A First Look at Third-Party Service Dependencies of Web Services in Africa

Aqsa Kashaf[1][0000−0002−8244−825X], Jiachen Dou[1], Margarita Belova[2], Maria Apostolaki[2], Yuvraj Agarwal[1][0000−0001−9304−6080], and Vyas Sekar[1]

[1] Carnegie Mellon University, USA
(akashaf,jiachend,yuvraja,vsekar)@andrew.cmu.edu
[2] Princeton University, USA
(margarita.bel, apostolaki)@princeton.edu

Abstract. Third-party dependencies expose websites to shared risks and cascading failures. The dependencies impact African websites as well *e.g.,* Afrihost outage in 2022 [15]. While the prevalence of third-party dependencies has been studied for globally popular websites, Africa is largely underrepresented in those studies. Hence, this work analyzes the prevalence of third-party infrastructure dependencies in Africa-centric websites from 4 African vantage points. We consider websites that fall into one of the four categories: *Africa-visited* (popular in Africa) *Africa-hosted* (sites hosted in Africa), *Africa-dominant* (sites targeted towards users in Africa), and *Africa-operated* (websites operated in Africa). Our key findings are: 1) 93% of the *Africa-visited* websites critically depend on a third-party DNS, CDN, or CA. In perspective, *US-visited* websites are up to 25% less critically dependent. 2) 97% of *Africa-dominant*, 96% of *Africa-hosted*, and 95% of *Africa-operated* websites are critically dependent on a third-party DNS, CDN, or CA provider. 3) The use of third-party services is concentrated where only 3 providers can affect 60% of the Africa-centric websites. Our findings have key implications for the present usage and recommendations for the future evolution of the Internet in Africa.

Keywords: DNS · Certificate authorities · Third-party dependency · Availability · CDN · Africa Internet.

1 Introduction

The websites we use everyday offload critical services such as name resolution (DNS), content distribution (CDN), and certificate issuance/revocation (CA) to third parties for key services *e.g.,* AWS Route 53 for DNS, Akamai for CDN, DigiCert for CA. As a result, the availability and security of these websites, and thus of our data and operations, depend on the availability and security of those third parties. The effects of such dependencies are routinely observed in the Internet today. For example, a dependency on DNS resulted in the downtime of multiple websites (more than 100K) for several hours together with their DNS provider (Dyn) which was attacked by a Mirai Distributed Denial of Service

(DDoS) attack [24]. Similarly, users of multiple websites lost access to their accounts for weeks, because a single CA issued an incorrect revocation of a certificate in 2016 [22].

To gauge the security risk that such dependencies entail, one needs to understand the prevalence of third-party dependencies across the websites that are important for users all over the world. While such studies exist [28, 25, 26, 47, 55, 33], their target users/websites are particularly skewed towards North America and Europe. The geographical bias of the datasets used in previous studies of third-party dependencies creates a critical gap as distinct regions exhibit unique characteristics, needs, and opportunities that are effectively ignored. Naively assuming that observations generalize across regions, entails risks as it underestimates the practicality of certain attacks and creates false assurance of the security of critical region-specific websites (*e.g.,* those related to government or health insurance in those countries). This is also recognized by the Internet Society's Measuring Internet Resilience in Africa (MIRA) project [46].

To bridge this gap, in this paper we study third-party dependencies of websites in Africa. Our study is motivated by the increasing number of DDoS attacks in Africa[21], the increasing popularity of third-party services, the low cyber readiness of African users and businesses [40]. These are exemplified by various recent attacks. For example, in July 2022, Afrihost, one of the major hosting and DNS providers in South Africa, went down for 30 hours due to load shedding which caused a cooling equipment failure in one of Afrihost's datacenters. Moreover, the relative scarcity of local providers urges website operators to rely often solely on global service providers such as Amazon, Akamai, and Cloudflare whose outages also affect users, and websites from Africa.

Beyond raising awareness of the unique security challenges that African users and operators face, our study contributes to the resilience of the Internet in Africa. Concretely, we aim to provide stakeholders and operators with more tailored insights, to help them avoid common pitfalls in using third-party dependencies, understand their attack surface, and optimize their defense strategies towards the most pressing needs.

To investigate third-party dependencies in African websites, we focus on websites which are *Africa-centric*: websites that are popular in Africa (*Africa-visited*), or predominantly targeted towards Africans (*Africa-dominant*), or are hosted in Africa (*Africa-hosted*), or are operated in Africa (*Africa-operated*). We investigate their dependencies using four measurement vantage points in Africa (Nigeria, Rwanda, South Africa, and Kenya). Specifically, our measurement study focuses on answering the following questions: First, how prevalent are third-party dependencies in the *Africa-visited*, *Africa-hosted*, *Africa-operated*, and *Africa-dominant* websites? Second, how centralized are third-party dependencies among providers used in *Africa-visited*, *Africa-hosted*, *Africa-operated*, and *Africa-dominant* websites? Finally, how does the dependence on third parties in Africa compare to the US? Since prior work [28] studies third-party dependencies from a US vantage point, hence, in this work, we use the US as a baseline.

Our main findings are as follows: First, third-party dependencies are 5% to 12% more prevalent in Africa as compared to the US. Moreover, for more popular sites, this gap increases up to 25%. Second, 93% of *Africa-visited*, 97% of *Africa-dominant*, 96% of *Africa-hosted*, and 95% of *Africa-operated* websites are critically dependent on a third-party DNS, CDN, or CA provider. Second, all vantage points in Africa are equally critically dependent on third-party DNS, CDN, and CA providers. Third, the top-three DNS, CDN or CA providers for *Africa-centric* websites serve as sole providers for up to 60% of the websites. Finally, the top providers for *Africa-visited* websites are mainly global providers (*e.g.,* Cloudflare, Amazon, etc.). However, for the hosted, dominant and operated sets, we observe some local providers among the top providers.

Our findings have key implications for the present usage and recommendations for the future evolution of the Internet in Africa. The high degree of centralization of providers and third-party dependencies make African websites vulnerable to various exploits, and availability attacks. While these dependencies mirror trends in other countries such as the US, there are some unique threats. First, Africa has unreliable Internet infrastructure which makes outages more commonplace [45, 10, 42] as observed in the Afrihost outage due to load-shedding [15]. Secondly, African website operators and service providers lack cyber expertise [40], due to which it can take longer for them to recover from an outage. By studying this issue in the African context, we highlight the need to build a more resilient Internet infrastructure in Africa.

2 Preliminaries

Before we formally define our measurement goals, we define a set of actionable metrics that we use throughout our analysis. These metrics have been taken from Kashaf et al. [28]. We also articulate several research questions, that we aim to answer in this study.

2.1 Dependency Metrics

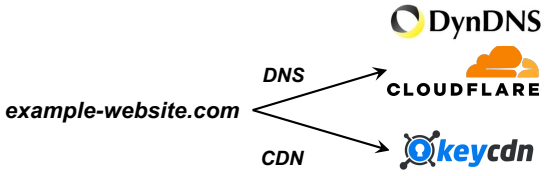


Fig. 1: *example-website.com* has a dependency on CloudFlare DNS and Dyn DNS. Moreover, it has a dependency on KeyCDN for CDN services. Since it uses a single CDN provider, it has a critical dependency on KeyCDN. However, it is redundantly provisioned with respect to DNS as it is using two DNS providers.

When a website uses another entity for a particular service (*e.g.,* DNS), we say that the website has a **third-party dependency** on that service provider, making it a third-party provider as opposed to having a private provider which belongs to the website itself as defined by Kashaf *et al.* [28]. We illustrate this in Figure 1. Here, *example-website.com* uses an entity other than itself for a particular service (here DNS and CDN). Therefore, *example-website.com* has a third-party DNS dependency on Cloudflare and Dyn DNS, and it has a third-party CDN dependency on KeyCDN. *example-website.com* in Figure 1 uses only a single CDN provider. Hence, it has a **critical dependency** on KeyCDN. However, since *example-website.com* uses two DNS providers, it is **redundantly provisioned** with respect to DNS and does not have a critical dependency on Cloudflare or Dyn DNS.

For DNS and CDN, we measure critical dependency by analyzing if a given website is redundantly provisioned or not. However, in the case of CA dependency, a website is critically dependent on a CA if it does not support Online Certificate Status Protocol (OCSP) stapling. If OCSP stapling is enabled, the user accessing a given website does not have to contact the OCSP server to check the website certificate for revocation. Instead, an OCSP response signed by the certificate authority comes stapled from the website server itself, thus removing the dependence on OCSP server [4].

Concentration of a service provider The number of websites dependent on a service provider gives the concentration of that service provider.

Impact of a service provider This gives the number of websites critically dependent on a service provider.

2.2 Taxonomy of Websites

To systematically study third-party service dependencies in Africa-centric web services, we create a taxonomy of websites (Table 1) based on *(i)* who visits them; *(ii)* who operates them; *(iii)* where are they hosted; and *(iv)* who are their dominant users. Below, we define these classes precisely at the granularity of a country.

Users of a website are the people who visit the website. A website may be used primarily by people from a single country (geolocation) or from multiple countries. We define u_C as the Internet user, who is geographically located in country C.

Owner/Operator of a website is the entity or person that builds and manages the website, makes security decisions, defines its privacy policy, etc. A website may have operators in a single country or in multiple countries. We define o_C as the website operator in country C.

Host of a website is the country (or countries) in which the servers running the website are. We use the notation h_C to specify that the hosting location of the website is in the country C.

Dominant country for a website is the country that has the majority traffic share for that website. We use d_C to denote the dominant country for a website.

