THE UNIVERSITY OF CHICAGO

MEASURING THE CONSOLIDATION OF DNS AND WEB HOSTING PROVIDERS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCE
IN CANDIDACY FOR THE DEGREE OF
MASTER OF SCIENCE

DEPARTMENT OF COMPUTER SCIENCE

BY
SYNTHIA WANG

CHICAGO, ILLINOIS
MAY 2023

Copyright © 2023 by Synthia Wang

All Rights Reserved

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Despite the Internet's continued growth, it increasingly depends on a small set of service providers to support Domain Name System (DNS) and web content hosting. This trend poses a variety of potential threats to the Internet, including susceptibility to outages, failures, and potential censorship by providers. Increasingly, an outage from a single organization can create widespread disruption across a large number of sites and services. Although the consolidation towards a small set of providers have been noted for several years, the degree of consolidation over time has not been fully quantified. As such, this paper aims to quantify consolidation in terms of popular domains' reliance on a small set of organizations for DNS and web hosting. We highlight the extent to which a set of relatively few platforms host the authoritative name servers and web content for the top 10,000 websites. The findings are surprising and somewhat alarming. We find that over 75% of the top 10,000 domains rely on only a single organization for hosting authoritative DNS resolution. Furthermore, two organizations, Cloudflare and Amazon, are the sole host of DNS name servers for over 37% of these popular domains. In terms of web hosting, we find that 61% of index pages and many external page resources for the top 10,000 websites are hosted by only five organizations: Cloudflare, Amazon, Akamai, Fastly, and Google.

CHAPTER 1

INTRODUCTION

The success of the Internet can be partially attributed to its distributed design. Indeed, distributing services over infrastructures across many parties has contributed to the relative security, resiliency, and accessibility of the Internet. Over the past several years, however, control of Internet infrastructure has begun to consolidate in fewer organizations than before. In particular, two critical aspects of Internet service—DNS and web hosting—were once naturally distributed but are now increasingly operated by relatively few providers. Various organizations, including the Internet Society, have expressed concern over the potentially negative effects of this so-called Internet consolidation: *"The fact that a few corporations dominate large parts of the Internet is not news. Today, a handful of actors play a significant role in our increasingly-connected societies. In this context it's important to consider what the implications of those trends are, not only from an economic perspective but also in terms of how they may shape the Internet in coming years."* (Society [2019])

These concerns over consolidation are more than a curiosity or idle concern; rather, they have concrete, and potentially wide-ranging, ramifications. One consequence is reduced resilience: Over the past several years, many websites have suffered considerable outages. For example, in October 2016, a distributed denial of service (DDoS) attack on Dyn, a major Domain Name System (DNS) provider, resulted in outages to more than 60 major Internet sites and services, including CNN, Etsy, GitHub, Netflix, the *New York Times*, Reddit, Slack, the *Wall Street Journal*, Yelp, and many others (Chacos [2016]). More recently, September 2021 alone saw several major outages. Notably, Facebook, WhatsApp, and Instagram were unreachable for five hours after an erroneous BGP update withdrew routing advertisements to the authoritative DNS name servers for these services (Martinho [2021]). In late September, a large portion of Amazon Web Services experienced a degradation that took Signal and Nest offline (Swinhoe [2021]). Although these incidents were caused by a failure within a single

organization, the magnitude of the disruptions was so pronounced because many Internet services depended on the functionality of that single organization. As people become more dependent on the availability of Internet websites and services for work, entertainment, and communication, the social and economic costs of these incidents increases.

The risks of consolidation also go beyond reliability. At play are issues of Internet censorship and platform control over online speech and marketplaces. Large content hosts such as Amazon and Cloudflare have, in the past, exercised discretion with shutting down websites (Lyons [2021], Prince [2019]); Cloudflare has also prevented certain web clients from reaching websites hosted on its platform (Fingas [2021]). The consolidation of hosting on a smaller number of platforms, particularly when any given site is hosted on *only one* of the platforms, thus poses grave risks along a number of dimensions.

This process of increasing control over Internet infrastructure and services by a small set of organizations has been defined as *Internet consolidation*, and has been defined with a relatively broad scope: *"The most visible aspects of this involve well-recognized Internet services, but it is important to recognize that the Internet is a complex ecosystem. There are many underlying services whose diversity, or lack thereof, are as important as that of, say, consumer-visible social networks. For instance, the diversity of cloud services, operating systems, browser engines is as important as that of application stores or the browsers themselves."* (Arkko et al. [2019]) This expansive definition of Internet consolidation raises questions about the *extent* and *evolution* of this phenomenon from a variety of facets. While some previous work has considered consolidation through the lens of DNS traffic (Moura et al. [2020a], Hounsel et al. [2020]), others have examined the economic and political implications of a few powerful companies dominating markets on the Internet (Society [2019]).

Despite some attention to this topic, however, little is still known about how consolidation trends are affecting the resiliency of Internet services. Specifically, we do not have precise measures about the websites and services that could be vulnerable to an outage of a single,

particular DNS service provider, content delivery network, or web hosting service. Of course, answering that "what if" question is challenging, due to the dynamic nature of Internet services and the nature of dependencies in complex systems such as the Internet; yet, we can begin to get some understanding of these vulnerabilities by studying how critical aspects of content delivery—namely, DNS hosting and web hosting—are consolidated for popular websites and services.

We study several aspects of consolidation. We first explore the extent to which a relatively small number of organizations control DNS hosting for popular websites (Section 4.1). We find that two organizations, Amazon and Cloudflare, are exclusively responsible for hosting the DNS name servers for over 37% of domains in the Tranco top 10K (i.e., 37.1% of domains exclusively host their DNS on one of these sites). Additionally, we find that over 89% of popular websites use name servers under the same primary domain, and over 75% of them use a single organization to host their name servers. We also study these phenomena for web hosting (Section 4.2) and find similar trends: in particular, five organizations—Cloudflare, Amazon, Akamai, Fastly, and Google—host about 60% of index pages in the Tranco top 10K, as well as the majority of external page resources for these sites.

These findings have significant implications for the design of current and future Internet services that are resilient to both accidental misconfiguration and overt shutdown. As others have noted, these consolidation trends also have economic implications, particularly as they relate to issues such as competition, barriers to entry, and permissionless innovation (Arkko et al. [2019], Bates et al. [2018]). It is thus important not only to report on consolidation at a particular moment in time, as we have done in this paper, but also to track how consolidation evolves over time. To facilitate ongoing measurements of consolidation, we have released all of our data and measurement metrics.[1]

1. The project website will be made public upon publication of the paper.

CHAPTER 2

BACKGROUND AND RELATED WORK

This section provides background on Internet consolidation, as well as on DNS and web hosting. We highlight particular previous work that has explored various aspects of consolidation in these two areas.

