Pushing the Frontier: Exploring the African Web Ecosystem

Rodérick Fanou,[1,2] Gareth Tyson,[3] Pierre Francois[4] and Arjuna Sathiaseelan[5]
[1]IMDEA Networks Institute, [2]Universidad Carlos III de Madrid,
[3]Queen Mary University of London, [4]Cisco Systems, Inc. [5]University of Cambridge
roderick.fanou@imdea.org, gareth.tyson@qmul.ac.uk,
pifranco@cisco.com, arjuna.sathiaseelan@cl.cam.ac.uk

ABSTRACT

It is well known that Africa's mobile and fixed Internet infrastructure is progressing at a rapid pace. A flurry of recent research has quantified this, highlighting the expansion of its underlying connectivity network. However, improving the infrastructure is not useful without appropriately provisioned services to utilise it. This paper measures the availability of *web content infrastructure* in Africa. Whereas others have explored web infrastructure in developed regions, we shed light on practices in developing regions. To achieve this, we apply a comprehensive measurement methodology to collect data from a variety of sources. We focus on a large content delivery network to reveal that Africa's content infrastructure is, indeed, expanding. However, we find much web content is still served from the US and Europe. We discover that many of the problems faced are actually caused by significant inter-AS delays in Africa, which contribute to local ISPs not sharing their cache capacity. We discover that a related problem is the poor DNS configuration used by some ISPs, which confounds the attempts of providers to optimise their delivery. We then explore a number of other websites to show that large web infrastructure deployments are a rarity in Africa and that even regional websites host their services abroad. We conclude by making suggestions for improvements.

Keywords

Content infrastructure, measurements, DNS, web

1. INTRODUCTION

Whilst Africa is surrounded by 12 major long haul submarine cables [37], the in-land optical fibre is still fragmented [38, 25]. Cellular provisions are expanding, but remain insufficient. Central Africa and the Sahel are the primary gaps and, as such, network operators are conscious that these gaps need to be filled. Thus, much of the immediate focus has been on IXP setups through the AXIS project [2], as well as on underlying edge connectivity, with projects such as Liquid Telecom, Project Loon and Google Link deploying both wireless and wireline connectivity [22, 13, 1],

Despite these positive steps, challenges still remain. Most prominently, several studies have revealed content as the dominant component of network traffic [11]. Yet the lack of service infrastructure (*e.g.*, web servers) in Africa means that both mobile and wireline users must often fetch content from the other side of the world [20]. Hence, we argue that researchers and engineers should begin to place more focus on both underlying connectivity and content infrastructure (*e.g.*, web servers, caches) in the region. With this in mind, many large companies have begun to deploy content infrastructure in Africa. Google is perhaps the most famous in this regard [1]. As of yet, however, there is little evidence of how these companies approach the challenge.

Several recent efforts have shed light on web infrastructures worldwide [23, 27, 40, 18, 14, 17, 42, 20]. But, they have not (*i*) focussed on underdeveloped countries/regions; or (*ii*) explored if worldwide results apply to underdeveloped regions. This leaves critical questions unanswered, largely driven by the unusual make-up of African Internet and web infrastructures when compared to more developed regions. First, the Internet in Africa is at a very different stage of its evolution; suboptimal topology and peering configurations can make communications (*e.g.*, protocol behaviour) very different [42]. Second, common practices used for content delivery (*e.g.*, placement of caches at IXPs) are difficult due to the lack of IXPs that fulfill the requirements for content delivery network (CDN) deployment [17, 15]. Third, hosting services are not as ubiquitous in Africa, potentially making the management of web content much more complex [20]. Fourth, due to the lower level of Internet penetration and disposable incomes [19], there are fewer (medium term) business incentives for optimising web delivery. Hence, it is critical that we begin to explore the African web ecosystem, so to better drive the design and deployment of technologies suitable for usage there.

In this paper, we aim to gain a holistic view of Africa's web infrastructure. The outcomes of this has serious ramifications for both mobile and wireline performance. We employ several measurement methodologies for exploring content provider and network operator configurations (§3). We start by exploring a large content provider, Google. We show that Google has made significant deployments in the region (§4). However, unlike their operations in Europe and the United States (US), they have primarily partnered with lo-

ACM 978-1-4503-4143-1/16/04.
http://dx.doi.org/10.1145/2872427.2882997

cal network operators to deploy their cache network. We find 1,094 functional caches in Africa hosted in 59 Autonomous Systems (ASes) and geolocated in 27 countries. Despite this achievement, the bulk of networks still rely on North America for access to Google content. By measuring redirections, we find that local network operators tend not to serve each other. Remarkably, we discover that significant inter-AS delays (caused by poor peering) mean that it is often actually more efficient to contact North America or Europe. This is particularly the case for Central African countries, which contain no Google Caches (GGCs). This creates a perverse incentive for providers to place servers *outside* Africa. We further investigate other reasons for suboptimal performance to find that various ASes have inefficient DNS configurations, using distant public resolvers that introduce significant delays to web fetches because of suboptimal redirects and high resolution delays (§5). Strangely, we notice that these DNS configurations are dictated by the network operators, rather than end users as previously assumed.