Who uses it?	Who operates it?	Who hosts it?	Who is it for?	Website sets
u_C	-	-	-	$W_C^{visited}$
-	h_C	-	-	W_C^{hosted}
-	-	o_C	-	$W_C^{operated}$
-	-	-	d_C	$W_C^{dominant}$

Table 1: We consider three sets (categories) of websites for our analysis which differ in the location of their users (usage), the location in which they are hosted (hosting), and their audience.

Using this taxonomy, we define the following website sets:

***Country-visited* websites $W_C^{visited}$:** This set is composed of websites that are used/visited by users u_C of country C. In other words, this includes websites that are popular in the country C. For example, *facebook.com* is among the top 1K websites in Kenya.

***Country-dominant* websites $W_C^{dominant}$:** These websites have the majority of their users in country C. They may be operated or hosted by single or multiple countries. These websites are specifically targeted toward a particular demographic. Studying this set is important because it includes websites that may not be very popular but are essential for African users such as government websites, and hospital websites. This set is different from the $W_C^{visited}$ websites. While the $W_C^{visited}$ set contains websites that are popular in a country, *e.g.,* *facebook.com* is popular in Kenya, however, the $W_C^{dominant}$ set contains websites which are primarily targeting the Internet users of Kenya. *facebook.com* is not primarily targeting Kenyans, while the website *kenyanmusic.co.ke* is primarily targeting Kenyans with its majority traffic from Kenya [3].

***Country-operated* websites $W_C^{operated}$:** This set comprises of websites operated by country C. These websites may have users from single or multiple countries and may be hosted by single or multiple countries. Studying this set facilitates investigating the implications of third-party dependencies from the perspective of African website operators.

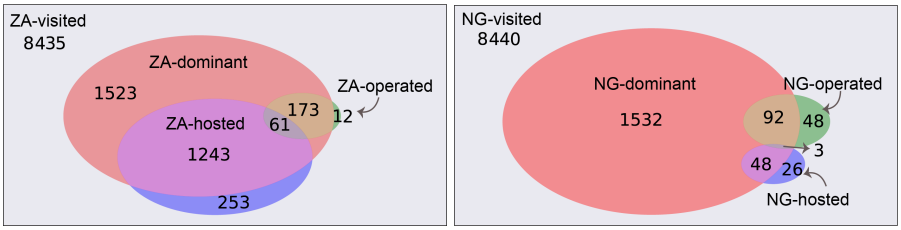
***Country-hosted* websites W_C^{hosted}:** This set comprises of websites that are hosted in country C. Each of these websites may have users from single or multiple countries and may be operated in single or multiple countries. Studying this set is important because it often contains sensitive websites which need to remain local such as banking websites, hospital websites, etc.

2.3 Research Questions

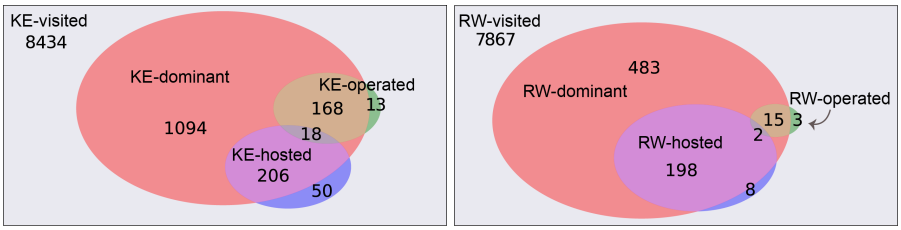
Given these definitions, we now define the main research questions that we answer in this paper.

– How prevalent are third-party critical CDN, DNS, and CA dependencies in Africa-centric websites?

[3] https://www.similarweb.com/website/kenyanmusic.co.ke/#geography



(a) South Africa (ZA) (b) Nigeria (NG)



(c) Kenya (KE) (d) Rwanda (RW)

Fig. 2: The figure shows the relationship between different website sets for all four countries. The visited set is the super-set of all the other sets according to our methodology described in Section 3

- How centralized are third-party dependencies among providers that serve Africa-centric websites?
- How does the state of third-party service dependencies in African countries compare to the US? We compare with the US to use it as a baseline, as the prior work [28] looks at the prevalence of third-party dependencies from a vantage point in the US.

3 Dataset

To perform measurements, first, we pick four vantage points located in Kenya (KE), Rwanda (RW), South Africa (ZA), and Nigeria (NG). We choose these countries as they provide us with a vantage point in each of South, East, West, and Central Africa. Moreover, we found it extremely hard to get physically located servers (using VPN or cloud providers) in more African countries. Many VPN providers do not have physically located servers [60], and cloud providers are largely scarce. Next, we prepare country-specific website sets for each country, and then use the same country as a vantage point to carry out measurements. For example, we study *NG-visited* websites from NG. This section explains our methodology for collecting websites for each country and sets of interest.

One could look at all existing websites that belong the categories we defined in Section 2. To make the sets more tractable and focus on the most impactful

websites, we start from the popular websites in each African country which constitutes the *country-visited* set $W_C^{visited}$. This helps in identifying websites that can impact the African Internet users, website operators, and the Internet economy of African countries the most.

We use the Chrome User Experience Report (CrUX) dataset [23] to get the top 10K popular websites in each selected African country. This dataset is curated monthly by aggregating browsing data of Chrome and Chromium users who have opted in for browser history and usage statistic reporting.This opt-in requirement may introduce bias and the list may not truly reflect popular websites in a region, however, prior work has evaluated the Google CrUX dataset and found it to be quite reliable with respect to popularity [53]. Moreover, Chrome and Chromium browsers constitute more than 80% traffic in our countries of interest [56]. CrUX is ranked by the number of completed page loads.

The CrUX dataset is aggregated by web origin (e.g., https://google.com). For DNS analysis, we need domain names, and using web origin may result in multiple entries for the same domain. Hence, we normalize this dataset by grouping web origins by domain names and choosing the smallest rank value as the rank for each domain. This same normalization technique has been previously done in prior work [53] and is shown to be accurate at capturing popular websites. [53] also shows that CrUX is better at capturing popular websites than other top lists as defined by visit and visitor metrics. In addition, most top lists only give popular websites in the world. However, for this analysis, we need regional popular websites and found that the CrUX dataset is a good source for that. We build our website sets based on the CrUX dataset for August 2022 and the definition of website sets can be found in Table 1.

Dataset for *country-visited* websites: We use the CrUX dataset of the top 10K websites, for NG, RW, KE, ZA, and US. We normalize this dataset for each country, by grouping web origins by domain names as mentioned above. This gives us the *country-visited* $W_C^{visited}$ dataset for each country.

The *country-dominant*, *country-hosted*, and *country-operated* website sets are built from this dataset with the relationship shown in Figure 2 for all countries. We describe our methodology below:

Dataset for *country-dominant* websites: As defined in Section 2, *country-dominant* websites are made for users located in the corresponding African country. A naive approach to collecting such a list is to filter websites by their country code top-level domain (ccTLD) [8]. However, this approach would result in many false positives because some domain registrars give .ccTLD domains to anyone. For example, `parse.ly` has the Libyan ccTLD, but the website is not made for or visited by Libyan users [4]. Therefore, we combine multiple heuristics to collect the *country-dominant* websites. Concretely, a website belongs to the *country-dominant* set, if it belongs to *country-visited* set which is the top 10K visited websites and satisfies one of three requirements. First, we pick websites with ccTLDs that belong to that particular country. Observe that this filtering is different from the previous heuristic as we require that the website is popular

[4] https://www.similarweb.com/website/parse.ly/#geography

in that country and has the ccTLD of the same country. Second, the website hostname contains Africa or the name of an African country. Again, while this heuristic would alone cause false positives *e.g.,*`ancient-egypt.org`[5] intersecting it with the popular sites in Africa considerably decreases those cases. Finally, we look at the website content of the landing page, and the website URLs referred to in the landing page to get the phone number associated with the website. We only consider a website to belong to a particular country if all the phone numbers mentioned on it have the country code of that country. This technique reduces false positives resulting from websites containing multiple phone numbers, not necessarily belonging to the website. For example, the website *viagogo.com* contains phone numbers of multiple countries including South Africa but its dominant country is actually Brazil[6]. To conclude, we define *country-dominant* websites $W_{dom-afr}$as:

$$W_C^{dominant} = (W_C^{ccTLD} \cup W_C^{substr} \cup W_C^{phone}) \cap W_C^{visited}$$

Dataset for *country-hosted* websites: To find websites hosted in an African country, we perform an IP geolocation lookup using the Maxmind GeoLite database [39], and the `ipinfo.io` [27] database for IPs missing in the MaxMind database. Instead of performing IP geolocation lookup on all existing websites, here also we only look at websites that are hosted in the corresponding country, and are also popular in it *i.e.,* are in the $W_C^{visited}$ set. IP geolocation databases have certain limitations. Particularly, these databases tend to erroneously geolocate IPs that belong to ASes with global presence and IPs that change ownership due to merger and acquisition as observed by prior work [37]. This may misclassify some websites as not being hosted in Africa.

Dataset for *country-operated* websites: To find websites operated in a given country, we look at the privacy policy and terms and conditions of websites to identify the country of interest. For example, in *murukali.com*, their terms and conditions page mentions, "These Terms of Service and any separate agreements whereby we provide you Services shall be governed by and construed in accordance with the laws of Rwanda." We use the Python Geograpy library [50] to extract geolocation mentions in the privacy policy or terms. We include only those websites for which we get a single country name, to decrease the number of false positives in classifying a website as being operated in a given country.

4 Methodology

We are interested in measuring the third-party dependencies of *Africa-centric* websites on authoritative Domain Name Servers, Content Delivery Networks, and Certificate Authorities for revocation information (OCSP servers and certificate revocation list (CRL) distribution points).