2.1 The Domain Name System (DNS)

DNS is responsible for translating domain names into the Internet Protocol (IP) addresses needed to communicate with the desired network endpoint, such as a website, service, or device. Assuming the DNS record is not cached at a recursive resolver, the client translates the domain by querying an authoritative name server for the domain's DNS A record. If a domain name's authoritative name servers are unreachable (and the A record is not cached), the client cannot communicate with the network endpoint. It is common for organizations other than the operator of the domain to host that domain's name servers, a trend we discuss in Section (4.1). As discussed in the next section, past work has observed that increasing centralization of third-party service providers and other aspects of the DNS can have consequences on the robustness and security of various Internet services.

2.1.1 DNS Consolidation.

Previous work has studied both the effects and extent of increased DNS centralization. The previous mentioned Dyn incident has been studied under the circumstance of DDoS attacks by Abhishta et al. (Abhishta et al. [2019]), who conclude that using multiple DNS providers is an effective countermeasure. Bates et al. also motivate their work by analyzing the Dyn incident and propose a metric for measuring the market share of organizations that provide DNS resolvers (Bates et al. [2018]). However, since the metric is designed based on antitrust

economics literature, which is not widely used in computer science literature, we do not apply it to our measurements. Others have also made attempts in quantifying the consolidation of DNS. Zembruzki et al. (Zembruzki et al. [2020])focus on DNS route tracing and found that up to 12,000 name servers used by websites in the Alexa top 1 million shared the same third-party infrastructure. As for public resolvers, Radu et al. found that the popularity of Google's public DNS resolver has increased tremendously over time, serving as the default resolver for over 35% of studied clients (Radu and Hausding [2020]). The popularity of public DNS resolvers might be explained by their lower response times in comparison to that of local resolvers, despite the advantage being dependent on client location (Doan et al. [2021]). To further investigate the centralized usage of cloud providers, Moura et al. found that more than 30% of DNS queries to two country-code top-level domains (ccTLDs) were sent from five large U.S. cloud providers: Google, Amazon, Microsoft, Cloudflare, and Facebook (Moura et al. [2020b]).

2.2 Web Hosting and Content Delivery Networks

2.2.1 Content delivery networks (CDNs)

can provide faster delivery of content (e.g., websites) to a global population of users. Many companies who operate CDNs also help to enhance security by providing services such as DDoS mitigation. Websites and services who do not rely on a CDN are subject to a variety of risks, including performance degradation during traffic surges ("flash crowds"), weak protection against DDoS attacks, higher Internet transit costs, and slower content delivery. These providers not only provide sophisticated infrastructure, but also make it easy for users to quickly establish hosting: whereas a decade ago, hosting a website entailed a significant amount of system administration on the part of the publisher, these CDNs also provide web hosting services that now make this process turnkey.

2.2.2 Web Hosting Consolidation.

Prior to our work, the consolidation of web hosting and CDNs has not been systematically studied. Recent advances in CDN identification techniques now allow us to perform a more extensive study of this question. In addition to commercial CDN identification websites, Ager at al. developed Web Content Cartography, which identify and classify CDN infrastructures (Ager et al. [2011]). However, to avoid uncertainties in our measurement, we do not explicitly measure CDNs for our dataset of domains. Instead, we measure the hosting of page content and index pages. In many cases, especially among the top 10K domains, CDNs are used to host the website. Thus, by measuring content hosting, we can capture the consolidation of CDNs in addition to other hosting options. In a preliminary, concurrent unpublished technical report, Moura et al. used the resource records in DNS zone files to identify up to 200 million domains and the owners of the autonomous systems that the domains belong to (Moura et al. [2021]). They found that one-third of the domains they studied were hosted by Google, GoDaddy, Cloudflare, and Amazon. Our work complements and extends this study, considering both popular domains, as well as both the DNS infrastructure (i.e., authoritative name servers) and the hosting infrastructure that hosts third-party, external resources of these popular websites, such as scripts.

CHAPTER 3

METHODS

In this section, we describe the methods that we used to measure two facets of Internet consolidation: DNS hosting and web hosting.

Overview. We study the authoritative name servers and hosting providers used by the 10,000 most popular domains from the September 2022 Tranco rankings (Le Pochat et al. [2019b]). We use Tranco since it aggregates the results of several other ranking methods to create an accurate and more robust list. For each domain, we determine:

1. the AS and organization hosting each of its **name servers**,
2. the AS and organization hosting its **index page**, and
3. the AS and organization hosting its **external page resources**.

We define *organization* as the company or entity that owns the autonomous system in which the server is found. In addition, we discuss the limitations of our approach.

3.1 DNS Consolidation

We identify the organization responsible for hosting the authoritative name servers using the pipeline illustrated in Figure 3.1. For each domain in the top 10,000, we query the whois server referred by the ICANN whois server for its name servers. For each name server, we query the most widely used public resolvers from Google and Cloudflare, `8.8.8.8` and `1.1.1.1`, respectively, both to increase the likelihood of discovering all the IPs associated with a given domain, and to simulate the results typical Internet users would receive. Next, we determine the AS of each IP using Team Cymru's cym IP-to-AS database. We consider the *organization name* for each AS as a distinct organization when analyzing consolidation,

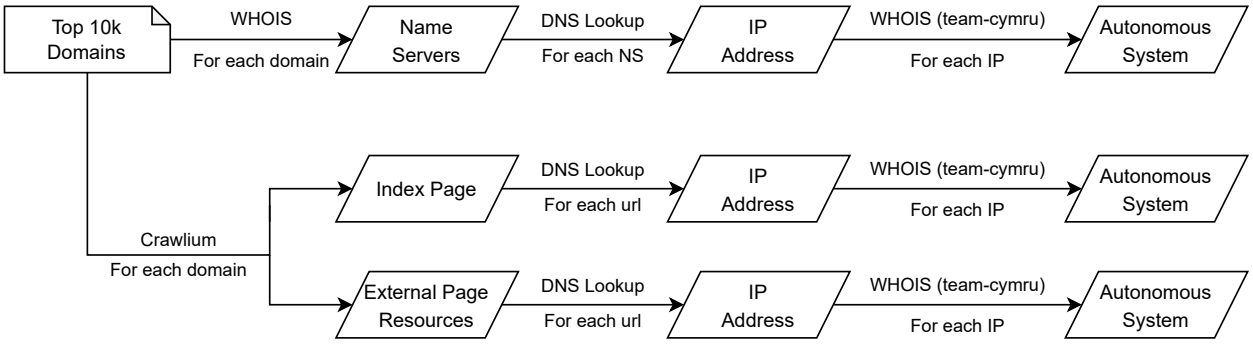

Figure 3.1: The measurement pipeline used to study each domain.

so that multiple AS numbers that share the same organization are considered as a single organization if the organization names match.

To eliminate any biases that might result from measuring from a single geographic location, we deploy our measurement from six vantage points on Amazon Web Service (AWS), namely California (US), Virginia (US), Tokyo (Japan), Mumbai (India), Frankfurt (German), and Cape Town (Africa). We chose these vantage points to get a global view within the AWS availability.

