Each at its Own Pace: Third-Party Dependency and Centralization Around the World

RASHNA KUMAR, Northwestern University, USA
SANA ASIF, Northwestern University, USA
ELISE LEE, Northwestern University, USA
FABIÁN E. BUSTAMANTE, Northwestern University, USA

We describe the results of a large-scale study of third-party dependencies around the world based on regional top-500 popular websites accessed from vantage points in 50 countries, together covering all inhabited continents. This broad perspective shows that dependencies on a third-party DNS, CDN or CA provider vary widely around the world, ranging from 19% to as much as 76% of websites, across all countries. The critical dependencies of websites – where the site depends on a single third-party provider – are equally spread ranging from 5% to 60% (CDN in Costa Rica and DNS in China, respectively). Interestingly, despite this high variability, our results suggest a highly concentrated market of third-party providers: three providers across all countries serve an average of 92% and Google, by itself, serves an average of 70% of the surveyed websites. Even more concerning, these differences persist a year later with increasing dependencies, particularly for DNS and CDNs. We briefly explore various factors that may help explain the differences and similarities in degrees of third-party dependency across countries, including economic conditions, Internet development, economic trading partners, categories, home countries, and traffic skewness of the country's top-500 sites.

CCS Concepts: • **Networks** → **Location based services**; **Network measurement**; **Public Internet**.

Additional Key Words and Phrases: centralization, third-party dependency, DNS, CDN, CA

ACM Reference Format:
Rashna Kumar, Sana Asif, Elise Lee, and Fabián E. Bustamante. 2023. Each at its Own Pace: Third-Party Dependency and Centralization Around the World. *Proc. ACM Meas. Anal. Comput. Syst.* 7, 1, Article 4 (March 2023), 29 pages. https://doi.org/10.1145/3579437

1 INTRODUCTION

There is a growing concern regarding consolidation trends in the Internet – the concentration of traffic, infrastructure, services and users on a handful of providers – given its economic, political and reliability implications [7, 14, 16, 34, 35, 38, 51, 70]. Centralization could amplify the impact of vulnerabilities and faults from the shared services, increase the risk of a captive market and significantly reduce user privacy by exposing a more complete user profile to a few, cross-market providers. The GlobalSign Certificate Revocation Error of 2016, the Amazon DNS DDoS attack in 2019, and the recent 2021 outage of Facebook and subsidiary companies, are just a handful of illustrating examples [2, 3, 5, 6, 8, 26, 50].

The Web provides a relatively accessible environment to characterize these centralization trends and, incidentally, to understand their concerning implications. Access to a website depends on a number of services provided by third-parties such as DNS, Content Distribution Networks (CDNs),

Authors' addresses: Rashna Kumar, rashnakumar2024@u.northwestern.edu, Northwestern University, USA; Sana Asif, sanaasif@u.northwestern.edu, Northwestern University, USA; Elise Lee, eliselee@cs.northwestern.edu, Northwestern University, USA; Fabián E. Bustamante, fabianb@cs.northwestern.edu, Northwestern University, USA.

2476-1249/2023/3-ART4
https://doi.org/10.1145/3579437

and Certificate Authorities (CAs). Accessing a site requires contacting at least one DNS authoritative nameserver to find the IP address of the webserver (or servers) hosting the site's content. These servers could be run by one or more CDNs, for increased reliability and performance. If these hosting servers use HTTPS, a client may also need to contact one or more CAs to verify the validity of the servers' SSL certificates. A popular website could rely on third-party providers for all these critical services.

Several recent studies have leveraged this observation to evaluate third-party dependencies in popular websites, typically from a single vantage point [19, 27, 38, 62] or only a few vantage points [64]. Kashaf et al. [38], for instance, characterizes the dependency of Alexa's most popular sites on third-party DNS, CA and CDNs, based on data collected from a vantage point in the US. The authors' analysis reveals, among other concerning findings, that between 78-97% of Alexa's top-100k websites depend on third-party DNS, CA or CDN, and that the use of third-party services is highly concentrated with the top-3 CDN, DNS or CA providers serving 50-70% of all websites they surveyed.

We build on these prior work ***to explore if, and to what extent, third-party dependency and centralization varies across countries and regions of the world***.

Our work is motivated by two simple observations (§2). *First*, while websites could potentially be accessed anywhere, not all websites are popular everywhere. Indeed, the popularity of websites, beyond a few top-ranked ones, is known to be region specific. *Second*, while many third-party services such as DNS and CDNs have been building global infrastructures, on and off-networks [29], their deployment is not (yet) omnipresent and their relative performance with respect to their competition varies across markets. This has served as motivation for brokering CDN [20] and, more recently, multi-cloud architectures [58, 68].

Motivated by these observations and concerns, our work explores third-party and centralization trends in the wider Internet. We report on a large-scale study we conducted in April 2021 and again in April 2022 based on regional top-500 popular websites accessed from vantage points in 50 countries, together covering all inhabited continents (a total of 16,774 unique websites).

We find 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. We explore various factors that may help explain the differences and similarities in levels of a third-party dependency across countries and regions, including economic conditions, Internet development, and economic trading partners.

In summary, we make the following contributions:

- We describe a methodology that builds on prior work [38, 64] to carry out a large-scale, longitudinal analysis of third-party dependency and centralization around the world (§3).
- We apply this methodology to characterize the dependencies of top regional sites in 50 countries, together capturing ≈ 78% of the Internet population. We find that third-party dependencies and critical dependency around the world vary widely, but market concentration is surprisingly high across all countries (§4).
- We present results from two consecutive years showing that these trends continue to increase for most services (§5).

- We report early results exploring various factors that may help explain the observed differences, such as economic conditions, Internet development, trading partners, website categories, websites' home country and ranking.(§6).
- We make available the collected dataset, including the list of top sites, analysis code and aggregated results for public use to enable replicability.

2 BACKGROUND AND MOTIVATION

Our work is motivated by recent reports about consolidation trends in the Internet [14, 29, 38, 49, 51, 70] and their economic, political and reliability implications [7, 16, 17, 34, 35, 45, 70]. The 2019 Global Internet Report [7] provides an overview of these trends in every aspect of the Internet economy, from access provision to service infrastructure and applications. It points out 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 (sometimes referred to as "GAFAs" – Google, Apple, Facebook, and Amazon, and "BATs"– Baidu, Alibaba, and Tencent) in control of much of the Internet's functionality and interoperability.

The Web offers an accessible context where to characterize and potentially monitor these trends, as sites build on third-party platforms for many of their key services from DNS and CDNs, to advertisements and user tracking, and a client visiting a website needs to interact with most of them. Recent studies have leveraged this observation to characterize these trends [19, 27, 38, 62, 64], typically from a single vantage point or a couple of regions.

In this work, we *investigate if, and to what extent, third-party dependency and centralization varies across countries and regions of the world*, building on a measurement approach that (1) focuses on regional top-ranked websites, and (2) relies on vantage points around the world.

Our work is motivated by two observations. First, the Web is made of 1.9 billion websites as of May 2022[1] and except for a few global sites, most are not universally popular. Second, while the infrastructure of large third-party services continues to expand, it is not yet omnipresent and may fare differently against their competition at different locales.

To illustrate the first point, we measure the overlap between the top-500 regional sites for each of the 115 countries in Alexa[2][3] and the top 100k sites from the global ranking. For each of these countries, we measured the overlap of the top-500 regional sites with the top 1k, 5k, 10k, 20k, 50k and 100k subsets of the global sites. As we extend the list of global sites considered, there is a higher probability of including a regional popular site in the global ranking and, thus, of finding a higher overlap between global and regional lists.