[5] https://www.similarweb.com/website/ancient-egypt.org/#geography
[6] https://www.similarweb.com/website/viagogo.com/#geography

To capture dependencies as observed by African users, we need to measure from multiple locations in Africa. Yet, accessing servers in various locations within Africa is challenging, due to the limited offered coverage from cloud service providers. To address this challenge, we combine the limited cloud presence with VPN services (PrivateVPN [48], ExpressVPN [20]) whose true location we diligently verified. We perform our measurements in October 2022 using four vantage points in Africa, scattered in East, West, South, and Central Africa. Particularly, our vantage points are in South Africa, Nigeria, Kenya, and Rwanda. For the South Africa vantage point, we use Amazon AWS, while all others are VPNs. To verify the location of the VPN server, we ping the server from different locations to identify the location with the smallest ping using online services like *ping.pe*, we also perform traceroute and we separately also reached out to the VPN provider to confirm the location of the VPN.

DNS Measurements: Given a website, we find out, 1)Does the website has a dependency on a third-party DNS provider? If so, 2) Is the website critically dependent on that DNS provider, or is it redundant? To find out the authoritative name servers, we use dig [17] (Domain Information Groper) which is a command-line tool to fetch the NS (nameserver) records which give the records for authoritative nameservers of a given website. To identify third-party nameservers, we follow the methodology documented in Kashaf *et al.* [28]. Particularly, we use top-level domain (TLD) matching [32], subject alternate name (SAN) lists [54], and start-of-authority (SOA) DNS record [6] to classify an NS as a third party. Particularly, we check if the second-level domain (SLD) and top-level domain (TLD) of the website and the NS match (*e.g.,* website *www.example.com* and its NS *ns1.example.com* have same SLD+TLD *i.e., example.com*) or if the SLD+TLD of the NS exists in the SAN list of the website, we classify it as private. If the SOA of the website and the NS do not match or if the concentration of the NS exceeds 50, we classify it as a third party. Kashaf *et al.* [28] shows that these heuristics can accurately classify providers as private or third-party. We get 12825 distinct (website, nameserver) pairs for *KE-visited*, 12287 pairs for *NG-visited* and 12792 pairs for *RW-visited*, and 14336 for *ZA-visited* websites of which 3% remain uncategorized as third-party or private for ZA, 4% for RW, 3% for NG and 2% for KE, and 3% from US. Hence, we conservatively exclude the websites involving them from our analysis. After identifying the third-party nameservers, we need to check if a website is redundantly provisioned. To do this, we group the nameservers of the websites by TLD and SOA records as documented in Kashaf *et al.*. Nameservers in the same group are considered to belong to the same provider. We observe 1010 distinct nameservers for KE, 1170 for NG, 1078 for RW and 980 for the ZA, and 1274 for US.

Certificate Revocation Measurements: Given a website, we are interested in knowing, 1)If the website has a dependency on a third-party CA provider. If yes, 2) Is the website critically dependent on that CA, or has it enabled OCSP stapling? We extract the CRL distribution points (CDP) and OCSP server information from the SSL certificate of the website. To fetch certificates, we first send a SYN on TCP port 443 to see if the website supports HTTPS. If we receive

a Connection Refused error, then it means the website does not support HTTPS. Next, we initiate an HTTPS connection with it and fetch the SSL certificates. In the *NG-visited* websites, 94.0% support HTTPS, 95.7% support HTTPS in *KE-visited*, and 94.3% support HTTPS in the *RW-visited*, 95.2% in *ZA-visited* and 94.6% in *US-visited*. We observed 22 distinct CAs for NG, 26 distinct CAs for RW, 24 distinct CAs for KE, 23 distinct CAs for ZA, and 23 distinct CAs for US. We classify the CAs as a third party, again using TLD matching, SAN list, and SOA records [28].

Certain private CAs issue certificates and provide revocation checking for their own domains only, e.g., Microsoft, etc. Therefore, we use the same heuristics as mentioned for DNS to classify whether OCSP servers and CDPs are private or third parties as in [28]. Particularly, we classify a CA as private if the SLD+TLD of the website matched the SLD+TLD of the OCSP server, or if the SLD+TLD of the OCSP server exists in the SAN list of the website. Moreover, we classify the CA as third-party if the SOA of the OCSP server and the website differ. Next, to see if a website has a critical dependency on OCSP servers, we check if it has enabled OCSP stapling using OpenSSL [61]. If enabled, the certificate's revocation status comes stapled from the webserver when a user visits the website, requiring no online revocation check from the OCSP server.

CDN Measurements: To find CDNs used by a website, we look at the canonical name (CNAME) redirects for the internal resources of a webpage. If the website is using a CDN for a particular resource, the CNAME of that resource will point to the CDN. First, we render the landing page of the website using Puppeteer [49] and record the URL of all the resources retrieved by a website. Then, if the SLD + TLD of the resource matches that of the website or it exists in the website's SAN list, we classify it as an internal resource [28]. Then, we query the CNAME record for all internal resources of the webpage and use the CNAME-to-CDN map from the prior work [28], which we verified and extended to include African CDNs. Then we classify a CDN as a private or third party by using the same SLD+TLD matching, SAN Lists, and SOA records as done in the case of DNS and CA and in [28]. We find that 18.5%, 23.9%, 19.6%, 22.0%, and 40.4% use CDN for *NG-visited*, *RW-visited*,*KE-visited*, *ZA-visited* and *US-visited*. We observe 56 CDNs for NG, 59 CDNs for RW, 59 CDNs for KE, 55 CDNs for ZA, and 60 CDNs for the US.

4.1 Limitations

- Our analysis considers only four vantage points in Africa. It is possible that the dependencies in countries for which we do not have more vantage points vary greatly. While we accept this limitation, however, getting vantage points in some of the African countries is extremely hard due to the lack of mature Internet infrastructure, including VPN server presence.
- We only look at popular websites. While this may overlook certain websites, studying all possible websites is not feasible. We argue that this is a reasonable compromise as popular websites will be the ones that impact African users and businesses the most.

– Our heuristics for *Africa-dominant* websites may have false positives and negatives. However, the correct way to find *Africa-dominant* websites would be to choose the websites which have the largest traffic share from Africa. We had to use these heuristics because the data for per country traffic share of a website is not available.
– We inherit the limitations of Kashaf *et al.* [28] as we use their methodology.
– We describe OCSP revocation checking as a critical dependency from a website's point of view. However, the online revocation behavior of browsers differs. For example, many existing browsers circumvent online revocation checking by using other mechanisms like CRLsets in Chrome [12]. Similarly, Safari performs online revocation checking in case of revoked certificates. Many browsers also consider failures in revocation checking as a soft-fail. Note that some browsers allow users to enable online revocation checking. Moreover, the system's TLS stack at the user in some cases always performs online revocation checking no matter what browser is used [12]. Hence, we focus on the dependency from the website side while keeping in mind that at the user end, there may be other accommodations that make the dependency not critical.

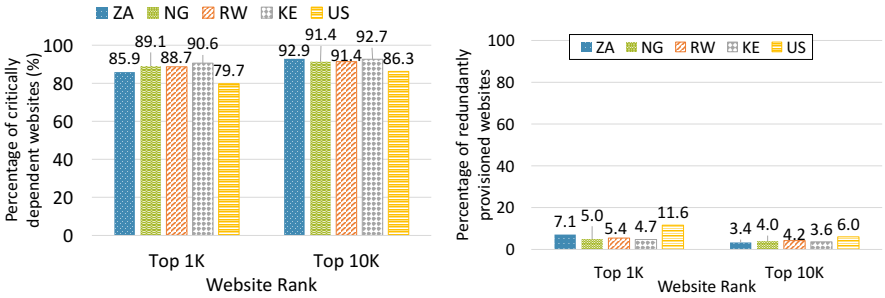
5 Findings

In this section, we first analyze third-party dependencies in *Africa-centric* websites and use the US as a baseline. Next, we analyze provider concentration to identify single points of failure in the African web ecosystem. Note that for all the findings below, we show comparisons using percentage point difference.

5.1 Third-Party Dependencies

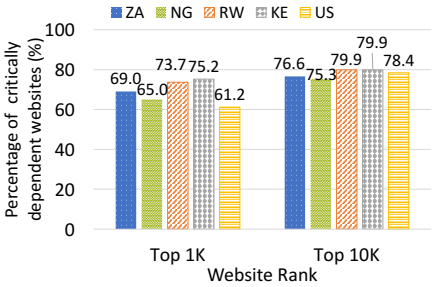
Observation 1: ***93% of Africa-visited websites are critically dependent on third-party CAs, CDNs, or DNS. In perspective, US-visited websites are up to 25% less critically dependent.***

Figure 3 illustrates the portion of *US-visited*, *ZA-visited*, *NG-visited*, *RW-visited* and *KE-visited* websites that are critically dependent or redundantly provisioned as a function of the particular service and as measured by vantage points that are in the corresponding region. Concretely, Figure 3a shows the percentage of critically dependent websites on DNS in African countries and the US. We observe that for DNS, critical dependency in *US-visited* is 5% to 7% less as compared to the African countries. Interestingly, when we look at more popular websites (top 1K), this gap further increases (6% to 10%) as shown in Figure 3a. This means that web users from the US are comparatively less vulnerable, especially if they are visiting more popular websites, as compared to African users. This result indicates that more popular websites in the US may

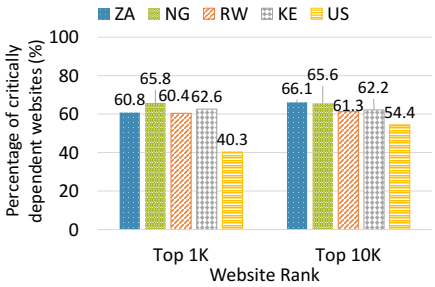


(a) DNS critical dependency

(b) DNS redundancy



(c) CDN critical dependency



(d) CA critical dependency

Fig. 3: (a) Critical DNS dependency for top 10K *US-visited* sites when measured from a US vantage point is 5% to 7% less than the top 10K *Africa-visited* websites. This gap in critical dependency increases to 6% to 10% in the more popular (top 1K) websites. (b) The percentage of websites that are redundantly provisioned is slightly higher (2%) in the *US-visited* websites as compared to the *Africa-visited* websites. However, when we look at more popular websites (top 1K), for *US-visited*, the percentage of redundantly provisioned websites is 5% to 7% higher than the *Africa-visited* websites. (c) Critical CDN dependency for the top 10K *US-visited* sites is similar to the top 10K *Africa-visited* websites. However, for more popular websites, *US-visited* sites are 4% to 15% less critically dependent than *Africa-visited* sites. (d) Critical CA dependency for the top 10K *US-visited* sites, when measured from a US vantage point, is 7% to 12% less than the top 10K *Africa-visited* websites. This gap in critical dependency increases to 20% to 25% in the more popular (top 1K) websites.

care more about availability as compared to the popular websites in African countries, making African Internet users more vulnerable.