3.2 Web Hosting Consolidation

We determine the organizations responsible for hosting web content on each domain by examining the index page and all external page resources on the domain's homepage. We consider both the index page and other resources because a domain may use different hosting providers for each. For example, Github (`https://github.com`) hosts its own index page but uses Fastly (`https://www.fastly.com`) to host other assets. We scrape the homepage of each domain using Crawlium (Arshad [2020]), an open-source web crawler used in prior works (Han et al., Bashir et al. [2019]), and extract information about each resource loaded on the page.

Index pages. We determine the host of each domain's index page using the pipeline illustrated in Figure 3.1. We cannot perform a DNS lookup directly on the domain because the domain may be different from the URL of the index page. This can occur when an HTTP request to the domain returns a 301 or 302 response, which means that the page has been permanently, or temporarily moved to another URL that it redirects to. For example, making a request to `nytimes.com` will respond with a redirect to `www.nytimes.com`, and `www.nytimes.com` and `nytimes.com` are translated to different IP addresses. To account for this, we use Crawlium to determine the URL of the index page by extracting the URL of the first HTTP request to return a 200 response, which means that the request has succeeded. We then resolve the URL of the index page and determine the organization from which it was loaded.

External page resources. Most modern websites rely on dynamic content fetched from third parties. Thus, in addition to the index page, we use Crawlium to collect information about every resource that is loaded by each domain. However, many of these resources may not be critical to the functioning of the page. Because we are interested in only the resources that are absolutely required by the page, we filter out ads, trackers, cookies, and certain social media elements using EasyList's block lists (eas [2020]). We then perform the same lookup process as we did for name servers to determine the organization from which each resource was loaded.

3.3 Data Analysis

A domain can become suddenly unavailable due to incidents such as attacks (dyn [2016]) and outages (Lawler [2021]) that compromise Internet services that the domain depends on. To quantify the effect of such incidents, we introduce the notion of domains being *affected* or *unreachable*. For our DNS analysis, if a domain uses a single AS to host all of its name servers,

then the domain would be *unreachable* if that AS becomes unavailable. If a domain uses an AS for at least one of its name servers, then the domain would be *affected* if the AS becomes unavailable. Similarly, for web hosting, we consider a domain to be *unreachable* when the AS that hosts its index page becomes unavailable. Finally, if the domain has external page resources hosted by an AS, then the domain is only *affected* if that AS becomes unavailable.

3.4 Limitations

We determine the name servers for domains but do not identify the name servers for all subdomains. These name servers are likely the same, but we acknowledge that in some cases these name servers may be different. If a subdomain uses different authoritative name servers, a subdomain may be reachable even if the domain is not. For example, if `nytimes.com` and `www.nytimes.com` have different name servers, the website may still be accessible even if the name servers for `nytimes.com` are unreachable.

Second, modern websites, especially popular ones, may use anycast DNS, which redirects each client to a nearby server in the CDN. As a result, a website may rely on several organizations to host web content, yet this diversity may not be reflected if we only measure from a single vantage point—and, naturally, our measurements are a function of the vantage points we choose. To mitigate these effects to some degree, we perform our measurements from six different vantage points and we have released our measurement pipeline as open-source software to allow other researchers to continually re-evaluate our results, over time and from a diversity of vantage points.

Finally, a website may have different hosting organization as a backup in case the primary host is unreachable, but only rely on that organization when the primary fails. If this situation occurs, our measurements would not necessarily observe the back up host and could overestimate the effects of consolidation for failure scenarios.

CHAPTER 4

FINDINGS

In this section, we present the results of our analysis. Results from the six vantage points are similar, so we present results measured from California; results from other vantage points are in Chapter 7 Appendix. We address the following questions: (1) Which AS and AS organizations host each domain's name servers? (Section 4.1) (2) Which AS or AS organizations host web content (including index pages and external page resources)? (Section 4.2) Our results show in both cases, hosting is dominated by a few large companies.

4.1 Which Organizations Provide Name Servers?

We first analyze which organizations host each domains' name servers.We characterize the potential consequences of organization outages on each domain in terms of two metrics: (1) unreachable and (2) affected, as defined in Section 3. In terms of our analysis of authoritative DNS name servers, a domain is unreachable if all of its name servers are hosted by the same organization and said organization is down. On the other hand, a domain is affected if at least one of the domain's name servers is hosted by an organization and that organization is down. Based on these definitions, the set of unreachable domains is a subset of the set of affected domains.

Two organizations are responsible for hosting the name servers for over 37% of domains. Amazon and Cloudflare exclusively host the name servers for 24.1% and 13.0% of domains, respectively. Table 4.1 shows the top ten most popular name server hosting providers by the percentage of domains that would be unreachable if that provider experienced an outage. Although these ten providers serve for over 50% of domains, there is a significant decline from Cloudflare to the next most common organization, Akamai. This trend likely results from the fact that Amazon and Cloudflare both offer lower-tier instant

services (aws) and enterprise level support, whereas Akamai does not offer the lower-tier hosting options (cdn [a]).

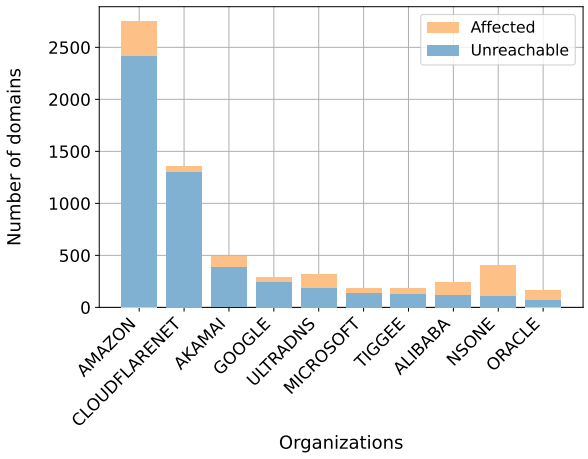

Figure 4.1: Number of domains that use exclusively one AS to host their name servers (unreachable) and use that AS at least partially (affected).

	% Unreachable	% Affected
Amazon	24.1	27.5
Cloudflare	13.0	13.6
Akamai	3.8	4.9
Google	2.4	2.9
UltraDNS	1.8	3.2
Microsoft	1.4	1.9
TIGGEE	1.3	1.8
Alibaba	1.2	2.4
NSONE	1.1	4.1
ORACLE	0.6	1.7
Total	50.7	64.0

Table 4.1: Percentages of domains that use exclusively one AS organization to host their name servers (unreachable) and use that AS at least partially (affected).