Figure 1 shows the percentage of overlap between these sets. The overall top-ranking websites are clearly dominated by just a few countries, even with the largest global ranking subset, with $\approx$ 25% of them having an overlap higher than 94%. On the other hand, half of the countries have less than $\approx$77% overlap even with the full top-100k list of websites.

The second observation has served as motivation for brokering CDN [20] and multi-cloud architectures [58, 68]. As no single infrastructure is omnipresent, websites popular in specific regions may choose to build on their own services or contract a well-provisioned local service provider. Even globally popular content providers (CP) are known to contract with different CDNs and DNSs in different part of the world based on connectivity, availability or cost [58]. For instance

[1] https://www.internetlivestats.com/total-number-of-websites/

[2] Despite concerns with the representativeness and stability of Alexa rankings, these are the longest lists of regional website rankings and considered the appropriate choice for end-user-based analysis [57].

[3] Alexa, however, retired as of May 1st, and therefore for future measurements we will use similarweb's [18] regional ranking.

Rakuten, the Japanese online retailing company, uses Akamai as its DNS provider for rakuten.com, but a private DNS for rakuten.jp. Similarly, wikipedia.org uses DigiCert as its CA provider in Germany and Sweden but IdenTrust Inc. in the US and Canada. Likewise, microsoft.com hosts content on both Akamai and Amazon Cloudfront in Canada, but only on Akamai in Sweden.

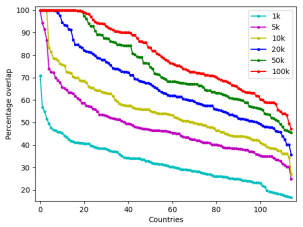


Fig. 1. Overlap between global and top regional sites. Countries are sorted based on their overlap and plotted left to right.

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

Fig. 2. Countries grouped by degrees of overlap between top-regional sites and the global ranking.

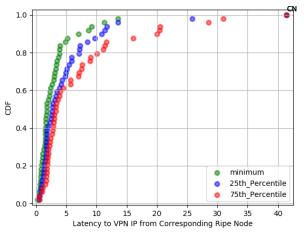


Fig. 3. Latency (ms) from RIPE Atlas Nodes to corresponding VPN Nodes.

3 METHODOLOGY AND DATASET

In this section, we describe a measurement methodology that builds on Kashaf et al. [38]'s to characterize service dependencies and centralization around the world. Our 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.

3.1 Country Selection

As Fig. 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 2 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 [66]. 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 [1]. To include China we use the Hotspot Shield VPN, since Nord and many other VPNs do not provide service in China [40].

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 [59].

Figure 3 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 [47] and IP2Location BD11.Lite [37]. Past work has shown the geolocation databases to be reliable at the country-level [54]. 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.

3.2 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.

3.2.1 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)[4] of the website and the nameserver with a match suggesting this is a *private* nameserver [42].

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*) [21]. 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 [38]. 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;

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

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 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).[5] 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 (§A) 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 (§A) 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.

3.2.2 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 [44, 63] 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 [10, 67]. 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 [46, 67]. 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 (§A).

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

[5]Following Kashaf et al. [38].

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 (§A) 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).

3.2.3 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.

3.2.4 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. 4.

3.3 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 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:

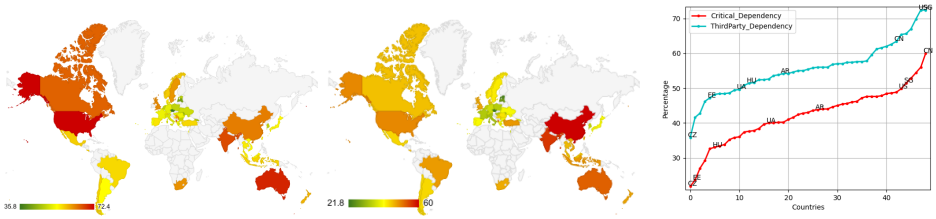
- How common is the third-party dependency of websites around the world?
- How much of this dependency is critical, dependent on a single third-party DNS or CDN provider?
- How concentrated is the market of third-party service providers within a country, region, and globally?

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

4.1 DNS Findings

Figure 4(a) and 4(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(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

Fig. 4. 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 5 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 6 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 (The map plots of these are shown in §A)). 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).

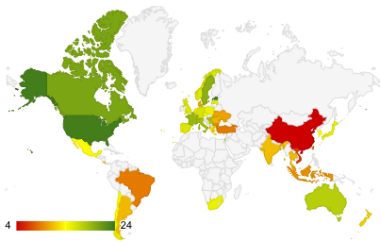


Fig. 5. Plot of DNS redundancy

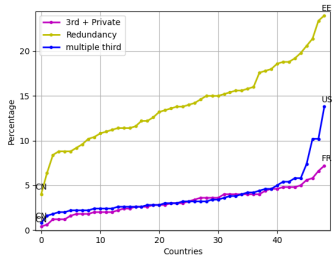


Fig. 6. Line plot of redundant, multiple third-party, and both third-party and private DNS.

We then identify the level of dependency of websites on the five most popular DNS providers across countries. Table 1 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 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 1. Top-5 DNS, CDN and CA providers across countries (average percentage of websites).

DNS providers and the Czech Republic has the least DNS centralization (60.9%). Table 2 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.

We compare our findings for the US top-100 and top-500 sites with Kashaf et al.'s [38] two years later. We observe an increase in third-party dependency, redundantly provisioned websites, and websites using multiple third-party providers but no change in critical dependency and the percentage of websites using both third and private DNS providers. The details of this analysis are included in (§A).

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 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.

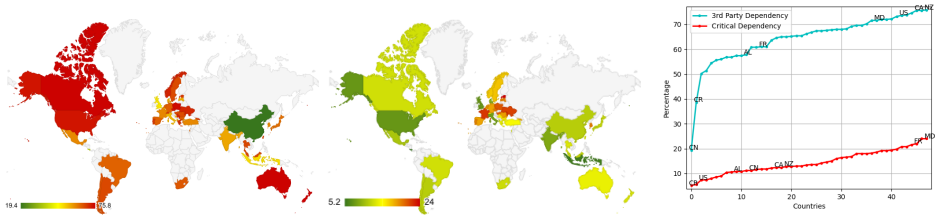
Region	DNS Providers	Avg	Std Dev	CA Providers	Avg	Std Dev	CDN Providers	Avg	Std Dev
The Americas	Amazon Cloudflare Akamai	79.1	4.2	DigiCert GlobalSign Comodo	69.1	2.7	Google, Akamai, Amazon Cloudfront	90.8	9.0
Europe	Cloudflare Amazon Akamai	74.6	6.5	DigiCert IdenTrust Inc. Comodo	70.6	5.3	Google, Akamai, 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 2. Top three DNS, CA and CDN providers with their corresponding market share per region (average percentage of websites).

4.2 CDN Findings

Figure 7(a) and 7(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 7(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 7 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

Fig. 7. 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 8 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 9 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 (The map plots of these are shown in §A)). 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%.

