis)Similarities

We conclude our evaluation of hosting trends by examining the similarities in governments' serving strategies across the countries in our study.

We use the same four categories of government hosting options and look at both URLs and bytes. The distribution of URLs and bytes across these sources creates a unique

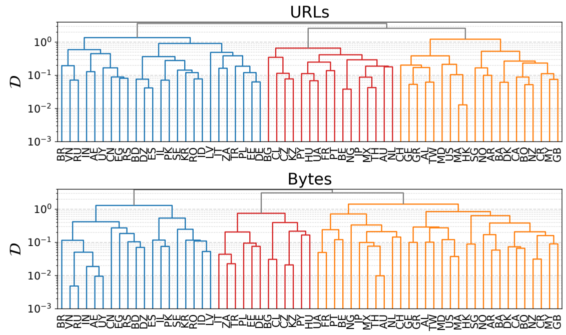


Figure 5.4. Similarities in governments' serving strategies across countries.

pattern, which represents the signature of a government's digital serving strategy. Our goal is to identify commonalities in these signatures across different countries.

We apply Hierarchical Agglomerative Clustering (HCA) using the Ward distance on a matrix that includes these four categories across countries, each represented by a row. This process results in the two three-branch dendograms shown in Figure 5.4. Each branch corresponds to the principal type of hosting sources (e.g., Govt&SOE).

The analysis shows the absence of strong regional patterns in government hosting strategies. For example, within the Southern Cone, Argentina, Brazil, and Chile each adopt a different approach, predominantly relying on 3P Global , Govt&SOE , and 3P Local , respectively. A similar diversity is observed in Southeast Asia, where Malaysia primarily depends on 3P Global in contrast to Indonesia's reliance on Govt&SOE . Even more remarkable is the situation within the European Union, a region bound by a common legislative framework yet displaying varied hosting preferences. For instance, Spain, Italy, and the Netherlands each show a distinct inclination, with major dependencies on Govt&SOE (64%), 3P Local (93%), and 3P Global (41%), respectively.

At the same time, it reveals similarities in the hosting strategies of countries from different regions despite having no apparent connections. For example, Brazil, Vietnam, and Russia share the same sub-tree due to their hosting similarities. We note, however, the challenges of generalizing from the observed trends and similarities. Apparent similar hosting practices may be driven by significantly different policies. In this case, Brazil's hosting choices may be the result of a comprehensive GDPR-like regulation, known as the LGPD [**125**], whereas Russia's [**160**] and Vietnam's [**126**] hosting models may respond to a focus on data localization and state control. France and Canada, though both predominantly rely on global providers (3P Global) for hosting, differ significantly in the extent of their reliance, with 42% and 79% of bytes, respectively, sourced from these providers. Likewise, Uruguay and Indonesia, primarily depending on government and state-owned enterprises (Govt&SOE), show considerable variance in their reliance, with 98% and 58% of bytes, respectively, attributed to government sources. These examples highlight the diverse approaches and degrees of dependency on specific hosting types, even among countries with similar strategies.

5.4. Hosting Registration and Server Locations

The previous section focuses on government preferences between on-premise and third-party hosting. Even when opting for third-party service, a government could have its content hosted within its jurisdiction. In this section, we explore this aspect of hosting, specifically answering: *What are the jurisdictions where the organizations serving government content are registered? What is the location of the servers hosting the content of government sites?*

We explore these starting with a global overview (§5.4.1), followed by a regional perspective (§5.4.2), and concluding with an analysis of cross-country dependencies (§8.2.1).

5.4.1. Global Trends

We examine the country of registration and the location of the servers hosting the government URLs in our dataset. Figure 5.1 categorizes this data globally into two distinct groups: (1) Domestic, and (2) International. While a majority of the URLs, to different extents, are served from servers located within the country (87%) and from addresses allocated to domestic organizations (77%), *23% of URLs are served from internationally registered organizations and 13% are served from servers located outside the country.* Note that foreign-registered organizations of domestically provided services may still need to comply with local legislation.

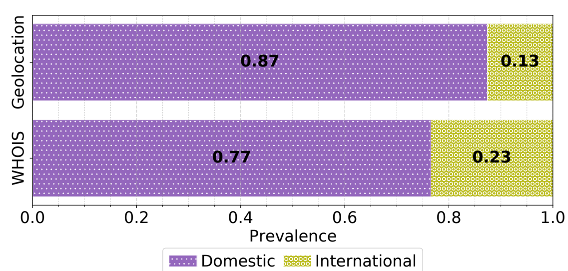


Figure 5.5. Fraction of Government URLs registered and served by Domestic or International Organizations.

Governments vs. Topsites. As in the previous section, we compare the hosting strategies of governments and popular websites, focusing on their use of domestic and international hosting solutions for the 14 selected countries.

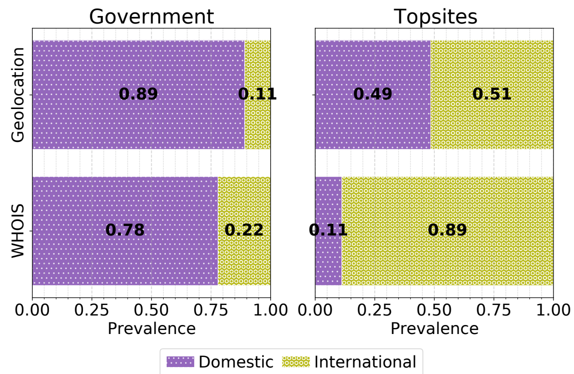
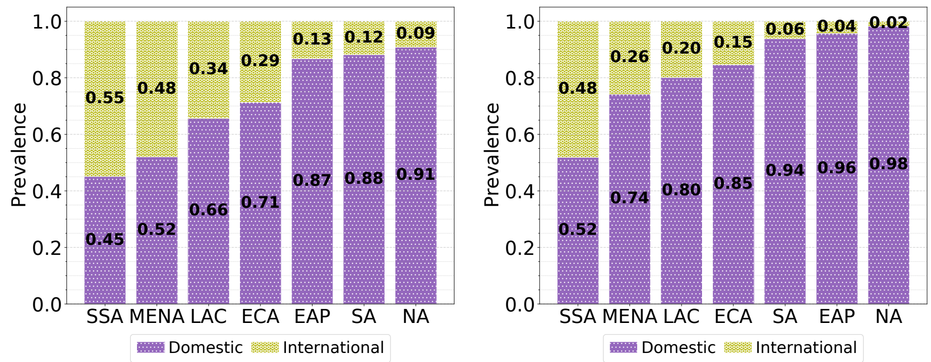


Figure 5.6. Comparison of domestic and international hosting between government websites and top sites within our selected subset of countries.

Figure 5.6 shows this comparison, displaying: (1) the country of registration of the organization and (2) the server locations serving the URLs in our dataset for this analysis. This comparison (Fig. 5.6) shows that governments predominantly opt for domestic hosting, with 78% of their URLs served by in-country registered organizations and 89% hosted within their borders. In contrast, popular websites prefer domestic hosting less; only 11% of their URLs are from domestically registered organizations, and just 49% of URLs are served from servers within their borders. This comparison highlights the different priorities between government entities, which favor control and jurisdictional autonomy, and popular websites that follow a more varied approach to digital service hosting.

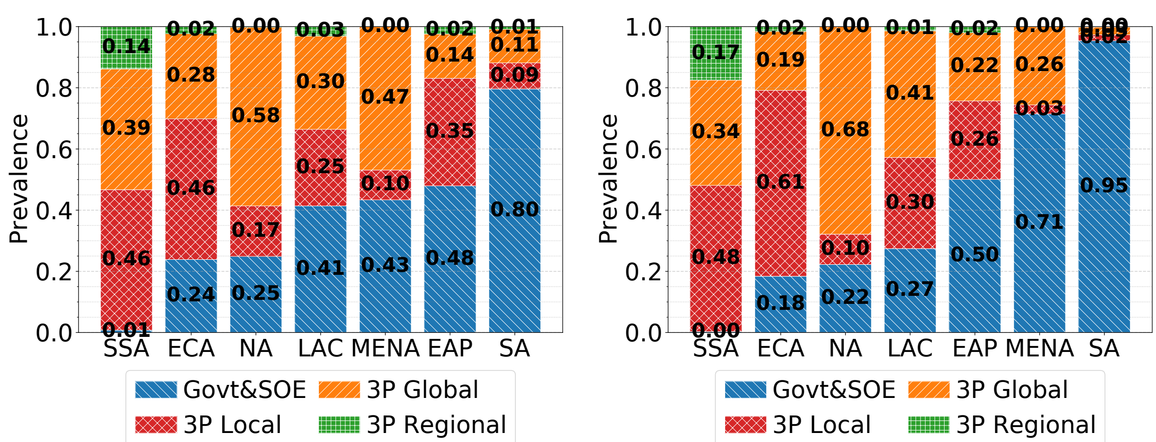
5.4.2. Regional Trends

At a regional level, we analyze the country of registration and the physical location of servers hosting government URLs in our dataset.



(a) Country of Registration (b) Server Location

Figure 5.7. Fraction of Government URLs registered and served by Domestic or International Organizations per region.



(a) URLs (b) Bytes

Figure 5.8. Fraction of URLs and Bytes served by each provider category per region.

Figure 5.7 presents this analysis, dividing organizations into two main categories: (1) Domestic and (2) International, with separate plots for their countries of registration (Fig. 5.7(a)) and server locations (Fig. 5.7(b)).

While most URLs in all regions are served from servers within their respective countries, the extent of this adoption varies significantly across regions. For example, in North America (NA), 98% of URLs are served domestically, compared to the Middle East and North Africa (MENA), where this drops to 74% and Sub-Saharan Africa (SSA) where the number of URLs hosted in the country drops to 52%. These variations are even more pronounced regarding the nationality of registrations. In North America, 91% of content is hosted by domestic companies, while in East Asia and the Pacific (EAP), Latin America and the Caribbean (LAC), Middle East and North Africa (MENA) and Sub-Saharan Africa (SSA), the percentages of URLs served by companies registered domestically are 87%, 66%, 52% and 45%, respectively. This may be partially explained by the maturity of digital markets in the US and Western Europe, where these third-party providers are registered.

5.5. Global providers and diversification

In the last section of our analysis, we focus on the networks responsible for serving government websites. The goal is to understand the role of Global Providers in this context (§5.5.1), and the degree of diversification among government providers (§5.5.2).

5.5.1. The Role of Global Providers

We have seen that governments are also engaged, if to a lesser extent, in the trend towards adopting third-party global providers for their digital services. In the following paragraphs, we characterize these providers, examining their global footprints, and analyzing countries' reliance on them.

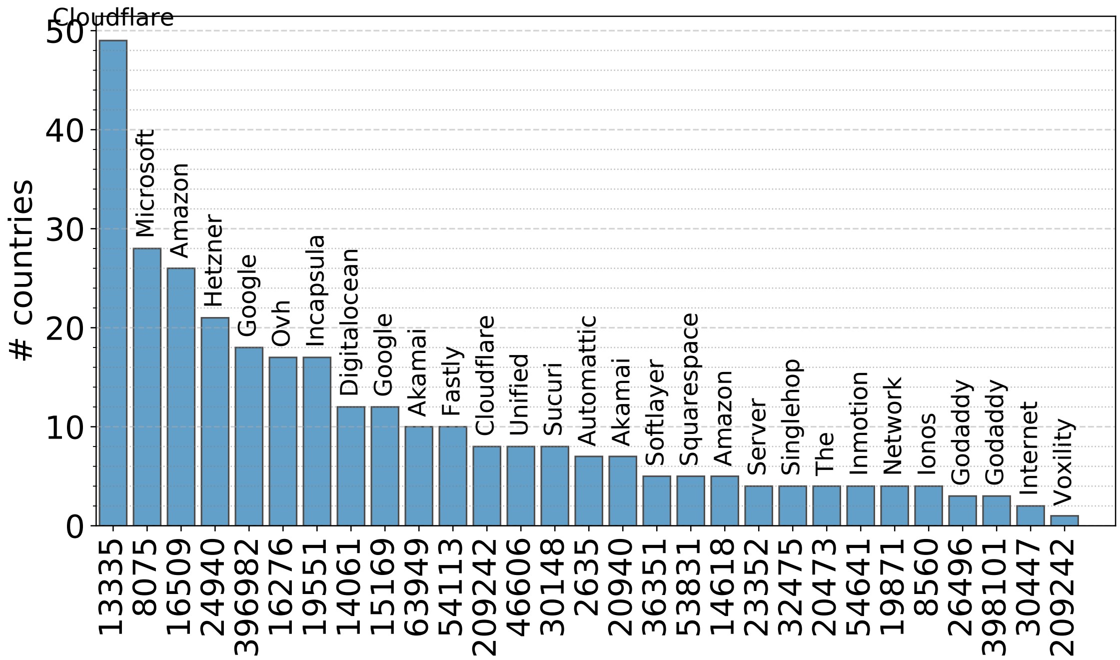


Figure 5.9. No. of countries that rely on Global Providers and CDF of Frac. of bytes served by Global Providers.

Figure 5.9 shows a histogram of the number of countries with government sites relying on one of the 28 global providers we identified. Cloudflare (AS13335) appears as the clear leader, serving content for 49 out of the 61 countries in our study. Cloudflare is followed by two other major cloud providers, AWS (AS16509, AS14618) and Azure (AS8075), hosting content for 31 and 28 countries, respectively.

To understand the degree of reliance on any given provider, we analyze the proportion of each country's data bytes served by each provider. At the top of the list, Amazon (AS16509) stands out by serving 97% of the bytes for an East Asian country, while Cloudflare (AS13335) is responsible for 72%, 58%, and 56% of the bytes for a country in Eastern

Europe, in South America, and a small Asian country, respectively. Additionally, Hetzner (AS24940) delivers 57% of the bytes for the government of a Scandinavian country.

5.5.2. Diversification of Hosting Providers

Diversification in hosting strategies can enhance the resilience of government services by reducing the risk of a digital shutdown caused by organizational failure. It also helps in creating isolation of data access across different domains. We explore whether governments tend to adopt more diversified hosting strategies and how this strategy correlates with their preference for using Govt&SOE ●, 3P Local ● or 3P Global ● for hosting their services.

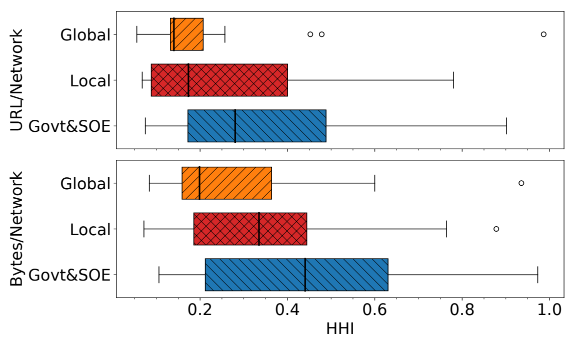


Figure 5.10. HHI distribution for the fraction of URLs and bytes served per hosting category.

To assess diversification in the networks serving government websites, we utilize the Herfindahl-Hirschman Index (HHI) [177], a common measure of market concentration. This index provides a score ranging from 0 to 1, indicating the level of network diversification, where a score closer to 0 indicates high diversification and a score closer to 1 indicates higher concentration. Figure 5.10 illustrates the HHI distribution for both the fraction of URLs and bytes served per network in each country. These are further categorized into three groups (Govt&SOE, 3P Local or 3P Global) based on the predominant source of bytes for each country.

While there is some overlap in the boxplots, governments mostly reliant on 3P Global tend to adopt more diversified strategies compared to those using 3P Local, and even more so than those relying on Govt&SOE. For example, while 63% (12 / 19) of the countries in the Govt&SOE category serve over 50% of their bytes from a single network, just 32% (8 / 25) of the countries in the 3P Global category depend on a single network for their bytes. Diversification is simpler with third-party providers, as it typically involves just contractual agreements. With on-premises hosting (Govt&SOE), on the other hand, diversification is more complex and may require significant capital investment.

5.6. Discussion

Our study is subject to a number of limitations. For starters, our compilation of government websites predominantly relies on self-reported information from governments (§5.1.1). We benefited from a global trend among governments towards developing data repositories to centralize government digital resources. In some countries, this process is part of legislative initiatives, such as in Brazil with the Digital Government Law, while in others, or efforts from the executive branch, such as Argentina's Ministry of Modernization and Spain's Ministry of Digital Transformation. This data is made available in different formats (e.g., HTML items, CSV files) and through different types of resources, from webpages to GitHub repositories, as in the case of the US Cybersecurity and Infrastructure Security Agency (CISA). Despite of this, the criteria for including services on these lists vary, often due to unique governmental administrative structures, legal frameworks, and cultural idiosyncrasies among other factors.

Our findings also reveal a lack of a standard convention for naming government domain names. While numerous countries adopt the ".gov" subdomain (or variations either in English or its equivalent in other languages) exclusively for government services, there are notable exceptions. For instance, state-owned enterprises rarely fall under this categorization and may use different domain structures. Furthermore, certain countries, including Germany, Poland, and the Netherlands, do not adhere to a specific subdomain convention for their government domains, indicating a varied approach to the digital identification of government entities across the globe which may impact our data collection effort.

In addition, our methodology focuses on public-facing services and excludes resources behind login portals, so it remains unclear if the same infrastructure supports publicly accessible and restricted resources. Despite recent advancements in understanding the potential use of Single-Sign-On (SSO) on top sites [**14**], these heuristics are not applicable to government sites that rarely accept third-party logins.

Finally, while we combine multiple approaches to minimize geolocation inaccuracies, we do not completely solve the problem. For instance, although active probing is the most accurate technique, it depends on factors such as server ICMP responsiveness and proximity of probes. In scenarios where active probing is not feasible, we resort to a multistage geolocation process, which can be costly. We opted for a conservative approach in our analysis, omitting (a small number of) IPs with geolocation from commercial databases that we could not validate.

5.7. Conclusion

We reported on the first comprehensive study exploring the hosting strategies of government digital services worldwide. Drawing from data collected across 61 countries spanning every continent and region, we examined preferred hosting models for public-facing government sites, and concentration in government services. These results provide an empirical basis for understanding how governments balance third-party expertise against sovereign control, and they inform policy agendas on digital sovereignty. Crucially, they also reinforce the dissertation's central methodological point: observed concentration is ambiguous without a comparative baseline. The same consolidation level may reflect ecosystem constraints in one setting and discretionary convergence in another, with different implications for exposure and resilience. Chapter 6 introduces the within-country, stratified comparison methodology that resolves this ambiguity, and Chapters 7–9 leverage that distinction to analyze resilience, exposure, and performance outcomes.

CHAPTER 6

Structural vs. Strategic Consolidation

Internet services increasingly depend on a small set of globally dominant infrastructure providers—including hosting platforms, authoritative DNS operators, and certificate authorities—creating shared failure domains and centralized control points. Prior work documents this concentration and its implications for resilience and sovereignty [**103**, **26**, **115**, **121**, **221**, **62**, **194**, **16**, **18**, **102**, **132**, **230**]. Governments are a critical case: as essential public services move online [**24**, **149**, **122**], they frequently rely on the same consolidated infrastructure. However, while existing studies (and the preceding chapters) primarily measure the extent of consolidation, they do not distinguish whether it reflects structural constraints in the surrounding ecosystem or strategic choices by governments, nor how these alternatives shape downstream outcomes.

Similar levels of consolidation can arise under different generative mechanisms. In some countries, governments may have little practical choice because viable domestic alternatives do not exist. In others, governments may centralize on a narrow set of providers despite the presence of alternatives, reflecting procurement decisions, governance structures, or policy preferences. Without separating these regimes, concentration metrics alone cannot reveal whether control and failure domains are the unavoidable consequence of the surrounding ecosystem, for example only one provider with domestic points of presence, or contingent on government decisions, for example centralized procurement, and therefore which resilience or performance interventions are realistically feasible.

Chapters 4 and 5 establish that consolidation is widespread across service layers and that governments often appear highly concentrated in the same provider markets as the commercial web. These measurements, however, leave a core ambiguity unresolved: the same concentration value can describe a government constrained by a thin domestic ecosystem or a government that centralizes despite available alternatives. The remainder of this chapter introduces a comparative baseline that resolves this ambiguity.

We recast consolidation as an inference problem and introduce a within country design that uses the domestic commercial web ecosystem as a counterfactual baseline. Comparing government deployments to popularity matched commercial sites embedded in the same national environment separates concentration consistent with shared ecosystem constraints, structural consolidation, from concentration specific to government decisions, strategic consolidation. To ensure the baseline reflects the broader market rather than only the capabilities of the largest commercial actors, we draw comparison sites from multiple popularity strata. Across 61 countries covering over 82% of the world's Internet population [25], we measure provider concentration across hosting, authoritative DNS, and certificate authorities.

6.1. Methodology

Our methodology is based on treating each country as a shared "structural block" that fixes much of the environment in which both government and commercial websites operate, including market conditions, regulation, and available infrastructure. Within each block, we compare government domains against commercial domains drawn from the same country and stratified by popularity to construct a within-country baseline. Similar consolidation patterns between government and commercial sites are consistent with consolidation driven by shared structural constraints, whereas systematic divergence indicates consolidation specific to government deployments. We describe how we (*i*) construct a multi-country corpus of government and commercial websites, (*ii*) extract each site's critical service providers from in-country vantage points, and (*iii*) quantify provider concentration and government–commercial differences.

6.1.1. Scope and Inputs

We analyze government and commercial websites across 61 countries spanning every continent and covering over 82% of the world's Internet population. Unless otherwise noted, all measurements and comparisons are performed independently within each country. The dataset includes all identifiable government websites and a popularity-stratified sample of commercial websites drawn from Chrome UX Report (CrUX) country lists. We focus on three service layers that directly affect web reachability, control, and trust: hosting, authoritative DNS, and certificate authorities (CAs).

6.1.2. Dataset Construction

Government Websites. We obtain government domains from a described corpus of government websites [**122**], drawn from national registries and authoritative government listings.

Commercial Websites. We start from the CrUX country-level list of size N and remove domains that appear in our government corpus to ensure the commercial strata are disjoint from government websites. We then stratify the remaining domains by popularity into three equal-sized buckets (Top, Middle, Bottom). From each bucket, we select 1,000 domains using systematic step sampling: within a bucket, we compute a step size $s = (N/3)/1000$ and select one domain every s entries in popularity order. Combining the samples yields 3,000 commercial domains, providing balanced coverage across the popularity distribution.

6.1.3. In-Country Measurements

To align inference with user experience, we collect measurements from an in-country vantage point using a country-local VPN endpoint [**153**]. [1] For each domain, we fetch the landing page and extract evidence of dependence on three classes of infrastructure services. We use a consistent browser and network configuration with uniform timeouts and retry logic across measurements to handle transient failures while preserving comparability.

[1]We validate the VPN locations by geolocating the VPN vantage point IP addresses using IPinfo [**110**].

Hosting. During page load, we record all contacted resources and map each resource's serving IP address to its origin AS using pfx2as [**38**]. We then label the hosting provider by mapping origin ASNs to organizations using CAIDA's AS-to-Organization dataset [**37**]. We treat these organizations as the hosting providers that the page depends on, capturing both first party hosting and embedded third party infrastructure that can materially affect availability and performance.

Authoritative DNS. We identify authoritative nameservers for each domain and map them to DNS providers using the AS Organization of the nameserver IP address, capturing control at the authoritative layer independent of recursive resolver behavior.

Certificate Authorities. For each TLS handshake, we parse the leaf certificate and extract the certificate chain. We label CA ownership using the CA Owner field from the Common CA Database (CCADB) [**36**] and treat the resulting organization as the site's certificate authority provider [**134**].

6.1.4. Quantifying Provider Concentration

For each service layer and popularity stratum, we quantify provider concentration using the Herfindahl–Hirschman Index (HHI) [**218**]. Let P denote the set of providers observed for a given service and stratum, and let s_p denote provider p's share, computed as the fraction of websites in that group that depend on p for the service. We compute:

$$HHI = \sum_{p \in P} s_p^2. \tag{6.1}$$

Higher HHI values indicate stronger concentration, reflecting greater dominance by a small number of providers.

6.1.5. Government-Commercial Comparison

While HHI summarizes concentration, it does not itself indicate whether government–commercial differences are statistically meaningful. To enable comparison, we use nonparametric bootstrapping to test whether government HHI values differ from those of commercial website (Top, Middle, Bottom). For each service and stratum, we resample domains with replacement (preserving the original sample size), recompute provider shares and HHI, and repeat this process for 5,000 iterations to obtain empirical sampling distributions.

From these samples, we construct bootstrap distributions of pairwise concentration gaps between the government and each commercial stratum (Top 1k, Middle, Bottom), e.g., $\Delta = HHI_{\text{gov}} - HHI_{\text{top}}$ (and analogously for Middle and Bottom), and assess significance using the 95% bootstrap confidence interval: if $0 \notin \text{CI}(\Delta)$, we treat the difference as statistically significant at the 0.05 level. Similar resampling-based inference has been used in prior work to interpret concentration indices beyond point estimates [**74, 189**].

6.1.6. Structural and Strategic Consolidation

We characterize consolidation patterns using the sign and statistical significance of ΔHHI. If ΔHHI is not statistically significant, government concentration is indistinguishable from the commercial baseline, consistent with *structural consolidation* under shared market conditions. If ΔHHI is statistically significant, government sites exhibit concentration that differs from commercial sites in the same environment, consistent with *strategic consolidation* specific to government deployments. Importantly, where significant, ΔHHI need not have the same sign across services or countries: governments can be more consolidated than commercial sites for hosting, yet less concentrated for certificate authorities.

We perform this analysis separately for hosting, authoritative DNS, and certificate authorities, and repeat it across popularity strata to distinguish divergence driven by market structure from divergence associated with procurement decisions or popularity-related effects.

6.2. Dataset Overview

We summarize the measurements collected in our study: in-country observations of government and commercial websites across 61 countries, spanning hosting, authoritative DNS, and certificate authority dependencies. Table 8.3 reports key dataset characteristics; the remainder of this section describes the scope of the websites, vantage points, and variables included in the analysis.

Dataset scope and interpretation. Our dataset represents a cross-sectional snapshot of web infrastructure dependencies. While deployments may evolve over time, the dataset captures consolidation patterns at scale across countries, services, and website types.

Web targets and services. For each of the 61 countries, we analyze three popularity buckets of commercial websites (Top 1k, Median 1k, Bottom 1k), comprising 56,999,

Component	Metric	Notes
Global reach	61	>82% of world Internet pop.
Web Targets	≈1.17M URLs	1,033,743 Govt, ≈174k Comm.
Infrastructure	≈8k Providers	5,382 Hosting, 2,424 DNS, 201 CA
Unique IPs	201,959	6.08% IPv6 adoption

Table 6.1. Summary of dataset characteristics. Counts represent unique URLs and providers.

58,803, and 58,262 URLs respectively (deduplicated per bucket across the country set). Separately, our government corpus contains 1,033,743 unique government URLs. Using per-domain identification results aggregated across all countries and buckets, we observe 5,382 unique hosting organizations, 2,424 unique DNS providers, and 201 unique CA providers.

Measurement targets. Our measurements use an IP target list containing 201,959 unique IP addresses; 12,279 are IPv6 (6.08%), and the remainder are IPv4.