We then expand our analysis to cover other web providers. Most are far behind Google in their support for African users (§6). We find that most other popular providers, including regional ones, have a very limited presence in Africa. Even the top local websites host their services outside of the continent. Using network path measurements, we show that these decisions have severe performance implications for all web providers under-study. Finally, we conclude by highlighting key lessons learnt, as well as suggesting recommendations for improving future deployments (§7).

2. RELATED WORK

There has been a recent surge of interest into the deployment of the Internet in Africa. This has partly been led by charities and organisations such as the African Union and the Internet Society [20, 26, 2]. However, beyond this, there has been an expanding push from prominent companies like Google and Facebook who see the huge economic potential of Africa. It is likely that this trend will continue, with further companies striving for dominance in the region. This dynamism is well reflected in the rapidly changing infrastructure witnessed on the continent. Of particular interest has been the use of Internet Exchange Points [17, 15], which are seeing expanding uptake. More general network performance measurements have also explored Internet provisions in Africa. For instance, Chetty *et al.* found that mobile throughput is often greater than fixed-line [10]. Zaki *et al.* [42] focussed on web performance (in Ghana), highlighting that key bottlenecks include slow DNS resolution and a lack of local caching. Remarkably, they found the actual time spent downloading content represents only a small fraction of the end-to-end page download time due to these problems. We take this as clear evidence of the limitations of solely provisioning better connectivity and not considering the upper layers.

There have also been a number of more general studies into content delivery infrastructures. Calder *et al.* [23] performed a detailed study of Google's web serving infrastructure and its expansion over 10 months. They enumerated the IP addresses of Google's infrastructure, finding their geographic locations, studying its growth, and matching users to clusters. Otto *et al.* [27] examined the role of DNS in the redirection process, exploring the potential of the EDNS0 client-subnet extension. Interestingly, by combining distributed DNS queries with EDNS0, we observe potential limitations of this past work. We note similar studies have been expanded to other CDNs such as EdgeCast and CacheFly [39].

Bischof *et al.* [8] explored the performance of end users around the world by analysing data collected over 23 months from end-hosts and residential gateways in 160 countries. They provide insight into the impact of broadband service market characteristics, *e.g.*, connection capacity, pricing, cost of increasing capacity and connection capacity on network usage. Our work is orthogonal to this, focussing on web infrastructure, rather than end user choices. Prominent works have further analysed redirection strategies to understand how CDNs map users to edge caches. For example, Su *et al.* found that Akamai primarily redirects clients based on active network conditions [40]. More recently, [14] performed the first evaluation of the dynamics of the mapping of network prefixes to front-ends from Google. They found high variance across the servers mapped to each location, with nearby clients often being redirected to clusters that are far apart.

Our focus differs from these works in that we target web deployments in Africa, which we discover are quite unusual. We also shed further light on the more general topic by improving existing methodologies through the combination of several measurement approaches. We take a broad perspective, looking at several different websites, content providers, and network operators. The rest of this paper explores this topic to understand the current state of content infrastructure in the region.

3. MEASUREMENTS METHODOLOGY

We begin by presenting our methodology used to analyse the nature and availability of content. It involves three key steps: (*i*) Collecting all IP prefixes for networks in Africa; (*ii*) Discovering all the content servers/caches that serve these African networks; (*iii*) Mapping the locations and underlying path characteristics between users and the content infrastructure. All our measurement data is public and available at [16].

3.1 AFRINIC Prefixes

To map content delivery infrastructure in Africa, it is clearly necessary to compile a comprehensive list of the IP addresses and networks within the continent. To achieve this, we parse the AFRINIC IPv4 assignment and allocation files from 2005 to 2015 [3]. These list the IP prefixes allocated by this Regional Internet Registry (RIR), as well as which countries they have been allocated to. Among 3,488 available IPv4 prefixes, 3,082 of diverse lengths are assigned or allocated as of April 30, 2015. These are the prefixes we consider in this study; we term them *AFRINIC prefixes.*

3.2 EDNS0 Client-Subnet Probes

Next, we collect a full list of all content caches that serve these *AFRINIC prefixes*. It would clearly be impossible to discover *every* cache, and therefore we focus on Google Caches (GGCs). Note that www.google.com is the top ranked website across the world and most African countries [4]. GGCs operate in a traditional CDN fashion: Whenever a user fetches a Google webpage he is simply redirected, via DNS, to a nearby GGC. As of yet, it is not obvious to what extent these have been introduced to Africa.

To measure this, we use the EDNS0 (Extension mechanisms for DNS version 0) client-subnet extension [27]. It has been developed to improve the accuracy of DNS-based redirections when a client is using a remote public resolver (*e.g.*, Open DNS). The extension allows clients to include their network prefixes in DNS queries (the prefix length is determined by the recursive resolver). By including the client's IP subnet, CDNs can redirect users to the correct server (rather than a location nearby to the public resolver).

We exploit this feature to launch EDNS0 queries with the client-subnet set to each of the *AFRINIC prefixes* (following a similar methodology to [23]). Through this, we can collect information on which GGCs users from across Africa are redirected to. We performed three EDNS0 crawls for www.google.com, using a variety of resolvers. First, we sent, every hour on March 06, 2015, EDNS0 queries through Google public DNS (8.8.8.8). Second, we directed our queries through their name servers ns1.google.com, ns2.google.com, and ns3.google.com (all support EDNS0) every hour on April 12, 2015. Third, we sent again EDNS0 queries through ns1.google.com from April 23, 2015 to May 09, 2015 every hour. This revealed 3,011 unique GGC IP addresses, which we term the *EDNS0 Probes* dataset.