The use of a single organization to host all name servers is prevalent however we define the top N domains. Figure 4.2 shows that the percentage of domains that use only a given organization to host its name servers in the top n domains is greater than 50% for a wide range of possible n. In other words, both popular and less popular domains rely on a single organization for DNS hosting. The same statement can be made when we look at ASes instead of AS organizations in Figure 4.3. However, the proportion of domains that rely on a single organization for DNS name server hosting increases by over 20% between the top 100 domains and the top 10,000 domains. Additionally, the popularity of each major organization peaks in different ranges of domains. For example, the proportions of domains using Google peaks in the top 100 and decreases with n. This is likely because Google hosts

its own domains but does not provide hosting services for other domains to the same degree as traditional hosting providers such as Cloudflare and Akamai. Cloudflare exhibits the opposite trend, hosting the lowest proportion of domains in the top 100 and increasing with n. The reason for this phenomenon is unclear, although one possibility is that Cloudflare is a less expensive option for hosting, and therefore more accessible, especially for the less popular domains (MSV [2020]).

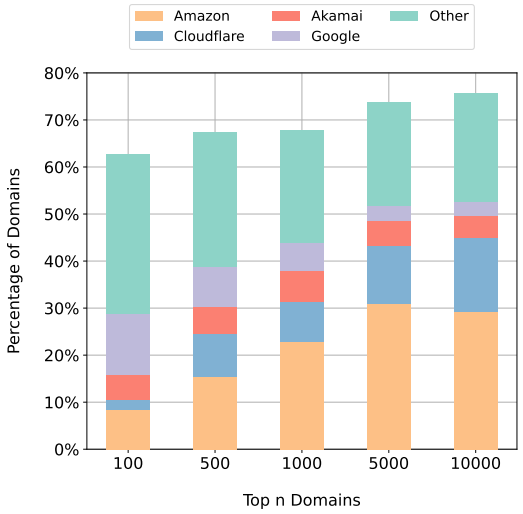

Figure 4.2: Percentage of domains that use exclusively one **AS organization** to host their name servers (unreachable).

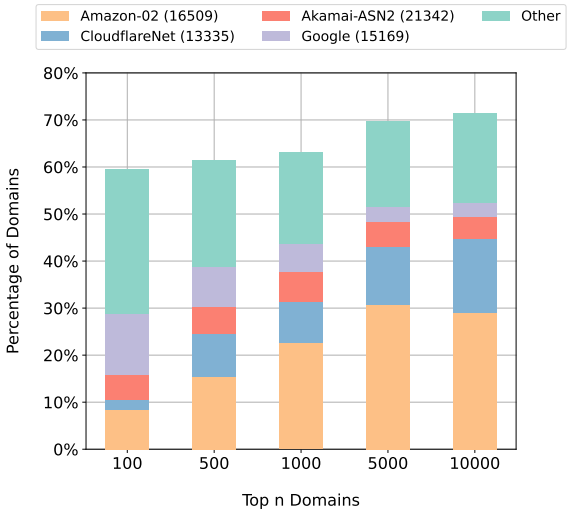

Figure 4.3: Percentage of domains that use exclusively one **AS** to host their name servers (unreachable).

4.2 Which Organizations Host Web Content?

We next study which organizations host the index page and the external page resources for the Tranco 10K domains. Of these domains, we were able to load the index page for 9,999; only 9,982 of these domains loaded external resources that were not filtered out by our block list (e.g., trackers).

Five organizations exclusively host the index page for a majority of domains. Figure 4.4 shows the percentages of index pages that are hosted exclusively by one organization. We find that Cloudflare, Amazon, Akamai, Fastly, and Google exclusively host over

	Threshold			
	0%	50%	75%	90%
Google	69.8	6.3	3.9	3.3
Amazon	56.1	13.7	6.8	3.8
Cloudflare	55.5	14.3	8.9	5.6
Akamai	42.9	8.9	4.8	3.1
Fastly	31.9	3.3	1.2	0.6
Total	-	46.5	25.6	16.4

Table 4.2: Percentage of domains that fetch greater than the indicated threshold of external page resources from a given organization. The 0% threshold is used to count domains that load any resources from a given organization.

61% of index pages of the domains. As with name server hosting, Google's representation peaks in the top 100 domains and Cloudflare in the top 10,000, indicating that each of the dominant organizations may offer service packages that are attractive to or even catered towards domains of similar popularity. With the exception of Fastly, the most popular index page hosts are the same as the most popular name server hosts, as presented in Section 4.1. In Figure 4.5 we look at individual ASes and found that the top five ASes come from Cloudflare, Amazon, Akamai, Fastly. We see a decreased number of domain in comparison to Figure 4.4, which is caused by domains using different ASes under the same organization.

The same five organizations host the majority of external page resources. A website may load external page resources from many organizations, yet in many cases there is a single dominant organization that hosts the majority of external resources for a given webpage. We identify the websites that fetch a majority of resources from a single organization. To do so, we count both the total number of resources and the number of resources loaded from each organization. Table 4.2 shows the five most popular sources of externally loaded content on each website's homepage using several thresholds. For example, 14.3% of domains loaded more than 50% of external page resources from Cloudflare.

More than 46% of websites fetch a majority of their external page resources from one of the five hosting providers, namely Cloudflare, Amazon, Akamai, Fastly, and Google.

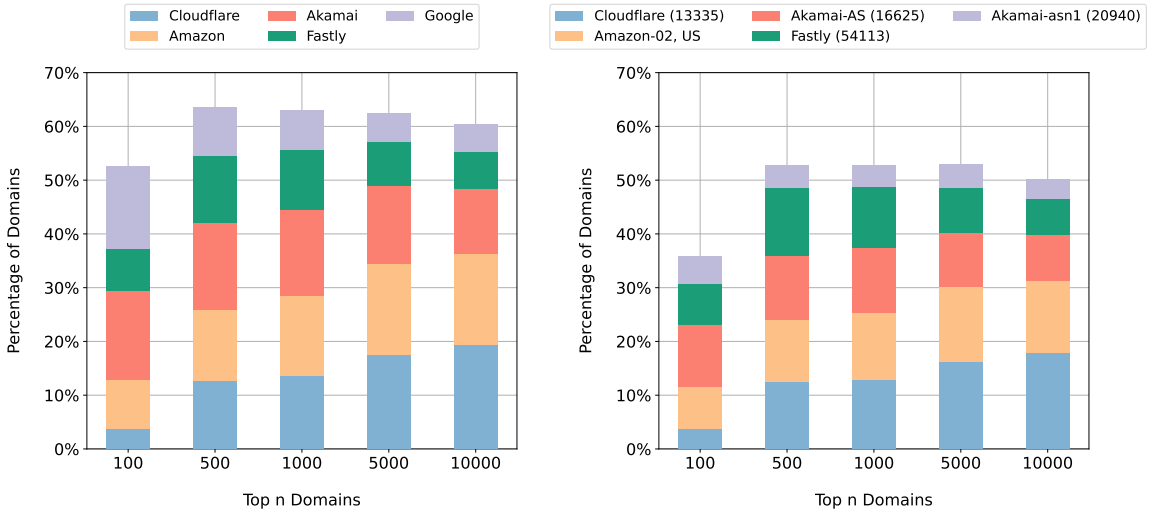

Figure 4.4: Percentage of domains that use exclusively one **AS organization** to host their index page (unreachable).