6.3. Structural vs. Strategic Government Consolidation

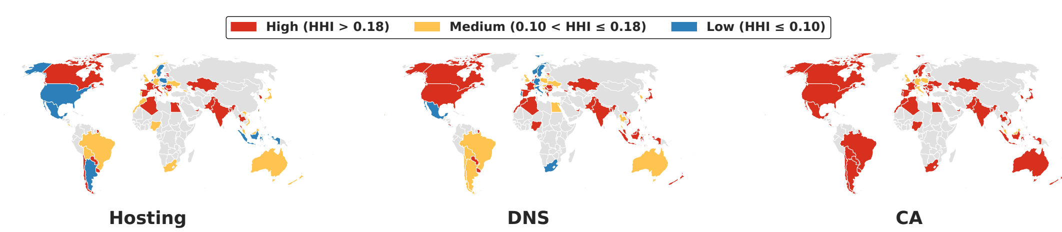


Figure 6.1. Government consolidation across hosting, DNS, and certificate authorities. Colors reflect HHI thresholds, highlighting service-specific differences in consolidation.

In this section, we apply our within-country, stratified framework to classify government consolidation as structural or strategic. Comparing government websites to popularity-matched commercial baselines across services reveals whether, for each country–service pair, observed consolidation reflects shared ecosystem constraints or government-specific deployment choices.

6.3.1. Government Consolidation by Services

We begin by summarizing observed levels of consolidation in government web infrastructure across three critical services: hosting, authoritative DNS, and CAs. This characterization establishes the empirical baseline needed to interpret whether government consolidation reflects shared ecosystem constraints or government-specific deployment choices. Each map in Fig. 6.1 shades countries by the government HHI for that service, with colors indicating standard concentration regimes. We adopt commonly used HHI thresholds (reported on the 0–1 scale; equivalently 0–10,000 in antitrust practice), where HHI < 0.10 denotes low concentration (HHI $< 1{,}000$), 0.10–0.18 moderate (1,000–1,800), and > 0.18 high ($> 1{,}800$) [**217**].

Across countries, hosting concentration in government infrastructure is typically moderate to high, and truly low hosting concentration is rare: only nine countries have hosting HHI < 0.10. Because our structural versus strategic analysis later conditions on countries with at least moderate government consolidation (HHI > 0.10), we briefly summarize the low-concentration cases here. This both documents the countries excluded from the subsequent split and highlights the limited settings in which government hosting remains unconcentrated. These cases are regionally clustered, with five in Europe (Hungary, Netherlands, Poland, Portugal, Sweden), three in the Americas (Costa Rica, Mexico, United States), and one in Asia (Indonesia).

Importantly, low aggregate concentration arises from qualitatively different configurations. In several European countries, low hosting HHI reflects dispersion across multiple domestic providers rather than reliance on a single global platform. Poland and Sweden illustrate this pattern: in Poland, multiple local hosts each capture nontrivial shares (Home.pl 21.6%, CyberFolks 11.5%, Nazwa.pl 8.8%), and Sweden shows a similar split across domestic providers (Oderland 14.8%, SiteVision 9.8%). Hungary follows the same theme, with a prominent government-adjacent operator (NISZ 11.1%), consistent with institutional hosting arrangements that retain a meaningful fraction of government deployments within nationally rooted providers. Comparable local operator signals appear outside Europe in Costa Rica, where ICE Telecom accounts for 12.8% of government hosting, and in Mexico, where reliance is split across several domestic network operators (Alestra 11.1%, UNINET 7.4%).

The United States presents a qualitatively different case. Hosting concentration remains low despite extensive hyperscaler participation, because government deployments are distributed across several large providers rather than collapsing onto a single dominant host (Amazon 24.1%, Akamai 8.6%, Microsoft 8.1%). These contrasts illustrate that similar hosting HHIs may reflect either domestic dispersion or competitive allocation among global platforms.

A similar pattern holds for authoritative DNS. DNS concentration is typically moderate to high across countries, and truly low concentration is even rarer than in hosting: only six countries have DNS HHI < 0.10 (Italy, Portugal, Sweden, Norway, Bosnia and Herzegovina, and Mexico). In the European cases, low DNS concentration is often supported by a nontrivial presence of domestic operators (e.g., Aruba and Vodafone Italia in Italy; Loopia and Netnod in Sweden; MEO and FCT in Portugal; Domeneshop in Norway; BH Telecom and QSS in Bosnia and Herzegovina). Mexico again appears as the sole non-European outlier, where low HHI is driven primarily by dispersion across multiple global platforms rather than a broad domestic operator base.

Certificate authorities exhibit the highest consolidation across countries. This is consistent with the structure of the global trust ecosystem: publicly trusted HTTPS certificates are effectively gated by inclusion in major platform root programs and compliance with CA/Browser Forum Baseline Requirements, which limits the viable set of issuers seen in practice [**148**]. In what follows, percentages denote the fraction of government sites whose server certificates chain to the corresponding CA. Domestic CAs appear as top issuers in only a small subset of countries; in most cases, government HTTPS converges on a small set of globally trusted issuers.

Where domestic issuers do appear, their presence coincides with explicit national trust infrastructures. Taiwan is the strongest example: CA concentration is high (HHI = 0.48), and domestic issuers account for most government deployments, led by TAIWAN CA (65.5%) and Chunghwa Telecom (18.2%), consistent with Taiwan's Government PKI rooted at the Government Root Certification Authority (GRCA) [**88**, **144**]. Other countries show meaningful but less extreme domestic presence, including France and Switzerland (HHI = 0.27 in both) and Spain (HHI = 0.26), where nationally supervised or state-owned certificate authorities remain prominent [**12**, **76**, **77**]. India exhibits a smaller but nontrivial local presence (HHI = 0.19) through eMudhra (17.0%), aligning with a nationally regulated CA regime [**65**, **56**].

These aggregate consolidation patterns are consistent with prior large-scale measurements of government web infrastructure, which document widespread reliance on third-party providers and high concentration across hosting and related services, but do not distinguish whether such consolidation reflects shared ecosystem constraints or government-specific deployment choices [**121**, **91**].

Taken together, these observations illustrate a central challenge for interpreting government consolidation: similar concentration levels can arise from different underlying configurations. In the next subsection, we address this ambiguity by directly comparing government deployments to popularity-matched commercial baselines within each country to distinguish structural from strategic consolidation regimes.

6.3.2. Classifying Strategic and Structural Consolidation

We operationalize the distinction between structural and strategic consolidation by directly comparing government concentration to commercial concentration within each country. We first assess, for each service in each country, whether government sites are meaningfully consolidated for that service (HHI greater than 0.1). When they are, we assign a service level label: structurally consolidated if government and commercial

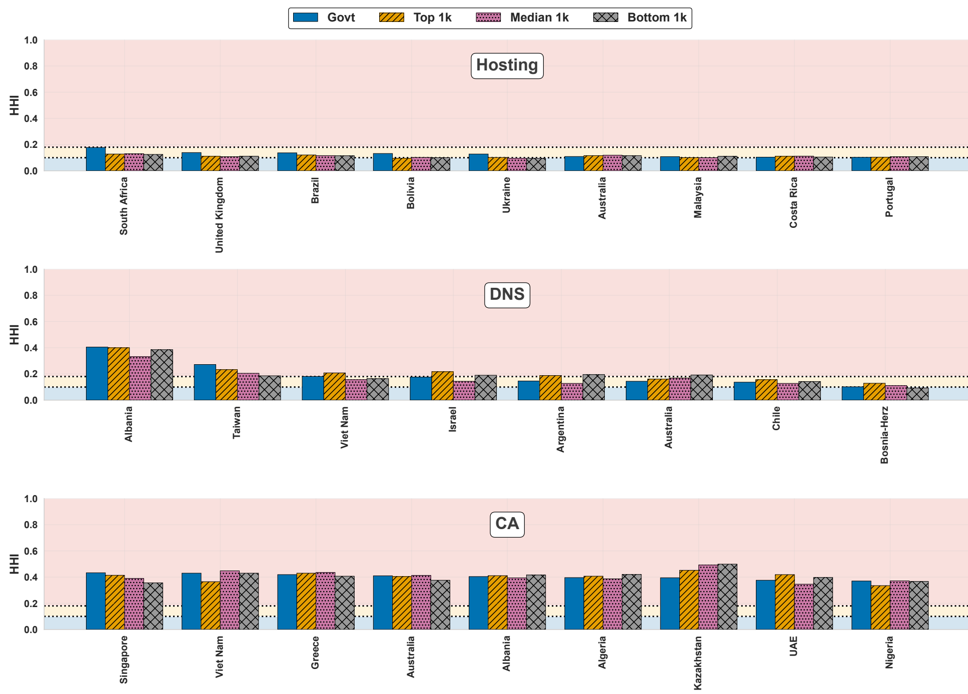


Figure 6.2. Structural consolidation cases by service. We show countries where government HHI exceeds 0.10 and aligns with concentration across commercial strata (Top 1k, Median 1k, Bottom 1k). Bars report HHI for government and commercial buckets per country; dotted lines and shaded bands denote low (≤0.10), moderate (0.10–0.18), and high (>0.18) concentration regimes.

concentration levels (HHIs) are comparable, and strategically consolidated if they differ statistically significantly in either direction.

This classification is based on concentration similarity rather than provider identity: structural consolidation can hold even when governments and commercial sites rely on different vendors, so long as the concentration is comparable.

6.3.3. Signatures of Consolidation Regimes

The resulting classification reveals two distinct consolidation regimes across countries and services. In structurally consolidated cases, government deployments closely mirror the concentration patterns of the surrounding commercial ecosystem, consistent with shared market constraints or institutionalized reliance on broadly available platforms. In strategically consolidated cases, we find that government concentration diverges sharply from commercial patterns, consistent with centralized platforms, procurement frameworks, or compliance regimes that narrow the feasible provider set for public-sector workloads.

Structural consolidation regimes. Several structurally consolidated countries illustrate mechanisms through which governments come to mirror the same provider concentration that characterizes their broader national web ecosystem.

Figure 6.2 shows countries classified as structurally consolidated for each service. For each panel (hosting, DNS, CA), we show countries whose government concentration is at least moderate and whose concentration closely tracks that of the commercial strata. Each group of bars corresponds to a country, with bars showing government and commercial HHIs; dotted reference lines and shaded bands indicate the concentration regimes used throughout the paper. The defining visual signature of structural consolidation is that the government bar closely follows the three commercial bars within the same country, indicating that government reliance on providers mirrors the broader ecosystem rather than reflecting a government-specific vendor choice.

Australia is a clear multi-service example. Across hosting, DNS, and certificate authorities, government and commercial sites rely on a nearly identical set of large, accredited providers, with only moderate variation in concentration across buckets. This alignment is consistent with Australia's Cloud First approach to government ICT and its multi-vendor procurement model [**66**], under which agencies are encouraged to adopt commercial cloud services from accredited providers rather than develop bespoke infrastructure. This pattern illustrates that structural consolidation does not require extreme concentration. Even at moderate concentration levels, government and commercial deployments can converge when policy actively encourages adoption of the same available platforms under shared market constraints.

The United Kingdom shows a comparable pattern at slightly lower concentration levels. Hosting HHIs are consistently moderate across all buckets, with government and commercial sites relying on the same dominant cloud and edge providers. In a mature market like the UK, this alignment is consistent with platformization rather than scarcity: the UK government has operated under a Cloud First policy since 2013 [**211**], and procurement mechanisms such as the Crown Commercial Service's G-Cloud framework standardize how public-sector bodies acquire cloud services [**58**].

South Africa provides a moderate but clearly structural hosting case. Hosting concentration is similar across buckets, yet the identities of the leading providers differ: government hosting is anchored in domestic firms such as Xneelo (35.6%) and Afrihost (11.1%), while commercial sites converge on hyperscalers and edge platforms. This illustrates that structural consolidation does not require identical vendors, only comparable concentration.

Vietnam shows a different but equally structural pattern, particularly in DNS. DNS concentration is consistently moderate to high and largely invariant across buckets. Government DNS is anchored by large domestic providers, while commercial strata converge on a small set of global managed DNS platforms. This bucket-invariant pattern is consistent with Vietnam's cybersecurity and data localization framework, including Decree 53 [**107**], which can compress provider choice across both public and private sectors.

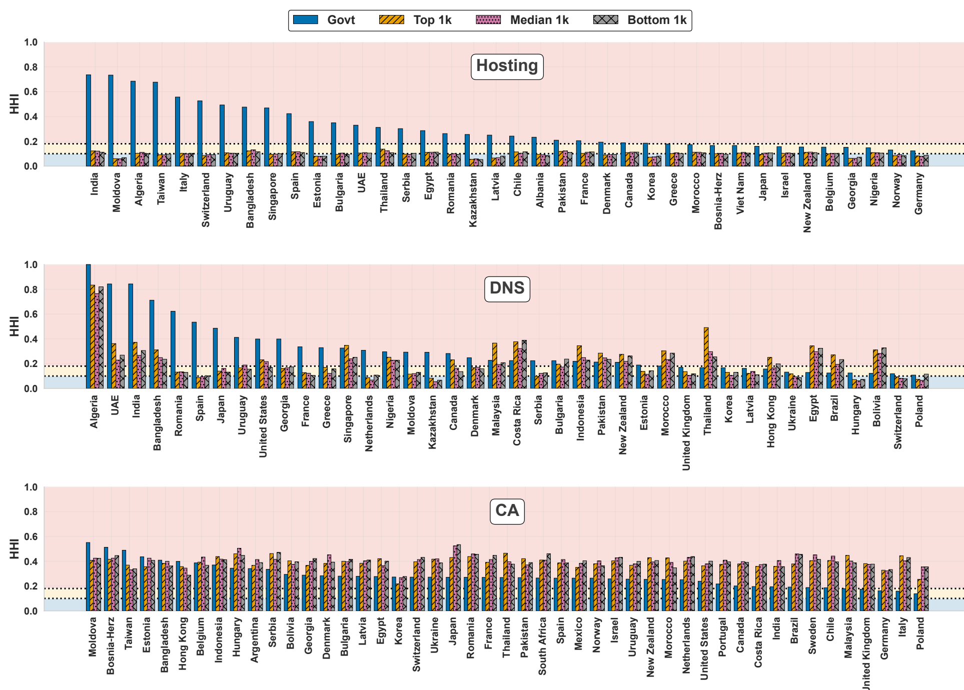


Figure 6.3. Strategic consolidation across services, sorted by government concentration (HHI). Countries shown here exhibit government consolidation patterns that diverge from the surrounding commercial ecosystem.

Strategic consolidation regimes. In contrast, a distinct set of countries exhibits clear strategic divergence. In these cases, government sites are substantially more or less concentrated than commercial sites for one or more services.

Unlike structural consolidation, which exhibits a relatively uniform signature across countries, strategic consolidation manifests through heterogeneous and service-specific mechanisms, reflecting differences in national platforms, procurement practices, and regulatory constraints.

Figure 6.3 plots strategic divergence cases by service, sorted by government HHI. Each country group shows the HHI for government sites and for the Top 1k, Median 1k, and Bottom 1k strata. We include only countries with at least moderate government concentration whose concentration diverges meaningfully from the surrounding commercial ecosystem.

The most consistent divergence appears in hosting, where government concentration exceeds commercial concentration across all strategically consolidated countries. DNS divergence is more heterogeneous, while CA divergence often runs in the opposite direction, with commercial strata as concentrated as—or more concentrated than—government, reflecting convergence of the broader web on a small set of globally trusted issuers.

6.4. Discussion

Our study operates under several measurement constraints that bound the scope of inference. We summarize these limitations and clarify how our design mitigates their impact.

First, our analysis is observational and does not establish causality; instead, we frame consolidation as a comparative inference problem and evaluate it relative to within-country commercial baselines.

Second, our characterization of dependencies focuses on three infrastructure services that directly affect reachability, availability, and trust and are repeatedly implicated in large-scale outages. Our analysis does not capture all forms of third-party dependence, such as client-side services or application-level integrations. By restricting attention to infrastructure-level services, we prioritize dependencies that are widely shared across sites and can be measured consistently at scale.

Third, provider attribution reflects infrastructure observed at measurement time, including IP-to-organization mappings, authoritative nameserver associations, and certificate chains. Deployments may evolve due to traffic engineering, migration, or operational changes. Accordingly, we interpret our results as a cross-sectional snapshot of consolidation patterns rather than a longitudinal characterization.

Finally, our classification of consolidation as structural or strategic is defined relative to the domestic commercial ecosystem within each country. In settings where the commercial web is itself highly constrained or globally dependent, this baseline may limit sensitivity to divergence. We mitigate this effect by stratifying commercial sites by popularity and comparing government deployments against multiple baselines rather than a single reference group.

Together, these choices enable robust comparative analysis of consolidation across countries and services, while bounding the scope of inference under partial observability.

6.5. Conclusion

Governments increasingly deliver essential public services over consolidated Internet infrastructure, but similar concentration levels can arise either from ecosystem constraints or from government deployment choices. This chapter shows that consolidation cannot be interpreted in isolation: without a relational baseline, concentration measurements underdetermine the source of dependence and the set of feasible interventions.

Using a within country, stratified comparison against popularity matched commercial sites, we distinguish structural from strategic consolidation across 61 countries and three critical service layers, hosting, authoritative DNS, and certificate authorities. Conditioning government deployments on whether they mirror or diverge from the surrounding commercial ecosystem makes observed variance interpretable and provides the foundation for Part II, which examines how consolidation regimes relate to resilience, exposure, and performance.

The structural versus strategic distinction is not only explanatory, it is necessary for interpretation. In Part II, we use it to test whether the consequences often attributed to consolidation differ depending on whether concentration is structurally induced or strategically chosen. We begin with resilience, asking whether strategic centralization buys meaningful practical fallback relative to what structurally constrained ecosystems already provide.

CHAPTER 7

Resilience Implications of Infrastructure Consolidation

This chapter begins Part II of the dissertation, which examines the user-facing implications of infrastructure consolidation. The analysis in this part focuses on three consequences of concentration: resilience, exposure, and performance. We begin with resilience because it provides the most immediate operational lens through which to test the distinction introduced in Chapter 6: if concentration reflects different underlying regimes, structural constraints versus strategic centralization, do these regimes produce meaningfully different fallback capacity in practice? As this chapter shows, the answer is largely no: when many services depend on the same providers and viable fallback is limited, failures can propagate broadly across otherwise unrelated systems regardless of consolidation type.

Real-world incidents underscore why resilience is a necessary complement to concentration metrics. In November 2025, an outage at Cloudflare, a dominant global CDN and DNS provider, disrupted access to major platforms including X and ChatGPT. In October 2025, an outage affecting Amazon's cloud infrastructure caused widespread disruption across websites and applications, including Reddit and Snapchat. Earlier incidents reveal the same pattern: in 2021, a misconfiguration at Fastly disrupted dozens of major websites, including Amazon, Reddit, *The New York Times*, Spotify, and the UK government's `gov.uk`; and in 2020, Let's Encrypt nearly revoked more than three million TLS certificates due to a software bug, which would have broken secure connections for a large number of dependent domains. Geopolitical events further reinforce this concern. During the invasion of Ukraine, attempts to reroute traffic through alternative upstream paths highlighted the extent to which transit dependencies can become points of coercion or disruption. Taken together, these episodes show that even highly capable providers can become single points of failure, but they do not by themselves reveal whether such fragility is structurally induced or exacerbated by strategic centralization. That is the question this chapter addresses.

These incidents establish why consolidation matters operationally, but they do not by themselves distinguish whether fragility reflects unavoidable ecosystem structure or avoidable centralization choices. That distinction matters for interpretation: if strategic consolidation is discretionary, one might expect it to buy better resilience through procurement, architecture, or fallback planning. This chapter evaluates that expectation by comparing provider redundancy across structurally and strategically consolidated settings.

7.1. Resilience as a Consequence of Consolidation

Consolidation is often described in terms of market share or provider dominance, but its operational consequences are best understood through the lens of resilience. In this dissertation, we use resilience to refer to the ability of web services to remain reachable when a provider experiences a fault, becomes unreachable, or is otherwise disrupted. Under this view, concentration matters not only because many sites depend on the same providers, but because such dependence can create shared failure domains across otherwise unrelated services. The comparative distinction introduced in Chapter 6 sharpens this question: if strategic centralization reflects discretionary concentration rather than ecosystem limits, one might expect it to be accompanied by stronger fallback arrangements. The remainder of this chapter evaluates whether that is in fact the case.

A key mechanism linking consolidation to fragility is limited practical fallback. Even when websites appear to use multiple endpoints or services, those dependencies may not correspond to independently operated alternatives that can continue to serve traffic during a disruption. We therefore evaluate provider redundancy as an operational measure of resilience, focusing on the number of organizationally distinct providers available for each critical service layer.

7.2. Provider Redundancy in Structural vs Strategic Consolidation

Consolidation is often framed in terms of dominant market share, but its operational impact is mediated by *redundancy*: the number of *independent* providers a site can fall back on if a primary provider fails, is blocked, or becomes unreachable. This section asks whether structurally and strategically consolidated countries differ on that operational dimension. We quantify redundancy by counting, for each domain, the number of distinct providers observed for each service and comparing the resulting distributions across structurally and strategically consolidated countries.

Provider counts are service-specific and require careful interpretation. For DNS, redundancy is meaningful only when authoritative nameservers are operated by *organizationally distinct* providers. We therefore compute DNS provider counts at the organization level, collapsing multiple nameservers operated by the same provider. This directly addresses a well-known pitfall in DNS configuration, where domains satisfy the two-nameserver guideline without achieving meaningful resilience, for example when both nameservers are operated by the same organization or depend on the same infrastructure.

For hosting, multiple observed providers do not necessarily imply effective redundancy. Multi-provider presence may reflect auxiliary infrastructure (e.g., analytics or third-party embeds), segmented deployments across subdomains, or geographically distributed footprints rather than interchangeable failover capacity. Certificate authorities are even more constrained in practice: domains typically present a single leaf certificate issued by one CA at a time, and parallel issuance from independent CAs is uncommon and constrained by the browser root store and CA compliance ecosystem. Accordingly, CA redundancy is effectively absent in practice, and we omit CAs from the provider-count comparison.

DNS exhibits minimal provider redundancy across consolidation types, with both the median and mean equal to 1. The distribution has only a small upper tail, reaching at most 10 providers in structurally consolidated countries and 9 in strategically consolidated countries. Consistent with prior work [**172**], we observe little to no organizational redundancy in authoritative DNS, with domains typically depending on a single independent DNS provider. Hosting shows higher provider counts, with both the median and mean around 3, and a heavier tail (maximum 38 in structurally consolidated countries versus 47 in strategically consolidated countries). However, these larger counts should not be conflated with guaranteed failover capacity given the multi-origin nature of modern web deployments.

Overall, structurally and strategically consolidated groups exhibit near-identical redundancy patterns for DNS and hosting, with matching medians and means. The central result is therefore not that the two regimes produce different resilience profiles, but that practical fallback remains limited in both. Strategic centralization does not systematically buy resilience: even when governments diverge from commercial market structure, independent fallback capacity remains scarce. Concentrated service ecosystems therefore remain operationally fragile, and policy-driven centralization does not appear to mitigate that fragility.

7.3. Discussion

These findings clarify an important distinction between concentration and resilience. Concentration describes how dependencies are distributed across providers, whereas resilience depends on whether those dependencies include viable, independent fallback options. In our data, redundancy is limited where it matters most: authoritative DNS shows near-zero organizational redundancy, and higher hosting provider counts often reflect deployment complexity rather than true failover capacity.

Interpreted through the framework of Chapter 6, this result is notable precisely because structural and strategic consolidation do *not* separate cleanly on resilience. Although the two regimes differ in origin, they converge on similarly limited practical fallback. This pattern helps explain why outages at major providers can produce disproportionate user-visible effects. When many services depend on the same organizations and few domains maintain independent alternatives, even routine failures can propagate widely. The resilience risks of consolidation therefore arise not only from provider dominance, but from the combination of dominance and limited practical fallback. While this chapter focuses on resilience through shared failure domains and practical redundancy, Chapter 8 turns to exposure, where the distinction between structural and strategic consolidation becomes more visibly consequential.

7.4. Conclusion

This chapter introduced Part II of the dissertation by examining the resilience implications of consolidation through the lens of failure amplification and provider redundancy. Recent outages illustrate how concentrated infrastructure creates shared failure domains, while our redundancy analysis shows that practical fallback is often limited, especially for authoritative DNS. Interpreted through the structural versus strategic framework of Chapter 6, the main result is that the two regimes do not meaningfully diverge on practical redundancy: strategic centralization does not buy stronger fallback than structurally constrained ecosystems already provide. The next chapter turns to exposure, where conditioning on consolidation type becomes more informative, including for foreign service dependencies and on-path intermediaries, before Chapter 9 examines performance implications through both infrastructure placement and steering.

CHAPTER 8

Exposure Implications of Consolidation

Exposure is a second major implication of consolidation because concentration can increase dependence on infrastructure outside local control. In this dissertation, we use exposure to refer to the extent to which access to web services depends on foreign jurisdictions, externally operated infrastructure, or shared intermediaries beyond domestic oversight. Exposure is also where the distinction introduced in Chapter 6 becomes especially informative: similar concentration levels need not imply similar external dependence. This chapter examines exposure across three layers. We begin with service endpoint exposure, comparing structurally and strategically consolidated countries using both provider registration and infrastructure geolocation across government and commercial site buckets. We then narrow to government hosting to analyze cross-border dependencies in where public-sector content is registered and served. Finally, we examine the Internet paths to government websites, showing how on-path transit and exchange dependencies create an additional and often hidden form of exposure that hosting location alone cannot detect.

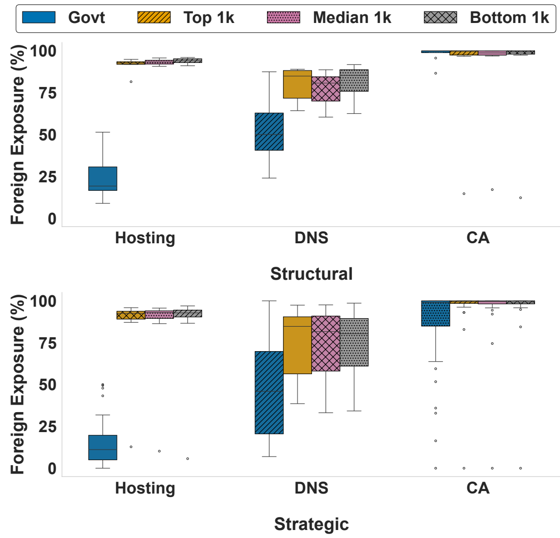
8.1. Exposure of Structural vs Strategic Consolidation

We start with foreign exposure, considering both provider registration and physical infrastructure location. This provides a cross-service and cross-sector view of exposure, allowing direct comparison across government and commercial site buckets under structural and strategic consolidation. The goal is not only to measure exposure, but to determine whether conditioning on consolidation regime changes how that exposure should be interpreted.