3.3 RIPE Atlas DNS Probes

A limitation of the above methodology is that we cannot be sure that the results returned via EDNS0 are equivalent to those that would have been returned to an actual client. To verify this, we augment our dataset with a second set of DNS measurements. Instead of adopting PlanetLab [31], which contains only 8 nodes in Africa, we use the RIPE Atlas infrastructure as it is currently the largest measurement infrastructure in the region. As of May 25, 2015, it has 379 vantage points deployed in 170 ASes across 45 African countries (out of 58) [33, 35].

We repeatedly launched, in parallel, 6 DNS requests of type A from all the available IPv4 RIPE Atlas probes in Africa to www.google.com. This was kept running for 7 days (from March 24, 2015 to March 30, 2015). The active probes performed the query 3 times each, roughly every 60s. We obtained 28,387,226 DNS queries results.

Since not all the probes were online during the whole measurement campaign, our DNS lookups involve a total of 225 probes hosted in 38 African countries. 988 ASes have been allocated by AFRINIC as of May 07, 2015. After removing all the requests that have been performed by probes in Africa hosted in non *AFRINIC prefixes*, our DNS probes cover 111 AFRINIC ASes (11.23%), and 146 *AFRINIC prefixes* (4.73%). This constitutes the widest vantage on Google's infrastructure in Africa available yet. From this campaign we obtained 1,917 GGCs IPs, which we term the *RIPE Atlas DNS* dataset.

3.4 Filtering Inactive Caches

In total, we discovered 3,428 GGC IPs via our RIPE Atlas DNS and EDNS0 campaigns (some IPs were in the outputs of both methods). Following the above, we performed 10 ICMP pings to each discovered cache to verify it was active. We also issued HTTP requests towards all GGCs to check which ones were alive. These tests have been performed from both Spain (ES) and the Great Britain (GB) over multiple runs to ensure correctness (on March 09, April 09, and 13, May 18, 2015). We discard IPs which did not respond to either pings or HTTP requests. 3,120 IPs remained. We call this set of IPs the *functional GGCs*.

3.5 Geolocating Caches

We next geolocate all GGCs. As discussed by Poese *et al.* in [32], geolocation databases are far from being as reliable as they claim. To avoid problems found with individual geolocation databases (DBs), we use a method similar to our prior work [15]. The main difference is the number of considered databases. We take 10 different geo DBs from different sources to find the location associated with each IP. These are: OpenIPMap (*OIM*) [34] whose entries are based on croudsourcing geolocation data from up to 25 operators; Reverse DNS lookups (*RDNS*) from which we infer the geolocation of an IP based on city, country codes (CCs) or airport names embedded in the reverse names; Team Cymru (*TC*) [41]; Whois (*WHOIS*); MaxMind GeoIP2City (*MM*) [24]; and the RIRs' assignment and allocation files for AFRINIC DB (*AF*) [3], RIPE NCC DB (*RP*) [36], APNIC DB (*AP*) [5], ARIN DB (*AR*) [6], and LACNIC DB (*LAC*) [21]. When all the DBs with an available entry for an IP give the same result, we use that country code. But when this is not the case, we choose 5 random RIPE Atlas probes in each of the possible countries and perform three user-defined ping measurements towards the considered IP. We assume that the IP is located in the country from which we record the lowest round trip time (RTT).

For 42% of GGC IPs, all the available DBs return the same country code. As for the remaining (1,812) IPs, only 1.1% show an inconsistency of 3 countries, whilst the rest have an inconsistency of 2 countries. The delay tie-breaking approach allows us to geolocate a further 57.6% of the GGCs. At the end of both steps, 99.5% of functional discovered GGCs are geolocated.

	3,105 GGCs IPs		**144 DNS resolvers**	
DB	**Coverage**	**Trust**	**Coverage**	**Trust**
OIM	0.45%	100%	0%	N/A
RDNS	8.27%	93.77%	0%	N/A
MM	98.29%	89.54%	100%	98.61%
RP	10.04%	75.32%	12.5%	88.89%
AF	35.81%	93.07%	81.25%	94.02%
AP	2.58%	100%	0.69%	100%
AR	10.66%	98.49%	22.91%	87.88%
LAC	0%	N/A	0%	N/A
TC	98.97%	90.34%	100%	95.13%
WHOIS	97.93%	47.41%	94.44%	8.82%

Table 1: *Comparison of Geolocation DBs for both GGCs' and DNS resolvers' IPs. N/A stands for Not Applicable.*

We summarise the results in Table 1. The coverage column shows the percentage of IPs for which a DB has given an answer (*i.e.* a valid CC). The Trust column shows the percentage of IPs for which the DB entry is equal to the country that we finally selected for that IP. Overall, the DBs are surprisingly accurate with many attaining a Trust above 0.9. That said, there are some significant outliers. *LAC* has no coverage, whilst some DBs such as *OIM*, *AP*, *RDNS*, *RP* and *AR* have a very low coverage (*e.g.*, 10% and below). *RP* and *WHOIS* are particularly poor. We notice, for instance, that 16.8% of the answers from *RP* are "EU", whilst the final location is either in Ghana (GH), Tunisia (TN), or the Netherlands (NL). Similarly, although it has a high coverage (97.93%), over half of the geolocations provided by *WHOIS* are inaccurate. These results highlight two key points. The first is that using these DBs in isolation would be very unwise in Africa. The second is that content providers using geolocation (for content adaptation or redirection) should use these DBs sorted in decreasing order of the average Trust: *OIM*, *AP*, *MM*, *RDNS*, *AF*, *AR*, *TC*, *RP*, and *WHOIS*.