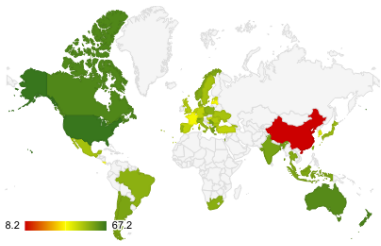


Fig. 8. Plot of CDN redundancy

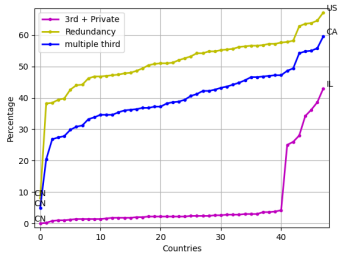


Fig. 9. Line plot of redundant, multiple third-party, and both third-party and private CDN.

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 1 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 2 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.

In the case of CDNs, we observe a decrease in critical dependency two years later with top-100 US Alexa sites compared to Kashaf et al. [38]. Whereas more websites are now dependent on a third-party provider, are redundantly provisioned, and use multiple third-party CDNs (§A).

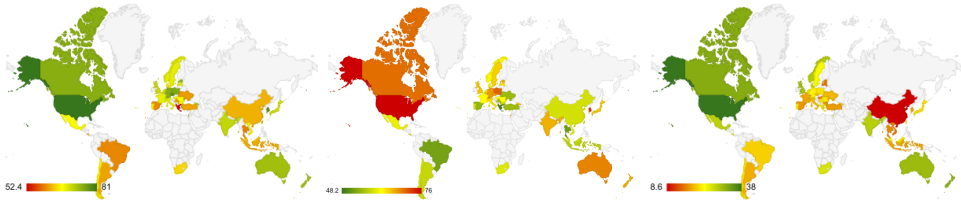
Summary. We observe that more than half of the countries have more than 68% of their regional sites dependent on a third-party CDN provider and more than 14% critically dependent on a third-party CDN. We see more redundancy (and therefore lower critical dependency) on CDN providers compared to DNS providers 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).

4.3 CA Findings

In the case of CA, Fig. 10 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 [24, 60]. 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 [24]. None of the websites in our set support OCSP must-staple.



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

Fig. 10. 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 1 shows the average percentage of websites relying on each CA provider across the different countries.

Table 2 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.

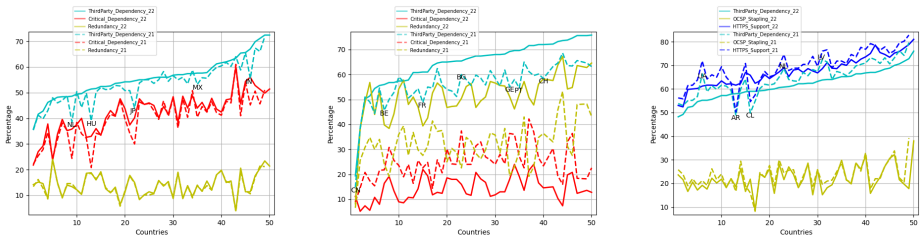
We observe that, for top-100 US Alexa sites, HTTPS support for CAs is the same as Kashaf et al.'s [38] report, even two years later. However, the support for OCSP stapling and third-party dependency has increased as shown in (§A).

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.

4.4 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 [13, 15, 25, 31].

5 SNAPSHOT MEASUREMENTS



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

Fig. 11. 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 11 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 11(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 11(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 11(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).

6 ANALYSING TRENDS

Our analysis reveals a wide range of third-party dependencies across the three services we look at: DNS, CA, and CDN. In this section, we evaluate several factors that may explain part of the variation observed. We specifically look at countries' level of economic development, Internet development, the set of economic trading partners along with the top websites' categories, home country and ranking. We analyze the correlation between a country's third-party dependency and these different factors, fully aware that no factor alone could explain the degree of third-party service dependency.

Consolidation of third-party services is often seen as an expected result of maturing markets and industries [7]. We thus focus our analysis on factors that capture economic development and Internet development. We adopt a subset of indices from Katz et al. [39], those for which current rankings are available for regional analysis. In addition, we look at correlations among trading partners under the hypothesis that the degree of third-party dependency on the web for a given country is closely related to that of its main trading partners, given their shared interests.

In our analysis, we compute the Pearson correlation coefficient and the p-value for each index with the mean of the DNS, CA and CDN third-party dependency. The coefficient gives a value between -1 and 1 with values between 0 and 0.3 indicating a weak correlation, between 0.3 and 0.7 indicating a moderate positive correlation, and above 0.7 indicating a strong positive correlation. Negative values similarly indicate negative correlations. We use a significance value (p-value) of 0.05. Table 3 (§A) summarizes the coefficients and p-value computed for each indicator across the regions.

6.1 Economic Development

To understand the relationship between the economic conditions of a country and the country's third-party dependency we look at Gross Domestic Product (GDP) Per Capita [41] and Economic Freedom [32] indicators. GDP is a measure of the country's economic output per person, while the Index of Economic Freedom aims to capture the ability of a society to take economic actions.

The value of the correlation coefficient between GDP and the third-party dependency, both in general and across regions, is moderate to strong positive except for Europe. Overall, we find a correlation coefficient of 0.47(all countries), but the correlation coefficient for the Americas is 0.90, for the Asia Pacific is 0.63 and for Europe is 0.30. All with p-values lower than 0.05 except for Europe.

In the case of Economic Freedom, we also find a moderate to strong correlation again across countries and regions with the exception of Europe. The correlation value for all countries is 0.46, while the correlation coefficient for the Asia Pacific is 0.80, for the Americas is 0.52 and for Europe is 0.30. However, the p-values for Europe and the Americas are higher than 0.05, indicating that the correlation between Economic Freedom and third-party dependency for these regions is not statistically significant.

6.2 Internet Development

A mature, developed Internet economy may partially determine the degree of third-party dependency and centralization among its most popular websites. To explore this, we use the World

Economic Forum's Networked Readiness Index (NRI) [52] and the International Telecommunications Union's ICT Development Index (IDI) [11] as indicators of the development of the Internet in a region. NRI also referred to as Technology Readiness measures the propensity of countries to utilize the opportunities provided by information and communication technology. On the other hand, IDI measures the digital divide and ICT performance across countries.

The correlation coefficient values between either factor, NRI and IDI, and third-party dependency are lower than those for the economic factors we analyze, with 0.42 for NRI and 0.35 for IDI. Per region, the correlation coefficient of NRI is 0.90, 0.49, and 0.32, for the Americas, Asia Pacific and Europe respectively. The correlation coefficient of IDI for the Americas is 0.77 and again lower for Asia Pacific (0.43) and Europe (0.27). For both these indicators, the p-values for Europe and the Asia Pacific are higher than 0.05, indicating that the correlation between these two regions is not statistically significant.

6.3 Trading Blocs

Trading blocs comprise countries within a specific geographical boundary that have relations to secure regional economic growth. We hypothesize that the degree of third-party dependency is closely related to that of its main trading partners, given their shared interests and potential common infrastructures.

To test this hypothesis, we explore the association between trading partnerships, as captured by regional trading blocks, and third-party dependencies. We focus on the seven major regional trading blocs [4] in the world economy and group countries in our dataset accordingly (Table 5).

Figure 12 shows the third-party dependency of the countries in each trading block. Countries in the NAFTA block (CA, MX, US) have the maximum dependence and countries in the MERCOSUR (AR, BR) and LAIA block (AR, BR, CL, MX) have the least third-party dependence. If trading partners were to have a significant role in determining a country's third-party dependency, we would expect little variability within the group. This is indeed the case for most trading blocks. The one exception is the APEC block, although we note that this block, which includes the US, China, Mexico, and Australia, has a particularly diverse set of countries in terms of economies, primary language and continents. For instance, the US has the maximum third-party dependency and high GDP and Economic Freedom in that bloc and China has the lowest third-party dependency and also low GDP and Economic Freedom explaining the high variability within that bloc.