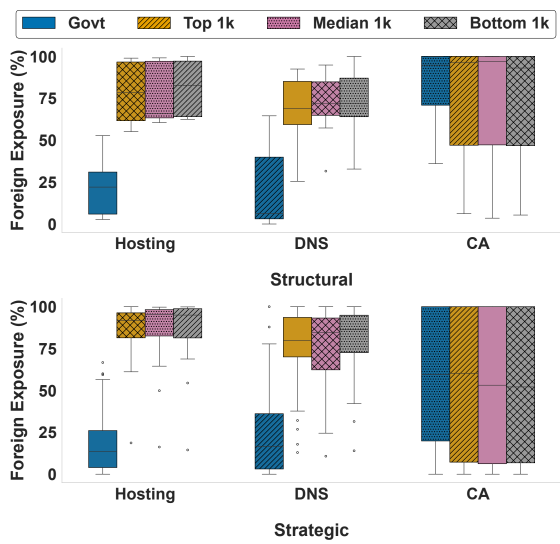
Throughout this section, we restrict our analysis to structurally and strategically consolidated countries whose service markets are moderately or highly concentrated.

Figures 8.1(a) and 8.1(b) summarize foreign exposure using two complementary lenses. In both figures, the x-axis groups results by service (Hosting, DNS, and CA), and within each service we show separate boxplots for each site bucket (Govt, Top1k, Median1k, Bottom1k). The two rows correspond to countries classified as structurally consolidated (top) and strategically consolidated (bottom) for that service. Each boxplot aggregates countries in the corresponding class and reports the distribution of foreign exposure values across countries (median and interquartile range). The y-axis differs across the figures: Figure 8.1(a) measures foreign exposure based on the WHOIS country associated with the observed service IPs, while Figure 8.1(b) measures foreign exposure based on inferred infrastructure geolocation for non-anycast endpoints, which helps isolate physical deployment effects from registration based signals.

Across services, structural consolidation yields relatively uniform exposure outcomes across countries, while strategic consolidation is associated with markedly higher cross-country dispersion. This contrast is central to the dissertation's argument: once concentration is interpreted through a comparative baseline, exposure no longer appears as a



(a) Foreign exposure by IP WHOIS country.



(b) Foreign exposure by geolocation for non-anycast IPs.

Figure 8.1. Foreign exposure under administrative (WHOIS) versus physical (geolocation) attribution.

simple consequence of concentration alone, but as an outcome shaped either by shared ecosystem constraints or by country-specific deployment choices.

8.1.0.1. Foreign Exposure by Provider Registration. We first quantify foreign exposure using the WHOIS country [59] associated with the service IP addresses we observe. For each service, we compute the fraction of sites in a country whose service IPs map to a WHOIS country outside that country, restricting to countries classified as structurally or strategically consolidated for that service.

Figure 8.1(a) shows foreign exposure by IP WHOIS country. For hosting, three patterns stand out. First, government sites are consistently less exposed than commercial sites across both consolidation types. Second, relative to structurally consolidated countries, both the government and commercial hosting distributions in strategically consolidated countries are shifted downward, indicating lower foreign exposure overall. Third, the variability differs by bucket: government exposure is more variable under structural consolidation (wider box) than under strategic consolidation, whereas commercial exposure is more variable under strategic consolidation than under structural consolidation. These patterns may reflect differences in incentives and constraints across sectors. Governments often have stronger procurement constraints and sovereignty or compliance motivations, which can keep exposure systematically lower than for commercial sites. The downward shift under strategic consolidation is consistent with cases where consolidation is driven by deliberate localization, affecting both government and, to some extent, the broader market. The variance asymmetry is consistent with heterogeneity in policy implementation: government localization can be applied relatively uniformly within a country, whereas commercial hosting decisions are more fragmented across firms and industries and differ widely across countries, producing greater dispersion under strategic consolidation.

DNS shows higher exposure across buckets, but remains systematically lower for government sites than for commercial sites in both consolidation types. The strategic DNS panel exhibits substantially greater dispersion, especially for the government bucket, indicating heterogeneous outcomes across strategically consolidated countries. A plausible explanation is that DNS exposure depends on which operators dominate authoritative DNS in a country: some strategically consolidated countries centralize government DNS on domestic incumbents or national platforms, while others consolidate onto a small set of global managed DNS providers, producing a wide cross-country spread.

For CA, foreign exposure is near saturated for commercial buckets in both panels, with tight distributions close to 100 percent. The clearest deviations appear as a handful of low exposure outliers, most prominently in the strategic government bucket, indicating that a small subset of strategically consolidated countries achieve measurably lower CA exposure for government sites. Overall, the strategic case shows greater variability than the structural case, and this increase in dispersion is much more pronounced for government than for commercial sites. This pattern is consistent with a global trust ecosystem in which most sites rely on a small set of internationally operated publicly trusted CAs, yielding uniformly high exposure for commercial domains, while government deployments occasionally follow country specific procurement or sovereignty driven requirements (e.g., mandated domestic issuance, national PKI, or preference for domestic CA operations) that can reduce exposure in some cases and produce greater cross country variation under strategic consolidation.

Across services, foreign exposure reflects both sectoral incentives and service-level constraints. Governments are consistently less exposed than commercial sites, strategic consolidation is associated with lower exposure but higher cross-country heterogeneity, and CA remains the least localizable service, exhibiting near-universal foreign exposure except for a small set of government-specific outliers.

8.1.0.2. Foreign Exposure by Infrastructure Geolocation. We next analyze foreign exposure based on the physical geolocation of service infrastructure. To avoid confounding effects introduced by global anycast deployments, we restrict this analysis to non-anycast endpoints. We identify anycast IPs using the longitudinal anycast census produced by Hendriks et al.'s LACeS system [**92**] and exclude endpoints whose observed service IPs are labeled as anycast in that dataset. We then geolocate the remaining non-anycast IPs using IPinfo [**110**][1]. Including anycast infrastructure would cause geolocation to closely mirror provider registration country, obscuring meaningful differences in physical deployment.

We summarize the prevalence of unique endpoints classified as non-anycast, by service and site category, for structurally and strategically consolidated countries in Table 8.1. While DNS endpoints are more often anycast, CA and hosting endpoints are predominantly non-anycast, providing broad coverage for our geolocation-based exposure estimates.

[1]Prior work by Darwich et al. [**61**] reports that, for 89% of targets, IPinfo's error is under 40 km, roughly within city level accuracy.

Bucket	Hosting (%)	DNS (%)	CA (%)
Structural			
govt	84.97	35.09	92.52
top_1k	67.89	29.50	97.19
median_1k	69.48	34.30	97.12
bottom_1k	70.46	27.07	96.23
Strategic			
govt	91.68	41.64	97.10
top_1k	74.68	53.10	98.04
median_1k	75.78	54.14	98.33
bottom_1k	75.83	46.98	97.63

Table 8.1. Share of unique endpoints classified as non-anycast, by service and site category, for structurally and strategically consolidated countries.

Among structurally consolidated countries, government infrastructure exhibits lower foreign geolocation exposure than commercial infrastructure for hosting and DNS, suggesting partial localization through deployment choices even when provider markets are constrained. Relative to provider registration signals, geolocation based estimates show greater physical localization across services, consistent with globally registered providers serving traffic from in country points of presence.

Strategically consolidated countries show a stronger and more heterogeneous localization pattern, reflecting uneven adoption of government specific localization policies across countries. Government hosting and DNS frequently exhibit lower foreign geolocation exposure than commercial sites, consistent with targeted localization in some countries. Compared to registration, geolocation yields substantially lower CA foreign exposure for both government and commercial sites, with the reduction most pronounced and most variable in strategically consolidated countries. This pattern is consistent with

CA-related endpoints often being physically served from in-country infrastructure even when the corresponding providers are registered abroad.

Notably, dispersion remains higher for strategically consolidated countries across services under the geolocation-based measure, indicating greater cross-country heterogeneity in where supporting infrastructure is physically deployed.

Together, these results characterize exposure at the service-endpoint level, showing how consolidation shapes foreign dependence through both provider registration and physical deployment. They also show why consolidation type matters for interpretation: structural consolidation is associated with more uniform exposure, whereas strategic consolidation produces more heterogeneous deployment outcomes. We next narrow the analysis to public-sector hosting, where these differences have direct implications for jurisdiction, accountability, and the delivery of essential state services.

The service-endpoint view reveals broad differences in how structural and strategic consolidation map onto foreign exposure, but it abstracts away from the specific public services that governments operate. To understand how these patterns manifest in state-facing infrastructure, we next narrow the analysis to government hosting. This allows us to examine not just whether exposure exists, but where cross-border dependencies are actually placed.

8.2. Government Hosting Exposure and Cross-Border Dependencies

To complement the service-endpoint analysis above, this section narrows to public-sector web infrastructure. We focus on government services because exposure in this setting has direct implications for sovereignty, accountability, and the secure delivery of essential public services. We examine both the country of registration of the organizations serving government content and the physical location of the servers hosting that content.

The previous section established broad patterns across service layers and site categories; here, we examine government hosting in greater detail to understand how those aggregate exposure patterns translate into concrete cross-border dependencies for state-facing services.

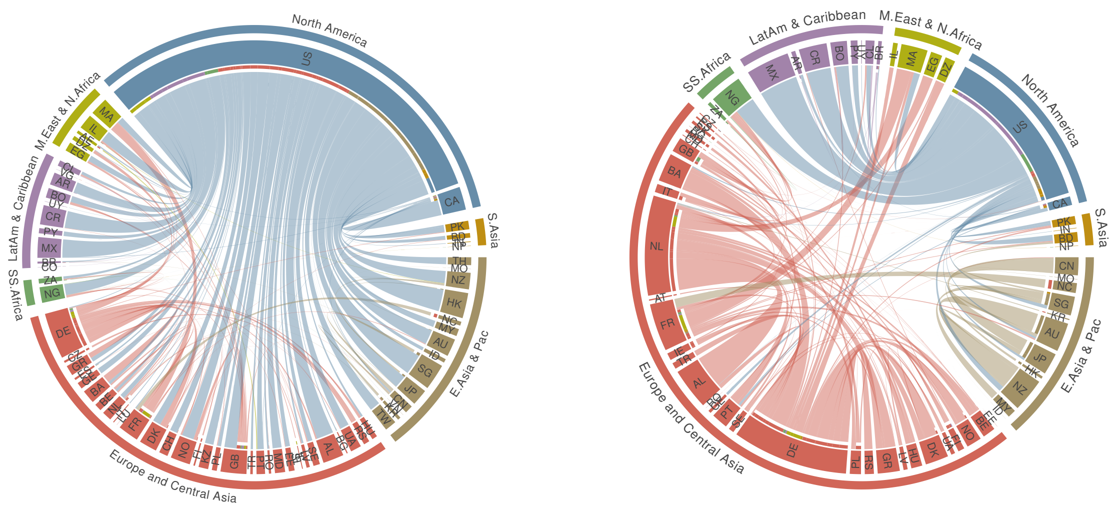
8.2.1. Cross-Border Dependencies

We explore the cross-border dependencies of government hosting to determine whether there are any preferences across the regions when selecting foreign countries from which this content is served.

Our analysis of cross-border dependencies examines both the country of registration and the location of servers from which governments' URLs are served.

Figure 8.2 presents this analysis through two circular Sankey diagrams, where countries are grouped using the World Bank's regional division, with one diagram showing the country of registration for these organizations (Fig. 8.2(a)) and the other showing the server locations (Fig. 8.2(b)). The plots reveal several interesting trends.

Inter-region dependency. This high-level analysis shows a clear trend with most governments largely relying on US-registered organizations in cases of foreign dependence. It also reveals that reliance on servers located abroad is generally confined to the same region; Table 8.2 shows this through interregional percentages.



(a) Country of Registration

(b) Server Location

Figure 8.2. **Cross Border Dependencies**. Flows represent the fraction of government resources that rely on a foreign country, either because the serving organization is registered there according to WHOIS records (Fig. 8.2(a)) or because the server itself is located there (Fig. 8.2(b)). Colors represent the region of the foreign country, while the color band connecting the flow to the foreign country represents the region of the source country that relies on it.

There are, however, some notable exceptions, such as the Middle East and North Africa (MENA) region relying on servers in Western European countries and Latin America and the Caribbean (LAC) predominantly depending on servers in the US.

In the case of Mexico and Costa Rica, we observe significant reliance on US-based servers with Mexico hosting 79.22% and Costa Rica 49.70% of their government URLs on servers in the US. In countries like Morocco, Egypt, and Algeria, the percentages of government URLs hosted on foreign servers are 48.38%, 21.1%, and 18.62%, respectively, similarly highlighting a pattern of dependence on international hosting solutions.

In sum, we observe that servers in North America and Europe serve 57% of government URLs crossing their respective country's borders. Brazil stands as the only exception in Latin America and the Caribbean, with only 1.78% of the URLs being served from the US, likely following Brazil's data regulation policy LGPD [**125**].

Region	%
Europe and Central Asia	94.87
East Asia and Pacific	80.79
North America	59.89
Latin America and Caribbean	3.41
Sub-Saharan Africa	2.95
Middle East and North Africa	0.00
South Asia	0.00

Table 8.2. Percentage of the cross-border dependencies that remain in the region.

Regional Affinity. When looking at cross-border dependencies within the same region – resources from other countries within the region – we find that South Africa hosts 100% of regional cross-border dependencies in Sub-Saharan Africa, Brazil hosts 85% in Latin America and the Caribbean (LAC), the US 83% in North America (NA), 76%, Japan hosts 60% in the East Asia and Pacific region and Germany accounts for 36% in Europe and Central Asia.

We also find some specific bilateral cases, such as New Zealand and Australia (with 40% of the URLs in New Zealand served from Australia). In general, we observe that 42%

of government URLs crossing their respective country's borders are served by servers within the same region.

GDPR Compliance. As part of our regional analysis, we explore compliance of government websites with the General Data Protection Regulation (GDPR) [**54**]. This EU regulation establishes that digital content within the European Union must be hosted on servers located within the member countries. Focusing on government websites, which might be more sensitive yet more likely to comply with their own regulations, we find a high level of compliance. Our analysis reveals that 98.3% (41,109 / 41,813) of URLs from EU countries are indeed served from servers within the EU's borders, indicating a strong alignment with GDPR requirements in the governmental digital sphere [**108**].

France and (former) colonies. We find interesting trends involving France with its historical and territorial connections. For instance, Morocco, which was a French protectorate from 1912 to 1956 [**219**], hosts 29.82% of its government URLs (that belong to 6 unique hostnames e.g., social.gov.ma) on servers located in France. On the other hand, 18.03% of the URLs of the French government are hosted on servers in New Caledonia, a French overseas territory in the southwest Pacific Ocean.

While New Caledonia is technically a part of France, its status is unique: it is not part of the European Union [**205**], it is an independent member of APNIC [**13**], listed by the UN as a non-self-governing territory [**213**], and has been engaged in long-standing discussions with France about independence [**206**]. Significantly, all URLs of the French government served from this territory are hosted by New Caledonia's state-owned provider, *Office des Postes et des Telecomm de Nouvelle Caledonie* (OPT-AS18200), and belong to the hostname *gouv.nc*. This highlights the complex interplay of historical, political, and technological factors in determining the hosting locations of government digital services.

China and India. China and India, two of the world's largest economies, show contrasting trends. Despite both countries predominantly depending on their domestic and government infrastructures, the extent of their reliance varies. For China, despite historical tensions with Japan [**201**, **151**], we find 26.4% of its URLs hosted by third-party providers in Japan. India, on the other hand, strongly prefers government hosting, with 99.3% of its URLs served domestically. This approach may relate to India's recent efforts to enhance data privacy, as reflected in the Digital Personal Data Protection (DPDP) Act passed in August 2023 [**105**].

Bilateral relationships and server deployments. The Dutch government adopts a singular approach to domain hosting, deploying servers abroad to support services linked to its bilateral relationships. For instance, *dutchculturekorea.com*, a cultural blog of the Embassy of the Kingdom of the Netherlands in Seoul, is hosted on a server located in Korea. Similarly, *nbso-brazil.com.br*, the website for The Netherlands Business Support Offices in Brazil, is served from a server within this South American country.

These hosting results show that jurisdictional exposure is shaped not only by whether governments use third-party infrastructure, but also by where the relevant organizations are registered and where servers are physically deployed. However, hosting location alone does not determine how citizens reach public services. While path exposure can arise under both structural and strategic consolidation, it is especially revealing in strategically consolidated settings, where government choices may shape not only where services are hosted, but also which intermediaries citizens must traverse to reach them. To capture this additional dimension of exposure, the next section examines the Internet paths to government websites and the on-path intermediaries that carry access traffic.

8.3. Hidden Path Exposure in Access to Government Services

The preceding sections characterize exposure through service endpoints and hosting infrastructure. We now extend that analysis to the network layer by examining the Internet paths to government websites. This focus is especially important for public-sector services because path-level dependencies determine which networks and jurisdictions can observe, influence, or disrupt citizen access to state-provided services. In the context of this dissertation, path exposure is a necessary complement to hosting-level analysis: even when concentration appears localized at the service endpoint, the actual route to that service may still traverse foreign-controlled infrastructure. Because these dependencies are not visible from hosting data alone, this section introduces a dedicated measurement and attribution framework to analyze on-path exposure and path centralization.

As governments rely on the Internet to deliver essential services, the structure and control of the networks that carry this traffic become matters of resilience, security, and sovereignty. Choices about whether to self-host or outsource delivery to commercial providers affect not only where services are deployed, but also which networks and jurisdictions mediate access to them. Self-hosting may increase control over operations, data governance, and regulatory compliance, while outsourcing offers scalability, global reach, and protection against threats such as DDoS attacks. Yet either approach can still leave access dependent on infrastructure beyond national oversight once the path itself is taken into account.

Hosting location, however, does not determine how citizens reach these services. Even when content is hosted domestically, access paths may traverse foreign transit networks or Internet Exchange Points (IXPs) before returning home. This "local-but-not-local" reachability, observed in studies of transnational routing detours [**73**], highlights a fundamental disconnect between the geography of servers and that of Internet paths.

The resulting routing dependencies matter because routing determines who can observe, influence, or disrupt access to government services [**158**, **127**]. Foreign transit networks may inspect or log traffic, as revealed by the NSA's MUSCULAR interception of inter-datacenter links [**89**, **64**], or shape reachability through filtering, throttling, or misconfiguration, as in Pakistan's 2008 YouTube hijack or the 2021 Akamai outage that disrupted government portals worldwide [**178**, **21**, **93**]. These risks are amplified in regions with limited domestic interconnection or where a small number of providers dominate routing.

As states pursue domestic hosting to assert "digital sovereignty", the goal of keeping critical public services under domestic jurisdiction and reducing foreign reliance, such hidden dependencies, where ostensibly local traffic detours through global hubs like Amsterdam or Singapore, may quietly undermine their efforts. Understanding sovereignty in practice thus requires visibility not only into where government services are hosted, but also into the Internet paths that carry users' traffic to them.

We address this gap through a global measurement of Internet routing to government websites. Using RIPE Atlas vantage points across 58 countries and a curated set of public-sector domains, we reconstruct traceroutes to map the Autonomous Systems (ASes) and IXPs that carry access traffic. We classify each hop by jurisdiction, domestic or foreign-controlled, and quantify two structural properties: cross-border exposure and path centralization. Methodologically, we introduce a conservative path reconstruction and attribution framework that infers jurisdictional exposure under incomplete and opaque routing, prioritizing coverage and risk sensitivity over single-path accuracy.

Our analysis reveals that cross-border exposure is widespread. In some regions, over 70% of paths to government services traverse foreign-controlled ASes and up to 40% rely on foreign IXPs, even when the destination is locally hosted. Conversely, in several countries routing is dominated by a small number of incumbents, creating potential chokepoints in national connectivity. We also find that countries with greater foreign transit exposure tend to exhibit weaker HTTPS adoption, suggesting that sovereignty and security risks often coincide.

Taken together, these results show that hosting domestically is not sufficient for digital sovereignty: the topology and control of network paths matter just as much as the location of servers. More broadly, they show that exposure cannot be inferred from concentration or hosting location alone. By exposing the unseen routes that connect citizens to their governments, this chapter highlights how sovereignty in practice is mediated by global interconnection.

8.3.1. A Framework for Inferring Jurisdictional Exposure

We develop a framework to measure cross-border dependencies in access paths to government websites. The framework integrates large-scale active measurements, path reconstruction, and jurisdictional attribution to identify how traffic reaches public-sector services and through which networks and countries. It directly operationalizes the three design requirements we outlined in the previous section: (*i*) broad *coverage* across countries and networks; (*ii*) *completeness* through reconstruction of incomplete traceroutes; and (*iii*) accurate *attribution* of each hop to its controlling organization and jurisdiction.

The measurement proceeds in three stages. We first select government web domains and identify vantage points within each country to capture representative user access. Second, we collect and reconstruct traceroutes to those domains, using a stitching process to infer end-to-end paths even when intermediate hops are missing. Finally, we map each hop to its originating Autonomous System (AS), classify ASes and Internet Exchange Points (IXPs) by country of registration, and compute jurisdictional exposure and path centralization.

These components provide the first global view of routing dependencies in access to public-sector services. To our knowledge, this is the first study to combine large-scale active measurements with jurisdictional attribution across more than fifty countries, enabling systematic quantification of both foreign exposure and infrastructure concentration.

8.3.1.1. Collecting Paths to Websites. The framework comprises three components: a set of *targets*, a set of *vantage points*, and the *paths* between them.

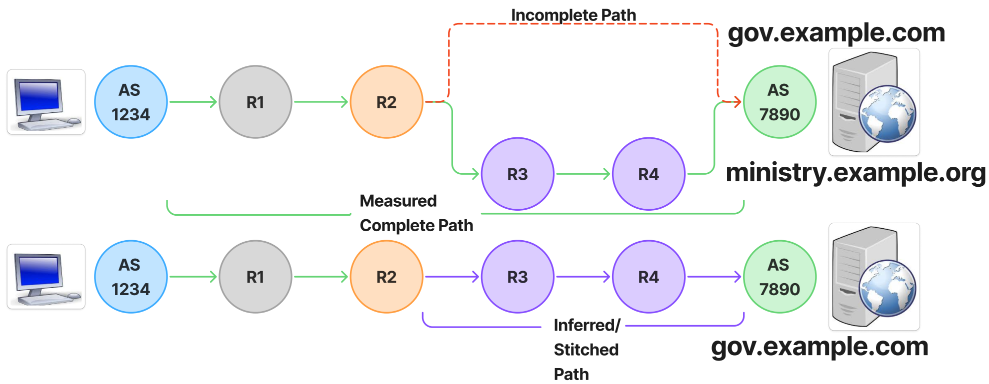


Figure 8.3. Example of traceroute stitching using a suffix from another trace to the same destination AS.

Targets. The input is a curated list of government web domains, which define the destinations for all measurements. Each domain represents a publicly accessible service operated or endorsed by a national government authority.

Vantage points. Although Internet routing is destination-based, end-to-end paths are shaped by source-specific policy and topology, and can therefore only be observed from the originating network. To approximate typical citizen access, we restrict vantage points to in-country locations, selecting probes from public measurement platforms such as *RIPE Atlas*. This choice ensures that observed paths reflect the perspective of domestic users rather than foreign vantage points or cloud resolvers.

Paths. From each vantage point to each target, we run traceroute to recover IP sequences, yielding the corpus of paths analyzed in this study. Section 9.3.2 details the dataset and implementation for the countries under study.

8.3.1.2. Complete and Incomplete Paths. As discussed in Sec. 3, many traceroutes fail to reach their destinations due to path opacity, including routers that suppress TTL-expired replies, MPLS tunneling, or filtered ICMP responses. To mitigate these gaps, we apply a stitching methodology that recovers missing hops by exploiting overlap among traces observed from the same or related vantage points.

Example. Figure 8.3 illustrates the intuition. A probe in AS1234 tracing a path to `gov.example.com` halts after router R2, three hops short of the destination ASN (AS7890). Another complete traceroute from the same source ASN to a different domain in AS7890 (`ministry.example.org`) includes R2 followed by additional hops to the destination. Because both traces share this intermediate router and terminate in the same destination ASN, we infer that the missing segment in the first trace can be completed using the suffix of the second. The resulting stitched path restores visibility into the full AS-level route from AS1234 to AS7890 by combining measured and inferred segments.

Methodology. Our approach builds on the path stitching technique introduced by *iPlane* [**136, 137**], which infers complete paths by matching suffixes from previously observed traceroutes that share router-level or BGP-prefix overlap. Whereas *iPlane* ranks candidate continuations to estimate a single most likely route, our objective is to conservatively characterize routing exposure under incomplete and policy-driven routing. Selecting a single inferred continuation risks underestimating third-party dependencies when multiple feasible paths traverse different transit networks or IXPs. We therefore retain all matching suffixes, favoring exposure completeness over precise single-path reconstruction. Each stitched path is later deduplicated when aggregating AS- and country-level statistics.

Tiered matching. We perform stitching in four progressively relaxed tiers:

(1) **Router-level match (same source ASN, same destination ASN):** Match the last responsive IP before the path becomes incomplete (e.g., R2 in Fig. 1) to a known router using CAIDA's ITDK dataset [**45**], then locate completed traceroutes from the same source ASN to the same destination ASN that traverse that router.

(2) **Prefix-level match (same source ASN, same destination ASN):** Match the BGP prefix of the last responsive IP and locate completed traceroutes from the same source ASN to the same destination ASN that include that prefix.

(3) **Router/prefix match (any source ASN, same domain):** Relax the source ASN constraint and match completed traceroutes from any source ASN to the same domain.

(4) **Router/prefix match (any source ASN, same destination ASN):** Match completed traceroutes from any source ASN to the same destination ASN.

At each stage, the algorithm appends matching suffixes from completed traceroutes to the truncated trace, beginning at the shared hop. Algorithm 3 summarizes this process.

8.3.1.3. Infrastructure Attribution. We enrich each traceroute with topology, organizational, and geographic metadata to attribute every hop to the network and jurisdiction responsible for carrying traffic.

Pseudocode 3 Stitching of Incomplete Traceroutes

```
Require: Incomplete traceroute T, corpus of completed traces C
Extract last responsive IP h, source ASN s, target domain d, destination ASN a
Get router ID r from ITDK for h
Get BGP prefix p for h
for each matching rule in order (stop after first match): do
    1. Traces in C from s to a that match r or p
    2. Traces in C from any ASN to d that match r or p
    3. Traces in C from any ASN to a that match r or p
    if match found then
        Append all matching suffixes (from shared hop to destination) to T
        return stitched T
    end if
end for
return original T                                        ▷ no match found
```

AS mapping. Each hop's IP address is mapped to its origin Autonomous System (AS) using CAIDA's prefix-to-AS dataset [**46**]. We then annotate every AS with its country of legal registration, enabling attribution of both domestic and foreign transit networks.

Country-level geolocation. To identify all countries traversed en route to a government resource, we combine multiple complementary geolocation techniques. For each hop, we first query its PTR record and extract embedded geographic hints using *The Aleph* [**208**], which has been shown to provide reliable location data when PTR records are available. If no PTR record is present or it lacks usable hints, we fall back to the commercial IPinfo database [**111**].

Because IP geolocation is prone to systematic error, we validate IPinfo's country-level assignments using active latency checks following the method of [**11**]. For each IP geolocated to a given country, we select five RIPE Atlas probes located there and issue ICMP pings. We consider the assigned location plausible if the minimum observed RTT is below a threshold derived from that country's maximum internal distance [**123**]. This validation step detects and discards implausible assignments due to common database inaccuracies.

IXP detection. We identify Internet Exchange Points along each path by comparing hop prefixes against PeeringDB's published IXP prefix list [**161**]. For every match, we record both the countries where the exchange operates facilities and its country of legal registration, allowing us to track the jurisdictions that interconnect government traffic.