3.6 Measuring path characteristics

The above provides a comprehensive set of geolocated GGCs in Africa. Alone, this does not provide insight into the path cost for users though. We therefore launched from February 18 to May 22, 2015 a paris-traceroute campaign from all the RIPE Atlas probes in Africa to each of the GGCs IPs. A traceroute between each probe and each GGC IP is issued at 5 randomly defined timestamps during the said period. We use UDP [12]. The measurement campaign resulted in total into 1,309,151 paris-traceroutes. For the purposes of this paper, we primarily focus on end-to-end delay measurements for these paths.

4. EXPLORING GOOGLE IN AFRICA

Due to its prominence, we begin by mapping out the Google infrastructure used by African networks. The statistics presented in this section are computed based on the redirection of *AFRINIC prefixes* to any functional GGC from both the EDNS0 and DNS campaigns.

4.1 Mapping Google

We discover in total 3,120 functional GGCs serving Africa. We first investigate which countries these GGCs are located in, shown in Figure 1. We colour code the locations: yellow markers represent GGCs hosted in RIPE NCC ASes, red ones are in ARIN, blue markers are in APNIC, and green ones are in AFRINIC ASes. The marker size is proportional to the number of IPs geolocated at that longitude and latitude. Table 2 also lists the top 10 ASes and countries in terms of cache numbers. The percentage between parentheses indicates the fraction of GGCs located in either the corresponding AS or country.

A diversity of ASes can be seen to host GGCs. We discover 80 ASes in total, most of which are not operated by Google: 70.2% of those ASes are allocated by AFRINIC, 22.6% by RIPE NCC, 5.9% by ARIN, and 1.1% are APNIC ASes. Clearly, African deployments have deviated heavily from Google's prior setup in developed regions, which has seen Google hosting most content within its own networks [23]. From our traces, only 41.9% of GGCs are hosted in Google ASes: 37.2% in AS15169 (Google) and 4.7% in AS36040 (YouTube Google). All other caches are spread thinly across third party networks. Africa is therefore quite different to the global profile of Google, where 90% of prefixes are mapped to caches in Google's own ASes [23].

We also find that many of the above ASes (about 30%) are based outside of Africa. Despite the conscious push of Google, a large number of foreign caches are still relied upon. 39.7% of all caches in our dataset are in the United States (US). As shown in Table 2, other popular countries include the Netherlands (NL), Malaysia (MY), and Germany (DE). Overall, 44 countries host a GGC: 27 in Africa, 12 in Europe, 2 in North America (US and Canada, CA), 2 in Asia (MY and Bahrain, BH) and 1 in Oceania (Australia, AU). Africa contains only 35% of all caches used by its residents. Most of them are located in South Africa (ZA), Egypt (EG), Mauritius (MU), Kenya (KE) and Nigeria (NG). The remainder are thinly spread across other African countries. A key reason for this is the lack of presence by Google's own network in Africa. Linking to the previous paragraph, we note that 89.9% of the caches within Google's ASes are actually based in the US (explaining the North American dominance). In other words, Google is yet to deploy its own wide area network reach into Africa, instead relying on third party ASes to host their content infrastructure.

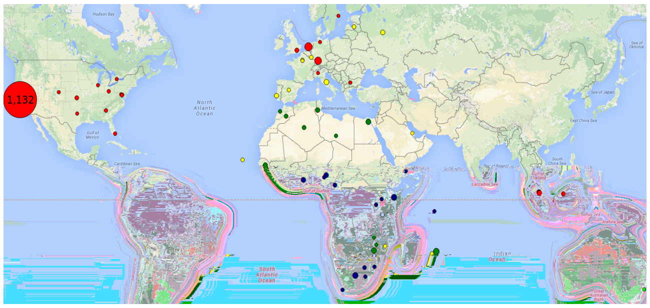


Figure 1: *Geolocation of GGCs serving AFRINIC prefixes according to our refined geolocation methodology. The marker size is proportional to the number of IPs geolocated at that longitude and latitude.*