Figure 4.5: Percentage of domains that use exclusively one **AS** to host their index page (unreachable).

Furthermore, over 16% of domains load over 90% of external page resources from these five organizations. These findings are similar to our study of index page hosts, in that many websites rely heavily on these five organizations for content hosting.

CHAPTER 5

DISCUSSION

Our results suggest that DNS and web hosting are concentrated around a few providers. We now discuss the implications of these findings.

5.1 Resilience and security

While some degree of consolidation in infrastructure can be beneficial for efficient security upgrades and protocol implementation, domains relying on a single organization introduces a single point of failure. As exhibited during the Dyn and AWS outages, a large-scale DDoS attack on a single organization can lead to many website outages. With this in mind, it is important to strike a balance between distributing reliance across many organizations and limiting redundancy.

Internet consolidation can also introduce the notion of "attractive surveillance targets" (Arkko [2019]), where too much information or traffic is controlled by a single entity. Companies such as Cloudflare, Amazon, and Akamai, who host content for many domains, have access to the network traffic to the domains, and therefore can collect more information, not only from the domains themselves, but also from users visiting those domains. Large amounts of user information concentrated within one system intensifies concerns of surveillance by both the controlling company and potential attackers.

5.2 Content moderation

Market concentration can give companies the power to selectively moderate content on platforms that they service, sometimes enforcing stricter restrictions than are currently legally required (the [2021]). As regulation continues to lag, these companies will have greater ability to both control online content and influence future policy. These implications are es-

pecially pertinent in light of ongoing debate on free speech online. Multi-national companies play an even greater role in content moderation, as they may mediate speech differentially across geo-political lines. In this work, we explain the critical role that name server and web hosts play in the content delivery process. Future work may continue to study how companies can exert control at other points of this process, including at DNS resolvers (Pearce et al. [2017]).

5.3 Standardizing metrics for consolidation

Since there is no standard metric for quantifying consolidation or guidelines on what kind of measurements should be collected for this purpose, future work could include proposing such metrics. Our findings on DNS hosting and content hosting only covers limited components of the Internet infrastructure. Therefore, future work could include more static attributes of domains such as domain registrars. The organizations with whom the domains are (or were originally) registered likely also illustrates a trend towards consolidation, with the result being a decrease in the number of companies managing the domain name registry and granting the purchasing or transferring of domain names. We also only base our study on IPv4 address, and the IPv6 address space remains unexplored.

CHAPTER 6

CONCLUSION

The consolidation of the hosting of various Internet infrastructure and services is an ongoing trend that potentially threatens Internet resilience, security, competition, and free and open communication. Although many organizations, from the Internet Society to the Internet Engineering Task Force, have expressed concern over this ongoing and evolving trend, we have, to date, only had a limited understanding of the extent and evolution of such consolidation. This paper aims to quantify two aspects of Internet consolidation—authoritative DNS name server hosting and web hosting. To do so, we analyzed the extent of consolidation of DNS authoritative name server providers and web content hosting providers for the most popular 10,000 Internet domains, as enumerated in the Tranco 10K.

Our analysis revealed that two organizations, Amazon and Cloudflare, are responsible for exclusively hosting the name servers for over 40% of domains in the Tranco top 10K, and that only five organizations—Cloudflare, Amazon, Akamai, Fastly, and Google—host about 60% of index pages in the Tranco top 10K, as well as the majority of external page resources for these sites. We also found that more than 75% of domains use only one AS organization for all of its name servers. These results suggest that, in the areas of DNS name server hosting and web hosting, consolidation is indeed significant. Our results nonetheless represent a single snapshot of the current state of affairs. Given the potential consequences of increasing consolidation, continual measurements could shed more light on these trends over time and provide insights into potential dependencies or points of vulnerability for modern Internet services. To facilitate such measurements, we will publicly release both our current measurements and measurement framework, and we will continue to perform and release updated measurements periodically.

REFERENCES

Amazon Route 53 Pricing. URL https://aws.amazon.com/route53/pricing/.

Akamai versus Cloudflare, a. URL https://www.cdnplanet.com/compare/akamai/cloudflare/.

CDN finder - CDN planet, b. URL https://www.cdnplanet.com/tools/cdnfinder/.

IP to ASN mapping. URL https://team-cymru.com/community-services/ip-asn-mapping/.

Build software better, together. URL https://github.com.

Cloudflare DDoS Protection & Mitigation, a. URL https://www.cloudflare.com/ddos/.

Fastly | the edge cloud platform behind the best of the web, b. URL https://www.fastly.com/.

Update Regarding DDoS Event Against Dyn Managed DNS on October 21, 2016, 2016. URL https://www.dynstatus.com/incidents/5r9mppc1kb77.

Easylist, 2020. URL https://easylist.to.

Credit-card firms are becoming reluctant regulators of the web, Oct 2021. URL https://www.economist.com/finance-and-economics/credit-card-firms-are-becoming-reluctant-regulators-of-the-web/21805450.

Abhishta Abhishta, Roland van Rijswijk-Deij, and Lambert J. M. Nieuwenhuis. Measuring the impact of a successful ddos attack on the customer behaviour of managed dns service providers. *SIGCOMM Comput. Commun. Rev.*, 48(5):70–76, jan 2019. ISSN 0146-4833. doi:10.1145/3310165.3310175. URL https://doi.org/10.1145/3310165.3310175.

H. B. Acharya, Sambuddho Chakravarty, and Devashish Gosain. Few throats to choke: On the current structure of the internet. In *2017 IEEE 42nd Conference on Local Computer Networks (LCN)*, pages 339–346, 2017. doi:10.1109/LCN.2017.78.

Bernhard Ager, Wolfgang Mühlbauer, Georgios Smaragdakis, and Steve Uhlig. Web content cartography. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference*, IMC '11, page 585–600, New York, NY, USA, 2011. Association for Computing Machinery. ISBN 9781450310130. doi:10.1145/2068816.2068870. URL https://doi.org/10.1145/2068816.2068870.

Inc. Alexa Internet. The top 500 sites on the web, 2021. URL https://www.alexa.com/topsites.

J Arkko. Centralised Architectures in Internet Infrastructure. *IETF Internet Draft*, 2019.

Jari Arkko, Brian Trammell, Mark Nottingham, Christian Huitema, Martin Thomson, Jeff Tantsura, and Niels ten Oever. Considerations on internet consolidation and the internet architecture. Technical report, Internet-Draft draft-arkko-iabinternet-consolidation-01, IETF Working Draft, 2019.

S. Arshad. Crawlium, 2020. URL https://github.com/sajjadium/DeepCrawling.

Sajjad Arshad, Amin Kharraz, and William Robertson. Identifying Extension-Based Ad Injection via Fine-Grained Web Content Provenance. In *Research in Attacks, Intrusions, and Defenses*, volume 9854, pages 415–436. Springer International Publishing, Cham, 2016. ISBN 978-3-319-45718-5 978-3-319-45719-2. doi:10.1007/978-3-319-45719-2_19. URL http://link.springer.com/10.1007/978-3-319-45719-2_19. Series Title: Lecture Notes in Computer Science.