Figure 3b illustrates the percentage of redundantly provisioned websites in DNS. We observe that there is not much difference (2%) between *US-visited* websites and *Africa-visited* websites. However, when we look at more popular websites (top 1K), the gap increases by 5% to 7% from 2%. At the same time, we find that the use of private DNS is only 3% to 4% higher in *US-visited* websites (not shown) and becomes 2% to 5% when we look at more popular websites (also not shown). This means that critical dependency in more popular *US-visited* websites is reduced because of an increase in redundancy instead of the

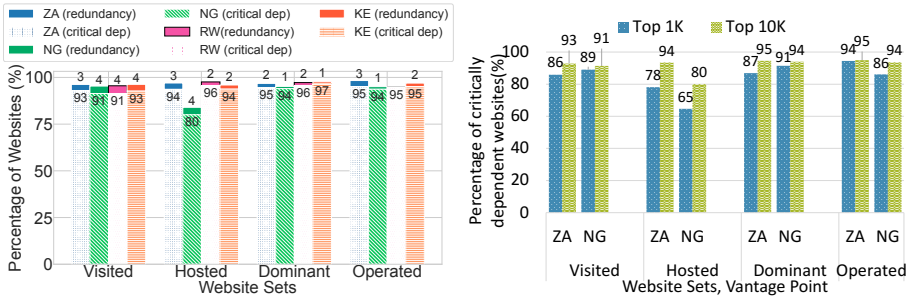
use of Private DNS. However, for *Africa-visited*, there is not much significant increase in redundancy for more popular websites, except South Africa.

In case of CDN dependency, 22%, 18%, 23% and 19% websites use a CDN in *ZA-visited*, *NG-visited*,*RW-visited* and *KE-visited* websites respectively, while in *US-visited*, 40% websites use CDN (not shown here). Figure 3c compares the critical CDN dependency in *US-visited* with *Africa-visited* websites. In the top 10K, critical CDN dependency in *US-visited* is comparable to the *Africa-visited* websites. We find the number of redundantly provisioned websites is also similar (not shown here). When we look at more popular websites (top 1K), the critical CDN dependency in *US-visited* is 4% to 14% less than *Africa-visited* websites while the CDN adoption in the top 1K websites is almost double in the US (44.6%) than African countries (20% to 27%). The use of private CDN remains negligible in *US-visited* and *Africa-visited* websites (not shown here). Moreover, the percentage of redundantly provisioned websites in the top 1K is 5% to 15% higher for the *US-visited* as compared to the *Africa-visited* websites. The reduced critical dependency as we move towards more popular websites in *US-visited* websites is because of an increase in redundancy.

Figure 3d shows the percentage of websites critically dependent on a CA in the *US-visited* and *Africa-visited* websites. The number of websites that support HTTPS is similar in *US-visited* and *Africa-visited* websites (not shown). Recall that for CAs, critical dependency is measured in terms of whether a website supports OCSP stapling or not. We find that *US-visited* websites are 6% to 12% less critically dependent on CAs compared to *Africa-visited*. Moreover, as we move to more popular websites (top 1K), the gap in critical dependency between *US-visited* and *Africa-visited* websites further increases to 20%-25%. This low adoption of OCSP stapling may be an indicator of low cyber readiness in Africa. Furthermore, in the US there have been many efforts to promote OCSP stapling, particularly by popular CDN providers such as Cloudflare, Amazon Cloudfront, and Akamai. Since the adoption of CDNs in *Africa-visited* websites is low, this could explain the lower adoption of OCSP stapling.

Observation 2: ***Critical DNS dependency in Africa-centric websites is extremely prevalent (92% to 97%), leaving users highly exposed. Third-party critical DNS dependencies are higher in more popular websites compared to less popular ones.***

To further investigate the results of Figures 3a and 3b, Figure 4a also shows critical dependency and redundancy of websites in a third-party DNS provider but distinguishes them between visited, hosted, dominant, and operated website sets. For the set of visited websites, the critical DNS dependency is very high 91% to 93%, and stable across countries. This shows that users in Africa from these countries are equally vulnerable to the side effects of DNS third-party dependencies. If we look at the hosted websites, the *NG-hosted* websites are less critically dependent compared to other African countries. Concretely, the third-party DNS dependency is only 84% in *NG-hosted* websites. This is due to



(a) (b)

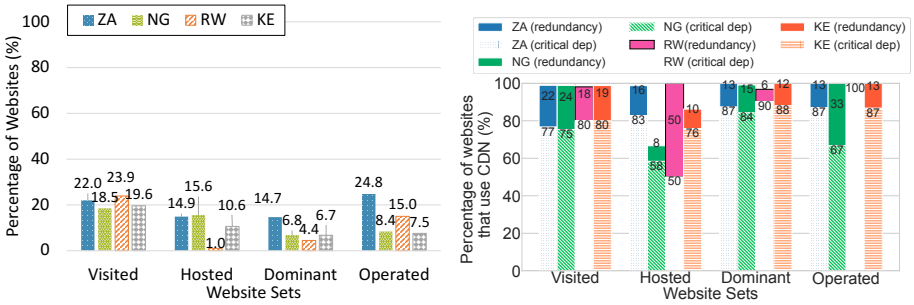
Fig. 4: (a)We show the percentage of critically dependent websites on third-party DNS providers with the percentage of redundantly provisioned websites stacked on it. The height of the bar stack shows the percentage of websites using a third-party DNS provider. Third-party critical DNS dependency is highly prevalent (more than 90%) in Africa-centric websites when measured from all four vantage points. (b) For each website set, we show how the critical dependency varies as we move from more popular (top 1K) websites to less popular (top 10K) ones for ZA and NG. Across all website sets, less popular websites are more critically dependent than more popular ones.

two key reasons. First, many websites belonging to Meta (*e.g.,* facebook.com, freebasics.com, whatsapp.com, etc.) are locally hosted in Nigeria, and these websites use private DNS. Indeed, we confirmed their hosting by pinging them from Nigeria. Second, the *NG-hosted* sets contain only a small number of websites (Figure 2) making the Meta associated domains statistically significant.

For dominant website sets, critical dependency for all African websites is very high, concretely 94% to 97%. In fact, the websites that predominantly target African Internet users are more vulnerable than *Africa-visited* websites, with a difference of 2% to 5%. There is almost negligible redundant provisioning in this set. For operated websites, again the critical dependency is 94% to 95% with negligible redundant provisioning. This trend in general shows that no matter where in Africa users are, or what they visit, they are highly vulnerable to the side effects of third-party DNS dependencies. Moreover, the fact that the trend persists across all African countries that we studied shows that the situation is dire for the entire continent. In fact, the countries for which we have results have relatively more developed Internet infrastructure.

Across countries, we observe that for ZA, critical dependency (though very high), and redundancy remain similar across different website sets. For NG and RW, with the exception of the *NG-hosted* websites, critical dependency in the specialized (dominant, hosted, operated) sets is larger than the corresponding *visited* set. In NG, this is because of a decrease in redundancy, while in RW this is because of a decrease in redundancy and a decrease in the use of private providers. KE has similar trends to RW and NG as shown in Figure 4a. All in all, specialized sets have reduced redundancy except for ZA.

Figure 4b shows the critical dependency for the top 1K and top 10K websites for ZA and NG. As we move towards more popular websites (top 1K), the critical



(a) (b)

Fig. 5: (a) We show the percentage of websites that use CDN in different website sets for each country. CDN usage is less in the specialized sets such as hosted, dominant, and operated as compared to the visited set except for ZA. (b) We show the percentage of critically dependent websites on third-party CDN providers with the percentage of redundantly provisioned websites stacked on it. The height of the bar stack shows the percentage of websites using a third-party CDN provider. Critical dependency on CDNs for Africa-centric websites is less prevalent as compared to critical DNS dependency.

dependency decreases across all website sets for both ZA and NG. This is partly because of an increase in the number of websites using Private DNS (not shown). For example, for ZA, third-party dependency decreases by 4% for *ZA-visited*, and *ZA-dominant*. For *ZA-hosted* it decreases by 8%, while for *ZA-operated* it remains the same. In addition to the increase in private DNS, we also observe an increase in redundantly provisioned websites. For example, in the case of ZA, redundantly provisioned websites increase from up to 4% in the top 10K, to 6%-12% in the top 1K. We observe a similar trend in NG, KE, and RW. While the increase in redundancy for more popular websites is encouraging, it is still far from ideal. Even for more popular websites, third-party dependencies are highly prevalent. Across different website sets, we see more encouraging trends. For example, the hosted websites in the top 1K are far less critically dependent than the other website sets. However, this trend is only for ZA and NG and does not appear in KE and RW where it is more similar to the other sets. In NG, this decrease in critical dependence is primarily because of the use of Private DNS. For ZA, however, this is because some of the websites using global providers are using multiple providers, and also because all the websites using TENET South Africa as DNS, are redundantly provisioned.