6.4 Websites' Content Category

It is possible that sites within the same category share a similar level of dependency on third-party services. To explore this, we rely on the McAfee SmartFilter Internet Database service [48, 64] to retrieve content categories of all the unique websites in our dataset. We identify 93 different categories in total. We then cluster websites by categories and identify those categories, per country, with the largest percentages of third-party dependencies. Figure 13 shows, on the x-axis, the subset of website categories with the highest fraction of third-party dependencies. The y-axis shows the 50 countries in our dataset, ordered by their third-party dependency (the highest in the US and the least in China). For each category and country, the size of the bubble represents the percentage of total sites of that country that belong to the given category and have a third-party dependency.

Previous work [62] has shown that News websites utilize more third parties (such as ad services) than other categories. We also see that the News websites along with the Business websites have higher third-party dependency across all countries. However, some countries show a different behavior, for instance in Portugal most third-party sites come from the Games category and in Taiwan from the Blogs category. In this analysis, we did not find a different pattern by dividing the countries based on regions.

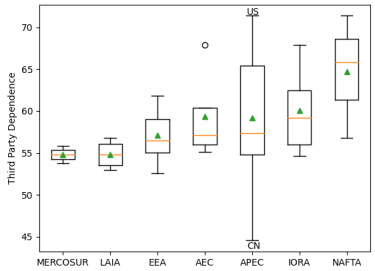


Fig. 12. Box plot of third-party dependency of countries in each trading bloc.

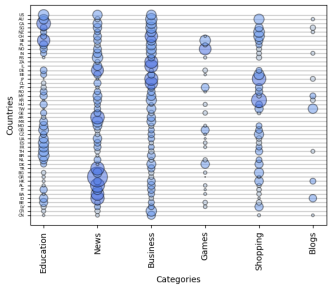
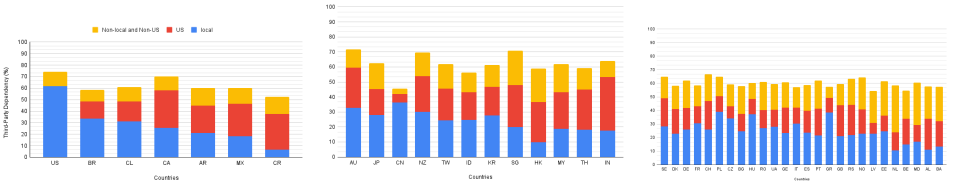


Fig. 13. Plot of the categories with the most third-party dependency across the countries.

6.5 Websites' Home Country

We also investigate if there is a relationship between the home country of a website and its third-party dependency. To find the home country of a website, we first use the country code top-level domain of the website and if that is not available we perform a WhoIs look-up to find the registrant organization's country [62, 65]. We find that the websites in our dataset come from 146 different countries.

For each of the 50 countries in our dataset, we then check if the websites of that country are homed locally, in the US, or non-locally[6]. We divide the countries into their specific regions, i.e. the Americas, Asia Pacific and Europe. Figure 14. shows for each country in the given region, the percentage of total sites from that country that have a third-party dependency and are homed locally, in the US, or non-locally. For most countries in the Americas, we see that the majority of the third-party sites come from the US. For all countries in Asia Pacifica, the majority of the third-party sites are either homed locally or in the US. Thus, for the Americas and the Asia Pacific, a higher third-party dependence can be attributed to the sites registered in the US. Whereas in the case of Europe most content is locally homed, only in a few countries it is dominated by non-local websites and only the UK has the highest percentage of third-party sites coming from the US. We also note that the countries listed in (§4.4) that have the highest third-party dependence across the three services, except for Japan, also have a high proportion of the third-party sites coming from the US.



(a) The Americas (b) Asia Pacific (c) Europe

Fig. 14. Plot showing the relation between the home of the websites and third-party dependency across regions.

[6]We refer to the websites as non-local if they are not homed in the country of their ranking or in the US .

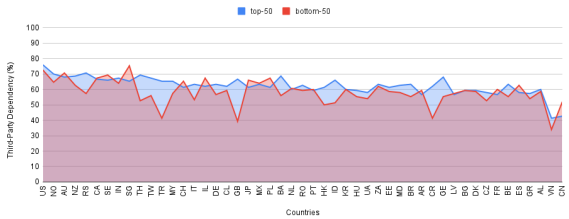


Fig. 15. Third-party dependency of top-50 and bottom-50 sites of each country.

6.6 Website Ranking

To determine if the traffic skewness of the top-500 sites corresponds to the adoption skewness of third parties, we use the ranking of the site as a proxy for the amount of traffic they receive. In this analysis, for each country, we measure the percentage of top-50 and bottom-50 sites, from the top-500 sites, that have a third-party dependency. Figure 15. shows that the top-50 sites for most countries have a relatively higher third-party dependence than the bottom-50 sites. We hypothesize that this could be attributed to the top sites receiving more traffic and so they are more likely to rely on third-party infrastructure for reliability.

Summary. In conclusion, as one would expect, we find that there is no single factor that could explain the differences across countries. However, our preliminary analysis shows that all of these factors partially explain some of the differences and similarities in third-party dependencies across countries. For instance, third-party dependency is strongly correlated with several factors such as GDP, NRI, and IDI in the case of the Americas, with low p-values. In the case of Asia, there is a strong correlation only with the Economic Freedom Indicator, while for Europe all indicators show a high p-value. To explore this, we divide Europe into four regions and compute the correlations and corresponding p-values, but find little improvements mostly for the case of Northern Europe. We also analyzed other indices such as the Global Connectivity Index, Digital Infrastructure, Knowledge Economic Index and Internet Penetration [9, 36, 56, 61]. All of these gave a high p-value for all regions and when looking at all countries, thus rejecting the null hypothesis that these indicators correlate to third-party dependency, except for the Americas where the correlation coefficient was approximately 0.9 for all these indicators with low p-values. Finally, we hypothesize that the consistent low correlation values between the different factors and third-party dependency in Europe can be explained by the low variability across all factors we study (e.g., similar values for GDP, degree of Economic Freedom and Internet development), which will naturally result in a low linear correlation.

7 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. [57] 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 [53] 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 [30]. 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 [57], 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. We leave this as future work.

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.

8 RELATED WORK

In recent years there has been a move towards Internet centralization and recent work has begun to analyze this in the context of DNS [14, 70], cloud providers [51], and the Web in general [27, 38]. A body of work has concentrated particularly on DNS centralization. This body of work studies the implications and proposes solutions to address the move toward centralization, which we show in this work as happening across services on the web and countries around the world to different degrees. Arkko [16] discusses the implications of the increasing trend towards DNS centralization and encourages that DNS issues be solved without relying on a single DNS service. Livingood [45] elaborates on the risks associated with the adoption of the latest secure variant of DNS i.e. DNS over HTTPS (DoH), because of its centralized nature and along with making recommendations to address that in the DoH implementation encourages further study before widespread adoption. Hoang et al.[33] propose a DNS resolution mechanism that shards DNS queries across multiple DoH resolvers to reduce the number of user queries going to a particular DNS resolver and therefore reducing their reliance on any single third-party DNS provider.