Sajjad Arshad, Amin Kharraz, and William Robertson. Include Me Out: In-Browser Detection of Malicious Third-Party Content Inclusions. In *Financial Cryptography and Data Security*, volume 9603, pages 441–459. Springer Berlin Heidelberg, Berlin, Heidelberg, 2017. ISBN 978-3-662-54969-8 978-3-662-54970-4. doi:10.1007/978-3-662-54970-4_26. URL http://link.springer.com/10.1007/978-3-662-54970-4_26. Series Title: Lecture Notes in Computer Science.

Muhammad Ahmad Bashir, Sajjad Arshad, and Christo Wilson. "Recommended For You": A First Look at Content Recommendation Networks. In *Proceedings of the 2016 Internet Measurement Conference*, pages 17–24, Santa Monica California USA, November 2016a. ACM. ISBN 978-1-4503-4526-2. doi:10.1145/2987443.2987469. URL https://dl.acm.org/doi/10.1145/2987443.2987469.

Muhammad Ahmad Bashir, Sajjad Arshad, Christo Wilson, and William Robertson. Tracing Information Flows Between Ad Exchanges Using Retargeted Ads. page 16, 2016b.

Muhammad Ahmad Bashir, Sajjad Arshad, Engin Kirda, William Robertson, and Christo Wilson. How Tracking Companies Circumvented Ad Blockers Using WebSockets. In *Proceedings of the Internet Measurement Conference 2018*, pages 471–477, Boston MA USA, October 2018. ACM. ISBN 978-1-4503-5619-0. doi:10.1145/3278532.3278573. URL https://dl.acm.org/doi/10.1145/3278532.3278573.

Muhammad Ahmad Bashir, Sajjad Arshad, Engin Kirda, William Robertson, and Christo Wilson. A Longitudinal Analysis of the ads.txt Standard. In *Proceedings of the Internet Measurement Conference*, pages 294–307, Amsterdam Netherlands, October 2019. ACM. ISBN 978-1-4503-6948-0. doi:10.1145/3355369.3355603. URL https://dl.acm.org/doi/10.1145/3355369.3355603.

Samantha Bates, John Bowers, Shane Greenstein, Jordi Weinstock, Yunhan Xu, and Jonathan Zittrain. Evidence of decreasing internet entropy: the lack of redundancy in dns resolution by major websites and services. Technical report, National Bureau of Economic Research, 2018.

Brad Chacos. Major DDoS attack on Dyn DNS knocks Spotify, Twitter, Github, PayPal, and more offline, Oct 2016. URL https://www.pcworld.com/article/410774/ddos-attack-on-dyn-knocks-spotify-twitter-github-etsy-and-more-offline.html.

I. Robert Chiang and Jhih-Hua Jhang-Li. Delivery consolidation and service competition among internet service providers. *Journal of Management Information Systems*, 31(3): 254–286, 2014. doi:10.1080/07421222.2014.995561. URL https://doi.org/10.1080/07421222.2014.995561.

Trinh Viet Doan, Justus Fries, and Vaibhav Bajpai. Evaluating public dns services in the wake of increasing centralization of dns. In *2021 IFIP Networking Conference (IFIP Networking)*, pages 1–9, 2021. doi:10.23919/IFIPNetworking52078.2021.9472831.

Shaun P Egan and Barry Irwin. High speed lexical classification of malicious URLs. page 3.

Jon Fingas. Tor accuses CloudFlare of blocking its anonymizing network, May 2021. URL https://www.engadget.com/2016-04-03-tor-and-cloudflare-fight-over-blocking.html.

Alex Fitzpatrick. Why amazon's move to drop parler is a big deal for the future of the internet. URL https://time.com/5929888/amazon-parler-aws/.

Vivek Ganti and Omer Yoachimik. Cloudflare recognized as a 'Leader' in The Forrester Wave for DDoS Mitigation Solutions, March 2021. URL https://blog.cloudflare.com/cloudflare-is-named-a-leader-in-the-forrester-wave-for-ddos-mitigation-solutions/.

Eduardo Graells-Garrido and Mounia Lalmas. Balancing diversity to counter-measure geographical centralization in microblogging platforms. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, HT '14, page 231–236, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450329545. doi:10.1145/2631775.2631823. URL https://doi.org/10.1145/2631775.2631823.

Catherine Han, Deepak Kumar, and Zakir Durumeric. On the infrastructure providers that support misinformation websites.

Austin Hounsel, Paul Schmitt, Kevin Borgolte, and Nick Feamster. Designing for tussle in (encrypted) dns. *arXiv preprint arXiv:2002.09055*, 2020.

Tobias Lauinger, Abdelberi Chaabane, Sajjad Arshad, William Robertson, Christo Wilson, and Engin Kirda. Thou Shalt Not Depend on Me: Analysing the Use of Outdated JavaScript Libraries on the Web. In *Proceedings 2017 Network and Distributed System Security Symposium*, San Diego, CA, 2017. Internet Society. ISBN 978-1-891562-46-4. doi:10.14722/ndss.2017.23414. URL https://www.ndss-symposium.org/ndss2017/ndss-2017-programme/thou-shalt-not-depend-me-analysing-use-outdated-javascript-libraries-web/.

Richard Lawler. Facebook explains the backbone shutdown behind its global outage on Monday, October 2021. URL https://www.theverge.com/2021/10/5/22710963/facebook-dns-bgp-outage-backbone-maintenance.

Anh Le, Athina Markopoulou, and Michalis Faloutsos. PhishDef: URL names say it all. In *2011 Proceedings IEEE INFOCOM*, pages 191–195. doi:10.1109/INFCOM.2011.5934995. ISSN: 0743-166X.

Victor Le Pochat, Tom Van Goethem, Samaneh Tajalizadehkhoob, Maciej Korczyński, and Wouter Joosen. Tranco: A research-oriented top sites ranking hardened against manipulation. In *Proceedings of the 26th Annual Network and Distributed System Security Symposium*, NDSS 2019, February 2019a. doi:10.14722/ndss.2019.23386.

Victor Le Pochat, Tom Van Goethem, Samaneh Tajalizadehkhoob, Maciej Korczyński, and Wouter Joosen. Tranco: A research-oriented top sites ranking hardened against manipulation. In *Proceedings of the 26th Annual Network and Distributed System Security Symposium*, NDSS 2019, February 2019b. doi:10.14722/ndss.2019.23386.

Li Linzhuo, Wu Lingfei, and Evans James. Social centralization and semantic collapse: Hyperbolic embeddings of networks and text. *Poetics*, 78:101428, 2020. ISSN 0304-422X. doi:https://doi.org/10.1016/j.poetic.2019.101428. URL https://www.sciencedirect.com/science/article/pii/S0304422X1830295X. Discourse, Meaning, and Networks: Advances in Socio-Semantic Analysis.