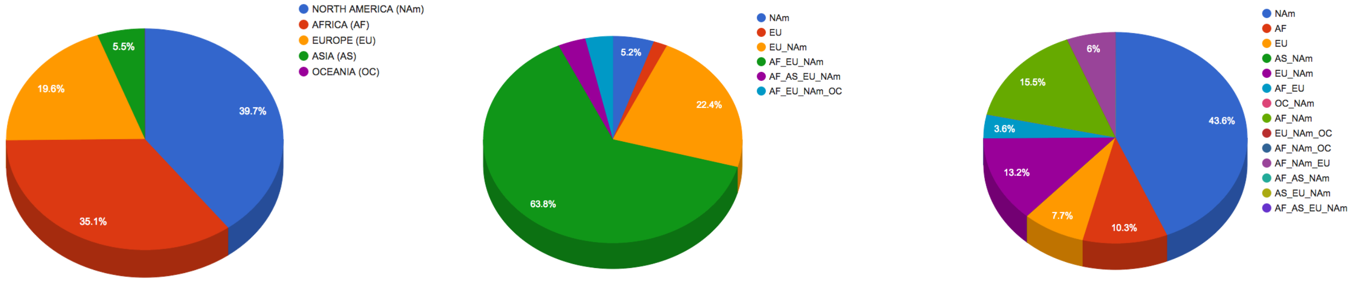
	AS		Country
1	GOOGLE, US (37.21%)	*1*	US (39.68%)
2	TMNET-AS-AP, MY (5.13%)	*2*	NL (5.26%)
3	YOUTUBE GOOGLE, US (4.74%)	*3*	MY (5.16%)
4	LEVEL3, US (2.56%)	*4*	DE (4.87%)
5	MEO-INTERNACIONAL, PT (2.05%)	*5*	ZA (4.45%)
6	RETN-AS, UA (1.98%)	*6*	EG (3.81%)
7	ROSTELECOM-AS, RU (1.53%)	*7*	MU (3.33%)
8	ETISALAT-MISR, EG (1.51%)	*8*	IT (2.24%)
9	TELECOM ITALIA, IT (1.5%)	*9*	KE (1.99%)
10	MTNNS-AS, ZA (1.47%)	*10*	NG (1.95%)

Table 2: *Top 10 ASes and countries hosting GGC IPs serving AFRINIC prefixes extracted from both DNS and EDNS0 methods. Parentheses contain the percentage of GGCs hosted.*

It is also worth noting that, despite recent findings regarding the routing of traffic via Europe [17, 15], Africa is not heavy reliant on Europe for Google content. This is despite the geographic closeness of Europe to Africa (compared to the US). Only 19.6% of caches are based in Europe. Worryingly, we also note that there are *no* caches in most central African countries, *e.g.*, Democratic Republic of Congo (CD), Congo (CG), Gabon (GA), Central African Republic (CF). Instead, caches are mostly pushed further to the edges of the continent (as shown in Figure 1). This is likely driven by the expanding Internet infrastructure in coastal areas that are within easy reach of submarine cables. In contrast, inner African areas still struggle to gain fibre connectivity, making cache deployment far more costly. That said, we find that even well meshed countries such as Angola (AO) and Namibia (NA) [38] have no GGCs. In other words, regulatory, governmental, and policy challenges still likely play a significant role.

4.2 Keeping Content in the Continent?

We next explore how self sufficient African content delivery is by inspecting how often requests are kept within the continent. For example, the presence of caches in the US is not important if they are only used occasionally. We



(a) *Locations of GGCs found in the dataset*

(b) *Percentage of African countries served by each continent*

(c) *Percentage of AFRINIC IPs jointly served by each continent*

Figure 2: *Statistics on Google redirections from AFRINIC prefixes extracted from EDNS0 and DNS query data.*

emphasise that we are *not* reverse engineering Google's redirection algorithm. Figure 2 presents (*i*) the proportion of caches found in each continent; (*ii*) the percentage of countries that are served by various combinations of continents; and (*iii*) the percentage of *AFRINIC prefixes* served by various combinations of continents.

Figure 2(a) reveals, as stated previously, that a significant number of GGCs are deployed in Africa. Nevertheless, 94.8% of African countries are served by the US at least once in our dataset. In fact, Figure 2(b) shows that 63.8% of countries spread their requests amongst Africa, Europe, and North America. This could be for many reasons, *e.g.*, using external caches to support "overflow", where demand exceeds capacity. That said, we observe that 22.4% of countries are exclusively served by North America and Europe. In fact, some countries namely, Togo (TG), Sao Tome (ST), and Mayotte (YT), are solely served by North America, indicating that this is not caused by the need for an "overflow". In these cases, we find none of the countries host their own GGC, forcing them into using external caches. Ideally, such users would be redirected to other nearby African countries but, clearly, certain reasons (later explored) prevent this. Thus, sadly, these countries suffer from both poor underlying connectivity, alongside distant web service hosting.

Comparing Figures 2(b) and 2(c) also highlights some interesting properties. Whereas the bulk of requests on a per country basis are redirected to North America, Europe, and Africa, this is not the case on a per network basis. Only 5.2% of *countries* solely use North American caches. In contrast, 43.6% of *networks* solely rely on North America. Further, whilst *no* country is exclusively served by African caches, we find that 10.3% of networks are. In other words, redirection is primarily based on specific networks rather than geographic regions (*i.e.* countries). This means that many networks fail to gain access to African caches, even though others in their country can do so. Choosing the "right" ISP therefore seems particularly important in this region.

4.3 Cache Sharing

We next inspect in what circumstances countries and networks share their caches with others. It is particularly pertinent in Africa, as recent work has highlighted that network operators often are resistant to cooperate [17]. Note that sharing is a product of both individual network policy and redirection strategies employed by Google. Figure 3 compares the number of caches within each country against the number of African countries that use those caches. Theoretically, if cache deployment were ubiquitous, each country should only need to serve requests from its own residents. In such a case, the number of countries mapped to a GGC should always be 1 (*i.e.* the blue line). Figure 3 shows, however, that this is not the case. In total, 60.6% of countries share their caches with at least one other country. Indeed, 57.9% of African countries (hosting GGCs) share their caches with other countries, whilst this percentage is 81.8% for those outside Africa.