8.3.1.4. Definitions. We next define key terms used throughout our analysis.

A government domain is classified as **domestically hosted** when the destination server is geolocated within the same country as the vantage point, using the methodology described in Section 8.3.1.3; it is **foreign hosted** otherwise. Hosting providers are identified by the AS delivering content to end users, capturing the effective delivery path rather than the authoritative origin. For Content Delivery Networks (CDNs), we report the location of the serving edge or origin server, even when placement reflects intentional cross-border optimization.

A **complete path** is any traceroute that either reaches the destination ASN directly or is reconstructed via the stitching technique described in Section 8.3.1. For incomplete traces, we retain all matching suffixes derived from previously observed paths that share router- or prefix-level overlap. We deliberately avoid ranking or filtering candidates to remain conservative, as discarding alternatives could underestimate cross-border exposure under opaque or unstable routing.

A **proper transit country** is any country appearing along a network path that is neither the vantage (origin) nor the hosting (destination) country, representing an intermediary jurisdiction through which government traffic transits.

Finally, we group countries using the World Bank regional framework [**207**]. Within Europe and Central Asia (ECA), we further distinguish EU member states (*ECA-EU*) from non-EU countries (*ECA-non-EU*) to contrast routing patterns under shared EU regulatory regimes with those in more heterogeneous national contexts.

8.3.2. Dataset

We summarize the dataset produced by our measurement campaign: traceroutes from 3,570 vantage points across 58 countries to 17,299 government domains. Table 8.3 reports key statistics; the remainder of this section details target selection, vantage criteria, the traceroute campaign, stitching yields, hosting breakdowns, and geolocation validation.

Table 8.3. Summary of dataset characteristics.

Category	Value	Category	Value
	Targets		**After Stitching**
FQDNs	17,299	Completed Paths	468,946
		IPs on Path	34,312
		ASes on Path	1,323
		IXPs on Path	367
	Vantage Points Countries (ASes)		**Paths to Content**
Total	58 (3,570)	Paths to Domestic-hosted Content	352,772 (75.2%)
		Paths to Foreign-hosted Content	116,174 (24.8%)
		Traceroutes	
Probes	3,570	Traceroutes	511,116
Dates	2025-03-25 – 2025-05-02	Complete Traceroutes	268,416

Government websites We use 17,299 federal-level government FQDNs from Kumar et al. [**123**] covering 58 countries. The list spans ministries, agencies, public services, and state-owned enterprises (government share >50%). While some sites are rebranded over time, manual spot-checks confirm representativeness during our collection window.

Vantage selection criteria We run traceroutes from RIPE Atlas probes geolocated within each country. To reduce bias toward dominant networks, we select at most one probe per AS. We analyze 58 countries with ≥3 probes, spanning all major regions: NA (2), LAC (8), ECA (29), MENA (3), SSA (1), SA (3), and EAP (12), together accounting for 78.1% of the global Internet population. A per-country breakdown appears in Table **??** (App. **??**).

Traceroute campaign We issued 511,116 traceroutes from 3,570 probes between March 25 and May 2, 2025.

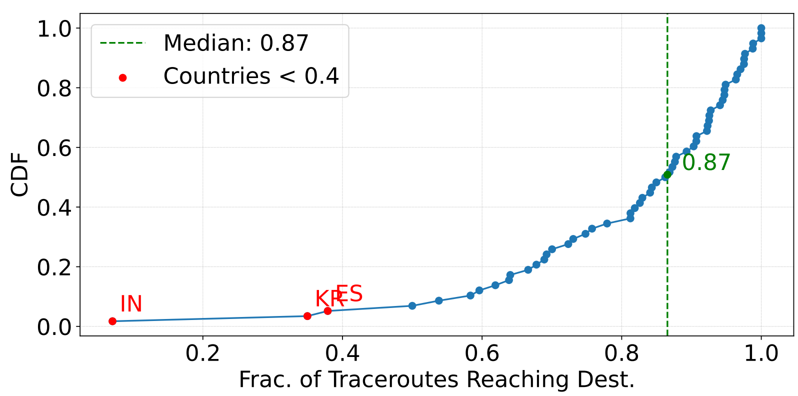


Figure 8.4. Per-country fraction of traceroutes that reach the destination ASN.

Fraction of stitched traceroutes Most countries exhibit high completion: in over half, >85% of traces reach the destination ASN (Fig. 8.4). A few countries (e.g., India, South Korea, Spain) show lower completion (<40%). We apply the stitching method of Sec. 8.3.1 to incomplete traces. Pre-stitch, 268,416 paths reach the destination ASN; post-stitch,

Table 8.4. Results of geolocation validation. We confirm 26 859 (78.3%) IPs via PTR and latency checks; 7,453 (21.7%) inconsistent entries are removed to ensure conservative exposure estimates.

Geolocation Validation Metric	Value
IPs with PTR Geo Hints	3,468
IPs with IPinfo Country Metadata	34,307
IPs Successfully Pinged	34,307
IPs Verified via PTR Geo Hints/Ping-Latency	26,859 (78.3%)
IPs Discarded	7,453 (21.7%)

we obtain 468,946 complete paths (an increase reflecting multiple valid suffixes per truncated trace, later deduplicated in aggregate analyses). The resulting paths traverse 34,312 unique on-path IPs, 1,323 unique on-path ASes (excluding source/destination ASes), 367 IXPs, and reach 780 unique destination ASNs. Validation with held-out complete traces (App. **??**) yields a mean Jaccard distance < 0.1 between stitched and original paths.

Domestic vs. foreign hosting Of 468,946 complete paths, 352,772 (75.2%) terminate at domestically hosted destinations and 116,174 (24.8%) at foreign-hosted destinations – i.e., roughly one quarter of paths reach government content hosted abroad.

Geolocation validation We combine PTR-derived hints, IPinfo metadata, and latency checks to validate country assignments. As summarized in Table 8.4, conflicting or insufficient signals are discarded, yielding conservative lower-bound exposure estimates.

8.3.3. Structural Path Dependencies

We analyze structural path dependencies along two orthogonal dimensions. The first, jurisdictional exposure, captures where government traffic travels and which foreign jurisdictions mediate access (§8.3.3.1–8.3.3.3). The second, path centralization, captures how access depends on a small set of networks or exchange points, whether routing concentration causes traffic to funnel through a few providers, creating chokepoints even when paths remain domestic. We return to this second dimension in §8.3.4.2, where we quantify such concentration using Herfindahl–Hirschman Index (HHI) and examine its operational implications.

Prior work has focused on where governments host their content; here, we shift attention to how users reach it. We quantify two dimensions of fragility: structural concentration, the degree to which connectivity depends on a few transit or IXP providers, and jurisdictional exposure, the share of those intermediaries registered abroad. Both factors affect performance, resilience, and digital sovereignty by introducing potential bottlenecks and external points of control. Appendix **??** and **??** extend this analysis by mapping the corporate jurisdictions of the ASes and IXPs observed along the paths.

8.3.3.1. IXP Dependencies. We begin with IXPs, critical junctures where networks interconnect to exchange traffic. IXPs provide a direct physical substrate for domestic routing and are a natural lens through which to assess whether locally hosted government traffic remains within national boundaries. By mapping IXP participation in end-to-end paths, we classify access routes into three categories: (*i*) those traversing domestic IXPs, (*ii*) those relying on foreign IXPs, and (iii) those that bypass IXPs entirely through private or transit interconnection. This breakdown reveals how national interconnection ecosystems shape exposure: strong domestic exchange correlates with path locality, while gaps in peering infrastructure or restrictive policies often force traffic through foreign exchanges, introducing avoidable dependencies.

Identifying IXPs on the Path. We detect IXP traversal by matching each traceroute hop against the IXP prefix list in PeeringDB. When a match is found, we verify that the IXP facility and the vantage point are located within the same country to distinguish domestic from foreign exchanges.

We exclude two edge cases. First, IXPs with facilities spanning multiple countries – 15 of 367 in our dataset (4.1%) – are omitted, as their multi-jurisdictional structure prevents clear national attribution. These account for only 0.01% of all paths. Examples include `NET.IX` and `NL.ix`, which advertise presence in hundreds of locations across more than 100 cities.[2] In contrast, operators such as `DE-CIX` and `AMS-IX` are listed as distinct entities per country in PeeringDB and are retained. A full list of excluded IXPs appears in Appendix **??**.

Second, we exclude paths traversing more than one IXP, which represent only 0.33% of the dataset. These multi-IXP routes typically reflect layered interconnection strategies—for example, domestic and global exchanges co-located within the same facility. Paths from Asia, Latin America, and Africa frequently include both `AMS-IX` and `NL.ix`, illustrating Amsterdam's role as a global aggregation hub. Similarly, global platforms such as `NET.IX` often appear alongside national IXPs, indicating hybrid peering configurations that combine local and international reach. Although excluded from our main analysis, such cases highlight complex multi-IXP connectivity and geographically distributed exchange ecosystems that warrant separate study.

[2] `NET.IX` operates over 220 facilities globally.

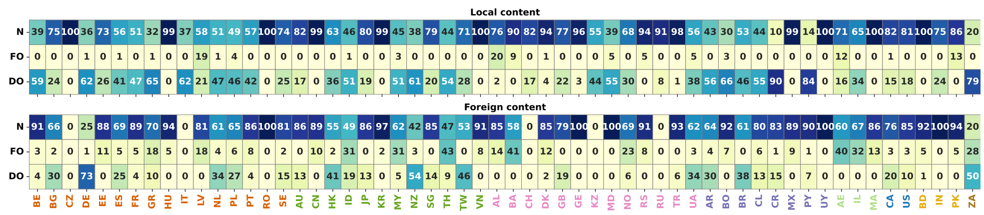


Figure 8.5. IXP usage on paths to government services by country and region. N: no IXP; DO: domestic IXP; FO: foreign IXP.

IXP Reliance. Figure 8.5 shows IXP dependence across regions for both domestically and foreign-hosted government content. Each subplot reports the share of complete paths falling into three categories—traversing domestic IXPs, traversing foreign IXPs, or bypassing IXPs entirely—grouped by vantage country and colored by region.

Across all regions, three dominant access patterns emerge. First, many countries exhibit strong localization, with most paths to domestically hosted government services passing through domestic IXPs. This suggests that interconnection and traffic-retention policies are often effective in keeping local traffic within national boundaries. Second, a substantial fraction of countries have paths that bypass IXPs altogether, typically where exchange points are sparse or insufficiently developed to attract major networks. Finally, some countries rely heavily on foreign IXPs, indicating persistent gaps in domestic peering or regional interconnection ecosystems.

Several nations route nearly all paths to locally hosted government services without traversing any IXP—including Vietnam (100%), Uruguay (100%), Morocco (100%), Bangladesh (100%), China (100%), and Mexico (99.5%). Strikingly, a similar no-IXP pattern also appears for foreign-hosted services in Uruguay (100%), India (100%), Georgia (100%), Bangladesh (92.2%), Bolivia (92.2%), Vietnam (91.7%), and Belgium (91.8%). These cases highlight countries where domestic or cross-border routing remains dominated by private or bilateral interconnection rather than public exchange fabrics.

At the opposite extreme, some countries offload a sizable share of locally hosted traffic through IXPs located abroad—such as Latvia (LV), Albania (AL), Pakistan (PK), and the United Arab Emirates (AE), each routing over 10% of domestic paths through foreign facilities.

Reliance on foreign IXPs is even greater for foreign-hosted government content. Thailand (TH), Bosnia and Herzegovina (BA), the United Arab Emirates (AE), Israel (IL), Indonesia (ID), Malaysia (MY), South Africa (ZA), and Norway (NO) route between 23% and 43% of such paths through exchanges outside their national borders. These patterns underscore how regional hubs, such as Singapore in Southeast Asia or Amsterdam in Europe, continue to mediate cross-border government traffic.

In contrast, some countries exhibit domestic consolidation rather than external dependence. Kazakhstan (KZ) hosts no government content abroad, and all IXP-traversing paths to domestic services pass through a single government-operated exchange, `KAZ-GOV-IX`, managed by the State Technical Service. This centralized design interconnects national ISPs and exemplifies a deliberate strategy to retain government traffic within sovereign infrastructure.

Overall, IXP participation reveals how structural and policy differences manifest in routing outcomes: dense local exchange ecosystems support domestic traffic retention, while their absence—or excessive centralization—creates dependencies that extend control beyond national borders.

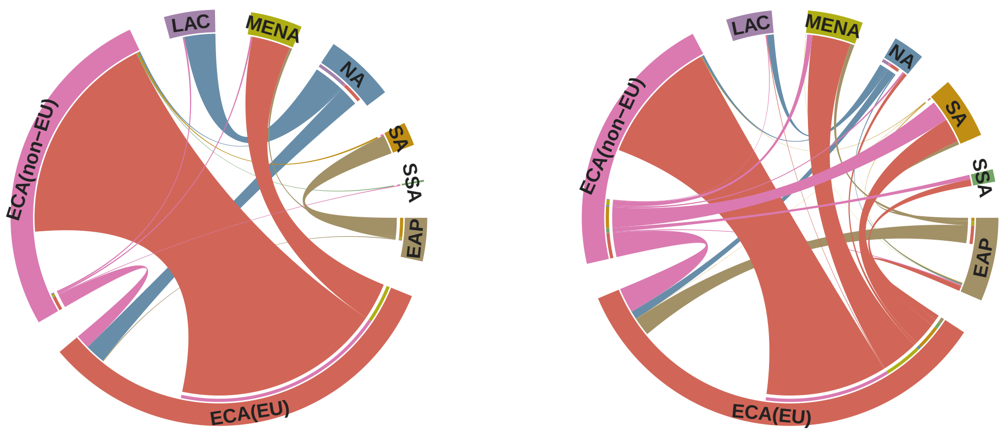


Figure 8.6. Cross-regional proper transit for government services. Local hosting remains regional; foreign hosting detours via Europe and East Asia.

8.3.3.2. Cross-Region Transit Dependencies. We analyze how user traffic to government websites traverses regional boundaries, identifying which regions act as transit hubs in the delivery of public-sector digital services. Specifically, we quantify the frequency of *proper transit*—traffic that crosses a third country or region distinct from both the user's vantage point and the content's hosting location—across both domestically and foreign-hosted government services.

Figure 8.6 visualizes these flows using chord diagrams based on World Bank regional groupings. The outer segments represent the origin regions (color-coded), while the inner bands denote the regions of transit countries encountered along the paths. The width of each arc reflects the share of paths from a source region that traverse a transit country in another region, excluding the vantage and hosting countries consistent with our definition of proper transit. Gaps in the diagrams indicate that paths remain within their region. Together, these visualizations reveal how traffic localizes within or leaks beyond regional interconnection ecosystems.

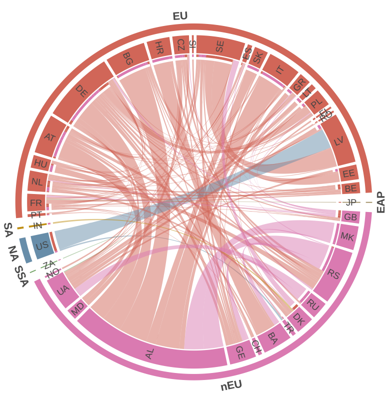
Regional differences are pronounced. **South Asia** shows the highest reliance on proper-transit countries, underscoring limited internal interconnection. For foreign-hosted services, an average of 15.6% of paths—and up to 72% in some countries—detour through hubs such as France, the United Kingdom, or Singapore. Even for locally hosted content, South Asian paths transit Singapore 23% of the time.

By contrast, the **European Union** exhibits strong regional containment. For locally hosted content, proper-transit reliance averages just 2.7%, reflecting Europe's dense IXP ecosystem and mature transit markets. A few outlier traces cross the Atlantic to the United States, but these remain rare.

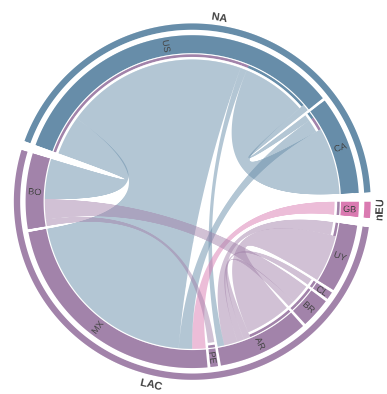
The **Middle East and North Africa (MENA)** region also displays notable external dependence. For foreign-hosted content, an average of 9.0% of paths—and up to 63% in the most extreme case—transit through Southern Europe (notably Spain, France, or Italy), mirroring long-standing geopolitical and commercial interconnection ties.

Sub-Saharan Africa (SSA) records the lowest exposure to proper-transit countries for domestically hosted services (average 1%), most often detouring through the United Kingdom. While this suggests improving local hosting and containment, paths to foreign-hosted services still depend on non-African transit hubs.

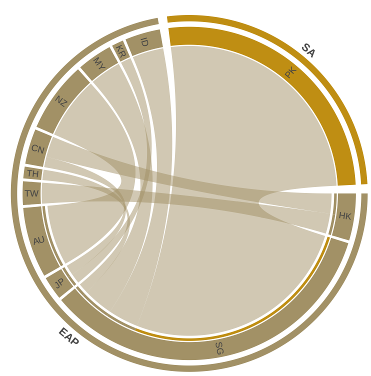
Finally, **Latin America** exhibits limited intra-regional transit and frequent reliance on North American intermediaries, highlighting persistent structural constraints in regional interconnection capacity. Appendix **??** provides additional per-country breakdowns.



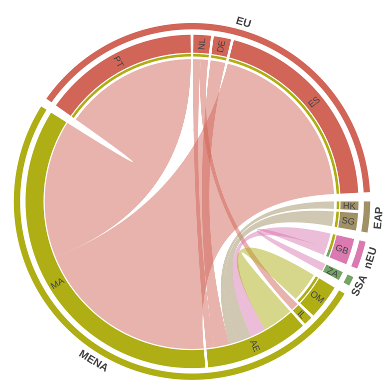
(a) Europe and Central Asia with and without EU.



(b) Latin America and Caribbean + North America

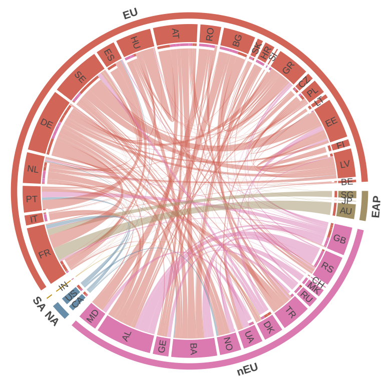


(c) East Asia and Pacific + South Asia

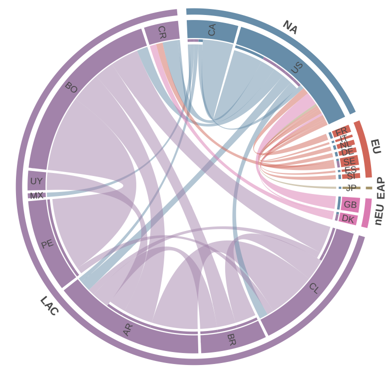


(d) Sub-Saharan Africa + Middle East and North Africa

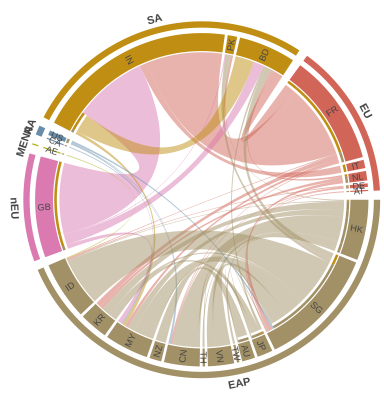
Figure 8.7. Cross-regional proper-transit flows for domestically hosted government services. Most paths remain within-region, but some countries route local government traffic through foreign transit hubs.



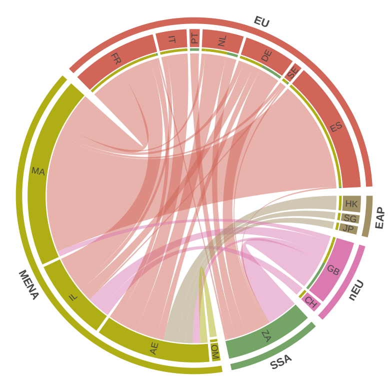
(a) Europe and Central Asia with and without EU.



(b) Latin America and Caribbean + North America



(c) East Asia and Pacific + South Asia



(d) Sub-Saharan Africa + Middle East and North Africa

Figure 8.8. Cross-regional proper-transit flows for foreign-hosted government services. Paths frequently traverse third-party regions, especially Europe and East Asia.

8.3.3.3. Transit Dependencies by Regions. We analyze country-level routing dependencies across major regions to identify how geography, infrastructure, and policy shape exposure to foreign transit. Figures 8.7 and 8.8 show proper-transit flows for domestically and foreign-hosted government services. Countries are grouped into four aggregates: Europe and Central Asia (ECA, split into EU and non-EU), Latin America and the Caribbean with North America (LAC+NA), East Asia and the Pacific with South Asia (EAP+SA), and Sub-Saharan Africa with the Middle East and North Africa (SSA+MENA).

Europe and Central Asia (ECA). ECA exhibits the strongest regional containment. For both domestically and foreign-hosted content, most paths remain within Europe, concentrating around major transit hubs in Germany, France, the Netherlands, and Italy. This reflects the region's dense IXP fabric and mature peering ecosystem. Long-haul detours outside Europe are rare but present, typically involving the United States or Asia, often tied to overseas territories or specialized hosting arrangements. Even for domestically hosted services, traffic frequently transits centralized regional hubs, indicating asymmetric routing and consolidation despite geographic proximity.

Latin America, the Caribbean, and North America (LAC+NA). LAC countries largely retain government traffic within the region, particularly for domestically hosted services. Brazil stands out as fully localized even for foreign-hosted content, while Canada shows systematic dependence on U.S. transit. Mexico is a notable exception within the region: over a quarter of paths to domestically hosted government services traverse U.S. networks. Smaller countries exhibit targeted long-haul dependencies driven by hosting choices, with traffic occasionally routed through Europe or Canada despite regional alternatives.

East Asia and the Pacific, and South Asia (EAP+SA). EAP demonstrates strong regionalization, anchored by Singapore as the dominant transit hub. Most countries retain both domestic and foreign-hosted traffic within the region, with limited reliance on extra-regional intermediaries. China's paths frequently traverse Hong Kong, reflecting its role as a functional national gateway. In contrast, South Asia shows substantial exposure to extra-regional transit: foreign-hosted government services often route through Europe or Singapore, reflecting limited regional interconnection capacity and the external orientation of hosting markets.

Sub-Saharan Africa and MENA (SSA+MENA). MENA maintains strong interconnection ties with both Europe and East Asia. Countries such as Morocco rely heavily on Southern European transit, while Gulf states exhibit eastward routing through Asian hubs. Even domestically hosted services in MENA frequently exit the region, underscoring persistent structural dependencies. SSA shows lower overall exposure for domestic hosting, but foreign-hosted services still rely disproportionately on European hubs, reflecting historical and commercial interconnection patterns.

Summary. Routing dependencies vary sharply by region. East Asia largely localizes government traffic, while South Asia, MENA, and parts of SSA depend heavily on extra-regional transit through a small set of global hubs such as Singapore and Amsterdam. Many countries either bypass IXPs entirely or rely on foreign exchanges, revealing gaps in domestic peering. Even where content is hosted locally, routing architecture, rather than hosting choice, ultimately determines jurisdictional exposure and control.

8.3.4. Risks

The structural dependencies described in Section 8.3.3 also translate into operational and sovereignty risks. We examine how routing exposure intersects with two additional dimensions: (i) protocol-level security, captured through HTTPS adoption, and (ii) control concentration among intermediary networks. Both can undermine user privacy, service availability, and national control over digital services. By combining these metrics, we highlight cases where routing fragility is compounded by weak encryption practices or excessive reliance on a single intermediary.

8.3.4.1. Exposure and Weak Encryption. We examine the most elementary — or "zero-order" — form of exposure: communications to government websites that leave the country without authentication or encryption. Although many other vulnerabilities exist, even within national borders, we focus here on this baseline risk, in which citizens' traffic travels in plaintext outside the government's jurisdiction and is thus exposed to interception or tampering along foreign segments of the route.

We begin by examining HTTPS adoption for 1,014,623 government URLs across 58 countries from the dataset of Kumar et al. [**123**], which includes government landing pages and their internal links obtained through recursive scraping up to seven levels. For each URL, we tested HTTPS support, either natively or through enforced redirects, by connecting via VPN providers [**152, 203, 94**] within the respective country to capture local accessibility. We verified the geographic locations of these VPN nodes using the methodology described in Section 8.3.1.3 to ensure that the vantage points operated from the countries they claimed to represent. Adoption rates vary widely, with a median of just

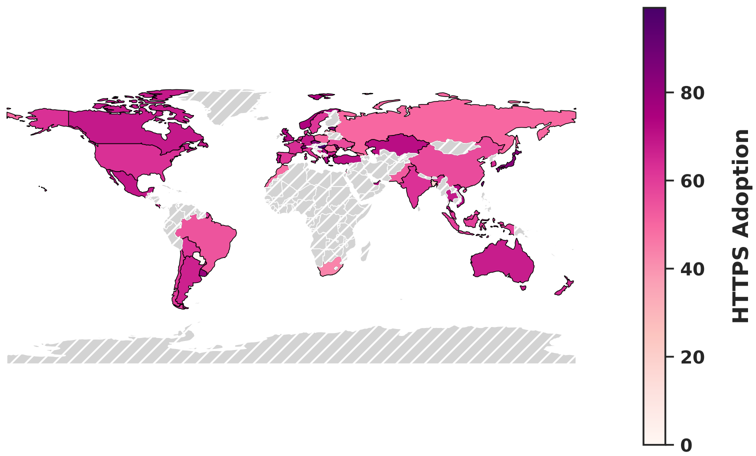
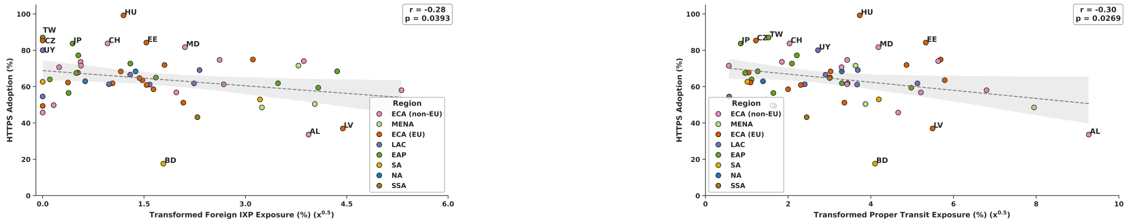


Figure 8.9. HTTPS adoption across government websites by country (median: 64%).

64.0%. Figure 8.9 presents a heatmap of HTTPS support—either natively or via enforced redirects—where darker shades indicate higher adoption. Some countries show strong deployment, such as Hungary (99.2%), Taiwan (87.0%), and Estonia (84.2%), while others fall well behind. Bangladesh (17.6%), Albania (33.6%), and Latvia (36.9%) rank among the lowest, revealing significant gaps in baseline web encryption.