Observation 3: ***Among the websites that use CDN, critical dependency is prevalent (75% to 80%); although less compared to DNS. Third-party critical dependencies in CDN are higher in more popular websites compared to less popular ones.***

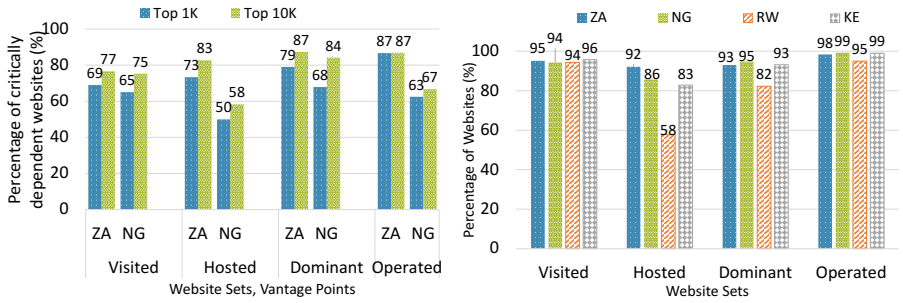
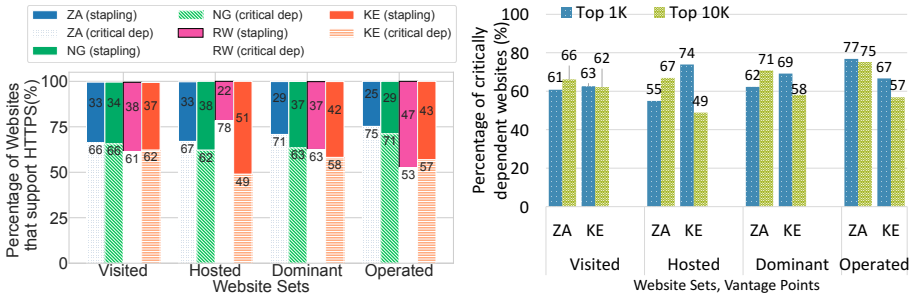


Fig. 6: (a) For each website set, we show the change in critical dependency as we move from more popular (top 1K) websites to less popular (top 10K) ones for ZA and NG. Critical CDN dependency is lower for more popular websites, as compared to the less popular ones. (b) The percentage of HTTPS support in websites is very high in Africa-centric websites, with the exception of the *RW-hosted* set.

Figure 5b shows the critical dependency and redundancy in websites that use CDN for different website sets. Here, the sum of critical dependency and redundancy gives the total third-party dependency. The number of websites using a CDN in each set is shown in Figure 5a. In the visited sites, 18.5% to 23.9% use a CDN. In general, we observe a decrease in critical dependency as compared to DNS. In the case of visited websites, *ZA-visited* and *NG-visited* are slightly less critically dependent and are slightly more redundantly provisioned as compared to *RW-visited* and *KE-visited* websites. The use of private CDN across all vantage points of the visited set is less than 1%, which is not surprising. For the hosted set, *NG-hosted* has a higher percentage of websites with private CDN (100-58+8). This is because the websites affiliated with Meta use a private CDN. We ignore the trend in *RW-hosted* websites, as only 1% (2 websites) use CDN. *KE-hosted* websites are less critically dependent than *KE-visited* websites; this is also because of the private CDN using Meta domains, which become statistically significant because not many *KE-hosted* websites use CDN. For *ZA-hosted* websites, the critical dependence is higher than *ZA-visited* websites. It is unclear why this is the case as the CDN providers for both sets are similar.

For the dominant website set, all the countries have more critical dependence compared to the visited set. This means websites that predominantly target African users are more vulnerable. However, as shown in Figure 5a, only a very small number of the dominant websites use a CDN. In the case of operated websites, the critical dependency is high for *ZA-operated*, *KE-operated*, and *RW-operated* websites. We do not see a specific reason for which the *NG-operated* are less critically dependent. The adoption of CDN in the specialized sets of hosted, operated, and dominant is very less to have a significant impact.

All in all, across countries, critical CDN dependency in the specialized set is higher than the visited set for ZA due to decreased redundancy. RW and KE follow the same trend with the exception of the hosted set. For NG, hosted, and



(a) (b)

Fig. 7: (a) We show the percentage of critically dependent websites on CA providers with the percentage of websites having stapling enabled stacked on them. Third-party CA critical dependency is less prevalent in Africa-centric websites as compared to DNS dependency. Moreover, KE and RW are less critically dependent in the hosted, dominant, and operated sets as compared to ZA and NG. (b)For each website set, we show the change in critical CA dependency as we move from more popular (top 1K) websites to less popular (top 10K) ones for ZA and KE. Increase in popularity does not reduce critical CA dependency in Africa-centric websites. In fact, for KE (and also NG and RW), critical CA dependency increases as we move towards more popular websites.

dominant sets have reduced critical dependency compared to the visited set, while the operated set has increased critical dependency.

Figure 6a shows the change in critical CDN dependency as we move from more popular (top 1K) websites to less popular websites (top 10K). For example, for ZA, the critical dependency for more popular websites is 8% to 10% lower than less popular ones (except the operated set). We observe a similar trend for RW and KE. This reduction in critical dependency for more popular websites is because they are more redundantly provisioned. The use of private CDN remains negligible for the top 1K and top 10K websites (not shown here).

Observation 4: ***In the case of CA critical dependency, 40% to 75% of the Africa-centric websites are critically dependent. For the hosted, dominant, and operated website sets, more popular websites are more critically dependent.***

Figure 6b shows the number of websites that support HTTPS. HTTPS adoption is in general very high in Africa-centric websites, which is encouraging. However, there are a few notable exceptions. For example, HTTPS adoption is low particularly in the *RW-hosted* websites. It is also low for *NG-hosted* and *KE-hosted* when compared to the visited websites. For RW, the *RW-dominant* website set also has lower HTTPS adoption as compared to other countries.

Figure 7a shows the percentage of critically dependent websites among all HTTPS-supporting websites. In general, critical dependency on CAs is less compared to DNS. In the visited website set, 33% to 38% of the websites that support

HTTPS, also support OCSP stapling. In the case of hosted websites, the trend remains largely similar for ZA and NG. For RW, which already has only 58% HTTPS (Fig 6b supported websites in *RW-hosted* website set, for the remaining websites, only 22% support OCSP stapling. Hence, the *RW-hosted* websites leave African users particularly vulnerable. More alarming is the fact that more than half of these critically dependent websites are government websites ending with *.gov.rw*. For KE, 51% of the *KE-hosted* websites support OCSP stapling, which is encouraging. OCSP stapling support in KE is in general better for all website sets as compared to other countries. In the case of RW, OCSP stapling support is also good except for the *RW-hosted* websites. OCSP Stapling support for ZA is not very encouraging compared to other African countries. The *ZA-operated* and *ZA-dominant* websites are particularly more vulnerable than the respective sets in other countries. This means that ZA Internet users are vulnerable to the side effects of third-party CA dependency. In the case of NG, the *NG-operated* websites are more vulnerable compared to other website sets for NG.

Overall, critical CA dependency in the specialized sets for ZA is higher than in the visited set. For KE, the trend is the opposite. For NG, all sets have a similar critical dependency with the exception of the *NG-operated* set. For RW, critical dependency is higher for the hosted and dominant set, while lower for the operated set when compared to the visited websites.

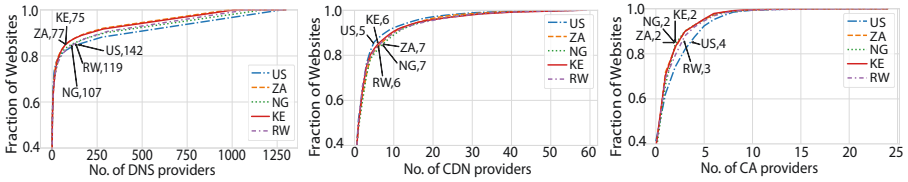
Figure 7b shows the change in critical dependency as we move from more popular (top 1K) websites to less popular (top 10K) websites. For ZA, the critical CA dependency follows the same trend as in the case of DNS and CDN, where more popular websites are less critically dependent (except for *ZA-operated* websites). However, for KE, critical dependency actually increases in more popular websites (top 1K). We observe a similar trend for NG and RW. It is unclear why this is the case. Nevertheless, it is not encouraging and implies that more popular hosted, dominant, and operated websites are more vulnerable to the side effects of third-party CA dependency including outages, performance degradation, etc.

6 Provider Concentration

In this section, we first look at the concentration among providers for *Africa-visited* websites and use *US-visited* websites as a baseline. Then we closely look at Africa, for different website sets.