Urban et al. [64] study the third parties embedded in a website and find that 93% of the websites are located in regions that might not be in line with the current legal framework. Similarly, Sørensen et al. [62] uses a subset of European and US websites from different website categories, known to

have either a higher or lower number of third-party URLs, and analyzes fluctuations in the use of third parties before and after GDPR. Both these works report the News category (e.g. ad services) to have the most third-party websites. We use additional heuristics beyond 2LD matching to identify third-party services, but as with these prior efforts, we find News, along with Business, websites to have the most third-party dependency. Moura et al. [51] measure centralization by analyzing the DNS traffic received by the two ccTLDs and B-root and the ASes responsible for generating that traffic. They find that 5 large Cloud Providers (CPs), from only 20 ASes, are responsible for 30% of the queries to the two ccTLDs over the 3-year time period they measure. These CPs include Microsoft, Google, Facebook, Cloudflare, and Amazon. Radu et al. [55] also provide an analysis of the trends in consolidation in the recursive DNS resolvers market and focus on consolidation shifts from 2016 to 2019. They report Google and Cloudflare as the top providers and report an increase in Google's coverage by 20.52% and Cloudfalre's by 14% between 2016 and 2019. Both analyses focus on centralization at the DNS resolver end, whereas our analysis focuses on server-side centralization (centralization around the DNS authoritative servers, CAs and CDNs used by the top websites). Bates et al. [19] analyze consolidation trends in DNS from 2011 to 2017 and the extent of redundancy at the authoritative DNS side given the current rise of cloud-based hosting and infrastructure. They find that following the Dyn incident in 2016, the number of domains using only a single nameserver reduced from 92.2% in 2016 to 87.3% in 2017. They also report an increase in the share held by top-4 providers from 17% to 49.8% and Dyn, Akamai, Amazon and Cloudflare as the top-4 providers. They finally noted that two-thirds of Dyn users switched to Amazon or NsOne between 2016 and 2017. Their study is based on a sample of the top-1000 US Alexa domains in a subset of Top-Level Domains (TLDs). We look instead at global trends from regional popular websites. Doan et al. [27] focuses on web consolidation around Content Delivery Infrastructures (CDI), observing that the fraction of webpages hosted on CDIs have increased by 83% between 2015 and 2020 (from roughly 8.2% to 15%), with higher penetration among the more popular pages. They report Google having the highest CDI (Content Delivery Infrastructures) penetration ratio and find Google, Akamai, Amazon and Cloudflare as the top providers with the highest CDI penetration but through a broader global analysis. In our study, we focus instead on regional trends toward third-party dependency and centralization across DNS, CDN and CA. Kashaf et al. [38] measure the degree of direct and indirect third-party dependence of top global 100k Alexa websites on DNS, CA, and CDN infrastructure. Our work partially builds on this prior effort, but focuses on trend variations across countries and regions of the world and explores factors that may help explain the observed differences.

In ongoing work, we are exploring the relationship between global trends toward third-party usage and centralization, and the deployment of third-party services infrastructure. There have been several attempts to study the infrastructure of DNS, CDN, and hosting services, including [12, 22, 23, 28, 43, 69], that have informed this effort.

9 CONCLUSIONS

We presented the first large-scale study of third-party dependency and centralization around the world. We used vantage points in 50 countries, across all inhabited continents, and focused on the most popular regional websites. We find a wide range of third-party service dependencies across countries, partially correlated to economic development, degree of economic freedom and Internet development. We also find that News and Business websites and higher-ranked websites have a higher degree of third-party dependence across most countries. Additionally, countries that have a higher percentage of third-party dependency also have more sites hosted in the US. Despite this high variability, our results suggest a highly concentrated market of third-party providers and, perhaps more problematically, increasing levels of dependency and centralization only a year

later. Beyond expanding our analysis, there are a number of promising directions for future work including the exploration of combined indices that may capture the complex socio-economic factors that may determine a country's third-party dependence. This work has shown that there is value in a country-level analysis of Internet infrastructure dependencies that a broader global analysis would miss. We have automated our analysis tool and plan to make available periodic checkpoints of these trends around the world using similarweb's regional website ranking for future analysis.

ACKNOWLEDGMENTS

The authors thank the shepherd Zihui Ge and anonymous reviewers for their valuable comments and feedback. We are also grateful to Marcel Flores, Esteban Carisimo, Hernan Galperin and Aqsa Kashaf for constructive conversations and feedback during the work. This work was partially supported by a Comcast Innovation Fund Award.

REFERENCES

[1] 2012. (2012). https://nordvpn.com
[2] 2016. Dyn Analysis Summary of Friday October 21 Attack. (October 2016). http://dyn.com/blog/dyn-analysis-summary-of-friday-october-21-attack/
[3] 2016. Globalsign certificate revocation issue. (October 2016). https://www.globalsign.com/en/status
[4] 2018. (2018). https://www.exportgenius.in/blog/10-major-regional-trading-blocs-of-the-world-236.php
[5] 2019. Azure global outage: Our DNS update mangled domain records, says Microsoft. https://www.zdnet.com/article/azure-global-outage-our-dns-update-mangled-\domain-records-says-microsoft/. (2019).
[6] 2019. How Verizon and a BGP Optimizer Knocked Large Parts of the Internet Offline Today. (June 2019). https://blog.cloudflare.com/how-verizon-and-a-bgp-\optimizer-knocked-large-parts-of-the-internet-offline-today/.
[7] 2020. Consolidation in the Internet Economy. (Feb 2020). https://future.internetsociety.org/2019/
[8] 2020. GoDaddy (all of it) went down this evening. (November 2020). https://domainnamewire.com/2020/11/17/godaddy-is-down/
[9] 2021. (Oct 2021). https://en.wikipedia.org/wiki/Knowledge_Economic_Index
[10] 2022. (May 2022). https://github.com/cisagov/findcdn
[11] 2022. (Mar 2022). https://en.wikipedia.org/wiki/ICT_Development_Index
[12] Bernhard Ager, Wolfgang Mühlbauer, Georgios Smaragdakis, and Steve Uhlig. 2011. Web Content Cartography.
[13] Akamai. 2022. Akamai CDN Deployment. (2022). https://seekingalpha.com/article/4379686-akamai-granddaddy-of-cdn-is-well-positioned-for-next-generation-applications
[14] Mark Allman. 2018. Comments on DNS Robustness.
[15] Amazon. 2022. Amazon CDN Deployment. (2022). https://aws.amazon.com/cloudfront/features/?whats-new-cloudfront.sort-by=item.additionalFields.postDateTime&whats-new-cloudfront.sort-order=desc
[16] Jari Arkko. 2019. Centralised Architectures in Internet Infrastructure. *IETF Internet Draft* (2019).
[17] Jari Arkko. 2020. The influence of Internet architecture on centralised versus distributed Internet services. *Journal of Cyber Policy* 5, 1 (2020), 30–45.
[18] Subin B. 2021. 7 Best Alexa.com Alternatives for Website Ranking and Traffic Analysis. (Dec 2021). https://beebom.com/best-alexa-com-alternatives/
[19] Samantha Bates, John Bowers, Shane Greenstein, Jordi Weinstock, Yunhan Xu, and Jonathan Zittrain. 2021. 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). https://doi.org/10.51685/jqd.2021.011
[20] Alexandros Biliris, Chuck Cranor, Fred Douglis, Michael Rabinovich, Sandeep Sibal, OliverSpatscheck, and Walter Sturm. 2002. *Computer Communications* 25 (March 2002). Issue 4.
[21] Michael Butkiewicz, Harsha V. Madhyastha, and Vyas Sekar. 2011. Understanding Website Complexity: Measurements, Metrics, and Implications.
[22] Matt Calder, Xun Fan, Zi Hu, Ethan Katz-Bassett, John Heidemann, and Ramesh Govindan. 2013. Mapping the Expansion of Google's Serving Infrastructure.
[23] Patricia Callejo, Rubén Cuevas, Narseo Vallina-Rodriguez, and Ángel Cuevas. 2019. Measuring the Global Recursive DNS Infrastructure: A View From the Edge. *IEEE Access* 7 (10 2019), 1–1. https://doi.org/10.1109/ACCESS.2019.2950325
[24] Taejoong Chung, Jay Lok, Balakrishnan Chandrasekaran, David Choffnes, Dave Levin, Bruce M. Maggs, Alan Mislove, John Rula, Nick Sullivan, and Christo Wilson. 2018. Is the Web Ready for OCSP Must-Staple? *Proceedings of the Internet Measurement Conference 2018* (Oct 2018). https://doi.org/10.1145/3278532.3278543