Several of these low-HTTPS-adoption countries also exhibit high infrastructure exposure, compounding the risks to user traffic. Albania stands out with the highest proper transit exposure (86.0%) and significant foreign IXP dependence (15.5%), while only a third of its government websites support HTTPS. Morocco routes 63.2% of paths through foreign transit ASes and has only 48.5% HTTPS adoption. Latvia shows dual exposure as well, with 30.2% of paths transiting foreign ASes, 19.7% involving foreign IXP use, and just 36.9% HTTPS coverage. The coexistence of insecure routing paths and limited HTTPS adoption highlights structural weaknesses in protecting government-bound traffic.



(a) Foreign IXP (b) Proper Transit

Figure 8.10. HTTPS adoption vs. foreign IXP and proper-transit exposure (negative correlation).

We next examine how HTTPS adoption relates to infrastructure exposure. In Fig. 8.10, we plot HTTPS adoption rates against square root–transformed measures of proper transit exposure and foreign IXP dependence (the transform addresses the heavy-tailed nature of these distributions). Each point represents a vantage country (colored by region), and the dashed line shows the fitted linear trend.

We compute the Pearson correlation between HTTPS support and the transformed exposure metrics, finding a moderate negative relationship in both cases: proper transit exposure ($r = -0.334$, $p = 0.0128$) and foreign IXP dependence ($r = -0.279$, $p = 0.0392$). In general, countries whose paths more frequently traverses foreign jurisdictions tend to have lower rates of HTTPS deployment.

We acknowledge that security risks are not limited to cross-border routing. Legacy equipment and unpatched systems can expose services to vulnerabilities, even when infrastructure is hosted and operated within national borders.

8.3.4.2. **Path Centralization and Failure Risks**. We now return to the second structural property, path centralization, and quantify the extent to which access to government services depends on a small number of intermediary networks or exchanges. We use concentration metrics (HHI) over transit ASes and IXPs as a proxy for path centralization, since high concentration implies that a large share of paths depend on the same intermediaries.

While foreign exposure highlights the geopolitical risks of interconnection, an equally critical risk lies in the concentration of control within a small number of intermediary networks. When paths to government websites are routed primarily through a single network or exchange, the entire platform becomes operationally dependent on that intermediary. A routing misconfiguration, policy change, or outage at that entity could disrupt access to large portions of government services.

Centralization Assessment To quantify the risk of path centralization, we compute the HHI [**216**] over the set of Internet exchanges and intermediate transit networks observed in each country's paths, excluding both source and destination networks. Following U.S. Department of Justice guidelines, we treat HHI values above 0.18 as highly concentrated and below 0.1 as unconcentrated. Countries exceeding these thresholds are flagged as relying on a narrow set of interconnection or transit providers.

Concentration in Transit and Exchange Dependencies Tables 8.5 and 8.6 summarize the HHI values for IXPs and intermediate transits across all countries, highlighting those classified as concentrated or highly concentrated.

IXP HHI values range from 0–0.28 (mean 0.08, median 0.05), while transit HHI values span 0.01–0.59 (mean 0.11, median 0.09). Only four countries exceed the 0.18 threshold for IXP-level concentration—Germany (0.28), New Zealand (0.26), Argentina (0.23),

Table 8.5. IXP HHI scores across countries for IXPs on the path (sorted by HHI). Countries with concentrated or highly concentrated markets are highlighted.

Country	HHI	Country	HHI	Country	HHI
DE	0.28	NZ	0.26	AR	0.23
KZ	0.19	PY	0.18	RO	0.17
ZA	0.16	CL	0.16	NL	0.15
BO	0.15	GR	0.14	MY	0.14
HU	0.14	CR	0.14	IT	0.13
MD	0.13	BD	0.12	CZ	0.12
RS	0.12	UA	0.10	BR	0.10
TH	0.10	PL	0.08	TW	0.08
ES	0.08	FR	0.07	HK	0.06
BG	0.06	IL	0.05	ID	0.05
BA	0.04	EE	0.04	BE	0.04
IN	0.03	LV	0.03	AL	0.02
PT	0.02	NO	0.02	CH	0.02
GB	0.02	AE	0.02	CA	0.02
DK	0.02	PK	0.01	JP	0.01
SG	0.01	MA	0.01	SE	0.01
AU	0.00	RU	0.00	KR	0.00
US	0.00	GE	0.00	VN	0.00
TR	0.00	MX	0.00	CN	0.00
UY	0.00				

Table 8.6. HHI scores across countries for transits on the path (sorted by HHI). Countries with concentrated or highly concentrated markets are highlighted.

Country	HHI	Country	HHI	Country	HHI
AE	0.59	KZ	0.33	AL	0.28
BA	0.21	TR	0.20	ES	0.20
BD	0.20	TH	0.19	EE	0.19
PK	0.19	LV	0.17	MA	0.16
DK	0.16	RU	0.16	MX	0.16
PT	0.13	PY	0.12	FR	0.12
IN	0.12	HK	0.11	SG	0.11
AR	0.10	PL	0.10	IL	0.10
ID	0.10	UA	0.09	CL	0.09
KR	0.09	RS	0.09	CR	0.09
BO	0.08	SE	0.08	BG	0.08
VN	0.08	MD	0.07	US	0.07
UY	0.06	NO	0.06	HU	0.06
CN	0.05	RO	0.05	GR	0.05
CH	0.04	BE	0.04	CA	0.04
MY	0.04	BR	0.03	GB	0.03
NL	0.03	GE	0.03	CZ	0.03
DE	0.03	ZA	0.02	IT	0.02
TW	0.02	NZ	0.01	AU	0.01
JP	0.01				

and Kazakhstan (0.19)—indicating heavily centralized interconnection ecosystems. Germany's paths converge on DE-CIX Frankfurt, which carries over half of all observed traffic; New Zealand depends similarly on AKL-IX Auckland (51.3%). Argentina's government traffic is concentrated around AR-IX Cabase (47.5%), and Kazakhstan routes 44% of paths through KAZ-GOV-IX, its government-operated exchange. Together, these cases illustrate how limited domestic peering diversity can create structural choke points in national infrastructure.

In contrast, ten countries exhibit high HHI values for transit dependencies, reflecting heavy reliance on a limited set of upstream ASes: the United Arab Emirates (0.59), Kazakhstan (0.33), Albania (0.28), Bosnia and Herzegovina (0.21), Turkey (0.20), Spain (0.20), Bangladesh (0.20), Thailand (0.19), Estonia (0.19), and Pakistan (0.19).

In most of these cases, concentration stems from incumbent telecom providers—often former state monopolies that remain dominant in their national markets, whether privatized or still government-controlled. We identify these incumbents using telecom-market reports, regulatory filings, and prior studies of state-owned networks [**43**].

Figure 8.11 shows the share of government paths traversing each incumbent's network. Dependence is most pronounced in the United Arab Emirates, where Etisalat (AS8966) carries over 76% of paths, and in Kazakhstan, where JSC Kazakhtelecom (AS50482, AS9198) handles more than 70%. Bangladesh displays similar concentration, with over 70% of paths routed through Bangladesh Data Center Company Limited (AS141773) and the Bangladesh Telegraph and Telephone Board (AS17494, AS45588). Pakistan relies primarily on PTCL (AS17557) and Telenor (AS24499), both state-linked providers, while in Turkey, Turk Telekom (AS9121) carries 42% of paths and retains a government "golden share."

In other countries, concentration results not from a single dominant provider but from a few intermediaries with collectively large shares. Spain, for example, shows a diversified provider landscape, yet Entidad Pública Empresarial Red.es (AS766) alone carries over one-third of observed paths.

Global backbone providers are another source of path concentration. Arelion (formerly Telia, AS1299), for example, appears in 20% of paths in Albania and 19% in Estonia, illustrating how large international carriers can shape national routing even without direct domestic presence.

Countries with low HHI values (<0.1) across both transit and IXP dependencies show a more balanced distribution of interconnection paths and limited reliance on any single network entity. Mature markets such as the United States, Canada, the United Kingdom, Switzerland, Sweden, and Australia benefit from competitive telecom sectors and strong participation in neutral IXPs. Smaller economies like Belgium, Norway, Bulgaria, and Taiwan achieve similar diversity through open peering and active engagement in regional exchanges. In China and South Korea, regulatory frameworks that enforce competition or structural separation contribute to reduced dependence on individual providers. Meanwhile, Vietnam, Georgia, and Uruguay appear to gain resilience from growing regional integration and participation in global peering ecosystems. Together, these cases illustrate how a mix of market competition, regulatory oversight, and infrastructural openness can mitigate centralization risks.

Observed path consolidation to government services may, however, reflect more than market dominance. In some cases, it arises from infrastructural constraints—such as limited domestic peering or centralized international gateways. In others, it results from deliberate policy or procurement choices about where to host services and which providers to trust. Both underscore that structural concentration can be a byproduct of governance as much as geography.

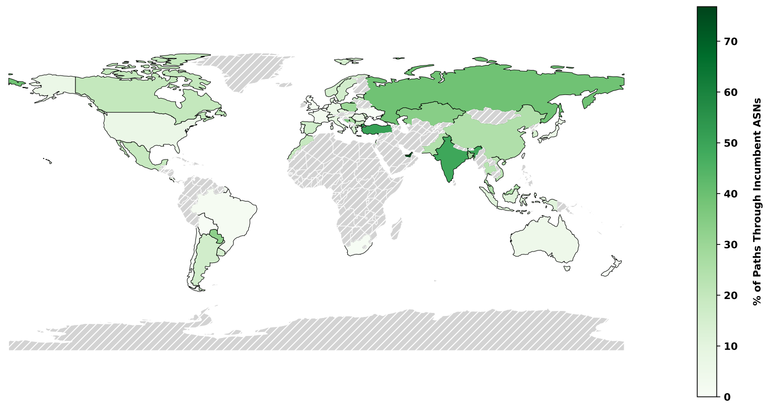


Figure 8.11. Share of government paths traversing incumbent transit ASes (peaks ¿70%).

Implications. Concentration and weak encryption act as complementary vulnerabilities. Countries where government traffic exits domestic jurisdiction and remains unencrypted expose citizens to extraterritorial surveillance. Where traffic is domestically contained but routed through a single operator, resilience and transparency are at risk. Mitigation thus requires parallel efforts: universal HTTPS enforcement to safeguard content, and diversified interconnection to preserve control.

Summary. Weak HTTPS adoption and routing exposure frequently co-occur, compounding surveillance and integrity risks when government traffic leaves national jurisdiction unprotected. At the same time, many countries exhibit heavy concentration on a small set of incumbents or IXPs, creating single-point dependencies even when traffic remains domestic. Sovereign resilience therefore depends jointly on encryption hygiene and interconnection diversity—not hosting location alone.

8.3.5. Limitations

Our analysis inherits limitations common to large-scale Internet measurement. While we apply conservative inference, validation, and careful interpretation, residual uncertainty remains due to routing opacity, incomplete coverage, and temporal scope.

Our inference relies on traceroute measurements, which are affected by ICMP filtering, per-flow load balancing, MPLS encapsulation, and routing dynamics. We mitigate these effects through conservative path stitching and aggregation across multiple vantage points, but transient behaviors and opaque hops may persist. Consequently, our estimates deliberately err toward underestimating, rather than overstating, jurisdictional exposure.

Our scope is limited to dependencies observable at the IP layer. Additional exposure may arise from lower-layer infrastructure, such as submarine cable routing or shared fiber ownership, that is not directly visible. Interconnection inference further depends on PeeringDB, a widely used but voluntary dataset; private or unlisted facilities may therefore be missed. As a result, observed IXP usage and physical-layer exposure should be interpreted as lower bounds.

Attribution accuracy depends on geolocation and vantage-point coverage. We mitigate geolocation errors through PTR-based validation and latency checks, discarding inconsistent mappings. Vantage-point coverage remains uneven across countries despite selecting diverse RIPE Atlas probes and limiting analysis to one probe per AS. We therefore restrict analysis to countries with sufficient coverage, though denser deployment would improve completeness.

Our measurements reflect routing behavior observed between March and May 2025 and capture steady-state configurations during that period. Short-lived policy changes, inactive failover paths, or responses to rare events may not be observed. Longitudinal measurements would be required to study routing dynamics under outages, policy shifts, or geopolitical disruptions.

Finally, our analysis is based on observed forwarding behavior. We do not infer contractual, economic, or regulatory drivers of routing, nor do we observe reverse paths, which may traverse additional intermediaries. Our results characterize technical exposure rather than operator intent.

Overall, these limitations stem from the Internet's inherent opacity and dynamism rather than experimental shortcomings. Our framework provides a conservative, lower-bound view of jurisdictional and structural dependencies in public-sector connectivity.

8.4. Conclusion

This chapter examined the exposure implications of consolidation across three layers. First, we compared service-endpoint exposure across structurally and strategically consolidated countries, showing that structural consolidation is associated with more uniform exposure outcomes while strategic consolidation is associated with greater cross-country variation. Second, we focused on government hosting and showed how cross-border dependencies persist in both provider registration and server location, with important regional differences. Third, we examined path-level exposure to government websites, showing that hosting domestically does not eliminate reliance on foreign transit and exchange infrastructure. Together, these results show that external dependence is not determined by concentration alone, but by how concentration is realized across services, deployments, and paths. The next chapter turns to performance implications, examining how consolidation shapes latency both through infrastructure placement and through content-delivery steering.

CHAPTER 9

Performance Implications of Consolidation

Performance is a third major implication of consolidation, but its effects cannot be understood from concentration alone. As argued throughout this dissertation, the same observed level of consolidation can reflect either structural constraints shared across an ecosystem or strategic deployment choices made by particular actors, and these two regimes need not produce the same user-facing outcomes. This chapter shows that performance is shaped through two linked mechanisms: directly, through where hosting, DNS, and certificate authority infrastructure are deployed; and indirectly, through how users are steered to content once requests enter opaque delivery systems. We first compare latency to key service endpoints under structural versus strategic consolidation, asking whether similar concentration levels impose different reachability constraints on domestic users. We then turn to content-delivery steering, showing that consolidation also matters through the control plane: CDN delivery increasingly depends on DNS-based replica selection, and as DNS resolution itself consolidates, resolver location becomes a weaker proxy for user location. Taken together, these analyses show that consolidation shapes performance not only by concentrating infrastructure, but by mediating who gets mapped where and with what latency consequences.

9.1. Latency in Structural vs Strategic Consolidation

Beyond exposure, consolidation can also shape end-user performance. Critically, the same level of concentration need not imply the same latency outcome: when consolidation is structural, latency may reflect shared ecosystem limits that constrain all sectors similarly, whereas when consolidation is strategic, it may reflect deployment or procurement choices that selectively improve or worsen domestic reachability.

Measuring latency to hosting, DNS, and CA endpoints complements the exposure analyses in Chapter 8 by capturing a direct user-facing consequence of consolidation: whether concentrated service markets translate into efficient domestic reachability, or instead impose performance penalties through longer paths and greater reliance on external networks. Framed through the structural/strategic distinction introduced in Chapter 6, this comparison asks whether performance is primarily governed by ecosystem-wide constraints that affect governments and commercial actors alike, or by deployment choices that allow governments to localize critical infrastructure and reduce latency for domestic users.

We analyze latency to three critical endpoint classes: hosting, DNS, and CA infrastructure, which directly influence page retrieval, name resolution, and TLS establishment.

For each endpoint IP, we issue three ICMP pings from in-country vantage points and use the minimum RTT per target; non-responsive targets are treated as loss and excluded. Overall, 76% of unique IPs respond.

Figure 9.1 shows ECDFs of RTT to hosting, DNS, and CA endpoints (log x axis), split by structural versus strategic consolidation; the dotted line marks the government median.

Structurally consolidated countries exhibit higher government latency (173.0 ms hosting, 155.0 ms DNS, 206.6 ms CA) with tight government–commercial overlap and ECDFs shifted right, whereas strategically consolidated countries show lower latency (121.0 ms hosting, 119.8 ms DNS, 117.5 ms CA) with curves shifted left relative to commercial.

Structural regimes exhibit tighter latency distributions, whereas strategic regimes show greater dispersion, indicating that performance under structural consolidation is largely governed by shared ecosystem constraints, while strategic deployments introduce wider variability.

These results provide the chapter's direct view of performance under consolidation. They show that similar levels of concentration can correspond to different latency and variability trade-offs depending on whether consolidation is structural or strategic, reinforcing the broader claim that observed concentration is ambiguous without a comparative baseline. Yet endpoint placement is only part of the story. Even when infrastructure is already deployed, user performance also depends on how requests are mapped to replicas in practice. We therefore next examine the indirect pathway through which consolidation shapes performance: content-delivery steering.

The endpoint results capture the performance consequences of consolidation at the level of infrastructure placement: where key services are located relative to domestic users. But user experience is not determined by placement alone. In practice, content

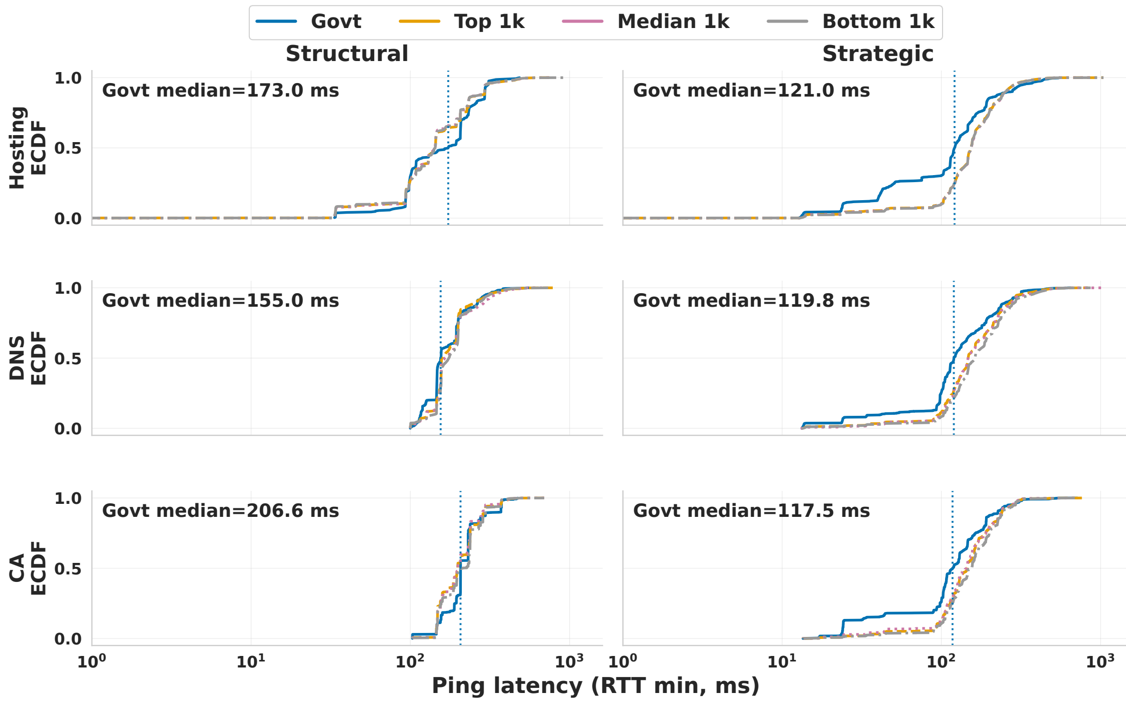


Figure 9.1. ECDF of minimum RTT (ping) to hosting, DNS, and CA endpoints from in-country vantage points. The dotted line marks the government median.

delivery also depends on how requests are mapped to specific replicas, often through steering decisions that are externally invisible. The next section therefore turns from the geography of infrastructure to the control logic that determines which infrastructure a user actually reaches.

9.2. Performance Through Content-Delivery Steering

The endpoint analysis above captures performance through the geographic and topological placement of infrastructure relative to users. A complementary mechanism is replica selection, through which content delivery networks (CDNs) choose which already-deployed server ultimately serves a user request [**48**, **41**, **128**]. Replica selection directly shapes latency, availability, and fault tolerance, but 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.

This lack of visibility is increasingly consequential under consolidation. Many CDNs rely on DNS-based redirection to map users to replicas [**159**, **222**], which makes steering behavior depend on a relatively small set of large, globally deployed DNS resolvers [**115**, **120**]. In other words, performance is shaped not only by consolidation in delivery infrastructure, but also by consolidation in the naming infrastructure that informs delivery decisions. As a result, replica selection may increasingly reflect resolver placement and policy rather than client location, particularly where regional infrastructure is sparse. 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 [**48**], IP anycast [**41**], and variants such as EDNS0 Client Subnet (ECS) [**55**] and regional anycast [**232**]. 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 [**138**, **159**, **75**, **40**, **7**, **97**, **113**, **202**, **210**, **167**, **183**, **1**, **83**, **222**]. 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.

To make this indirect performance pathway observable, 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, allowing us to study how opaque steering behavior mediates the performance consequences of consolidation.

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

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

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. This finding is central to the chapter's broader argument: if popular content is primarily delivered through DNS-based steering, then performance increasingly depends on the quality and location of the DNS infrastructure that stands between users and replicas. We therefore in the next section examine the user-facing DNS ecosystem itself, asking how resolver consolidation may weaken the very signal on which this dominant steering model depends.

This prevalence makes DNS-based steering more than a CDN design choice; it makes it a dependency on the broader DNS ecosystem. If resolver location is used to approximate client location, then consolidation among resolvers can directly weaken steering quality even when CDN deployments remain unchanged. The next section examines that dependency explicitly by analyzing how modern DNS resolution increasingly relies on a concentrated set of third-party providers.

9.2.1. Finding CDN's Redirection Technique

We now describe a methodology to experimentally identify a CDN's redirection technique. The methodology uses RIPE Atlas probes [**20**] as clients and a set of strategically selected DNS resolvers. These clients resolve CDN hosted resources, using the different DNS resolvers, and collect latencies to the assigned replicas. The CDN replica selection approach is identified based on the relative differences between the collected latency distributions.

As a first step, clients perform resolutions of a CDN hosted resource in a given locale, using a set of five (5) selected DNS recursive resolvers. These five resolvers cover different geographic *scopes*, namely: (i) the same metro area as that of the client, (ii) a different metro area in the same country as the client, (iii) a different country within the same region as the client, (iv) neighboring region of the client, and a (v) a non-neighboring region (Fig. 9.2).

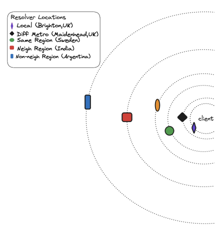


Figure 9.2. DNS resolvers at different distances from the client.

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

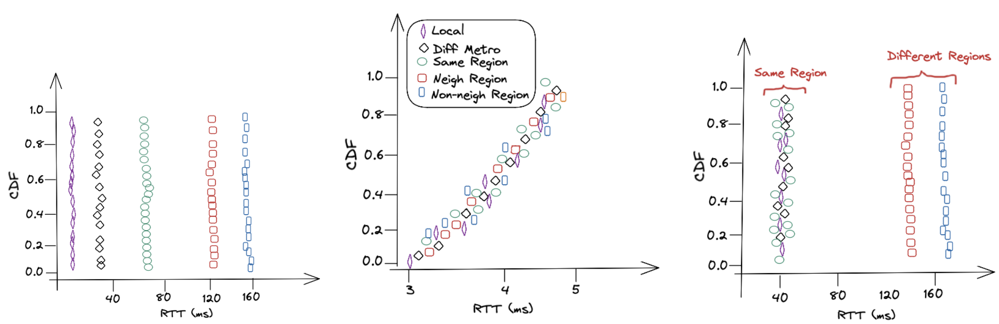
Figure 9.3. Internet Population represented by our vantage points.

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

This process is repeated for a large set of resources hosted on the target CDN. Clients collect latencies to the assigned CDN replicas with different DNS resolvers. The resulting latency distributions, one per DNS resolver scope, are subsequently used to infer the CDN redirection technique.

9.2.1.1. Canonical Examples for Illustration. Intuitively, if the replica selection approach of a CDN takes into account the specific metro location of a client (Brighton in Fig.9.2), the set of replicas it assigns should be different from (and closer than) what it would assign if the client's DNS resolver were located in a different country (Sweden) or a different region (India).

On the other hand, if the CDN's replica selection approach does not differentiate between clients in different regions, then the set of replicas it assigns when using a client's local DNS resolver, and a DNS resolver in any other neighboring or non-neighboring region (Argentina) should be similar.



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

Figure 9.4. Expected CDF sets for DNS-based ($CRV[i] \approx 1$ for all i), Anycast ($CRV[i] \approx 0$) and Regional-Anycast replica selection ($CRV[1] \approx 0$, $CRV[3] \approx 1$).

Before detailing our approach, it is useful to visualize a potential set of latency distributions to the replicas assigned by prototypical versions of replica selection approaches. Figure 9.4 show these latency distributions for DNS-based, Anycast and Regional-anycast.

If a CDN uses DNS-based replica selection, the set of replicas assigned by it should be further away from the client as the DNS resolver used also moves away. Correspondingly, the distributions of latencies to the assigned replicas for these resolvers should be spaced and away from the client (i.e., higher RTTs). Figure 9.4(a) illustrates this case, with five distributions of latencies to assigned replicas using the five different DNS resolvers (local, different metro, etc).

When a CDN uses Anycast for replica selection, the location of the client's DNS resolver is not a factor for the selection. Consequently, the distribution of latencies to replicas assigned using the different resolvers should largely overlap. Figure 9.4(b) illustrates this case.

Last, if a CDN uses regional anycast for replica selection, we expect the distributions of latencies to assigned replicas within the same anycast region to be similar, but to the left of (i.e., shorter than) those associated with DNS resolvers in different anycast regions. Figure 9.4(c) shows this case, with the latency distributions of the first three DNS resolver locations – local, in a different metro, and in a different country of the same region – clustered together, while those corresponding to different – neighboring and non-neighboring region – set clearly apart.

9.2.1.2. Identifying Redirection Techniques. To identify three of the main CDN redirection approaches – DNS-based, Regional Anycast and Anycast – we compare the distribution of latencies to the replicas assigned with different resolvers.

We capture the distance between the latency distributions in the *coefficient of Regionalization Vector (CRV)*. 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. For instance, $CRV[0]$ captures the distance between the latency distributions of the local resolver and a resolver in a different metro area, while $CRV[3]$ measures the distance between the local resolver and a resolver in a non-neighboring region.

To compute the distance between two latency distributions we use the Kolmogorov–Smirnov (KS) distance, the maximum difference being taken over all values of x of the absolute differences between two latency distributions F and G ($D = max|F(x) - G(x)|$). [2] Using the cumulative distribution function (CDF) of latencies, each $CRV[i]$ ranges between 0 and 1, being equal to 0 when both distributions are identical and 1 when they are completely different.

We rely on CRV to identify three of the main CDN redirection approaches: DNS-based, Regional Anycast and Anycast. For the prototypical cases of DNS-based replica selection, we expect that at least $CRV[1]$ and $CRV[3] \approx 1$. Other entries in the CDN's CRV could range between 0 and 1 depending on the scale of deployment of the particular CDN. For instance, the CDFs of a CDN like Akamai using DNS-based replica selection with a

[2]We choose KS distance as it does not assume any specific distribution, making it particularly useful for latency data, which often do not follow a normal distribution.

large-scale deployment of over 170,000 servers in 1,300 ISPs [233], should look similar to our prototypical case (Fig. 9.4(a)) and have $CRV[i] \approx 1$ for all i. The Anycast-based replica selection approach, on the other hand, should result on all CDFs overlapping, stretching on the latency range from the closest to the furthest away replica, and a $CRV[i] \approx 0$.