Observation 5: ***The concentration of providers in Africa-visited websites is slightly higher than US-visited websites for DNS and CA.***

Figure 8 shows the cumulative fraction of websites for a given number of DNS, CDN, and CA providers. To compare the degree of concentration between *Africa-visited* and *US-visited* websites, we plot the fraction of websites served by a given number of providers. We label the number of providers that cover 85% of the websites for each country. In general, we observe a similar degree of concentration in *US-visited* and *Africa-visited* websites. Figure 8a shows the fraction



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

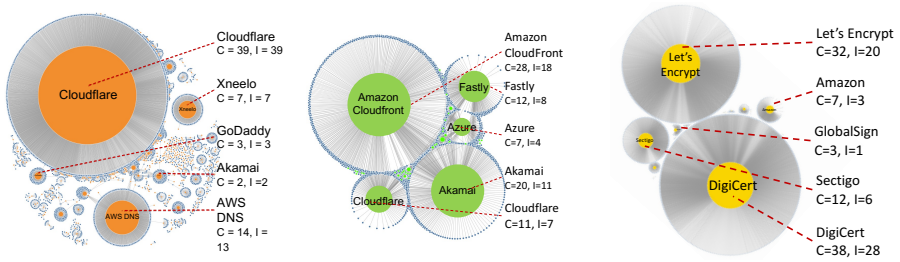
Fig. 8: The CDF of websites against the number of DNS, CDN, and CA providers for African countries and the US is shown. (a) Concentration of DNS providers in *ZA-visited* and *KE-visited* is slightly higher than *RW-visited*, *NG-visited* and *US-visited* websites. (b) The concentration of CDN providers in *Africa-visited* and *US-visited* websites is largely similar, with the concentration in *US-visited* websites being slightly higher. (c) The concentration of CA providers in *Africa-visited* websites is slightly higher than the *US-visited* websites.

of websites served by a given number of DNS providers. For *ZA-visited* and *KE-visited* websites, the concentration is slightly higher than *US-visited* websites. In general, a single DNS provider critically serves more than 40% of *Africa-visited* websites, while in the case of *US-visited*, a single provider critically serves 34% of the websites. Interestingly, the top 5 providers for *US-visited*, *NG-visited*, *RW-visited*, and *KE-visited* websites are the same global DNS providers (Amazon, Cloudflare, GoDaddy, NS1, Akamai). However, for *ZA-visited* websites, we do find local providers like Xneelo and Afrihost.

Figure 8b shows the fraction of websites served by a given number of CDN providers. We observe high concentration in *Africa-visited* as well as *US-visited* websites. Moreover, top CDN providers in *Africa-visited* and *US-visited* websites are also the same and are all global providers. Although *US-visited* websites have higher CDN adoption, the concentration among providers remains the same, which means websites are using the same few CDN providers. Figure 8c shows the fraction of websites served by a given number of CA providers. CA providers are more concentrated for *Africa-visited* as compared to *US-visited* websites. While the top providers in *US-visited* and *Africa-visited* websites are similar, we observe some minor differences. For example, Let's Encrypt and Sectigo are more popular in *Africa-visited* websites as compared to *US-visited* websites where Amazon is more popular. In general, DigiCert is the major provider in all.

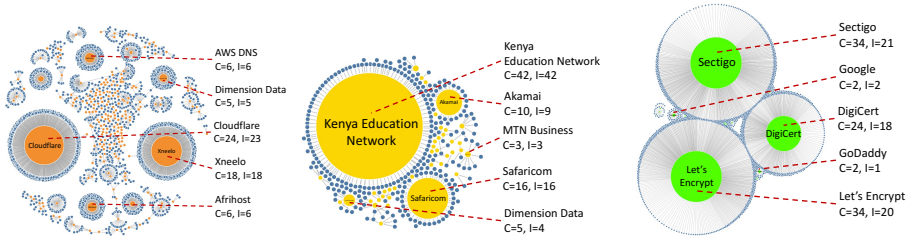
Overall, we observe that African users are as vulnerable to the side effects of third-party dependencies as US users. Note that this is not encouraging or alleviating because Africa faces more challenges with respect to cyber security expertise, reliable infrastructure, etc., and hence single points of failure in Africa can have more severe consequences.

Observation 6: ***Approximately 60% of the total African-visited sites are critically dependent on the top 3 DNS, CDN, or CA providers.***



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

Fig. 9: Figure 9a shows the dependency graph of the *ZA-visited* websites on third-party DNS providers, Figure 9b shows the dependency graph of *NG-visited* websites on third-party CDNs, and Figure 9c shows the dependency graph of *KE-visited* websites on third-party CAs. The size of a node in the dependency graph is proportional to its in-degree (signifying a dependency on the provider). We label the concentration C and impact I of the top 5 providers in terms of the percentage of total websites. (a) Cloudflare and Amazon serve most of *ZA-visited* websites and have higher concentration and impact than other third-party DNS providers. (b) Amazon Cloudfront and Akamai have a slightly higher concentration and impact as CDN providers for *NG-visited*. (c) DigiCert and Let's Encrypt serve the largest number of *KE-visited* websites and have a higher concentration and impact than other CA providers.



(a) *ZA-dominant* $\rightarrow$ *DNS* (b) *KE-hosted* $\rightarrow$ *DNS* (c) *NG-dominant* $\rightarrow$ *CA*

Fig. 10: (a) Africa local providers like Afrihost and Xneelo show up in the top 5 DNS providers for *ZA-dominant* websites. (b) Kenya Education Network provides DNS service for the largest number of *KE-hosted* websites. (c) Sectigo, Let's Encrypt, and DigiCert provide CA services to almost the same number of *NG-dominant* websites. The three providers also have similar concentration and impact.

Figure 9 shows the dependency graph for *Africa-visited* websites. The size of a node is proportional to its in-degree which is the number of websites dependent on it. We also label the concentration (C) and impact (I) of each provider as described in Section 2 in terms of the percentage of websites. Figure 9a shows the dependency graph for DNS providers for *ZA-visited* websites. We find that Cloudflare alone critically serves 39% of the *ZA-visited* websites. In general, the top 3 DNS providers critically serve 59% of the websites. We observe the same trends for other countries. For example, for *RW-visited* websites, the top 3 DNS providers critically serve 60% of the websites, 62% for *NG-visited*, and 58% for

KE-visited websites. Moreover, the top 2 providers in all the countries are the same, namely Cloudflare and AWS DNS. For NG, we do not observe any local DNS provider in the top 10 DNS providers. For Kenya, we observe the Kenya Education Network (KENET) as one of the major DNS providers. For Rwanda, we observe AOS.rw as one of the major local providers. For ZA, we observe many local DNS providers in the top 10, namely Xneelo, Dimension Data, DiaMatrix, and Afrihost. For more popular websites (top 1K), the local DNS providers also come in the top 3 providers, for example, Kenya Education Network (KENET) for top 1K *KE-visited* websites, AOS.rw for top 1K *RW-visited* websites, and Dimension Data for top 1K *NG-visited* websites. However, websites using these local providers have almost zero redundant provisioning.

In the case of CDN providers, Figure 9b shows the CDN dependency graph for *NG-visited* websites. The top 3 providers in *NG-visited* websites critically serve 37% of the websites that use CDN. We observe similar trends for other countries. For example, for *RW-visited* websites, the top 3 CDN providers critically serve 47% of the websites, 39% for *ZA-visited*, and 44% for *KE-visited* websites. Importantly, we find no local CDN provider being used by our African websites. Moreover, the top CDN providers remain similar for all African countries, even for more popular websites (top 1K).

In the case of CA providers, Figure 9c shows the CA dependency graph for *KE-visited* websites. The top 3 CA providers critically serve 54% of the *KE-visited* websites. We observe similar trends for other countries. For example, for *NG-visited* and *ZA-visited* websites, the top 3 CDN providers critically serve 57% of the websites that support HTTPS, and 51% for *RW-visited* websites that support HTTPS. The top CA providers across all countries remain the same. There are again no local providers.

For the *Africa-dominant* websites, many local providers dominate. For example, Figure 10a shows the DNS dependency graph for *ZA-dominant* websites. The concentration of DNS providers is evident: the top 3 DNS providers critically serve 47% of the *ZA-dominant* websites. More importantly, the top providers include many local providers such as Afrihost, Xneelo, and Dimension Data. *KE-dominant* websites have similar trends in DNS dependency, where KENET, Safaricom, and Kenya Web Experts are among the top providers. Similarly, for RW, local providers such as AOS.rw, Kaneza, and Afriregister are among the top providers. However, for *NG-dominant*, we do not see any local DNS provider.

Overall, there is concentration in *Africa-dominant* websites across all services. For example, in the top 3 DNS providers for *Africa-dominant*, the concentration remains between 48% to 58%. In the case of CDN dependency, the concentration of top 3 CDN providers for *Africa-dominant* websites remains around approximately 50% to 63%. Similarly, for CA dependency in *Africa-dominant* websites, the concentration of top 3 CA providers for *Africa-dominant* websites remains around approximately 52% to 62%. In the case of CDN and CA dependency, we do not see any local providers across all website sets. For example, Figure 10c shows the CA dependency graph for *NG-dominant* websites.

In the case of *Africa-hosted* websites as well, there is concentration across all services. Figure 10b shows the DNS dependency graph for *KE-hosted* websites. A large number (42%) of these websites are served by Kenyan Education Network (KENET), which is a not-for-profit service provider that primarily serves universities, research institutes, government websites, and hospitals. Overall, the top 3 DNS in *Africa-hosted* websites critically serve 42% for ZA, 44% for NG, 68% for KE, and 91% for RW. For RW, only a single DNS provider AOS.rw critically serves 87% of the *RW-hosted* websites. In the case of CDN, only 3 CDN providers critically server 56% to 58% of *Africa-hosted* websites. Similarly, only 3 CA providers critically serve 45% for KE, 49% for NG, 75% for RW, and 60% for ZA in the hosted websites. For Rwanda, Digicert alone serves 63% of the *RW-hosted* websites, and for ZA, Let's Encrypt alone serves 42% of the *ZA-hosted* websites.