[25] Cloudflare. 2022. Cloudflare. (2022). https://www.cloudflare.com/network/
[26] David Coldewey. 2020. Cloudflare DNS goes down, taking a large piece of the Internet with it. TechCrunch Blog. (July 2020). http://tcrn.ch/3pbDJzL
[27] Trinh Viet Doan, Roland van Rijswijk-Deij, Oliver Hohlfeld, and Vaibhav Bajpai. 2022. An Empirical View on Consolidation of the Web. *ACM Transactions on Internet Technology* 22, 3 (Aug 2022), 1–30. https://doi.org/10.1145/3503158
[28] Rodérick Fanou, Gareth Tyson, Eder Leao Fernandes, Pierre Francois, Francisco Valera, and Arjuna Sathiaseelan. 2018. Exploring and Analysing the African Web Ecosystem. *ACM Trans. Web* (2018), 26. https://doi.org/10.1145/3213897
[29] Petros Gigis, Matt Calder, Lefteris Manassakis, George Nomikos, Vasleois Kotronis, Xenofontas Dimitropoulos, Ethan Katz-Bassett, and Georgios Smaragdakis. 2021. Seven years in the life of Hypergiants' off-nets.
[30] Google. 2022. Chrome User Experience Report | Chrome UX Report |Google Developers. (2022). https://developers.google.com/web/tools/chrome-user-experience-report
[31] Google. 2022. understanding-google-cloud-network-edge-points. (2022). https://cloud.google.com/blog/products/networking/understanding-google-cloud-network-edge-points
[32] Heritage. 2022. Economic Freedom. (2022). https://www.heritage.org/index/explore?view=by-region-country-year&u=637879938517599600
[33] Nguyen Phong Hoang, Ivan Lin, Seyedhamed Ghavamnia, and Michalis Polychronakis. 2020. K-resolver: towards decentralizing encrypted DNS resolution. *arXiv preprint arXiv:2001.08901* (2020).
[34] Geoff Huston. 2019. DNS Resolver Centrality. APNIC Blog. (September 2019). https://labs.apnic.net/?p=1260
[35] Geoff Huston. 2021. CDN and centrality. APNIC Blog. (July 2021). https://blog.apnic.net/2021/07/02/opinion-cdns-and-centrality/
[36] IMD. 2021. World Digital Competitiveness Rankings - IMD. (2021). https://www.imd.org/centers/world-competitiveness-center/rankings/world-digital-competitiveness/
[37] IP2Location. [n. d.]. Free IP Geolocation Database. ([n. d.]). https://lite.ip2location.com/
[38] Aqsa Kashaf, Vyas Sekar, and Yuvraj Agarwal. 2020. Analyzing Third Party Service Dependencies in Modern Web Services: Have We Learned from the Mirai-Dyn Incident?
[39] Raul Katz and Fernando Callorda. 2018. Accelerating the development of Latin American digital ecosystem and implications for broadband policy. *Telecommunications Policy* 42, 9 (Oct 2018), 661–681. https://doi.org/10.1016/j.telpol.2017.11.002
[40] Mohammad Taha Khan, Joe DeBlasio, Geoffrey M. Voelker, Alex C. Snoeren, Chris Kanich, and Narseo Vallina-Rodriguez. 2018. An Empirical Analysis of the Commercial VPN Ecosystem *(IMC '18)*. Association for Computing Machinery, New York, NY, USA, 443–456. https://doi.org/10.1145/3278532.3278570
[41] Avery Koop. 2021. Mapped: GDP per Capita Worldwide. (July 2021). https://www.visualcapitalist.com/mapped-gdp-per-capita-worldwide/
[42] Deepak Kumar, Zane Ma, Zakir Durumeric, Ariana Mirian, Joshua Mason, J. Alex Halderman, and Michael Bailey. 2017. Security Challenges in an Increasingly Tangled Web.
[43] Zhenyu Li, Donghui Yang, Zhenhua Li, Chunjing Han, and Gaogang Xie. 2018. Mobile Content Hosting Infrastructure in China: A View from a Cellular ISP. *Passive and Active Measurement Lecture Notes in Computer Science* (2018), 100–113.
[44] Mozilla Public Suffix List. [n. d.]. Public Suffix List. ([n. d.]). https://publicsuffix.org/
[45] J. Livingood, M. Antonakakis, B. Sleigh, and A. Winfield. 2019. Centralized DNS over HTTPS (DoH) implementation issues and risks. (2019).
[46] Srdjan Matic, Gareth Tyson, and Gianluca Stringhini. 2019. PYTHIA: a Framework for the Automated Analysis of Web Hosting Environments. *The World Wide Web Conference* (2019).
[47] Maxmind. 2022. MaxMind Server IP Addresses. (2022). https://dev.maxmind.com/geoip/geolite2-free-geolocation-data
[48] Mcafee. 2022. Mcafee sitelookup. (2022). https://sitelookup.mcafee.com/en/feedback/url
[49] Foivos Michelinakis, Hossein Doroud, Abbas Razaghpanah, Andra Lutu, Narseo Vallina-Rodriguez, Phillipa Gill, and Joerg Widmer. 2018. The Cloud that Runs the Mobile Internet: A Measurement Study of Mobile Cloud Services. https://doi.org/10.1109/INFOCOM.2018.8485872
[50] Chance Miller. 2021. PSA: Facebook, Instagram, Messenger, and WhatsApp went down for 6 hours; here's why [U]. (Oct 2021). https://9to5mac.com/2021/10/04/instagram-facebook-whatsapp-down/
[51] Giovane Moura, Sebastian Castro, Wes Hardaker, Maarteb Wullink, and Cristian Hesselman. 2020. Clouding up the Internet: how centralized is DNS traffic becoming?
[52] NRI. 2022. Network Readiness Index. (2022). https://networkreadinessindex.org/countries/#ranking-wrapper
[53] Victore Le Pochat, Tom Van Goethem, Samaneh Tajalizadehkhoob, Maciej Korczynski, and Wouter Joosen. 2019. Tranco: A Research-Oriented Top Sites Ranking Hardened Against Manipulation.