Finally, if the CDN uses Regional-anycast for replica selection we would expect the CDFs of latencies to replicas in the same region to be the same, $CRV[1] \approx 0$, and different from the CDFs of latencies to replicas associated with non-neighboring regions, $CRV[3] \approx 1$. Note that the regional divisions used by the measurement and the underlying infrastructure of the CDNs may not match exactly, as the CDN's regions are determined by factors such as the underlying infrastructure and Internet penetration. For instance, while some CDNs differentiate between northern and southern Europe, most CDNs treat all of South America as a single region. This mismatch is however mitigated by our selection of multiple resolver scopes, allowing the system to account for varying regional divisions, while capturing the overall behavior of each system.

We choose to use the distances between latency distributions to the assigned replicas to identify redirection approaches, instead of comparing the set of IP prefix of these replicas. While the latter approach might initially seem simpler, its complexity quickly escalates when addressing necessary details such as determining the right prefix length to use and establishing the correct similarity thresholds for identification. Additionally, although techniques like MAnycast2 [196] can determine whether the IP prefixes used by CDNs are Anycast, they do not provide insights into the redirection model employed by a CDN in a given country.

9.2.1.3. Checking for ECS Support. Besides relying on one (or more) of the above replica selection approaches, a CDN may support ECS to make per-client decisions and avoid the client-LDNS mismatch problem. For ECS to function, both the client's LDNS and the CDN must implement the ECS specification. Calder et al. [**40**] examines ECS adoption in LDNSes using DNS queries captured over one month from Microsoft's Azure Cloud platform.

We detect support for ECS by a CDN by issuing ECS resolution queries for its domains and the CNAME chain, using the Google Public DNS resolver – known to support ECS – and the nameserver of the domain. We check for EDNS scope greater than zero to determine if the domains supports ECS. We issue three queries and compare the assigned CDN servers with two distant subnets. If the two vectors do not match, and include responses aligning with the geolocation of the provided subnet, as indicated by geolocation databases, it suggests that the given CDN in fact uses the eDNS0 subnet extension for replica assignment.

9.2.2. Detecting CDN Replica Selection

We now describe the application of our methodology to experimentally identify the replica selection approaches used by CDNs to deliver Web content around the world. In the following section (§9.2.3) we validate our approach contrasting the known approaches used by three large CDNs with what our methodology identifies.

We follow the United Nations geo scheme [**223**] and rely on published statistics on Internet penetration [**199**] to select vantage points for our study. We ensure that the set of countries included in our analysis (*i*) capture a sufficiently large fraction of the Internet user population, and (*ii*) host VPN vantage points whose locations can be verified.

We place vantage points in 19 countries that span all inhabited continents and together capture 66% of the world's Internet user population. Table 9.3 shows the regions and their corresponding Internet user population captured by the countries included in our study. For each continent, the selected set of countries account for $\geq$ 50% (and up to $\approx$ 89%) of the continent's Internet population.

For each of these countries, we collect the top 1,000 sites based on the Google CrUX dataset [**86**]. These popular sites should serve as a good proxy of the most commonly accessed content by users in each country, while the 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 [**152**, **94**, **203**] 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 obtain the set of CDNs from our self-populated CNAME-CDN map [**50**, **228**]. Additionally, for validation and to identify the CDNs hosting resources without a CNAME redirect, we compare the autonomous system number (ASN) of the resource with those of popular CDNs [**140**, **228**]. We validate the claimed locations of our VPN providers, by geolocating the VPN vantage points' IPs using two popular geolocation databases: MaxMind GeoLite2 [**141**] and

IP2LocationBD11.Lite [**109**]. Past work has shown that geolocation databases are reliable at the country level [**168**].

We use RIPE Atlas probes as proxies of clients in each of the selected countries. Next, we select resolvers in each geographic scope that our client will use to resolve the CDN hosted resources. For the local resolver scope, we use the client probe's resolver and for the resolvers in other geographic scopes, we select Regional DNS resolvers [**67**] in each of those scopes. [3] We validate the locations of the selected DNS resolvers following a similar approach as with the VPN servers. 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.

9.2.3. Known CDNs for Validation

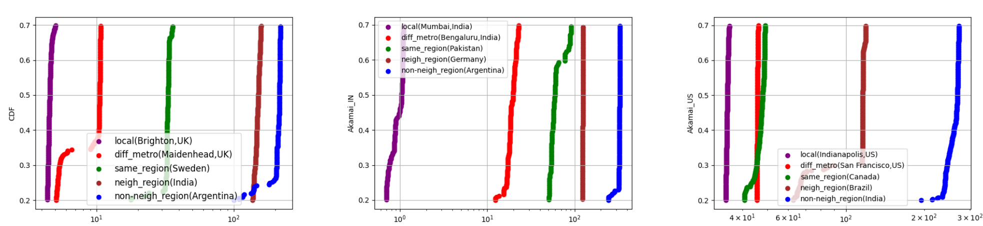
Before presenting our analysis of CDNs' replica selection approaches around the world, we confirm the findings of our methodology when applied to a set of large CDNs – Akamai, Cloudflare, and Edgio – with known replica selection approaches. We confirm the redirection models used by these CDNs by checking their websites, previously published works, personal communication, and third-party websites such as CDN Planet [**163**]. We

[3] A sensitivity analysis using various combinations of cities within the client's vantage country, different metropolitan areas and countries for DNS resolvers, and different DNS resolvers within each geographical scope, revealed no variations in the identified redirection models.

observe that the redirection models we infer from our analysis coincide with those reported by these CDN providers.

9.2.3.1. Replica Selection by Akamai. We focus first on the replica selection approach identified by our methodology for Akamai. We run our data collection and analysis using all collected resources hosted by this CDN, and computed the CRV vectors for each country.

Figure 9.5 shows the CDFs of latencies to the set of Akamai replicas assigned for each resolver scope in the UK, India, and the US. We plot the middle percentile for all CDNs across different countries as the tail may contain edge cases, including non-cacheable content (retrieved from the origin). The clean, non-overlapping CDFs at most resolver scopes resembles that of our prototypical DNS-based approach of Fig. 9.4(a). Correspondingly, we find values for $CRV[1]$ and $CRV[3]$, ranging between 0.9 to 1.0, across these countries as shown in Fig. 9.8(a).



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

Figure 9.5. Akamai using DNS-based redirection technique in India, United Kingdom, and the US.

Our analysis clearly shows that Akamai, as reported [**48**, **154**, **162**], relies on DNS-based replica selection across all countries in our dataset. In addition, the Akamai CDN supports ECS for replica selection as we show in Sec.9.2.4.6.

The CDFs associated with resolvers in the *different metro* and *same region* in the US are interesting, showing the occasional mismatch between infrastructure and geographic scopes (Fig. 9.5(c)). With a vantage point in Indianapolis, US, the replicas assigned with a resolver in Canada are at a similar latency than those assigned with a resolver in San Francisco, the same country but more than 2,200 mi away.

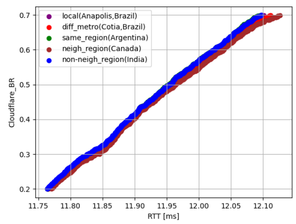
9.2.3.2. Replica Selection by Cloudflare. The second CDN we use for validation is Cloudflare. Cloudflare is known to rely on anycast for replica selection [**53**, **164**]. As in the case of Akamai, we run our data collection and analysis using all collected resources hosted by Cloudflare and computed the CRV vectors for each country.

Figure 9.6 plots the latency CDFs to the replicas assigned for Cloudflare-hosted resources in Brazil, the UK, and India. Across the three countries, the figure shows similar latencies to replicas obtained at each resolver scope with overlapping CDFs. The three figures are nearly equivalent to those of our prototypical Anycast case (Fig. 9.4(b)).

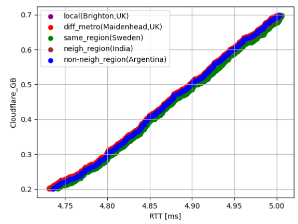
Figure 9.8(b) shows the $CRV[1]$ and $CRV[3]$ values for Cloudflare in the same countries, each with the expected values ≈0. Our analysis confirms Cloudflare's use of anycast for replica selection.

9.2.3.3. Replica Selection by Edgio. Last we examine Edgio, one of the few CDNs using regional anycast for replica selection [**142**, **165**].[4]

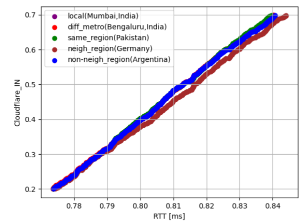
[4]EdgeCast (AS15133) and Limelight (AS22822) became Edgio in 2022; we focus on measurements of the EdgeCast Network.



(a) Cloudflare-Brazil



(b) Cloudflare-United Kingdom

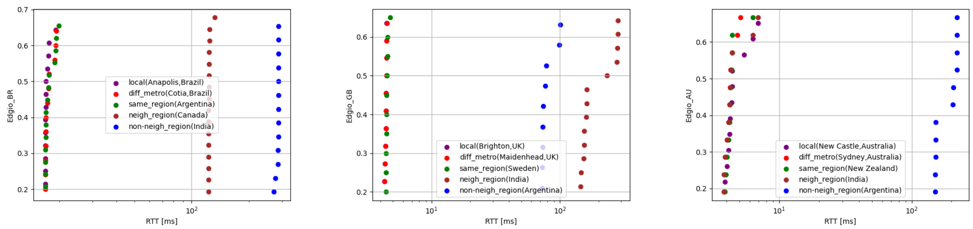


(c) Cloudflare-India

Figure 9.6. Cloudflare using Anycast redirection technique in Brazil, United Kingdom, and India

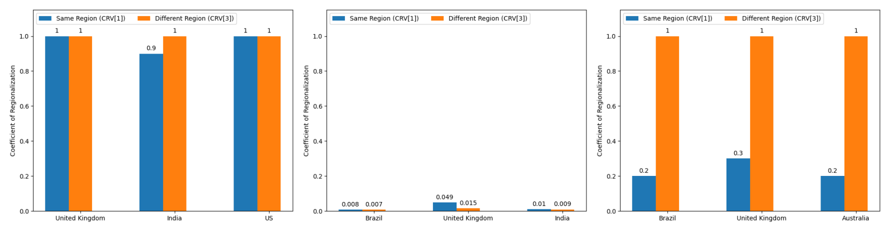
Figure 9.7 shows CDFs of latencies to Edgio replicas from clients in Brazil, the UK, and Australia. The figure shows similar latencies for all resolver locations within the same region and non-overlapping CDFs for resolver scopes in a different, non-neighboring region. Figure 9.7(c) for Australia illustrates an interesting case of mismatch between the geographic divisions and the underlying CDN infrastructure. The CDFs associated with a DNS resolver from the same country, but different metro (Sydney) overlaps with the one associated with a DNS resolver in a neighboring region (India). Further examining the responses revealed that for 81% of the Edgio objects we measure, the neighboring region responses received the same IP address as the different metro, suggesting the responses are part of the same anycast infrastructure region. Personal communication with operators confirmed our findings in this case.

Figure 9.8(c) shows the $CRV[1]$ and $CRV[3]$ for Edgio in these three countries. As expected, $CRV[1]$ has a lower value ranging between 0.2 to 0.3, and $CRV[3]$ has a higher value of 1, confirming that, as reported, Edgio uses Regional Anycast as its CDN replica selection approach.



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

Figure 9.7. Edgio using Regional Anycast redirection technique in Brazil, United Kingdom, and Australia



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

Figure 9.8. Coefficient of regionalization for same ($CRV[1]$) and different regions ($CRV[3]$) for three larger CDNS.

9.2.4. Global Analysis of CDN Replica Selection

After confirming the findings of our methodology with those advertised by the CDNs, we now turn to the remaining CDN providers observed in our scan of each measured country's top sites. We analyze the geographical trends in CDN redirection techniques used around the world and report the redirection technique predominantly used by the CDNs in our dataset, based on our measurements. We close the section reporting on our analysis of ECS support.

9.2.4.1. World CDNs Replica Selection. We identify the replica selection approaches of 14 other CDNs around the world, including large CDNs such as Amazon CloudFront, and Fastly, as well as regional providers such as Azion in Brazil, NGENIX in Russia, and Taobao and Tencent in China. The Google CDN presents an interesting challenge and we discuss it separately. Due to space restrictions, we focus our detailed discussion on a subset of these CDNs.

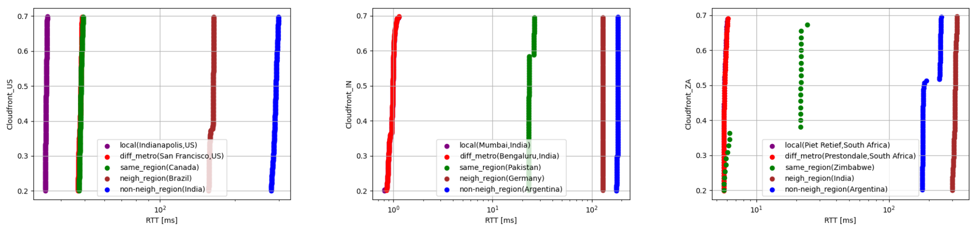
9.2.4.2. Global CDNs. Amazon CloudFront is a global CDN operated by Amazon Web Services with servers located in Western Europe, Asia, Australia, South America, and Africa, and several major metro areas in the US. According to the company site, CloudFront uses a global network of over 450 PoPs in more than 49 countries.

Figure 9.9 shows CDFs of latencies to Amazon CloudFront servers assigned at each resolver scope in the US, India and South Africa. Although different in scale than Akamai, the CDFs for different resolver scopes are mostly non-overlapping indicating the use of DNS-based replica selection. The overlap in the narrower resolver scopes (local and different metro in Russia, and the metro and same region scopes from the US vantage point) indicates the difference in scale: both are directed to the same or similar replicas.

The South Africa CDF for same region (Zimbabwe) is split in two largely vertical but separate lines, where nearly 40% of the selected replicas are in the same latency range as those assigned with DNS resolvers in other metro areas within South Africa, indicating domain dependent behavior for the region.

To confirm our identification of Amazon CloudFront's replica selection approach, we look at the corresponding CRV. CloudFront shows high values for $CRV[1]$ and $CRV[3]$, ranging between 0.8 to 1, in all three countries (Fig. 9.11(a)).

Furthermore, we also checked for the presence of prefixes associated with Amazon's services beyond CloudFront [9], aiming to investigate potential variations in redirection models employed by Amazon across different services and applications. However, we found minimal number of such prefixes in our collected data.



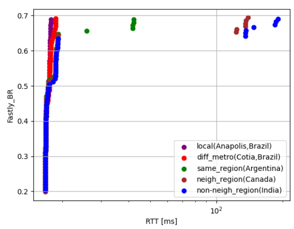
(a) Amazon-US (b) Amazon-India (c) Amazon-South Africa

Figure 9.9. Cloudfront matches DNS-based redirection technique, as shown in the US, India and South Africa.

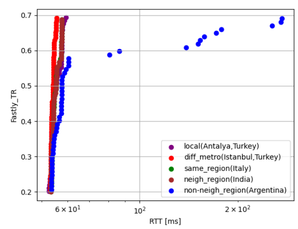
Fastly is another global CDN, with ≈ 80 PoP on 6 continents, known as a real-time CDN it offers a range of other services including streaming media and private CDN.

Figure 9.10 shows the different CDFs of latencies for all DNS resolvers' scope and clients located in Brazil, Turkey and South Africa. It presents, predominantly, the patterns of Anycast replica selection with a large overlap of all CDFs, across DNS resolver scopes and relatively low coefficients of regionalization. We see a few Fastly responses showing DNS-based replica selection, although with a deployment scale significantly smaller than Akamai's. These modes may reflect different customers or configurations.

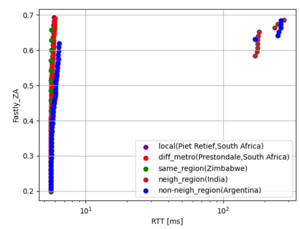
Fastly's $CRV[1]$ and $CRV[3]$ range between 0.2 to 0.4, across these countries, as shown in Fig. 9.11(c). We observe that the coefficient of regionalization in the case of Fastly is much lower than Akamai but not as low as Cloudflare.



(a) Fastly-Brazil

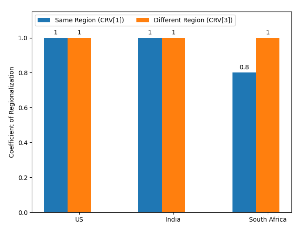


(b) Fastly-Turkey

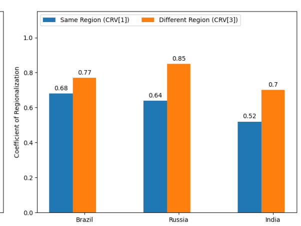


(c) Fastly-South Africa

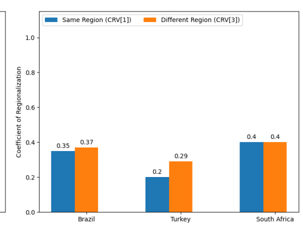
Figure 9.10. Fastly predominantly matches Anycast redirection, as shown in Brazil, South Africa and Turkey.



(a) CloudFront



(b) Google



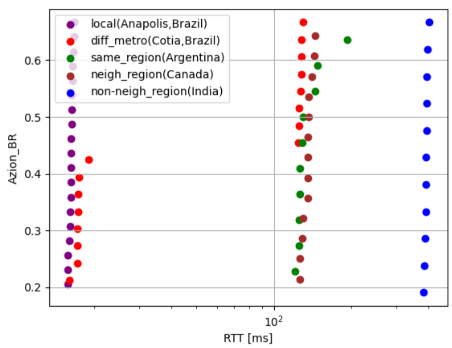
(c) Fastly

Figure 9.11. Coefficient of regionalization for same ($CRV[1]$) and different regions ($CRV[3]$) for other CDNs.

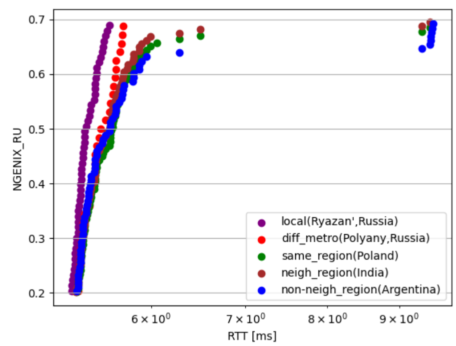
9.2.4.3. Regional CDNs. We focus now on regional CDNs, targeting small, perhaps region or country-specific, markets with correspondingly smaller presence in the top-ranked websites. While not the goal of this work, a regional focused study with larger number of websites in the region could reveal other small, regional CDNs and/or customer-specific replica selection approaches.

Figure 9.12 shows CDFs for two of the regional CDNs covered in our study: Azion in Brazil and NGENIX in Russia. The latency CDFs for Azion shows the clear separation of DNS-based replica selection, and with $CRV[1]$ and $CRV[3]$(≈1).

NGENIX, on the other hand, predominantly matches the Anycast model with overlapping CDFs across resolver scopes (for most hosted content) and low values of $CRV[1]$ and $CRV[3]$(≈0.3).



(a) Azion-Brazil



(b) NGENIX-Russia

Figure 9.12. Redirection techniques used by regional CDNs.

9.2.4.4. The Challenge of Mixed Approach. The Google CDN presents a particular challenge to our approach. Figure 9.13 shows the CDFs of latencies to Google replicas in Brazil, Russia and India, for different resolver scopes. These CDFs show two distinct patterns, where a portion of responses (e.g., ≈ 50% in Brazil) shows the pattern of Anycast-based replica selection while another segment shows a clear DNS-based pattern.

Correspondingly, the coefficient of regionalization between the resolver scopes show a mid-range of values, with $CRV[1]$ and $CRV[3]$ ranging between 0.52 and 0.85 for these countries (Fig. 9.11(b)).

The graphs and the associated coefficients of regionalization suggest that Google adopts a mixed approach, depending on the customer or resource. Separating resources based on the approaches used shows a clear distinction between Anycast (Fig. 9.14) and

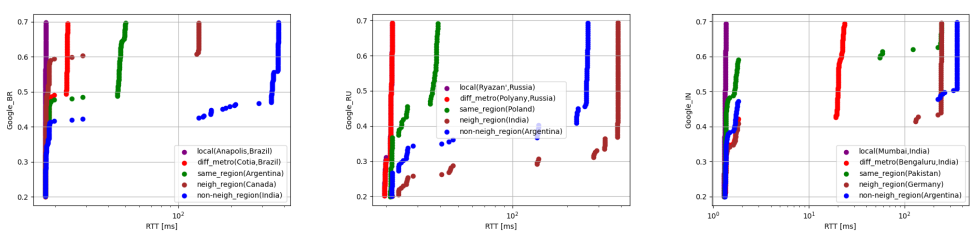
Country	DNS(%)	Anycast(%)
France	43.8	56.1
Germany	42.1	57.9
Russia	63.7	36.3
Spain	63.9	36.1
Turkey	58.1	41.9
United Kingdom	44.1	55.9
US	45.4	54.6
Argentina	47.4	52.6
Brazil	53.3	46.7
China	46.4	53.6
India	45.3	54.7
Indonesia	62.9	37.1
United Arab Emirates	60.2	39.8
Australia	42.4	57.6
Algeria	52.0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

Table 9.1. % of Google resources that match DNS and Anycast.

DNS-based replica selection (Fig. 9.15). Each of these sub-views clearly follow the expected patterns and the corresponding coefficients of regionalization are similarly consistent, with $CRV[1]$ and $CRV[3]$ of ≈0 for the Anycast-based case and $CRV[1]$ and $CRV[3]$ of 1 for DNS-based replica selection.

For a more detailed view of the redirection models used by Google, we calculated the percentage of Google resources that are assigned CDN replicas based on the Anycast or DNS-based models for every country and region in our study. Table 9.1 shows that there is an approximately equal split between Google content served using the DNS-based redirection model and the Anycast-based redirection model. Looking at resource domains, we find that the key services of the Google parent company, such as google.com, youtube.com, googleadservices.com and googlesyndication.com rely on DNS-based replica selection. Whereas, apparently external domains, such as snapchat.com and chess.com, are served using the anycast model.

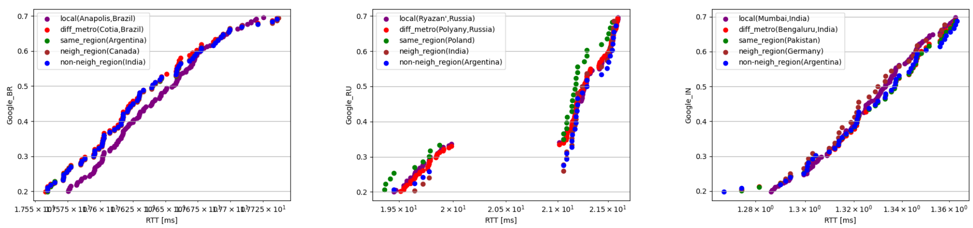
Additionally, we were also curious if Google uses different redirection techniques for different applications. Since Google publicly announces the prefixes it uses for the cloud applications [87], we used them to examine the Google cloud replicas. We found that the CDN redirection model Google uses for its cloud applications is indeed exclusively anycast with $CRV[1]$ and $CRV[3]$ of ≈ 0 , whereas for the native applications it uses a mixed approach as shown in Fig. 9.13.



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

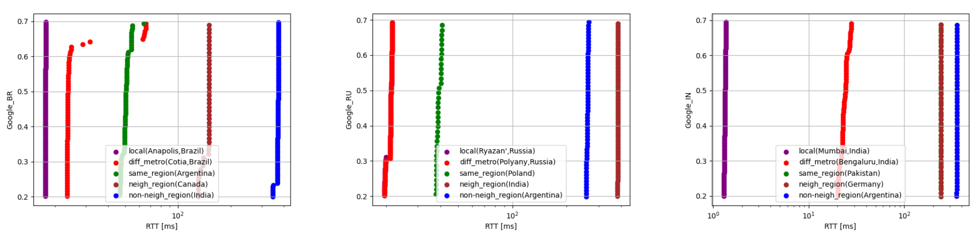
Figure 9.13. Google uses a combination of redirection techniques that matches the DNS-based for some content and Anycast redirection model for other, as shown in Brazil, Russia, and India

9.2.4.5. Geographic Trends. We close our analysis of main redirection approaches by looking at at their adoption by CDNs across countries and regions in our dataset. Table 9.2 lists the countries, the percentage of world Internet users they represent, and the percentage of resources in each country that match each redirection technique. The table also aggregates the resources from the countries in each region to show the percentage



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

Figure 9.14. Google content using Anycast redirection, as shown in Brazil, Russia, and India



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

Figure 9.15. Google content using DNS-based redirection, as shown in Brazil, Russia, and India

of resources delivered using each redirection technique, along with the percentage of Internet users potentially affected per region. For instance, for the US (highlighted) 36.8% of resources use DNS-based redirection model, 32.8% use a combination of DNS-based and Anycast, 23.8% use Anycast, and only 6.5% use Regional Anycast.

It shows that in the US and Australia, the majority of resources use DNS-based redirection model, while in much of Europe and China, the percentage of resources using

DNS and Anycast redirection models are roughly similar. In the case of all African countries, as well as Russia, Turkey, India, Indonesia, and UAE, a significant percentage of resources rely on Anycast.

Resources, however, are not all of equal size. Table 9.3 shows the percentage of the bytes of content in each country that use each redirection technique. We calculate this by aggregating the total bytes of each resource in the set. Here, interestingly we observe that most content from the countries we measure worldwide (40.9%), and more specifically in North America(34.7%), South America(35.5%), Africa(56.7%) and Oceania(40.1%) uses DNS-based redirection model.

Differences in the percentage of content served using a given replica selection approach may be attributed to multiple factors, many of them interrelated. Every CDN relies, predominantly, on one replica selection approach and captures different market shares in different regions. Large geographic areas with rich infrastructure deployment would probably lean towards a different replica selection approach than regions with smaller infrastructure footprints and/or fewer users.

9.2.4.6. ECS Support. ECS-based replica selection support is, in a sense, orthogonal to the above approaches in that a given CDN could adopt regional anycast as it primary replication approach while, at the same time, respond to clients' ECS requests. For every CDN in our dataset, we also report the number of domains hosted on the CDN that support ECS, i.e. they respond to an ECS-equipped query with a response scope greater than 0. We also check if those domains use ECS when mapping clients, i.e. give a different response based on the client subnet provided in the query.

Location	Internet Users(%)	DNS-based(%)	Anycast(%)	Regional Anycast(%)	Combination(%)
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 9.2. Percentage of resources using a particular redirection approach. For instance, 36.8% resources in the US use DNS-based, while 32.8% use a combination of DNS-based and Anycast.

Most CDNs in our study have over 80% of the domains that support ECS. Given the client subnets of two distant regions, all CDNs except for the 7 ones shown in Table 9.4 give different responses corresponding to the provided subnets, for at least a subset of domains. We confirm the non-random nature of this behavior by verifying that the geolocation of the response matches that of the client subnet provided.