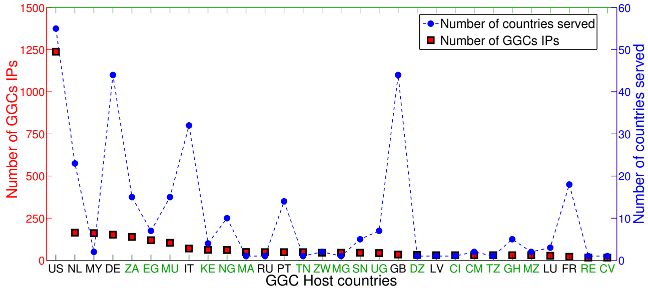


Figure 3: *Distribution of Google caches (GGCs) serving AFRINIC prefixes across countries. Includes percentage of other countries that the GGCs are shared with. African GGCs host countries are in green, whilst GGCs host countries on other continents are in black.*

Unsurprisingly, the most extreme is the US (1,238 caches), which serves almost all African countries (55). Similarly, in Europe, the number of African countries served by DE (152 caches) and GB (34 caches) is 44 each. Most importantly, countries outside Africa share their caches, on average, with three times as many other countries as African ones (25 vs. 7). In Africa, sharing is largely performed by more developed states, *e.g.*, ZA (serves 15 countries with 139 caches) and KE (serves 4 countries with 62 caches). In contrast, many less developed countries have very different trends. There are countries that host a large number of caches, yet only serve one other country, *e.g.*, Zimbabwe (ZW), which contains 45 caches, Mozambique (MZ) 30, and Cameroon (CM) 30. Meanwhile countries such as MA, TN, MG, Algeria (DZ), Tanzania (TZ) and the Ivory Coast (CI) never serve a user in another country.

Of course, this lack of sharing is actually driven by individual networks, rather than entire countries. 15.1% of the networks containing our RIPE Atlas probes host a cache (and hence they can access a cache in their own networks). Only 63.1% ever share their caches with others. For instance, AS37183 Utande Internet Services (ZW), AS36914 Ubuntunet (TZ), AS21042 Gulfsat AS (MG), and AS24835

Rank	CC	% Caches hosted	EDNS0 queries	DNS queries	Total queries
1	US	39.68%	45.85%	49.35%	49.29%
2	NL	5.26%	2.09%	1,02%	1.04%
3	MY	5.16%	0.05%	0.02%	0.02%
4	DE	4.87%	2.86%	12.89%	12.72%
5	ZA	9.41%	10.29%	5.55%	5.62%
6	EG	3.81%	3.03%	0.002%	0.06%
7	MU	3.33%	5.07%	1.07%	1.14%
8	IT	2.24%	1.34%	2.35%	2.33%
9	KE	1.99%	2.68%	0.13%	0.17%
10	NG	1.95%	6.19%	0.004%	0.11%

Table 3: *Percentage of total redirections towards GGCs in top 10 countries hosting caches. Computed based on outputs from EDNS0 probes from all AFRINIC prefixes; DNS queries from RIPE Atlas probes; and union of both methods (labelled Total queries).*

RAYA Telecom (EG) never serve other networks. To study the extremity of this, Table 3 compares the percentage of GGCs in a country against the percentage of requests redirected to that country (last column). A proportional and cooperative redirection strategy would result in these two percentages being identical. Clear trends can be seen, with 39% of caches in the US receiving 49% of our requests from Africa. As for caches in Europe, we notice that those in DE (4.8%), receive 12.7% of requests. Caches in these countries therefore serve a disproportionately large number of requests. In contrast, African countries host 35% of caches, yet only serve 16% of requests. Hence, despite wide deployment, African caches do not receive a fair proportion of requests.

4.4 Understanding Disincentives for Sharing

The above raises many questions about *why* most European and North American networks share their caches, whilst a considerable amount of African ones do not. Our analysis suggests that a key reason for not sharing caches is that many networks in Africa still remain disconnected from regional Internet eXchange Points (IXPs) [15]. Sharing cache capacity would therefore generate transit costs, suffer from high inter-AS delay and, consequently, reduce the probability of a CDN redirection algorithm selecting a non-peered neighbour. In order to explore this, we collect information on IXP peering from IXP websites, PeeringDB and Packet Clearing House (PCH) [30, 29].

We first note that *no* GGC IPs belong to an IXP prefix. Instead, often networks hosting caches are *peered* at IXPs. For example, 99.9% of the requests served by DE caches are actually redirected to networks peering at DE-CIX in Hamburg; all redirects to GB go to Google's own AS peered at the LONAP IXP; and 99.7% of redirects to NL go to third party networks peering at AMSIX. Similarly, 99.9% of redirects to the US go to peers of one of 33 US IXPs.

In these cases, sharing cache capacity is straightforward, as IXP membership allows low-delay, low-cost interactions between networks. Similar principles should apply in Africa. Thus, we use our paris-traceroute dataset to check if the African networks sharing their caches are those peered at IXPs. We find that all African ASes connected to an IXP share their caches. The top 2 networks for sharing are in ZA (MWEB and Internet Solutions). Unfortunately, however, we discover only 18.6% of African ASes hosting a GGC are peered at an IXP. This means that for approximately 80% of ASes, sharing their caches would generate transit costs and likely produce high inter-AS delays, therefore creating strong disincentives (both for network operator policy and Google redirection algorithms). Nearly all redirects that stay within Africa are between network peered together at an IXP. This strong correlation suggests that the main barrier to unlocking significant web performance improvements in Africa is actually to enable cache sharing via peering. This would dramatically increase the positive impact that a new cache could have on a region.