In addition to this, the providers for *Africa-operated* websites are also highly concentrated. For example, for *NG-operated* websites, Cloudflare serves as a DNS provider for more than half of the websites. We observe similar trends in CDN and CA providers and across countries. The high degree of concentration in the specialized sets also points towards the vulnerability of African users to single points of failure. Moreover, the existence of local providers in the specialized sets while encouraging also raises questions about the resilience of these websites. The high concentration among these local providers makes them single points of failure, where their expertise to defend against attacks and security incidents is not determined as compared to global providers like Amazon.

7 Discussion

In light of our findings, now we present some implications and recommendations for African users, website operators, and service providers.

High Concentration: We find that there is a great degree of concentration in the use of third-party DNS, CDNs, and CAs in the Africa-centric websites. This high concentration creates even more single-points-of-failure which are already prevalent in Africa [45]. Naturally, the third-party dependencies in combination with the problematic intermittent connectivity [15, 45, 42] hinder the growth of the digital economy in Africa, which would require reliable communication among users and businesses. Hence, it is of paramount importance that the websites are redundantly provisioned so that the outage of service providers does not affect the websites and that the website operators are trained to effectively handle outages and recover from failures.

Highly prevalent third-party dependencies: While the concentration of third-party dependency in Africa-centric websites risks their availability, it also creates opportunities. Indeed, third-party providers have certain benefits such as better quality of service, higher capacity, better security expertise, etc. which small websites cannot afford on their own. Hence, using third-party providers is not necessarily bad, but critically depending on it is.

Sparse local providers: We find that on all Africa-centric websites, the number of local providers is very small, except for South Africa. This is problematic in two ways. First, the lack of local providers questions the cyber-autonomy of Africa-centric websites and reduces the diversity of providers available to Africa-centric websites. Indeed, governments could and have tried to rectify that. For example, in Rwanda, with the help of Korea Telecom, the Government of Rwanda created a service provider AOS.rw that serves many Rwanda-centric websites. Even, not-for-profit initiatives like KENET, and South Africa TENET which provide DNS, and web hosting services among others to websites, are often supported by the government. Second, the use of non-local providers in some cases can also increase the cost of Internet access in Africa, if it implies content loading from outside Africa. Africa has one of the highest transit costs [51], hence accessing remote content also makes Internet access expensive for Africans. In our data, we find that most websites are hosted outside of Africa. Therefore, there is an incentive for policymakers to promote local hosting of content so that local providers and infrastructure are promoted.

Higher critical dependency in the specialized sets: In our analysis, we find that the prevalence of third-party critical dependency is higher in the specialized website sets, which are the *hosted*, *dominant*, and *operated* sets, as compared to the *visited* set. This is particularly more evident in the *dominant* and *operated* set for all services and countries. This is not an encouraging trend. This indicates that websites targeting Africans (dominant set) and websites being operated in Africa (operated set) are not paying enough attention to reliability, making them more vulnerable to the side effects of third-party dependencies.

8 Related Work

A huge body of work exists that performs dependency analysis. Some of those analyze dependencies on the country, or/and ISP. For example, Simeonovski *et al.* analyzes dependencies with respect to global scale threats where bad actors can be a country, an autonomous system, or a service provider like an Email server, DNS *etc.* [55]. Similarly, NSDMiner discovers network service dependencies such as ISPs, from passively observed network traffic [43]. Zembruzki *et al.* [62] looks at centralization among hosting providers. Hsiao *et al.* [25] analyzes the cyber-autonomy of government websites of the G7 countries. Dell *et al.* [16] studies third-party DNS dependency using a passive DNS dataset. WebProphet measures the internal backend infrastructure of websites for performance [35]. Similarly, Ikran *et al.* studies dependency chains in third-party web content [26].

Many studies try to understand CDNs and hosting infrastructure [31, 57, 11, 1, 38]. These are complementary to our work. Other work analyzes the critical paths to understand how content affects the page load time (e.g., [59]), or focuses on the privacy implications of the tracking services (e.g., [52, 34, 30]). However, our work is orthogonal as it focuses on the infrastructure services at a higher level than individual websites. Kumar et al. [32] study HTTPS adoption and Podins et al., [47] measure the implementation of Content Security Policy, among third-

party web content. Other efforts (e.g., [44, 41]) analyze third-party web content for attacks. Ager *et al.* identifies and classifies content hosting and delivery infrastructures across the world [1]. Zmap [19] and Censys [18] present tools to scan the Internet at scale to find vulnerabilities like heartbleed. Our focus on web infrastructure is complementary to this work. Other work has analyzed the use of TLS, the certificate ecosystem, and the use of Certificate Revocation in the wild (e.g., [13, 58, 36, 14, 63, 29, 32]). These suggest potential attacks that could be executed via the third party services we analyze here.

There have been many efforts to understand the African Internet Ecosystem. For example, Akanho *et al.* measures the EDNS and TCP compliance in the nameservers for African websites [2]. Chavula *et al.* analyzes the location of cloud hosting providers in Africa for latency [9]. Calandro *et al.* analyzes the hosting of African news websites [7] to determine the fraction of local content. Similalry Brinkman *et al.* [5] discusses the interweaving connection in the Internet due to dependencies and tries to seek what constitutes "African websites", which we provided a definition for in our work. Arouns *et al.* looks at the DNS landscape for African ccTLDs [3]. Our work is complementary to these efforts as we also try to understand the resilience of the Internet in Africa.

9 Conclusion

In this work, we analyze third-party DNS, CDN, and CA dependencies in Africa-centric websites in an effort to bridge the gap between previous works, and offer region-specific actionable insights to African users and operators. Particularly, we study the prevalence of third-party dependencies on *Africa-visited*, *Africa-dominant*, *Africa-operated*, and *Africa-hosted* websites. We find that *Africa-centric* websites are highly vulnerable to the side effects of third-party dependencies. In addition, we find that there is a high degree of concentration in the use of third-party service providers, meaning that a handful of providers serve a large portion of the websites. Our findings have implications for the current usage and recommendations for the future evolution of the Internet in Africa.

10 Availability

Our code is publically available[7]. Our work does not raise any ethical concerns.

11 Acknowledgements

We thank Amreesh Phokeer for their feedback and insights. Furthermore, we greatly appreciate the anonymous reviewers and our shepherd Oliver Gasser for their feedback and comments. We would like to acknowledge the support from the Bill and Melinda Gates Foundation through the Upanzi network. This work was also partially supported by NSF Awards TWC-1564009 and SaTC-1801472 and the Carnegie Mellon CyLab Security and Privacy Institute.

[7] https://github.com/AqsaKashaf/Analyzing-Third-party-Dependencies

References

1. Ager, B., Mühlbauer, W., Smaragdakis, G., Uhlig, S.: Web content cartography. In: Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference. pp. 585–600 (2011)
2. Akanho, Y., Alassane, M., Houngbadji, M., Phokeer, A.: African nameservers revealed: Characterizing dns authoritative nameservers. In: International Conference on e-Infrastructure and e-Services for Developing Countries. pp. 327–344. Springer (2020)
3. Arouna, A., Phokeer, A., Elmokashfi, A.: A first look at the african's cctlds technical environment. In: International Conference on e-Infrastructure and e-Services for Developing Countries. pp. 305–326. Springer (2020)
4. Bock, H.: The problem with ocsp stapling and must staple and why certificate revocation is still broken, 2017. URL https://blog. hboeck. de/archives/886-The-Problem-with-OCSP-Stapling-and-Must-Staple-and-why-Certificate-Revocation-is-still-broken. html (2017)
5. Brinkman, I., Merolla, D.: Space, time, and culture on african/diaspora websites: a tangled web we weave. Journal of African Cultural Studies **32**(1), 1–6 (2020)
6. Butkiewicz, M., Madhyastha, H.V., Sekar, V.: Understanding website complexity: measurements, metrics, and implications. In: Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference. pp. 313–328 (2011)
7. Calandro, E., Chavula, J., Phokeer, A.: Internet development in africa: a content use, hosting and distribution perspective. In: International Conference on e-Infrastructure and e-Services for Developing Countries. pp. 131–141. Springer (2018)
8. Country domains: A comprehensive ccTLD list. https://www.ionos.com/digitalguide/domains/domain-extensions/cctlds-a-list-of-every-country-domain/
9. Chavula, J., Phokeer, A., Calandro, E.: Performance barriers to cloud services in africa's public sector: A latency perspective. In: International Conference on e-Infrastructure and e-Services for Developing Countries. pp. 152–163. Springer (2018)
10. Chege, K.G.: Measuring internet resilience in africa (Nov 2020), https://www.internetsociety.org/blog/2020/11/measuring-internet-resilience-in-africa/
11. Choffnes, D., Wang, J., et al.: Cdns meet cn an empirical study of cdn deployments in china. IEEE Access **5**, 5292–5305 (2017)
12. Chromium, G.: Crlsets, https://www.chromium.org/Home/chromium-security/crlsets/
13. Chung, T., Liu, Y., Choffnes, D., Levin, D., Maggs, B.M., Mislove, A., Wilson, C.: Measuring and applying invalid ssl certificates: The silent majority. In: Proceedings of the 2016 Internet Measurement Conference. pp. 527–541 (2016)
14. Chung, T., Lok, J., Chandrasekaran, B., Choffnes, D., Levin, D., Maggs, B.M., Mislove, A., Rula, J., Sullivan, N., Wilson, C.: Is the web ready for ocsp must-staple? In: Proceedings of the Internet Measurement Conference 2018. pp. 105–118 (2018)
15. Comment, D.S.: Load shedding in south africa causes cooling system failure at mtn data center (Jul 2022), https://www.datacenterdynamics.com/en/news/load-shedding-in-south-africa-causes-cooling-system-failure-at-mtn-data-center/