9.2.4.7. **CDNs and Their Predominant Redirection Model**. Table 9.4 lists the set of CDNs found in our dataset and the predominant redirection model they match based on our experiments. We see that 12 out of the 17 CDNs in our dataset match the DNS-based

Location	Internet Users(%)	DNS-based(%)	Anycast(%)	Regional Anycast(%)	Combination(%)
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 9.3. Percentage of total bytes of the resources in each country, region, and worldwide that use the corresponding redirection techniques.

redirection model. Cloudflare, Stackpath, Fastly, and NGENIX predominantly match the Anycast redirection model. Edgio and Yahoo are the only two CDNs that match Regional Anycast. Google matches DNS-based for some customers and Anycast for others, whereas for Fastly we see a only few responses showing DNS-based replica selection. We note that some CDNs may use different redirection techniques in different regions. Since some CDNs did not have sufficient resources in some regions for a measurement analysis,

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

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 9.4. The predominant model CDNs match based on our measurements[5]and if they respond to ECS.

as a result the redirection models we observe are for the regions we measured. Finally, we find that 10 out of the 17 CDNs in our dataset use ECS.

9.2.5. 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.

9.2.5.1. 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. 9.9(b)), likely reflecting limited infrastructure [**224**]. 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 [**196**, **41**] 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.

9.2.5.2. 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.

9.3. DNS Resolver Consolidation

Read together with the previous section, this creates a cross-layer chain: consolidation in content delivery makes steering mechanisms performance-critical, and the dominance of DNS-based steering makes performance increasingly sensitive to consolidation in DNS resolution. This is precisely the kind of indirect consequence that would be missed if consolidation were studied one layer at a time.

After showing that DNS-based replica selection is widely used, we next examine consolidation on the user-facing side of DNS resolution. This shift is essential to the chapter's indirect performance argument. DNS-based steering assumes that resolver location is a meaningful proxy for user location, but as resolver infrastructure consolidates onto a small set of third-party providers, that proxy can weaken. The result is a cross-layer performance penalty: concentration in DNS resolution can degrade the quality of CDN steering, thereby increasing the latency users experience when accessing content.

DNS, originally designed to map human-readable names to network addresses, has evolved into a cornerstone of modern Internet infrastructure, supporting not only naming but also scalability and security [**145**]. Today, DNS is a key determinant of user experience and a major observation point for user activity. It directly affects performance because loading a single website often requires many DNS resolutions [**33, 30, 31**]. It also affects performance indirectly because many CDNs still rely on DNS for replica selection, assuming that a client's resolver location is a good proxy for the client's own location [**95, 39**].

Traditionally, the process of DNS resolution starts with a stub resolver that queries a pre-configured recursive resolver, which then retrieves the answer by querying one or more authoritative DNS servers. As DNS has taken on an increasingly critical role, prior work has shown that the once simple, textbook model of DNS resolvers has also evolved into a complex infrastructure. This includes ingress or forwarding resolvers, hidden resolvers, and egress resolvers, sometimes organized in cooperating pools [**186**, **4**].

Despite this complexity, a common simplifying assumption persists: that the client-side DNS path is effectively managed by a single organization, typically the client's ISP or a third-party DNS provider. In practice, however, resolution may traverse multiple resolvers operated by different organizations, creating hidden dependencies and potential mismatch between users and the infrastructure making decisions on their behalf.

ISP-provided resolvers are not always optimal in terms of resolution time [**2**, **182**], and slower DNS lookups can materially degrade user experience. They can also introduce reliability, privacy, and censorship concerns [**180**, **29**, **130**]. In response, a third-party DNS ecosystem has grown over time, but it is operated by only a handful of providers such as Google, Cloudflare, and IBM, reinforcing broader trends toward service consolidation and centralization [**193**, **100**, **220**, **146**].

Earlier studies have highlighted consolidation in public DNS services [**99**, **171**] and DNS traffic market share [**147**]. However, the modern DNS resolution process is more complex, often involving multiple recursive or forwarding resolvers, increasingly managed by third-party providers. In this chapter, we investigate the degree to which end-users, either directly or through their ISPs, depend on third-party DNS recursive resolvers and the resulting consolidation of this service. Within the dissertation's broader framework, the goal is not only to document centralization in DNS as a standalone phenomenon, but to show how consolidation in one infrastructure layer can propagate performance consequences into another.

Using all available RIPE Atlas probes we conducted a large-scale measurement campaign to capture clients' ingress and egress DNS resolvers from 803 ISPs around the world. We also complemented our dataset with a crowdsourced experiment that added 243 users in 77 ISPs. The extended dataset comprised 880 ISPs in 113 countries around the world. Our analysis extends beyond the multilayer structure of DNS resolution to consider additional complexities, such as resolvers located in different networks and owned by separate organizations. We explore various dimensions of this mismatch, including the physical distance between clients and their ingress/egress resolvers and cases where resolvers at the country level differ from the clients' location.

Our findings reveal that approximately 47% of clients use an egress resolver outside the autonomous system (AS) of their configured DNS resolver, and nearly all of these egress resolvers (97%) belong to third-party providers. We also observed a long tail of third-party ingress resolvers, with 145 providers in total. Among these, Google emerged as the dominant provider, accounting for 5.6% of ingress resolvers, followed by Cloudflare at 3.3%. An even higher consolidation trend was evident for third-party egress resolvers, where 489 distinct providers were identified, with Google leading at 19.3% and Cloudflare following at 13.1%. These findings highlight the growing centralization of DNS resolution services.

9.3.1. Methodology

For our analysis we select all RIPE Atlas probes [**20**] and expand the AS coverage leveraging Amazon Mechanical Turk (AMT) [**10**] and a tool we created to capture ingress/egress resolvers of a volunteer.

Clients resolve a subdomain whose authoritative server we control. For each client, we formulate a unique subdomain that encapsulates the information about the client. This subdomain includes the IP address, country, and ID of the client and the timestamp of the request. We obtain all these fields of the client probe from the RIPE Atlas API. In the case of crowdsourced measurements, we ask the client to enter their ingress resolver's IP on our tool's website by running a simple command, which is then embedded within the above URL as well. For privacy reasons, we embed the /24 prefix of the client's IP in our subdomain.

Upon receiving the query at our authoritative nameserver, we record the query including the client's IP, the ingress resolver's, and the IP of the egress resolver that directly contacts our nameserver. Subsequently, we use RouteViews BGP dumps [**181**] to find the AS of the client's IP, the ingress resolver's IP, and the egress resolver's IP for each client. This information is then used to map each AS to its organization using the IPInfo API [**106**] and report if there is any organizational level mismatch between the three entities.

Finally, we examine how often DNS resolutions take place in a country different than the client's country. For this, we leverage IPMap [**71**] geolocation. To gain confidence, we also use commercial IP geolocation databases [**106**, **141**] which are known to be accurate at the country level [117]. We consider a resolver to be in a different country than the client if both geolocation methods agree.

9.3.2. Dataset

In this section, we briefly describe the dataset we collected using both RIPE Atlas and a crowdsourced experiment run using Amazon Mechanical Turk.

We use a total set of 12,000 connected RIPE Atlas probes and resolve the sub-domain that we control. We were able to identify the AS of all relevant entities for more than 10,000 instances, i.e. unique (client,ingress and egress) tuples and the organization of the client, ingress resolver, and egress resolver for 5,331 probes.

We use this dataset to understand the percentage of probes that use different third-party egress resolvers as opposed to the ingress resolvers selected by the client. We further investigate their probable location and their potential implications. Our Amazon Mechanical Turk experiment added 262 users in 77 ISPs and 80 additional ASes not represented by RIPE Atlas in 7 countries.

Overall, our dataset includes the perspective of 10,432 clients in 880 ISPs, spread over 113 countries around the world.

In addition to information on the client side of DNS, we use RIPE Atlas to gather latency measurements (3 pings) to probes' ingress and egress resolvers to explore the potential overhead of third-party, egress resolvers. We collected latency measurements for about 4,000 clients, ≈40% of those in our dataset.

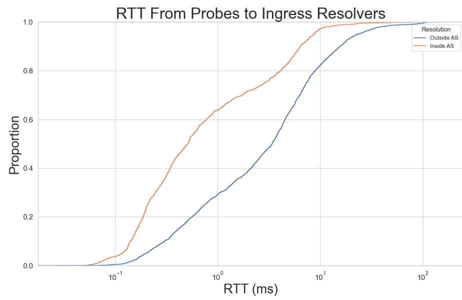
Finally, using our implementation of IPMap [**71**] and the public geolocation database IPInfo, we obtained geolocation results for the 7829 pingable IP addresses. There were 5523 pingable egress resolvers and 2306 pingable ingress resolvers. If both of these methods agree on locating the resolver outside the country, we consider the resolver to be outside of the country of the client. We find that this occurs in about 11% of cases. To gain confidence in our geolocation, we also checked our results against a commercial geolocation database Maxmind and found that our results match for all cases.

9.3.3. Analysis

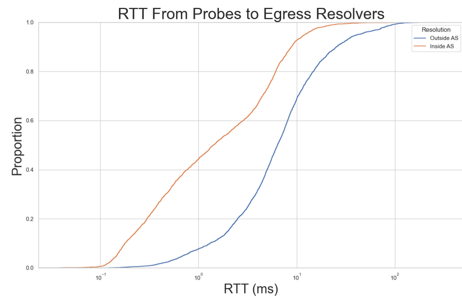
We now present our analysis of the additional aspects of complexity within the multi-tiered DNS infrastructure.

Third Party Resolution. Table 9.5 presents the AS and organizational relationship between clients, ingress, and egress resolvers. We split clients based on the relation between the ASes and organizations of the client and their ingress resolver and that of the ingress and egress resolver. We say that an ingress or egress resolver is a third party if it is in a different AS and organization than the probe.

We observe that the case which is commonly assumed to be the majority, that the client-side DNS infrastructure all belongs to the same organization, only happens in 38% of cases. *In 5% of cases the client, ingress, and egress resolvers all belong to different AS and 47% of the times client and ingress are in the same AS but the egress is in a different AS.* We show that the organizational relationships match the AS relationships at a very high rate as well.



(a) Ingress Resolver



(b) Egress Resolver

Figure 9.16. CDF of the min RTT from probes to their resolvers.

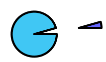
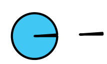
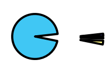
C %	C/I ASes	I/E ASes	Organizational Relationship
38%	=	=	All in the same organization
47%	=	≠	 "C=I=E : 4.4%" "C=I≠E : 95.6%"
10%	≠	=	 "C=I=E : 1.1%" "C≠I=E : 98.9%"
5%	≠	≠	 "C=I=E : 1.8%" "C≠I≠E : 95%" "C=I≠E : 2.8%" "C≠I=E : 0.4%"

Table 9.5. AS and organizational relationship between clients (C), ingress (I), and egress (E) resolvers.

We observe a long tail distribution of usage of third party ingress resolvers with a total of 145 providers. Among these, Google stands out as the most popular choice, accounting for 5.6%, trailed by Cloudflare at 3.3%. Similarly, in the case of third-party egress resolvers, we observed 489 third-party providers, with Google again being the most common choice (19.3%), closely followed by Cloudflare (13.1%). In the cases of ingress and egress mismatch, 12% of the time the ingress resolver belonged to Google and 4% of the times to Cloudflare. Notably, AS47583 (Hostinger) exclusively routes its queries through

Google. Deutsche Telekom and Vodafone Germany distribute their queries over Cloudflare and Google, occasionally using their own egress resolvers. Meanwhile, Comcast predominantly uses its own resolver but occasionally diverts its queries to Google or Cloudflare.

Latency Analysis. We analyze the latencies to the ingress and egress resolvers from the clients to understand the impact of using a third-party service. Figure 9.16(a) and Fig. 9.16(b) plot CDFs of ping latencies from the clients to the ingress/egress resolvers when they are in the same AS and outside of the AS of the client. About 5% of the probes using a third party ingress resolver have an RTT value to their resolver of more than 50ms. 15% of clients using a third party egress resolver have an RTT of higher than 50ms to their egress resolver. Overall, both figures show that the RTTs to resolvers are markedly higher when the client is using a third party resolver. However on average, egress resolvers are slightly further away from clients than ingress resolvers. Using the idea that the speed of light in fiber is equivalent to $c_f = 2/3 * c$ and from the data in Fig 9.16, we can say that for egress resolvers, 15% of them may be more than 10,000 km away from their client. For ingress resolvers, about 5% of them may be more than 10,000 km away from the client.

Geographic Analysis. We used our implementation of IPMap as well as a commercial geolocation database to confirm out-of-country resolution. If both of these methods agree that the resolver is in a different country than the client, we consider the resolver to be outside of the client's country. We find that this occurs in about 11% of cases.

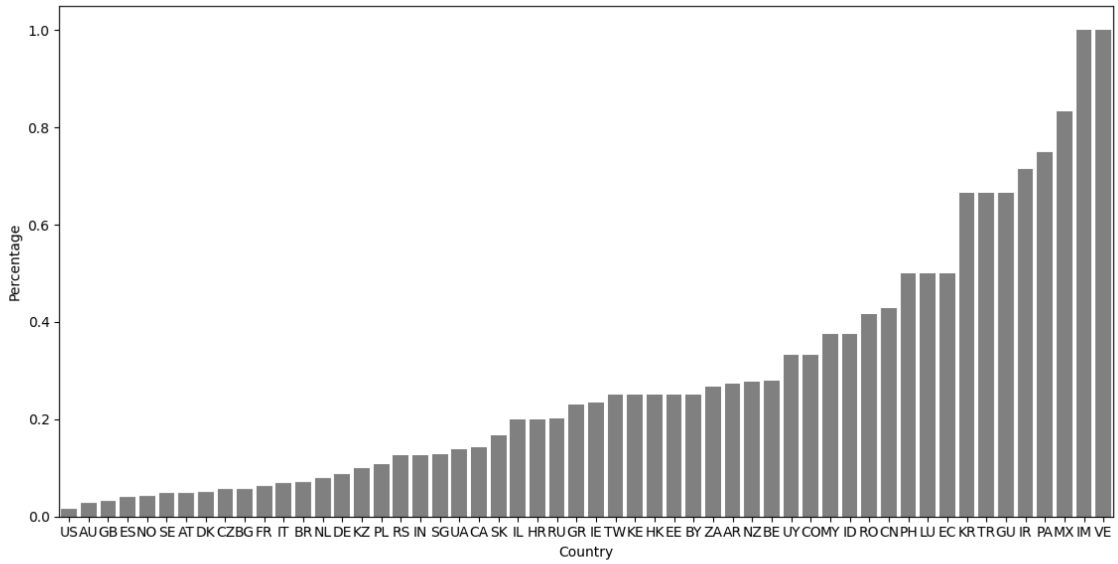


Figure 9.17. Percentage of clients using egress resolvers in other countries.

Figure 9.17 plots the percentage of clients using egress resolvers in another country. Vietnam and the Isle of Man have the highest percentage of out-of-country resolution. According to the OONI observatory, Vietnam ISPs appear to primarily implement censorship through DNS which may partially explain our findings [**157**]. The Isle of Man is a self-governing British Crown dependency and relies largely on UK Internet infrastructure. As one would expect, large countries with extensive DNS infrastructure like the US have very few resolutions taking place outside of the country. We found several interesting cases where probes were having their DNS resolutions done in countries with which the country of origin has a tense relationship. For example, out of 92 resolvers we were able to geolocate for probes located in Russia, 17 of them were being resolved outside of the country. The most popular destinations were Finland and Germany, with some instances of resolution being performed in the US, Sweden, the Netherlands, Poland, the

Czech Republic, and Japan. In the case ofArgentina, several resolutions were being performed in Brazil and Chile. We also found cases where the resolution was taking place on another continent, for instance in the case of Iran majority of resolutions were taking place in Germany and for clients in Singapore, several resolutions taking place in the Netherlands.

9.4. Conclusion

Performance is not shaped by consolidation in a single way, and this chapter reinforces the dissertation's central claim that observed concentration is ambiguous without a comparative baseline. At the service-endpoint level, structurally consolidated regimes tend to impose higher domestic latency with tighter government commercial overlap, consistent with shared ecosystem constraints, while strategically consolidated regimes often achieve lower median latency but with greater dispersion, consistent with uneven localization and deployment choices. Thus, even similar concentration levels can carry different performance consequences depending on why consolidation arises.

The chapter also shows that performance is mediated indirectly through steering. Replica selection acts as a hidden steering wheel of the Internet: every request is mapped to a server, influencing latency and reliability in ways that are rarely visible to users or policymakers. To make this control plane observable, we introduced a lightweight methodology that infers whether a CDN relies on DNS-based steering, global anycast, or regional anycast using latency fingerprints across resolvers. Applied at global scale, the analysis shows that DNS-based steering is the predominant approach used to deliver popular content.

That result matters because it links content-delivery performance directly to the client-side DNS ecosystem. If CDN steering depends on resolver location, then consolidation in DNS resolution becomes performance-relevant even when the content infrastructure itself is unchanged. We show that DNS resolution infrastructure is increasingly complex and consolidated, often involving hidden resolver layers operated by different organizations, networks, and even countries. In this setting, resolver location can become a poor proxy for user location. As resolver consolidation grows, DNS-based steering becomes more vulnerable to user-resolver mismatch, creating a concrete pathway through which concentration in one layer degrades performance in another.

Taken together, these findings show that the performance implications of consolidation arise through two linked channels: where infrastructure is deployed, and how users are steered across that infrastructure through opaque intermediaries. In doing so, this chapter extends the dissertation's broader argument from resilience and exposure to performance: distinguishing structural from strategic consolidation clarifies not only where concentration exists, but how and when it matters for end users. More broadly, the results suggest that replica selection should be treated not only as a performance optimization problem, but as a sociotechnical control point with implications for resilience, accountability, and governance. They also reinforce a recurring theme of the dissertation: when critical control planes become opaque, assessing the true consequences of consolidation becomes correspondingly harder.

CHAPTER 10

Conclusion: Consolidation, Opacity, and Future Directions

This dissertation examined how the Internet's growing reliance on third-party infrastructure is reshaping web service provision, patterns of dependency, and the consequences of concentration. Across content hosting, authoritative DNS, and certificate authorities, it showed that consolidation is widespread across both commercial and public-sector web infrastructure, but also that observed concentration is ambiguous when interpreted in isolation: similar concentration levels can arise from very different underlying conditions and need not imply the same consequences. To address this, the dissertation introduced a comparative framework that distinguishes structural consolidation from strategic consolidation by interpreting government infrastructure choices relative to the surrounding commercial ecosystem. This distinction is the dissertation's central contribution. It shows that concentration is not only a matter of how much infrastructure is concentrated, but also of why it is concentrated, and that these different regimes produce measurably different implications for resilience, exposure, and performance.

Taken together, these results support a broader conclusion: consolidation is not only a market-structure phenomenon, but an operational and governance condition whose consequences depend on its origin. Once concentration is interpreted through the structural/strategic distinction, it becomes possible to explain why apparently similar levels of consolidation can produce different outcomes in control, fallback capacity, exposure, and user experience. In resilience, the key issue is not only provider dominance but the combination of shared dependencies and limited practical fallback, regardless of whether concentration arises from ecosystem-wide constraints or deliberate choices. In exposure, the central concern is not only foreign hosting but the wider infrastructure through which users reach services, including on-path intermediaries that may sit outside domestic jurisdiction, and here the distinction between structural and strategic regimes helps explain why some countries exhibit uniformly high exposure while others exhibit much greater variance. In performance, the dissertation showed that concentration can affect latency directly through service localization and indirectly through opaque steering mechanisms, especially when DNS-based replica selection interacts with resolver consolidation.

These findings also highlight a deeper challenge that extends beyond any single service layer. The Internet now functions as civilization-scale infrastructure, underpinning communication, public services, commerce, and critical operations. Yet the systems that support it are increasingly difficult to assess independently. Cloud platforms, CDNs, transit networks, and trust infrastructures operate as tightly coupled but individually opaque systems. Proprietary control planes, hyperscale orchestration, and limited external visibility into routing and service logic reduce the externally observable surface of the Internet even as dependence on it continues to grow. This opacity matters because it limits the ability to distinguish between concentration that is structurally difficult to avoid and concentration that reflects strategic choices and could, in principle, be changed. As a result, resilience is often treated as an assumed property of digital infrastructure rather than a demonstrated one, and consolidation is often described without a clear account of what its observed form actually means.

This dissertation contributes to narrowing that gap by making important aspects of consolidation and its consequences observable through Internet-scale measurement and comparative analysis. Its comparative design is especially important in this respect: by evaluating government dependencies relative to the commercial ecosystems in which they are embedded, the dissertation shows how external measurement can support explanation rather than description alone. At the same time, it underscores the limits of what can be directly measured. Many of the mechanisms that shape resilience, exposure, and performance are only partially visible from the outside. Path measurements are incomplete, service attribution is imperfect, and control-plane decisions such as replica selection are often opaque. This does not make measurement less important. It makes it more urgent to develop methods that explicitly account for partial observability, inference limits, and uncertainty when assessing the consequences of consolidation.

More broadly, the dissertation suggests that the central challenge is no longer only measuring concentration, but interpreting concentration under conditions of incomplete visibility. The structural/strategic distinction helps address that challenge by providing a comparative basis for explanation, but it also reveals how much remains hidden from direct observation. This tension between explanation and observability motivates the future research agenda that follows.

A promising direction for future work is to move toward a science of observable resilience for Internet infrastructure: one that treats resilience not as an assumed attribute, but as a property that can be meaningfully defined, bounded, measured, and evaluated under realistic observability constraints. The need for such a science follows directly from the dissertation's central claim. If observed concentration is ambiguous without a comparative baseline, and if the systems producing that concentration are increasingly opaque, then future research must develop stronger ways to infer not only how much consolidation exists, but what kind of consolidation is being observed and what consequences can be justified from the evidence. Such a research agenda would require new methods to characterize what can and cannot be inferred from external measurements, how those inference limits change as infrastructure evolves, and how to validate resilience claims without requiring privileged access to proprietary systems. It would also require stronger integration across Internet measurement, systems design, network operations, security, and governance, since the consequences of consolidation now cut across technical and institutional boundaries.

Several concrete directions follow from this dissertation. First, future work can extend comparative consolidation analysis across additional service layers and control planes, including recursive DNS, identity and authentication services, and cloud control infrastructure, to determine whether the structural/strategic distinction generalizes beyond the services studied here. Second, resilience assessment can be strengthened by developing stress-testing methodologies that operate within external measurement limits, for example by combining passive observations, controlled perturbation, and privacy-preserving provider disclosures. Third, longitudinal measurement can be used not only to track consolidation itself, but also to track how regimes shift over time, including whether strategic choices increasingly converge toward structural constraints or whether new forms of strategic dependence emerge as ecosystems mature. Fourth, longitudinal analysis can also be used to track the observability gap itself: how much of the infrastructure remains inferable from public vantage points as deployment and control become more centralized and opaque. Finally, there is a need for governance and standards conversations that do not assume transparency by default, but instead ask what evidence should be available to assess the resilience of systems that now mediate essential public and economic life.

The central message of this dissertation is that observed consolidation cannot be interpreted at face value. The same concentration level may reflect structural constraints shared across an ecosystem or strategic choices made by particular actors, and these regimes carry different implications for resilience, exposure, and performance. By measuring consolidation across services and sectors, introducing a comparative framework to distinguish structural from strategic dependence, and tracing the consequences of that distinction for user-facing outcomes, this work provides a foundation for more accountable and evidence-based assessment of Internet infrastructure. It also argues that such assessment must now grapple directly with opacity: as digital infrastructure becomes more critical and more difficult to observe, the ability to infer, bound, and evaluate resilience from partial external evidence will become an increasingly important scientific and public challenge.

Bibliography

[1] Abley, J., and Lindqvist, K. Operation of anycast services. *www.rfc-editor.org* (2006).

[2] Ager, B., Mühlbauer, W., Smaragdakis, G., and Uhlig, S. Comparing dns resolvers in the wild. In *Proc. of IMC* (2010).

[3] Akamai. Akamai cdn deployment, 2022.

[4] Al-Dalky, R., Rabinovich, M., and Schomp, K. A look at the ecs behavior of dns resolvers. IMC '19.

[5] AlJahdali, H., Albatli, A., Garraghan, P., Townend, P., Lau, L., and Xu, J. Multi-tenancy in cloud computing. In *Proc. of IEEE SOSE* (2014).

[6] Allman, M. Comments on DNS robustness. In *Proc. of IMC* (2018).

[7] Alzoubi, H. A., Lee, S., Rabinovich, M., Spatscheck, O., and der Merwe, J. V. Anycast cdns revisited. In *Proc. of the WWW* (2008).

[8] Amazon. Amazon cdn deployment, 2022.

[9] Amazon. Amazon aws, 2023.

[10] Amazon. Amazon mechanical turk, 2023.

[11] Anderson, S., Salamatian, L., Bischof, Z. S., Dainotti, A., and Barford, P. igdb: connecting the physical and logical layers of the internet. In *Proceedings of the 22nd ACM Internet Measurement Conference* (2022).

[12] ANSSI (Agence nationale de la sécurité des systèmes d'information). French trusted list, 2026.

[13] APNIC. APNIC Service Region, 2024. `https://www.apnic.net/about-apnic/corporate-documents/documents/corporate/apnic-service-region/`.

[14] Ardi, C., and Calder, M. The prevalence of single sign-on on the web: Towards the next generation of web content measurement. In *Proc. of IMC* (2023).

[15] Arkko, J. Centralised architectures in internet infrastructure. *IETF Internet Draft* (2019).

[16] Arkko, J. Centralised architectures in internet infrastructure, 2019.

[17] Arkko, J. The influence of Internet architecture on centralised versus distributed Internet services. *Journal of Cyber Policy* (2020).

[18] Arkko, J. The influence of internet architecture on centralised versus distributed internet services.

[19] Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R. H., Konwinski, A., Lee, G., Patterson, D. A., Rabkin, A., Stoica, I., et al. Above the clouds: A berkeley view of cloud computing. Tech. rep., 2009.

[20] ATLAS, R. Atlas console, 2023.

[21] Ballani, H., Francis, P., Zhang, X., and Mahajan, S. A study of prefix hijacking incidents. In *Proc. of ACM SIGCOMM* (2007).

[22] Bank, T. W. The world by region. `https://datatopics.worldbank.org/sdgatlas/archive/2017/the-world-by-region.html`, 2017.

[23] Bank, T. W. Brief: Digital government for development, 2024.

[24] Bank, W. Brief: Digital government for development, 2024.

[25] Bank, W. Individuals using the internet (% of population) (it.net.user.zs), 2026.

[26] Bates, S., Bowers, J., Greenstein, S., Weinstock, J., Xu, Y., and Zittrain, J. Evidence of decreasing internet entropy: The lack of redundancy in dns resolution by major websites and services. *SSRN Electronic Journal* (2018).

[27] Bates, S., Bowers, J., Greenstein, S., Weinstock, J., Xu, Y., and Zittrain, J. Evidence of Decreasing Internet Entropy: The Lack of Redundancy in DNS Resolution by Major Websites and Services. *Journal of Quantitative Description: Digital Media 1* (2021).

[28] Biliris, A., Cranor, C., Douglis, F., Rabinovich, M., Sibal, S., OliverSpatscheck, and Sturm, W. *Computer Communications* (2002).

[29] Bortzmeyer, S. DNS censorship (DNS lies) as seen by RIPE Atlas. RIPE Labs, December 2015. https://labs.ripe.net/Members/stephane_bortzmeyer/dns-censorship-dns-lies-seen-by-atlas-probes.