4.5 GGC Performance

The above has shown that sharing caches in Africa is difficult because of a lack of peering between networks. On the one hand, this introduces transit costs for networks sharing caches (likely resulting in network policies that prohibit it). Whilst, on the other hand, poor peering produces high inter-AS delays, driving CDNs away from such redirects. Hence, we briefly wish to quantify this using our path probes (§3.6), which measure the delay between all RIPE Atlas probes and all potential GGCs. To measure the best case scenario, we extract the minimum RTT for each probe.

Figure 4(a) shows a CDF of the minimum RTTs to the GGCs measured over each probe in our dataset. Remarkably, the web requests to North American caches actually attain the lowest delay, with an average RTT of just 35.7 ms (median of 13.7 ms) compared to 104.4 ms for African caches (median of 60.5 ms) (see Figure 4(b)). This confirms our remarks earlier that a key element of our observations is the presence of high delay inter-African AS communications. CDN redirection therefore naturally takes users away from such links, regardless of their geographical closeness.

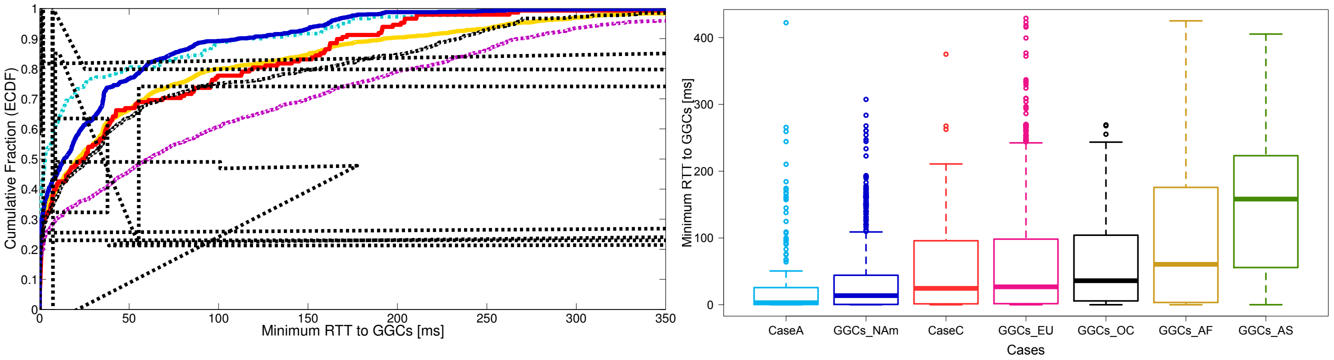
We similarly witness low delay to Europe (with an average of 71.7 ms and a median of 26.83 ms), with only Asia (153 ms average and 158.2 ms median) performing worse than Africa, as presented in Figure 4. It is also important to note that for African networks that have their own cache, this is not the case. Networks that host a GGC unsurprisingly benefit from lower delay redirects, with an average minimum RTT of 32.9 ms (the best median 3.05 ms) compared to 57 ms for those without (median of 24.5 ms). We later expand upon this to inspect HTTP fetches from local caches. This confirms that the suboptimality found in African topologies [23] impacts the ability for caches to be locally used/shared within a reasonable delay bound.

5. DNS IN AFRICA

A critical component in the above exploration is DNS, which drives the redirection process. Now, we examine DNS configuration in Africa in relation to the above findings.

5.1 Mapping DNS Resolvers

We start by mapping out the locations of DNS resolvers used by ISPs in Africa. We collect the IP addresses of all (239) default resolvers used by our RIPE Atlas probes. 70 are private addresses (*e.g.*, 10.0.0.1); we discard these for the rest of this section, leaving the remainder that have public IP addresses. We then geolocate the resolvers using our methodology presented in §3.5. Figure 5 presents the results; the marker size is proportional to the number of ASes using the corresponding resolver in that location. It can be seen that the bulk of the retained resolvers are located within



(a) *CDF of the minimum RTT distribution*

(b) *Boxplot of the minimum RTT distribution*

Figure 4: *Delay distribution from different sets of RIPE Atlas probes to serving GGCs. Minimum RTTs per probe are considered. The cases listed in 4(b) correspond to those in the legend of figure 4(a) and the colors are the same.*

Africa, as one would expect. However, surprisingly, we also find resolvers operating in both the US and Europe.

It is well known that non-local resolvers can have an impact on CDN performance [27]. Hence, we compute the number of DNS resolvers that exist on the same AS as their RIPE Atlas probe. 83.8% of resolvers are hosted within the same network as the probe. From the perspective of CDN redirection, these 83.8% of networks are ideal, as the CDN would be able to effectively locate the client (using the DNS resolver's IP address). However, 16.2% of unique resolvers are hosted within different networks. Further, 34.6% of all the probes share these resolvers located in different networks, showing that many ISPs utilise third party resolvers by default. Note that this is different to the typical assumption that the use of public resolvers is dictated by end users. Instead, we find that many ISPs use DHCP to automatically configure clients to use third party resolvers. The reason for ISPs doing this is generally easier management. For emerging network providers in Africa, this is clearly attractive.