Kim Lyons. Amazon is kicking Parler off its web hosting service, Jan 2021. URL https://www.theverge.com/2021/1/9/22222637/amazon-workers-aws-stop-hosting-services-parler-capitol-violence.

Celso Martinho. Understanding How Facebook Disappeared from the Internet, Oct 2021. URL https://blog.cloudflare.com/october-2021-facebook-outage/.

Giovane C. M. Moura, Sebastian Castro, Wes Hardaker, Maarten Wullink, and Cristian Hesselman. Clouding up the internet: How centralized is dns traffic becoming? In *Proceedings of the ACM Internet Measurement Conference*, IMC '20, page 42–49, New York, NY, USA, 2020a. Association for Computing Machinery. ISBN 9781450381383. doi:10.1145/3419394.3423625. URL https://doi.org/10.1145/3419394.3423625.

Giovane C. M. Moura, Sebastian Castro, Wes Hardaker, Maarten Wullink, and Cristian Hesselman. Clouding up the Internet: How Centralized is DNS Traffic Becoming? In *Proceedings of the ACM Internet Measurement Conference*, IMC '20, page 42–49, New York, NY, USA, 2020b. Association for Computing Machinery. ISBN 9781450381383. doi:10.1145/3419394.3423625. URL https://doi.org/10.1145/3419394.3423625.

Giovane C. M. Moura, Raffaele Somesse, and Mattijs Jonker. Hosting industry centralization and consolidation, 2021. URL https://arxiv.org/abs/2109.01187.

Janakiram MSV. Cloudflare Takes On AWS Lambda@Edge With A Cheaper And Faster Serverless Platform, July 2020. URL https://www.forbes.com/sites/janakirammsv/2020/07/27/cloudflare-takes-on-aws-lambdaedge-with-a-cheaper-and-faster-serverless-platform/. Section: Cloud.

Johanna E. Möller and M. Bjørn von Rimscha. (de)centralization of the global informational ecosystem. *Media and Communication*, 5(3):37–48, 2017. ISSN 2183-2439. doi:https://doi.org/10.17645/mac.v5i3.1067.

Paul Pearce, Ben Jones, Frank Li, Roya Ensafi, Nick Feamster, Nick Weaver, and Vern Paxson. Global measurement of {DNS} manipulation. In *26th {USENIX} Security Symposium ({USENIX} Security 17)*, pages 307–323, 2017.

Matthew Prince. Why we terminated daily stormer. URL https://blog.cloudflare.com/why-we-terminated-daily-stormer/.

Matthew Prince. Why We Terminated Daily Stormer, Aug 2019. URL https://blog.cloudflare.com/why-we-terminated-daily-stormer/.

Roxana Radu and Michael Hausding. Consolidation in the dns resolver market – how much, how fast, how dangerous? *Journal of Cyber Policy*, 5(1):46–64, 2020. doi:10.1080/23738871.2020.1722191. URL https://doi.org/10.1080/23738871.2020.1722191.

Claude E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948. doi:10.1002/j.1538-7305.1948.tb01338.x.

The Internet Society. Consolidation in the Internet Economy. Technical report, The Internet Society, 2019.

Dan Swinhoe. AWS suffering EC2 and EBS performance issues in Northern Virginia, Sep 2021. URL https://www.datacenterdynamics.com/en/news/aws-suffering-ec2-and-ebs-performance-issues-in-northern-virginia/.

Meng-Han Tsai, Kai-Chi Chang, Chang-Cheng Lin, Ching-Hao Mao, and Huey-Ming Lee. C amp;c tracer: Botnet command and control behavior tracing. In *2011 IEEE International Conference on Systems, Man, and Cybernetics*, pages 1859–1864. doi:10.1109/ICSMC.2011.6083942. ISSN: 1062-922X.

Matthias Wählisch, Thomas C. Schmidt, Markus de Brün, and Thomas Häberlen. Exposing a nation-centric view on the german internet – a change in perspective on AS-level. In Nina Taft and Fabio Ricciato, editors, *Passive and Active Measurement*, Lecture Notes in Computer Science, pages 200–210. Springer. ISBN 978-3-642-28537-0. doi:10.1007/978-3-642-28537-0_20.

Xueyang Xu, Z. Morley Mao, and J. Alex Halderman. Internet censorship in china: Where does the filtering occur? In Neil Spring and George F. Riley, editors, *Passive and Active*

Measurement, Lecture Notes in Computer Science, pages 133–142. Springer. ISBN 978-3-642-19260-9. doi:10.1007/978-3-642-19260-9_14.

Luciano Zembruzki, Arthur Selle Jacobs, Gustavo Spier Landtreter, Lisandro Zambenedetti Granville, and Giovane C. M. Moura. : Measuring centralization of dns infrastructure in the wild. In Leonard Barolli, Flora Amato, Francesco Moscato, Tomoya Enokido, and Makoto Takizawa, editors, *Advanced Information Networking and Applications*, pages 871–882, Cham, 2020. Springer International Publishing. ISBN 978-3-030-44041-1.

CHAPTER 7

APPENDIX

This section presents data obtained from other vantage points where we deployed our measurements.

7.1 Percentage of domains that fetch greater than the indicated threshold of external page resources from a given organization

	Threshold			
	0%	50%	75%	90%
Google	69.9	6.2	3.9	3.3
Amazon	56.1	13.8	6.8	3.8
Cloudflare	55.5	14.2	8.8	5.6
Akamai	42.7	8.6	4.7	3.0
Fastly	32.0	3.3	1.2	0.6
Total	-	46.1	25.4	16.3

Table 7.1: Virginia

	Threshold			
	0%	50%	75%	90%
Google	69.8	6.2	3.8	3.3
Amazon	56.2	14.0	7.0	3.9
Cloudflare	55.5	14.3	8.8	5.6
Akamai	43.1	9.0	4.8	3.1
Fastly	31.8	3.3	1.2	0.6
Total	-	46.8	25.6	16.5

Table 7.2: Tokyo

	Threshold			
	0%	50%	75%	90%
Google	69.8	6.3	4.0	3.4
Amazon	56.1	13.9	6.9	3.9
Cloudflare	55.1	14.2	8.8	5.6
Akamai	36.2	7.8	4.3	2.9
Fastly	33.5	3.2	1.1	0.5
Total	-	45.4	25.1	16.3

Table 7.3: Mumbai

	Threshold			
	0%	50%	75%	90%
Google	69.7	6.1	3.9	3.3
Amazon	56.1	13.7	6.7	3.9
Cloudflare	55.5	14.0	8.6	5.5
Akamai	43.0	8.9	4.9	3.1
Fastly	31.6	3.1	1.1	0.5
Total	-	45.8	25.2	16.3

Table 7.4: Frankfurt

	Threshold			
	0%	50%	75%	90%
Google	69.7	6.0	3.8	3.2
Amazon	56.2	14.1	7.1	4.1
Cloudflare	55.5	14.1	8.8	5.6
Akamai	34.7	7.1	4.1	2.7
Fastly	31.1	3.0	1.1	0.6
Total	-	44.3	24.9	16.2