[30] Bottger, T., Cuadrado, F., Antichi, G., Fernandes, E. L., Tyson, G., Castro, I., and Uhlig, S. An empirical study of the cost of DNS-over-HTTPS. In *Proc. of IMC* (2019).

[31] Bozkurt, I. N., Aguirre, A., Chandrasekaran, B., Godfrey, P. B., Laughlin, G., Maggs, B., and Singla, A. Why is the Internet so slow?! In *Proc. of PAM* (2017).

[32] Bustamante, F. E., Doyle, J., Willinger, W., Fayed, M., Alderson, D. L., Low, S., Savage, S., and Schulzrinne, H. Towards re-architecting today's internet for survivability: Nsf workshop report. *SIGCOMM Comput. Commun. Rev.* (2024).

[33] Butkiewicz, M., Madhyastha, H. V., and Sekar, V. Understanding website complexity: Measurements, metrics, and implications. In *Proc. of IMC* (2011).

[34] Butkiewicz, M., Madhyastha, H. V., and Sekar, V. Understanding website complexity: Measurements, metrics, and implications. In *Proc. of IMC* (2011).

[35] Butkiewicz, M., Madhyastha, H. V., and Sekar, V. Understanding website complexity: Measurements, metrics, and implications. In *Proc. of ACM SIGCOMM* (2011).

[36] CA Certificates in Firefox. Ca certificates in firefox report (ccadb), 2023.

[37] CAIDA. As to organizations mappings, 2023.

[38] CAIDA. Routeviews prefix to as mappings dataset for ipv4 and ipv6, 2023.

[39] Calder, M., Fan, X., and Zhu, L. A cloud provider's view of EDNS client-subnet adoption.

[40] Calder, M., Fan, X., and Zhu, L. A cloud provider's view of edns client-subnet adoption. In *Proc. of TMA Conference* (2019).

[41] Calder, M., Flavel, A., Katz-Bassett, E., Mahajan, R., and Padhye, J. Analyzing the performance of an anycast cdn. In *Proc. of IMC* (2015).

[42] Carisimo, E., Gamero-Garrido, A., Snoeren, A. C., and Dainotti, A. Identifying ases of state-owned internet operators. In *Proc. of IMC* (2021).

[43] Carisimo, E., Gamero-Garrido, A., Snoeren, A. C., and Dainotti, A. Identifying ases of state-owned internet operators. In *Proc. of IMC* (2021).

[44] cdnplanet. Cdn finder, 2023.

[45] Center for Applied Internet Data Analysis (CAIDA). Macroscopic internet topology data kit (itdk), 2024.

[46] Center for Applied Internet Data Analysis (CAIDA). Routeviews prefix to as mappings dataset for ipv4 and ipv6, 2024.

[47] Chander, A., and Le, U. P. Breaking the web: Data localization vs. the global internet. *SSRN Electronic Journal* (2014).

[48] Chen, F., Sitaraman, R. K., and Torres, M. End-user mapping: Next generation request routing for content delivery. In *Proc. of ACM SIGCOMM* (2015).

[49] Chung, T., Lok, J., Chandrasekaran, B., Choffnes, D., Levin, D., Maggs, B. M., Mislove, A., Rula, J., Sullivan, N., and Wilson, C. Is the web ready for ocsp must-staple? *Proceedings of the Internet Measurement Conference 2018* (2018).

[50] Cisagov. findcdn, 2023.

[51] Cloudflare. Cloudflare, 2022.

[52] Cloudflare, 2023.

[53] Cloudflare. Cdn · cloudflare reference architecture docs, 2023.

[54] Consulting, I. General data protection regulation (gdpr), 2013.

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

[56] Controller of Certifying Authorities (CCA), G. o. I. Root certifying authority of india (rcai) certificate policy and certification practice statement (cp/cps), 2011.

[57] Coppolino, L., D'Antonio, S., Mazzeo, G., and Romano, L. Cloud security: Emerging threats and current solutions. *Computers & Electrical Engineering* (2017).

[58] Crown Commercial Service. G-cloud framework, 2025.

[59] Daigle, L. Whois protocol specification, 2004.

[60] Darwich, O., Rimlinger, H., Dreyfus, M., Gouel, M., and Vermeulen, K. Replication: Towards a publicly available internet scale ip geolocation dataset. In *Proc. of IMC* (2023).

[61] Darwich, O., Rimlinger, H., Dreyfus, M., Gouel, M., and Vermeulen, K. Replication: Towards a publicly available internet-scale ip geolocation dataset.

[62] Dell'Amico, M., Bilge, L., Kayyoor, A., Efstathopoulos, P., and Vervier, P.-A. Lean on me: Mining internet service dependencies from large-scale dns data.

[63] Demchak, C., and Dombrowski, P. Cyber westphalia: Asserting state prerogatives in cyberspace. *Georgetown Journal of International Affairs* (2013).

[64] DeNardis, L. *The Global War for Internet Governance.* Yale University Press, 2014.

[65] Digital India, Government of India. Controller of certifying authorities (cca), 2026.

[66] Digital Transformation Agency, Australian Government. Secure cloud strategy, 2021.

[67] DNS, P. Public dns server list, 2023.

[68] Doan, T. V., van Rijswijk-Deij, R., Hohlfeld, O., and Bajpai, V. An empirical view on consolidation of the web. *ACM Transactions on Internet Technology 22*, 3 (Aug 2022), 1–30.

[69] Dobson, C. Achieving equity in digital government services, 2023.

[70] Douzet, F., Pétiniaud, L., Salamatian, L., Limonier, K., Salamatian, K., and Alchus, T. Measuring the fragmentation of the internet: The case of the border gateway protocol (bgp) during the ukrainian crisis. In *2020 12th International Conference on Cyber Conflict (CyCon)* (2020).

[71] Du, B., Candela, M., Huffaker, B., Snoeren, A. C., and claffy, k. Ripe ipmap active geolocation: Mechanism and performance evaluation. *SIGCOMM Comput. Commun. Rev.* (2020).

[72] Dönni, D., Machado, G., Tsiaras, C., and Stiller, B. Schengen routing: A compliance analysis.

[73] Edmundson, A., Ensafi, R., Feamster, N., and Rexford, J. Nation-state hegemony in internet routing. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS)* (2018).

[74] Efron, B., and Tibshirani, R. J. *An Introduction to the Bootstrap.* 1993.

[75] Fan, X., Katz-Bassett, E., and Heidemann, J. Assessing affinity between users and cdn sites. In *Proc. of TMA Conference* (2015).

[76] Federal Office of Information Technology, Systems and Telecommunication (FOITT), Switzerland. Swiss government pki: Cp/cps, 2025.

[77] FNMT-RCM (Fábrica Nacional de Moneda y Timbre - Real Casa de la Moneda). Digital certification (ceres project), 2026.

[78] for Federal Government, A. C. C. Cloud computing for federal government, 2023.

[79] for US Government, A. Azure for us government — microsoft azure, 2023.

[80] Fruhlinger, J. The opm hack explained: Bad security practices meet china's captain america, 2020.

[81] Fund, I. M. STATE-OWNED ENTERPRISES: THE OTHER GOVERNMENT. `https://www.imf.org/~/media/Files/Publications/fiscal-monitor/2020/April/English/ch3.ashx`, 2020.

[82] Gigis, P., Calder, M., Manassakis, L., Nomikos, G., Kotronis, V., Dimitropoulos, X., Katz-Bassett, E., and Smaragdakis, G. Seven years in the life of hypergiants' off-nets. In *Proc. of ACM SIGCOMM* (2021).

[83] Goel, U., Wittie, M. P., and Steiner, M. Faster web through client-assisted cdn server selection. In *Proc. of ICCN* (Oct 2015).

[84] Google. Chrome user experience report — chrome ux report —google developers, 2022.

[85] Google. understanding-google-cloud-network-edge-points, 2022.

[86] Google. About crux, 2023.

[87] Google. Obtain google ip address ranges, 2023.

[88] Government Root Certification Authority (GRCA), Taiwan. Government public key infrastructure (gpki) - introduction, 2002.

[89] Greenwald, G., and MacAskill, E. Nsa secretly intercepts google and yahoo data centers worldwide. *The Guardian* (2013).

[90] Grossman, R. L. The case for cloud computing. *IT professional* (2009).

[91] Habib, R., Ruth, K., Akiwate, G., and Durumeric, Z. Formalizing dependence of web infrastructure.

[92] Hendriks, R., Luckie, M., Jonker, M., Sommese, R., and van Rijswijk-Deij, R. Laces: An open, fast, responsible and efficient longitudinal anycast census system. In *Proceedings of the 2025 ACM Internet Measurement Conference* (2025).

[93] Hicks, M. Outage analyses: Seven outages that shook up 2021. CISCO/ThousandEyes Blog, 2022.

[94] hotspotshield. Hotspotshield vpn, 2023.

[95] Hounsel, A., Borgolte, K., Schmitt, P., and Feamster, N. D-DNS: towards re-decentralizing the DNS.

[96] HOUSE, T. W. Fact sheet: Building digital experiences for the american people — omb, 2023.

[97] Huang, C., Wang, A., Li, J., and Ross, K. Measuring and evaluating large-scale cdns. In *Proc. of IMC* (2008).

[98] Huitema, C., Huston, G., Hs, E., Leer, G., and Zhang. *Draft Report of DINRG Workshop on Centralization in the Internet.* 2021.

[99] Huston, G. DNS resolver centrality, 2019.

[100] Huston, G. DNS resolver centrality. APNIC Blog, 2019.

[101] Huston, G. CDN and centrality. APNIC Blog, 2021.

[102] Huston, G. Cdn and centrality, 2021.

[103] Huston, G. Looking at centrality in the dns, 2022.

[104] Ibrahim, S., He, B., and Jin, H. Towards pay-as-you-consume cloud computing. In *Proc. of Conference on Services Computing* (2011).

[105] India's MINISTRY OF LAW AND JUSTICE. India's Digital Personal Data Protection (DPDP) Act, 2023. https://www.meity.gov.in/writereaddata/files/Digital%20Personal%20Data%20Protection%20Act%202023.pdf.

[106] Info, I. Ipinfo, 2023.

[107] International Trade Administration, U.S. Department of Commerce. Vietnam: Cybersecurity data localization requirements, 2022.

[108] Iordanou, C., Smaragdakis, G., Poese, I., and Laoutaris, N. Tracing Cross Border Web Tracking. In *Proc. of IMC* (2018).

[109] IP2Location. Free IP Geolocation Database, 2023.

[110] ipInfo. ipinfo, 2023.

[111] IPinfo. Ipinfo ip address data api, 2025.

[112] Jansen, W., Grance, T., et al. Guidelines on security and privacy in public cloud computing.

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

[114] Kashaf, A., Dou, J., Belova, M., Apostolaki, M., Agarwal, Y., and Sekar, V. A first look at third-party service dependencies of web services in africa.

[115] Kashaf, A., Sekar, V., and Agarwal, Y. Analyzing third party service dependencies in modern web services: Have we learned from the mirai-dyn incident? In *Proc. of IMC* (2020).

[116] Khan, M. T., DeBlasio, J., Voelker, G. M., Snoeren, A. C., Kanich, C., and Vallina-Rodriguez, N. An empirical analysis of the commercial vpn ecosystem. IMC '18.

[117] Koch, T., Katz-Bassett, E., Heidemann, J., Calder, M., Ardi, C., and Li, K. Anycast in context: A tale of two systems. In *Proceedings of the 2021 ACM SIGCOMM 2021 Conference* (2021), pp. 398–417.

[118] Kumar, D., Ma, Z., Durumeric, Z., Mirian, A., Mason, J., Halderman, J. A., and Bailey, M. Security challenges in an increasingly tangled web. In *Proc. of the WWW* (2017).

[119] Kumar, R., Asif, S., Lee, E., and Bustamante, F. E. Each at its own pace: Third-party dependency and centralization around the world. *Proc. ACM Meas. Anal. Comput. Syst.* (2023).

[120] Kumar, R., Asif, S., Lee, E., and Bustamante, F. E. Each at its own pace: Third-party dependency and centralization around the world. In *Proc. of ACM SIGMETRICS* (2023).

[121] Kumar, R., Asif, S., Lee, E., and Bustamante, F. E. Each at its own pace: Third-party dependency and centralization around the world. *Proc. ACM Meas. Anal. Comput. Syst.* (2023).

[122] Kumar, R., Carisimo, E., De Angelis Riva, L., Buzzone, M., Bustamante, F. E., Qazi, I. A., and Beiró, M. G. Of choices and control - a comparative analysis of government hosting.

[123] Kumar, R., Carisimo, E., De Angelis Riva, L., Buzzone, M., Bustamante, F. E., Qazi, I. A., and Beiró, M. G. Of choices and control – a comparative analysis of government hosting. In *Proc. of IMC* (2024).

[124] Lakshmanan, R. Let's encrypt revoking 3 million tls certificates issued incorrectly due to a bug, 2020.

[125] Law, B. G. D. P. Lgpd brazil - general personal data protection act, 2023.

[126] Le, T. V. Vietnam - data protection overview, Nov 2019.

[127] Leyba, K. G., Edwards, B., Freeman, C., Crandall, J. R., and Forrest, S. A. Borders and gateways: Measuring and analyzing national AS chokepoints. In *Proc. of ACM COMPASS* (2019).

[128] Li, Z., Levin, D., Spring, N., and Bhattacharjee, B. Internet anycast: Performance, problems, & potential. In *Proc. of ACM SIGCOMM* (2018).

[129] List, M. P. S. Public suffix list.

[130] Liu, Y., Song, H. H., Bermudez, I., Mislove, A., Baldi, M., and Tongaonkar, A. Identifying personal information in Internet traffic. In *Proc. of ACM COSN* (2015).

[131] Livingood, J., Antonakakis, M., Sleigh, B., and Winfield, A. Centralized dns over https (doh) implementation issues and risks, 2019.

[132] Livingood, J., Antonakakis, M., Sleigh, B., and Winfield, A. Centralized dns over https (doh) implementation issues and risks, 2019.

[133] Luckie, M., Huffaker, B., Marder, A., Bischof, Z., Fletcher, M., and claffy, k. Learning to extract geographic information from internet router hostnames. In *Proc. of CoNEXT* (2021).

[134] Ma, Z., Mason, J., Antonakakis, M., Durumeric, Z., and Bailey, M. What's in a name? exploring ca certificate control.

[135] MacAskill, E., Dance, G., Cage, F., Chen, G., and Popovich, N. Nsa files decoded: Edward snowden's surveillance revelations explained, 2013.

[136] Madhyastha, H. V., Isdal, T., Piatek, M., Dixon, C., Anderson, T., Krishnamurthy, A., and Venkataramani, A. iplane: An information plane for distributed services. In *Proc. of USENIX NSDI* (2006).

[137] Madhyastha, H. V., Katz-Bassett, E., Anderson, T. E., Krishnamurthy, A., and Venkataramani, A. iplane nano: Path prediction for peer-to-peer applications. In *Proc. of USENIX NSDI* (2009).

[138] Mao, Z. M., Cranor, C. D., Douglis, F., Rabinovich, M., Spatscheck, O., and Wang, J. A precise and efficient evaluation of the proximity between web clients and their local dns servers. In *Proc. of USENIX ATC* (2002).

[139] Matic, S., Tyson, G., and Stringhini, G. Pythia: a framework for the automated analysis of web hosting environments. *The World Wide Web Conference* (2019).

[140] Matic, S., Tyson, G., and Stringhini, G. Pythia: A framework for the automated analysis of web hosting environments. In *The World Wide Web Conference* (2019).

[141] Maxmind. Maxmind server ip addresses, 2022.

[142] Mcquistin, S., Uppu, S., and Flores, M. Taming anycast in the wild internet. In *Proc. of IMC* (2019).

[143] Michelinakis, F., Doroud, H., Razaghpanah, A., Lutu, A., Vallina-Rodriguez, N., Gill, P., and Widmer, J. The cloud that runs the mobile internet: A measurement study of mobile cloud services.

[144] Ministry of Digital Affairs (MODA), Taiwan. Government digital infrastructure: Operations (gpki and grca), 2022.

[145] Mockapetris, P. Domain names – concepts and facilities. Tech. rep., IETF, 1987.

[146] Moura, G. How centralized is dns traffic becoming?, November 2020.

[147] Moura, G., Castro, S., Hardaker, W., Wullink, M., and Hesselman, C. Clouding up the internet: how centralized is dns traffic becoming? In *Proc. of IMC* (2020).

[148] Mozilla. Mozilla root store policy, 2025.

[149] Nations, U. Un e-government survey 2022: The future of digital government, 2022.

[150] Nations, U. Egovkb — united nations ¿ about ¿ overview ¿ e-government development index, 2023.

[151] News, C. Explainer: What's behind strained china-japan relations, 2022.

[152] nordvpn. Nord vpn, 2023.

[153] NordVPN. Nordvpn: Vpn service for online privacy and security, 2026.

[154] Nygren, E., Sitaraman, R., and Sun, J. The akamai network: A platform for high-performance internet applications. *ACM SIGOPS Operating Systems Review 44*, 3 (July 2010).

[155] of California Department of Justice, S. California consumer privacy act (ccpa), 2023.

[156] Office, U. G. A. Solarwinds cyberattack demands significant federal and private-sector response (infographic), 2021.

[157] OONI. MAP State of Internet Censorship Report 2022 - Vietnam. `https://ooni.org/post/2022-state-of-internet-censorship-vietnam/`, 2023.

[158] Oppenheimer, H. Digital interdependence and power politics. *British Journal of Political Science 55* (2025), e93.

[159] Otto, J., and Bustamante, M. S. J. R. F. E. Content delivery and the natural evolution of DNS: remote DNS trends, performance issues and alternative solutions. In *Proc. of IMC* (2012).

[160] Overview, R. D. P. Russia - data protection overview, 2020.

[161] PeeringDB. Peeringdb, 2025.

[162] Planet, C. Akamai cdn, 2023.

[163] Planet, C. Cdn planet, 2023.

[164] Planet, C. Cloudflare cdn, 2023.

[165] Planet, C. Edgio cdn, 2023.

[166] Pochat, V. L., Goethem, T. V., Tajalizadehkhoob, S., Korczynski, M., and Joosen, W. Tranco: A research-oriented top sites ranking hardened against manipulation. In *Proc. of NDSS* (2019).

[167] Poese, I., Frank, B., Ager, B., Smaragdakis, G., Uhlig, S., and Feldmann, A. Improving content delivery with padis. *IEEE Internet Computing* (2012).

[168] Poese, I., Uhlig, S., Kaafar, M. A., Donnet, B., and Gueye, B. Ip geolocation databases: Unreliable? *ACM SIGCOMM CCR* (2011).

[169] Pohlmann, N., Sparenberg, M., Siromaschenko, I., and Kilden, K. *Secure Communication and Digital Sovereignty in Europe.* 2014.

[170] Programme, U. N. D. Human development index, 2023.

[171] Radu, R., and Hausding, M. Consolidation in the DNS resolver market – how much, how fast, how dangerous? *Journal of Cyber Policy* (2020).

[172] Ramasubramanian, V., and Sirer, E. G. The design and implementation of a next generation name service for the internet. In *Proceedings of the ACM SIGCOMM 2004 Conference* (2004).

[173] Reuters. Fastly blames software bug for major global internet outage, 2021.

[174] Reuters. Russia reroutes internet traffic in occupied ukraine to its infrastructure, 2022.

[175] Reuters. Amazon says aws cloud service back to normal after outage disrupted thousands of sites, 2025.

[176] Reuters. Cloudflare outage cuts access to x, chatgpt and other web platforms for thousands, 2025.

[177] Rhoades, S. A. The herfindahl-hirschman index, 1993.

[178] RIPE NCC. Youtube hijacking: a ripe ncc ris case study, 2008.

[179] RIPE NCC. IPmap, 2024. `https://ipmap.ripe.net/`.

[180] Roberts, P. F. Comcast suffers DNS outage, denies pharming link, Apr 2005. `https://www.networkworld.com/article/2318771/comcast-suffers-dns-outage--denies-pharming-link.html`.

[181] routeviews. RouteViews IPv4 Prefix to AS mappings - coalesced. `https://catalog.caida.org/dataset/routeviews_ipv4_prefix2as_coalesced.`, 2023.

[182] Rula, J. P., and Bustamante, F. E. Behind the curtain – cellular dns and content replica selection. In *Proc. of IMC* (2014).

[183] Saroiu, S., Gummadi, K. P., Dunn, R. J., Gribble, S. D., and Levy, H. M. An analysis of internet content delivery systems. In *Proc. of USENIX OSDI* (2002).

[184] Scheitle, Q., Hohlfeld, O., Gamba, J., Jelten, J., Zimmermann, T., Strowes, S. D., and Vallina-Rodriguez, N. A long way to the top. *Proc. of IMC* (2018).

[185] Schneider-Petsinger, M., Wang, J., Jie, Y., and Crabtree, J. *US-China Strategic Competition The Quest for Global Technological Leadership*. 2019.

[186] Schomp, K., Callahan, T., Rabinovich, M., and Allman, M. On measuring the client-side dns infrastructure. In *Proceedings of the 2013 conference on Internet measurement conference* (2013), pp. 77–90.

[187] Selenium. Seleniumhq browser automation, 2024.

[188] Shad, D. M. R. Cyber threat in interstate relations: Case of us-russia cyber tensions. *Policy Perspectives* (2018).

[189] Simko, I. We are the champions. the index for evaluating concentration of championships using a sliding window approach. *Heliyon* (2022).

[190] Singanamalla, S., Jang, E. H. B., Anderson, R., Kohno, T., and Heimerl, K. Accept the risk and continue: Measuring the long tail of government https adoption. In *Proc. of IMC* (2020).

[191] Singh, R., Dunna, A., and Gill, P. Characterizing the deployment and performance of multi-cdns. In *Proc. of IMC* (2018).

[192] Singla, A., Chandrasekaran, B., Godfrey, P. B., and Maggs, B. The internet at the speed of light. In *Proc. of HotNets* (2014).

[193] Society, I. Consolidation in the internet economy, 2020.

[194] Society, I. Consolidation in the internet economy, 2020.

[195] Sommese, R., Bertholdo, L., Akiwate, G., Jonker, M., van Rijswijk-Deij, R., Dainotti, A., Claffy, K., and Sperotto, A. Manycast2: Using anycast to measure anycast. In *Proc. of IMC* (2020).

[196] Sommese, R., Bertholdo, L., Akiwate, G., Jonker, M., van Rijswijk-Deij, R., Dainotti, A., Claffy, K., and Sperotto, A. Manycast2: Using anycast to measure anycast. In *Proceedings of the ACM Internet Measurement Conference* (2020).

[197] Stadnik, I. Control by infrastructure: Political ambitions meet technical implementations in runet. *First Monday* (2021).

[198] Stapling, O. The problem with ocsp stapling and must staple and why certificate revocation is still broken - hanno's blog, 2017.

[199] Stats, I. W. World internet users statistics and 2019 world population stats, 2022.

[200] Stats, I. W. World internet users statistics and 2019 world population stats, 2023.

[201] Stokes, B. Hostile neighbors: China vs. japan, 2016.

[202] Su, A.-J., Choffnes, D. R., Kuzmanovic, A., and Bustamante, F. E. Drafting behind akamai (travelocity-based detouring). In *Proc. of ACM SIGCOMM* (2006).

[203] surfshark. Surfshark vpn, 2023.

[204] Sørensen, J., and Kosta, S. Before and after gdpr: The changes in third party presence at public and private european websites. *The World Wide Web Conference on - WWW '19* (2019).

[205] The European Commission. Overseas Countries and Territories, 2024. `https://international-partnerships.ec.europa.eu/countries/overseas-countries-and-territories_en`.

[206] The French Republic. Accord sur la Nouvelle-Caledonie signe a Noumea le 5 mai 1998, 2024. `https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000555817`.

[207] The World Bank. World bank country and lending groups, 2025.

[208] Thiagarajan, K., Carisimo, E., and Bustamante, F. E. The aleph: Decoding dns ptr records with large language models. In *Proceedings of the 2025 ACM Conference on Emerging Networking Experiments and Technologies (CoNEXT)* (2025).

[209] Timlib. Webxray domain owner list.

[210] Triukose, S., Wen, Z., and Rabinovich, M. Measuring a commercial content delivery network. In *Proc. of the WWW* (2011).

[211] UK Government, Government Digital Service. Cloud first policy, 2013.

[212] United Nations. UN E-Government Survey 2022 – The Future of Digital Government.

[213] United Nations. Non-Self-Governing Territories, 2024. https://www.un.org/dppa/decolonization/en/nsgt.

[214] United Nations Department of Economic and Social Affairs. United nations e-government survey 2024: The future of digital government, 2024.

[215] Urban, T., Degeling, M., Holz, T., and Pohlmann, N. Beyond the front page: Measuring third party dynamics in the field.

[216] U.S. Department of Justice. Herfindahl-hirschman index, 2018.

[217] U.S. Department of Justice, A. D. Herfindahl-hirschman index, 2024.

[218] U.S. Department of Justice and Federal Trade Commission. Horizontal merger guidelines, 2010. Updated 2023.

[219] US Department of State. A Guide to the United States' History of Recognition, Diplomatic, and Consular Relations, by Country, since 1776: Morocco, 2024. https://history.state.gov/countries/morocco.

[220] Wang, S., MacMillan, K., Schaffner, B., Feamster, N., and Chetty, M. A first look at the consolidation of dns and web hosting providers.

[221] Wang, S., MacMillan, K., Schaffner, B., Feamster, N., and Chetty, M. Measuring the consolidation of dns and web hosting providers, 2024.

[222] Wang, Z., Huang, J., and Rose, S. Evolution and challenges of dns-based cdns. *Digital Communications and Networks 4*, 4 (Nov 2018), 235–243.

[223] Wikipedia. United nations geoscheme, Jul 2022.

[224] Wikipedia. Amazon cloudfront, May 2023.

[225] Workd Bank Group. Digital government, April 2025.

[226] world population. internet-users-by-country, 2022.

[227] Xue, J., Choffnes, D., and Wang, J. Cdns meet cn an empirical study of cdn deployments in china. *IEEE Access 5* (2017), 5292–5305.

[228] Xue, J., Choffnes, D., and Wang, J. Cdns meet cn an empirical study of cdn deployments in china. *IEEE Access 5* (2017), 5292–5305.

[229] Yeganeh, B., Durairajan, R., Rejaie, R., and Willinger, W. A first comparative characterization of multi-cloud connectivity in today's internet. In *Proc. of PAM* (2020).

[230] Zembruzki, L., Jacobs, A. S., Landtreter, G. S., Granville, L. Z., and Moura, G. dnstracker: Measuring centralization of dns infrastructure in the wild.

[231] Zembruzki, L., Jacobs, A. S., Landtreter, G. S., Granville, L. Z., and Moura, G. `dnstracker`: Measuring centralization of dns infrastructure in the wild. In *Proc. of AINA* (2020).

[232] Zhou, M., Zhang, X., Hao, S., Yang, X., Zheng, J., Chen, G., and Dou, W. Regional IP Anycast: deployments, performance, and potentials. In *Proc. of ACM SIGCOMM* (2023).

[233] Zhu, G., and Gu, W. User mapping strategy in multi-cdn streaming: A data-driven approach. *IEEE Internet of Things Journal* (2021), 1–1.

[234] Zhu, J., Vermeulen, K., Cunha, I., Katz-Bassett, E., and Calder, M. The best of both worlds. In *Proc. of IMC* (2022).