[54] Ingmar Poese, Steve Uhlig, Mohamed Ali Kaafar, Benoit Donnet, and Bamba Gueye. 2011. IP Geolocation Databases: Unreliable? 41, 2 (2011).
[55] Roxana Radu and Michael Hausding. 2020. Consolidation in the DNS resolver market – how much, how fast, how dangerous? *Journal of Cyber Policy* 5 (02 2020), 1–19. https://doi.org/10.1080/23738871.2020.1722191
[56] Global Rankings. 2022. Connectivity Index. (2022). https://www.mobileconnectivityindex.com/#year=2021&globalRankings=overall&globalRankingsYear=2021
[57] Quirin Scheitle, Oliver Hohlfeld, Julien Gamba, Jonas Jelten, Torsten Zimmermann, Stephen D. Strowes, and Narseo Vallina-Rodriguez. 2018. A Long Way to the Top. (2018).
[58] Rachee Singh, Arun Dunna, and Phillipa Gill. 2018. Characterizing the Deployment and Performance of Multi-CDNs.
[59] Ankit Singla, Balakrishnan Chandrasekaran, P. Brighten Godfrey, and Bruce Maggs. 2014. The Internet at the Speed of Light.
[60] OCSP Stapling. 2017. The Problem with OCSP Stapling and Must Staple and why Certificate Revocation is still broken - Hanno's blog. (2017). https://blog.hboeck.de/archives/886-The-Problem-with-OCSP-Stapling-and-%5CMust-Staple-and-why-Certificate-Revocation-is-still-broken.html
[61] Statistics. 2022. Internet Pentrations. (2022). https://www.statista.com/statistics/227082/countries-with-the-highest-internet-penetration-rate/
[62] Jannick Sørensen and Sokol Kosta. 2019. 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). https://doi.org/10.1145/3308558.3313524
[63] Timlib. [n. d.]. WebXray Domain Owner List. ([n. d.]).
[64] Tobias Urban, Martin Degeling, Thorsten Holz, and Norbert Pohlmann. [n. d.]. Beyond the Front Page: Measuring Third Party Dynamics in the Field. ([n. d.]). https://doi.org/10.1145/3366423.3380203
[65] Whois. 2022. Whois. (2022). https://whois.icann.org/en
[66] world population. 2022. internet-users-by-country. (2022). https://worldpopulationreview.com/country-rankings/internet-users-by-country
[67] Jing'an Xue, David Choffnes, and Jilong Wang. 2017. CDNs Meet CN An Empirical Study of CDN Deployments in China. *IEEE Access* 5 (2017), 5292–5305. https://doi.org/10.1109/ACCESS.2017.2682190
[68] Bahador Yeganeh, Ramakrishnan Durairajan, Reza Rejaie, and Walter Willinger. 2020. A First Comparative Characterization of Multi-cloud Connectivity in Today's Internet.
[69] Hao Yin, Bo Qiao, Yan Luo, Li Ruyue, and Y. Yang. 2015. Demystifying commercial content delivery networks in China. *Concurrency and Computation: Practice and Experience* 27 (06 2015). https://doi.org/10.1002/cpe.3464
[70] Luciano Zembruzki, Arthur Selle Jacobs, Gustavo Spier Landtreter, Lisandro Zambenedetti Granville, and Giovanne Moura. 2020. `dnstracker`: Measuring Centralization of DNS Infrastructure in the Wild. In *Proc. of AINA*.

A APPENDIX

A.1 Ethics

This work does not raise any ethical issues.

A.2 Countries and Country Codes

Set of countries included in our analysis, their two-letter code and their percentage of the world's Internet population.

Country Code	Country	Internet Pop (%)
AE	United Arab Emirates	0.19
AL	Albania	0.05
AR	Argentina	0.72
AU	Australia	0.45
BA	Bosnia and Herzegovina	0.05
BE	Belgium	0.22
BG	Bulgaria	0.10
BR	Brazil	3.43
CA	Canada	0.73
CH	Switzerland	0.17
CL	Chile	0.32
CN	China	21.22
CR	Costa Rica	0.08
CZ	Czech Republic	0.18
DE	Germany	1.67
DK	Denmark	0.12
EE	Estonia	0.02
ES	Spain	0.91
FR	France	1.28
GB	United Kingdom	1.40
GE	Georgia	0.07
GR	Greece	0.17
HK	Hong Kong	0.14
HU	Hungary	0.16
ID	Indonesia	4.56
IL	Israel	0.15
IN	India	17.89
IT	Italy	1.08
JP	Japan	2.50
KR	South Korea	1.06
LV	Latvia	0.03
MD	Moldova	0.07
MX	Mexico	1.91
MY	Malaysia	0.54
NL	Netherlands	0.34
NO	Norway	0.11
NZ	New Zealand	0.10
PL	Poland	0.74
PT	Portugal	0.16
RO	Romania	0.27
RS	Serbia	0.13
SE	Sweden	0.21
SG	Singapore	0.10
TH	Thailand	0.78
TR	Turkey	1.33
TW	Taiwan	0.47
UA	Ukraine	0.63
US	United States	6.7
VN	Vietnam	1.56
ZA	South Africa	0.68

A.3 Regions and Countries

Regions used in this work and countries from our dataset in those regions.

Region	Country Codes
The Americas	US CA BR AR MX CR CL
N.Europe	GB SE DK NO LV EE
W.Europe	DE NL FR CH BE
E.Europe	PL CZ BG HU RO UA MD GE
S.Europe	IT ES RS PT GR AL BA
Asia Pacific	AU NZ SG JP HK CN TW VN ID MY KR TH IN
Africa and Middle East	ZA IL TR AE

A.4 Heuristics for Third-party Dependency Analysis

Algorithm 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 and 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
```

Algorithm 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) and 2ld(service.url) in SANList then
    service.type ← private
else if SOAProvider(w) ≠ SOAProvider(service.url) then
    service.type ← third
end if
return service.type
```

A.5 Correlation of sites using HTTPS and Third-Party CA

Support for HTTPS and use of a third-party CA are strongly positively correlated, as seen in Fig. 16. On average, nearly all (96.8%) of HTTPS-supporting websites also use a third-party CA, with a standard deviation of only 2% and no outliers.

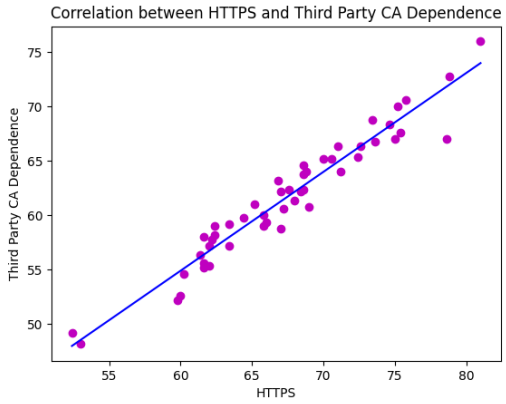
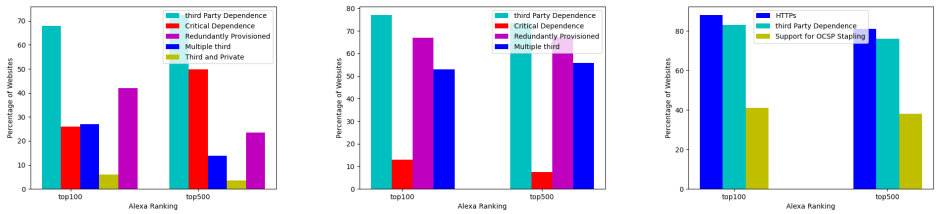


Fig. 16. Scatter plot of countries top-500 website using HTTPS (x-axis) and percentage sites using a third-party CA (y-axis).

A.6 US Dependency and Centralization



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

Fig. 17. DNS, CDN and CA metrics for top-100 and top-500 sites in the US

We look at dependency and centralization with the top-100 and top-500 US Alexa sites and compare our findings to those Kashaf et al. [38], and analyze the trends in dependency two years later. Looking at both top-100 and top-500 makes it easier to compare our findings to those Kashaf et al. [38] as they use top-100 and top-1000 sites in their analysis.