16. Dell'Amico, M., Bilge, L., Kayyoor, A., Efstathopoulos, P., Vervier, P.A.: Lean on me: Mining internet service dependencies from large-scale dns data. In: Proceedings of the 33rd Annual Computer Security Applications Conference. pp. 449–460 (2017)
17. dig: dns lookup utility. https://linux.die.net/man/1/dig
18. Durumeric, Z., Adrian, D., Mirian, A., Bailey, M., Halderman, J.A.: A search engine backed by internet-wide scanning. In: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. pp. 542–553 (2015)
19. Durumeric, Z., Wustrow, E., Halderman, J.A.: Zmap: Fast internet-wide scanning and its security applications. In: Presented as part of the 22nd USENIX Security Symposium (USENIX Security 13). pp. 605–620 (2013)
20. ExpressVPN: High-speed, secure and anonymous vpn service | expressvpn (2016), https://www.expressvpn.com/
21. Global, S.: Latest research shows DDoS attacks up by 300% in Africa since 2019. https://seacom.com/media-centre/latest-research-shows-ddos-attacks-300-africa-2019/
22. Globalsign certificate revocation issue. https://www.globalsign.com/en/status (October 13, 2016), accessed: May 23, 2020
23. Google: Chrome ux report, https://developer.chrome.com/docs/crux/
24. Hilton, S.: Dyn analysis summary of friday october 21 attack. http://dyn.com/blog/dyn-analysis-summary-of-friday-october-21-attack/ (Oct 26, 2016), accessed: May 23, 2020
25. Hsiao, H.C., Kim, T.H.J., Ku, Y.M., Chang, C.M., Chen, H.F., Chen, Y.J., Wang, C.W., Jeng, W.: An investigation of cyber autonomy on government websites. In: The World Wide Web Conference. pp. 2814–2821 (2019)
26. Ikram, M., Masood, R., Tyson, G., Kaafar, M.A., Loizon, N., Ensafi, R.: The chain of implicit trust: An analysis of the web third-party resources loading. In: The World Wide Web Conference. pp. 2851–2857 (2019)
27. Comprehensive IP address data, IP geolocation API and database - IPinfo.io. https://ipinfo.io/
28. Kashaf, A., Sekar, V., Agarwal, Y.: Analyzing third party service dependencies in modern web services: Have we learned from the mirai-dyn incident? In: Proceedings of the ACM Internet Measurement Conference. pp. 634–647 (2020)
29. Kotzias, P., Razaghpanah, A., Amann, J., Paterson, K.G., Vallina-Rodriguez, N., Caballero, J.: Coming of age: A longitudinal study of tls deployment. In: Proceedings of the Internet Measurement Conference 2018. pp. 415–428 (2018)
30. Krishnamurthy, B., Wills, C.: Privacy diffusion on the web: a longitudinal perspective. In: Proceedings of the 18th international conference on World wide web. pp. 541–550 (2009)
31. Krishnamurthy, B., Wills, C., Zhang, Y.: On the use and performance of content distribution networks. In: Proceedings of the 1st ACM SIGCOMM Workshop on Internet Measurement. pp. 169–182 (2001)
32. Kumar, D., Ma, Z., Durumeric, Z., Mirian, A., Mason, J., Halderman, J.A., Bailey, M.: Security challenges in an increasingly tangled web. In: Proceedings of the 26th International Conference on World Wide Web. pp. 677–684 (2017)
33. Kumar, R., Asif, S., Lee, E., Bustamante, F.E.: Third-party service dependencies and centralization around the world (2021). https://doi.org/10.48550/ARXIV.2111.12253, https://arxiv.org/abs/2111.12253
34. Lerner, A., Simpson, A.K., Kohno, T., Roesner, F.: Internet jones and the raiders of the lost trackers: An archaeological study of web tracking from 1996 to 2016. In: 25th USENIX Security Symposium (USENIX Security 16) (2016)

35. Li, Z., Zhang, M., Zhu, Z., Chen, Y., Greenberg, A.G., Wang, Y.M.: Webprophet: Automating performance prediction for web services. In: NSDI. vol. 10, pp. 143–158 (2010)
36. Liu, Y., Tome, W., Zhang, L., Choffnes, D., Levin, D., Maggs, B., Mislove, A., Schulman, A., Wilson, C.: An end-to-end measurement of certificate revocation in the web's pki. In: Proceedings of the 2015 Internet Measurement Conference. pp. 183–196 (2015)
37. Livadariu, I., Dreibholz, T., Al-Selwi, A.S., Bryhni, H., Lysne, O., Bjørnstad, S., Elmokashfi, A.: On the accuracy of country-level ip geolocation. In: Proceedings of the Applied Networking Research Workshop. pp. 67–73 (2020)
38. Matic, S., Tyson, G., Stringhini, G.: Pythia: a framework for the automated analysis of web hosting environments. In: The World Wide Web Conference. pp. 3072–3078 (2019)
39. Maxmind, L.: Geoip country database
40. Moyo, A.: Africa found wanting on cyber crime preparedness. https://www.itweb.co.za/content/4r1lyMRoaVAqpmda (Dec 2019)
41. Mueller, T., Klotzsche, D., Herrmann, D., Federrath, H.: Dangers and prevalence of unprotected web fonts. In: 2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM). pp. 1–5. IEEE (2019)
42. Mutiso, R., Hill, K.: Why hasn't africa gone digital? Scientific American (Aug 2020), https://www.scientificamerican.com/article/why-hasnt-africa-gone-digital/
43. Natarajan, A., Ning, P., Liu, Y., Jajodia, S., Hutchinson, S.E.: NSDMiner: Automated discovery of network service dependencies. IEEE (2012)
44. Nikiforakis, N., Invernizzi, L., Kapravelos, A., Van Acker, S., Joosen, W., Kruegel, C., Piessens, F., Vigna, G.: You are what you include: large-scale evaluation of remote javascript inclusions. In: Proceedings of the 2012 ACM conference on Computer and communications security. pp. 736–747 (2012)
45. Phokeer, A.: The gambia's internet outage through an internet resilience lens (Jan 2022), https://pulse.internetsociety.org/blog/the-gambias-internet-outage-through-an-internet-resilience-lens
46. Phokeer, A., Chege, K., Chavula, J., Elmokashfi, A., Gueye, A.: Measuring internet resilience in africa (mira). Internet Society (2021)
47. Podins, K., Lavrenovs, A.: Security implications of using third-party resources in the world wide web. In: 2018 IEEE 6th Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE). pp. 1–6. IEEE (2018)
48. PrivateVPN: Privatevpn: The world's most-trusted private vpn provider, https://privatevpn.com/
49. Puppeteer. Puppeteer (May 2022)
50. Rakshit, S.: geograpy3: Extract countries, regions and cities from a url or text (Oct 2022), https://pypi.org/project/geograpy3/
51. Rao, N.: Bandwidth costs around the world (Aug 2016), https://blog.cloudflare.com/bandwidth-costs-around-the-world/
52. Roesner, F., Kohno, T., Wetherall, D.: Detecting and defending against third-party tracking on the web. In: Presented as part of the 9th USENIX Symposium on Networked Systems Design and Implementation (NSDI 12). pp. 155–168 (2012)
53. Ruth, K., Kumar, D., Wang, B., Valenta, L., Durumeric, Z.: Toppling top lists: Evaluating the accuracy of popular website lists. In: Proceedings of the 22nd ACM Internet Measurement Conference. pp. 374–387 (2022)
54. SAN Certificates: Subject Alternative Name – Multi-Domain (SAN). https://www.digicert.com/faq/subject-alternative-name.htm

55. Simeonovski, M., Pellegrino, G., Rossow, C., Backes, M.: Who controls the internet? analyzing global threats using property graph traversals. In: Proceedings of the 26th International Conference on World Wide Web. pp. 647–656 (2017)
56. SimilarWeb: Top browsers market share - most popular browsers in august 2022 | similarweb, https://www.similarweb.com/browsers/
57. Singh, R., Dunna, A., Gill, P.: Characterizing the deployment and performance of multi-cdns. In: Proceedings of the Internet Measurement Conference 2018. pp. 168–174 (2018)
58. VanderSloot, B., Amann, J., Bernhard, M., Durumeric, Z., Bailey, M., Halderman, J.A.: Towards a complete view of the certificate ecosystem. In: Proceedings of the 2016 Internet Measurement Conference. pp. 543–549 (2016)
59. Wang, X.S., Balasubramanian, A., Krishnamurthy, A., Wetherall, D.: Demystifying page load performance with wprof. In: Presented as part of the 10th USENIX Symposium on Networked Systems Design and Implementation (NSDI 13). pp. 473–485 (2013)
60. Weinberg, Z., Cho, S., Christin, N., Sekar, V., Gill, P.: How to catch when proxies lie: Verifying the physical locations of network proxies with active geolocation. In: Proceedings of the Internet Measurement Conference 2018. pp. 203–217 (2018)
61. Young, E.A., Hudson, T.J., Engelschall, R.: Openssl: The open source toolkit for ssl/tls (2011)
62. Zembruzki, L., Sommese, R., Granville, L.Z., Jacobs, A.S., Jonker, M., Moura, G.C.: Hosting industry centralization and consolidation. In: NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium. pp. 1–9. IEEE (2022)
63. Zhu, L., Amann, J., Heidemann, J.: Measuring the latency and pervasiveness of tls certificate revocation. In: International Conference on Passive and Active Network Measurement. pp. 16–29. Springer (2016)