Table 7.5: Cape Town

7.2 Percentage of domains that use exclusively one AS to host their index page (unreachable).

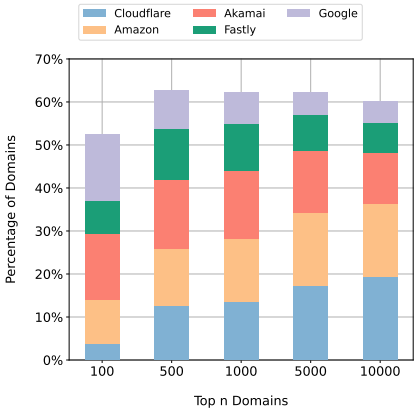

Figure 7.1: Virginia

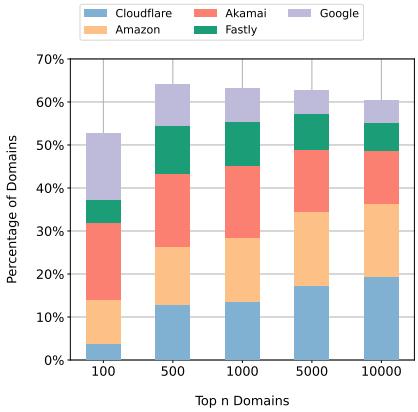

Figure 7.2: Tokyo

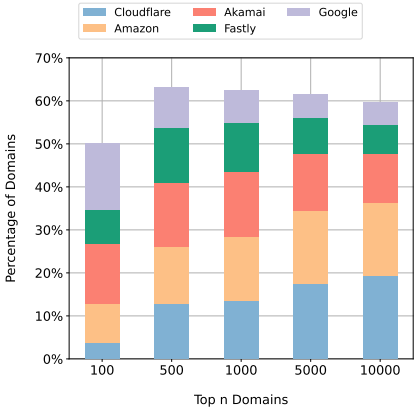

Figure 7.3: Mumbai

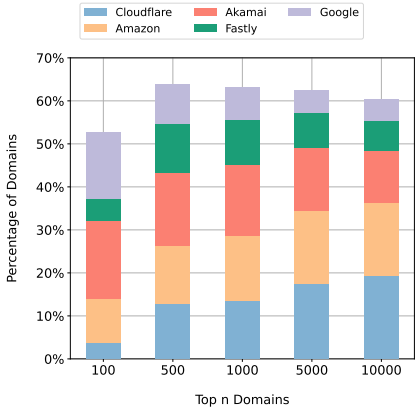

Figure 7.4: Frankfurt

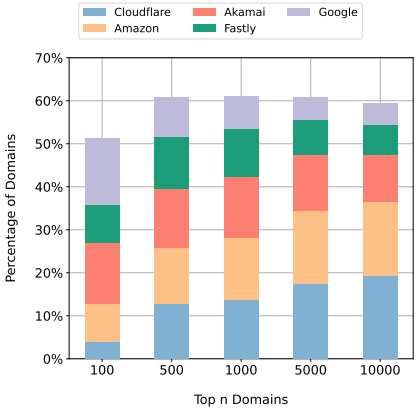

Figure 7.5: Cape Town

7.3 Percentage of domains that use exclusively one AS organization to host their index page (unreachable).

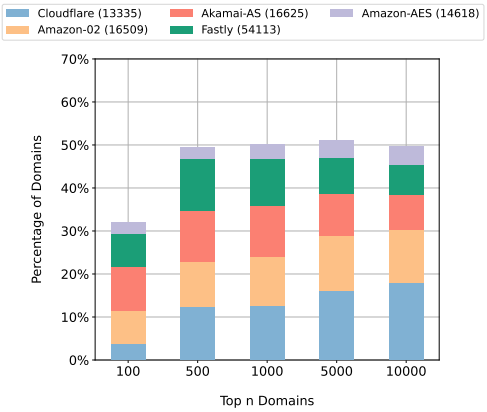

Figure 7.6: Virginia

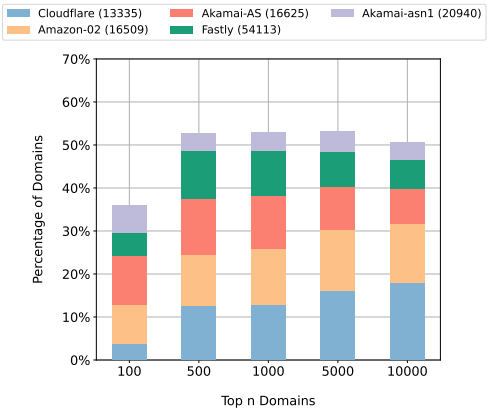

Figure 7.7: Tokyo

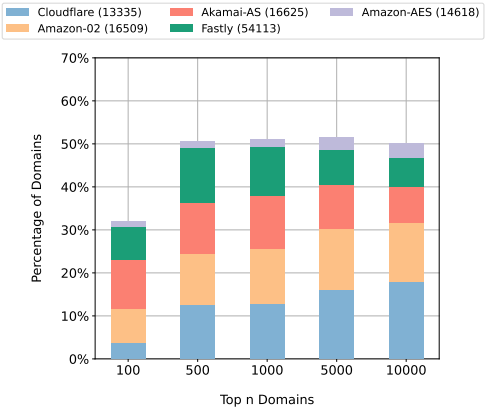

Figure 7.8: Mumbai

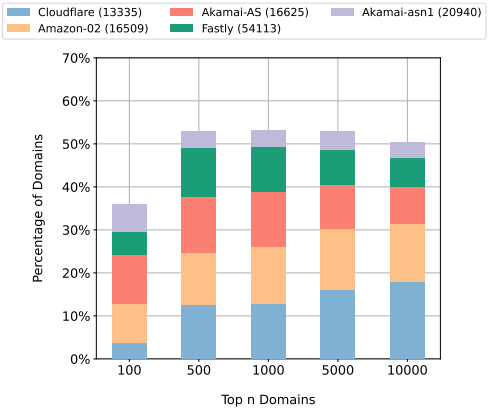

Figure 7.9: Frankfurt

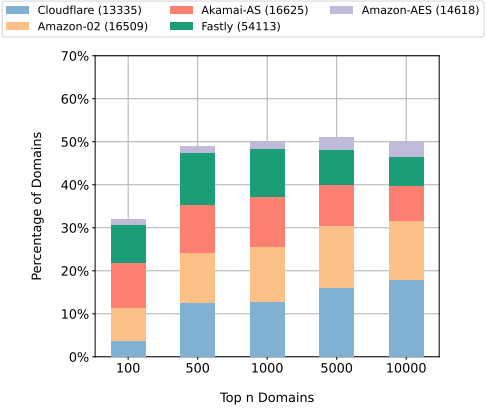

Figure 7.10: Cape Town