Figure 17(a) is a bar graph of dependency and redundancy in DNS for the top-100 and top-500 websites. We observe an increase in third-party dependency from 48% to 68% in the top-100 sites. However, the percentage of websites being redundantly provisioned by DNS providers has increased from 21% to 42%. There is no change in critical dependence and the percentage of third and private DNS providers. We see similar trends between top-100 to top-500 as Kashaf et al. [38] shows from top-100 to top-1000 Alexa sites, i.e. third-party dependency increases(68% to 72.4%) and critical dependence increases (26% to 49.8%), whereas we see a decrease in websites redundantly provisioned(42% to 23.4%) and served by multiple third-party providers(27% to 13.8%).

Figure 17(b) shows third-party and critical dependencies, and the fraction of those redundantly provisioned and using multiple third-party CDNs among the top-100 and top-500 websites. For the top-100 websites, we observe more websites are now redundantly provisioned by CDNs (32% to 67%) and more use multiple third-party CDNs(32% to 53%). There is a slight increase in third-party

dependency (76% to 77%) but a significant decrease in critical dependence (43% to 13%) two years later.

Lastly, Fig. 17(c) shows the degree of HTTPS support, third-party dependency and support for OCSP stapling for the top-100 and top-500 sites in the US. In this case we observe that our measurements for HTTPS support (88%) match that of Kashaf et al.'s [38] paper for Alexa top-100 sites, however third-party dependency and the support for OCSP stapling has increased two years later (71% to 83% and 21% to 41%).

A.7 Indirect Dependencies

Apart from direct dependence of websites on third-party infrastructures like DNS, CA and CDN, these services also rely on other third-party services. So even if a website relies on a private DNS, CA or CDN service but if, for instance, a private CDN uses a third-party DNS then the website is also now (indirectly) dependent on a third-party service.

To measure these transitive dependencies, like Kashaf et. al [38], we look at (*i*) CDN→DNS dependence, (*ii*) CA→CDN dependence and (*iii*) CA→DNS dependence across our countries for the top-500 regional sites.

To measure CDN→DNS dependence, we collect a list of CDNs used by the top-500 regional sites in each country and find their CNAMEs using the CDN-CNAME map. Then we find the nameserver of each cname and to classify the CDN as private or third-party, we use the heuristics described in Algorithm 2 (§A), where the cname is passed as *w* and each nameserver is passed as *service.url*.

We do the same for CA→DNS dependence, we collect all CAs used by the top-500 regional sites in each country and find their CA urls. Then we find the nameserver of each CA url and classify the CA as private or third-party. To do this, we again use the heuristics described in Algorithm 2, where the CA url is passed as *w* and each nameserver is passed as *service.url*.

Finally, for CA→CDN dependence, we use the CAs collected and first find the CNAMEs of the CA urls. We then find the CDNs from the CNAMEs using the CDN-CNAME map and classify them as private or third using the technique in (§3.2.2)

When measuring transitive dependencies between
CDN− >DNS infrastructure, we find 60 CDNs, on average, across all our vantage points and 30.3% of them, on average, are dependent on a third-party DNS. With this indirect dependency, 26 more websites are now dependent on third-party DNS.

Additionally, we find a total of 68 CAs across all countries and an average of 26 CAs per country. Measuring CA− >DNS dependency, shows that on average 17% of CAs depend on third-party DNS and measuring CA− >CDN dependency shows 26.3% of these CAs depend on a third-party CDN.

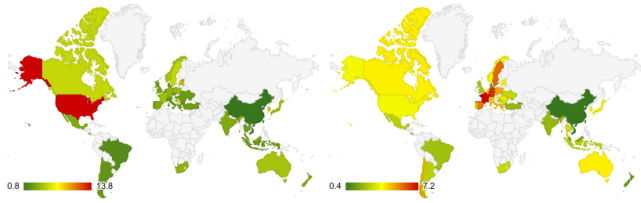
Given our smaller dataset across countries, the number of additional websites depending on third-party services due to the indirect dependecies of CA on DNS and CDN did not increase significantly.

A.8 Redundancy Maps

Figure 18 plots a map showing the distribution of top-500 sites using multiple third-party and third-party and private DNS. Figure 19 plots a map showing the distribution of top-500 sites using multiple third-party and third-party and private CDNs.

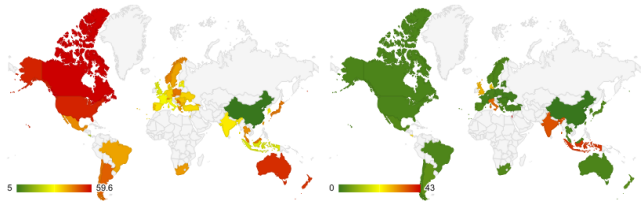
A.9 Correlation Coefficients and P-Values

Table 3 summarizes the coefficients and p-value computed for each socio-economic indicator across the regions and Table 4 provides a summary for each region of Europe .



(a) Multiple Third Party (b) Third-party and Private

Fig. 18. Plot of multiple third-party, and both third-party and private DNS.



(a) Multiple Third Party (b) Third-party and Private

Fig. 19. Plot of multiple third-party, and both third-party and private CDN.

	EconomicFreedom		GDP		NRI		IDI	
	Co-eff	**P-Value**	**Co-eff**	**P-Value**	**Co-eff**	**P-Value**	**Co-eff**	**P-Value**
Americas	0.52	0.226	0.90	0.005	0.90	0.005	0.77	0.042
Asia Pacific	0.80	0.005	0.63	0.027	0.49	0.125	0.43	0.187
Europe	0.30	0.141	0.33	0.095	0.32	0.112	0.27	0.188
All Countries	0.46	0.001	0.47	0.001	0.42	0.003	0.35	0.016

Table 3. Correlation Coefficients and P-Values of the correlation of each index with the average third-party dependency across the regions.

	EconomicFreedom		GDP		NRI		IDI	
	Co-eff	**P-Value**	**Co-eff**	**P-Value**	**Co-eff**	**P-Value**	**Co-eff**	**P-Value**
N.Europe	0.65	0.162	0.79	0.061	0.84	0.0362	0.40	0.426
W.Europe	0.06	0.853	-0.03	0.932	-0.14	0.665	0.22	0.498
E.Europe	0.29	0.492	0.33	0.431	0.33	0.419	-0.02	0.965
S.Europe	-0.55	0.199	-0.61	0.272	-0.33	0.466	-0.31	0.502

Table 4. Correlation Coefficients and P-Values of the correlation of each index with the average third-party dependency across the European regions.

A.10 Trading Blocs

Table 5 lists the Trading Blocs and the countries within each Trading Bloc.

Received August 2022; revised October 2022; accepted December 2022

Trading Bloc	Country Codes
AEC	ID, MY, SG, TH, VN
APEC	AU, CA, CL, CN, HK, ID, JP, KR, MX, MY, NZ, SG, TH, US, VN
EEA	BE, BG, DE, DK, EE, ES, FR, GR, HU, IT, LV, NL, NO, PL, PT, RO, SE
IORA	AE, AU, FR, ID, IN, MY, SG, TH, ZA
LAIA	AR, BR, CL, MX
MERCOSUR	AR, BR
NAFTA	CA, MX, US

Table 5. Trading Blocks and countries in our dataset.