However, this comes at the cost of performance degradation for CDNs [9], since these clients would appear as if they were in a different network (where the resolver is). In 32.5% of cases, the third party DNS resolver used by the probe is not even in the same country. 13.6% are located outside of Africa, with 4.7% in the US. This is reflected in the geographic distances observed between our probes and the resolvers. On average, the third party resolvers are 13,690 km away from the probes they serve (maximum of 18,116 km and a minimum of 996 km). In contrast, ISPs using local resolvers have distances ranging from just 0.07 km to a maximum of 3,554 km (average of 325 km). Interestingly, however, in Africa (unlike elsewhere [27]) this has limited impact on the CDN redirections due to the existing propensity to select caches in Europe and the US anyway. Instead, the greater concern is the extra delay introduced by the high resolution round trip times, as discussed next.

5.2 DNS Resolver Performance

The above has highlighted that many African ASes are using resolvers in different networks, and even different countries. We find significant resolution delays introduced by these third party resolvers hosted in other countries, with an average delay of 129 ms compared to just 25 ms for resolvers hosted by the ISP. We therefore next split the DNS queries into two categories: Those sent to resolvers in the same country (67.5%) and those sent to resolvers in different countries (32.5%).

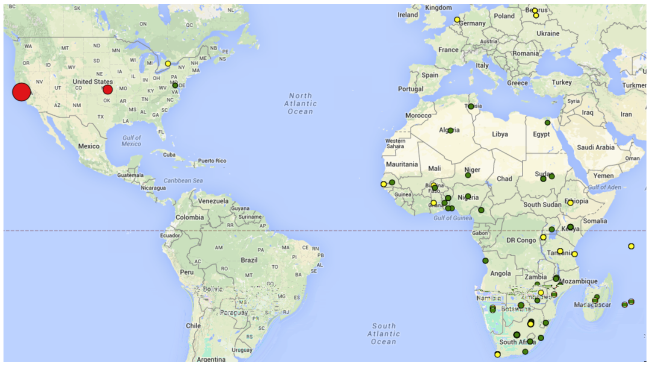


Figure 5: *Locations of DNS resolvers used by Atlas probes. The marker size is proportional to the number of ASes that use them. They are coloured given their types: Google DNS in red, Open DNS in blue, Open resolvers in yellow and ISPs' resolvers in green.*

The first category is composed of (*i*) ISP resolvers located in the same country (86.1%); and (*ii*) Open resolvers in the same country (13.9%). The second category is composed of DNS queries sent to (*i*) Open DNS resolvers (0.8%); (*ii*) open resolvers in different countries (4.1%); (*iii*) ISP resolvers located in different countries (15.1%); and (*iv*) Google DNS (80%). Figure 6 presents the resolution delay distributions.

The average response time (including RTT and DNS query treatment) of third party resolvers in different countries is 132 ms (with a maximum of 11,977 ms and a minimum of 2 ms). Meanwhile, the average response time of local resolvers in the same country is 25 ms (with a minimum of 0.7 ms and a maximum of 18,440 ms). Thus, unsurprisingly, we find that the lowest delay is provisioned by DNS resolvers hosted by the local ISP in the same country (some ISPs have a presence in multiple countries that share resolvers). Marginally worse performance is provided by third party resolvers that are located in the same country. The most significant drop in performance is introduced by public resolvers such as Google DNS. Although they are presented as methods to improve performance, this does not work in Africa. This is primarily caused by the lack of public resolver infrastructure on the continent; for instance, Google

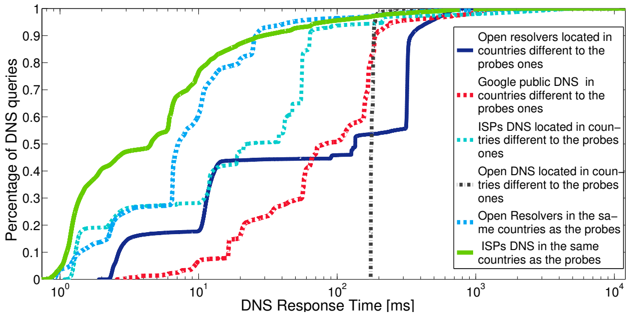


Figure 6: *Cumulative distribution of DNS resolution delays.*

DNS anycast routes all our African queries to US resolvers. Use of such distant resolvers adds in excess of 100 ms delay for approximately 50% of Atlas probes. Although only 16.2% of resolvers are third party, this is a critical issue for their customers. African operators are therefore outsourcing not only the hosting of web content but also the operation of key infrastructure such as DNS, with damaging consequences.

5.3 EDNS0 vs. DNS

During our measurements, we gained a unique insight into Google's redirection strategy using our two techniques: (*i*) DNS queries from RIPE Atlas probes, and (*ii*) EDNS0 queries. We therefore briefly explore the differences between these two mechanisms. To do this, we subset our data to leave only prefixes for which we have data from both RIPE Atlas probes and EDNS0. We also remove all RIPE Atlas probes that use a public resolver. This is because these would naturally have different results between DNS and EDNS0, since the measurement devices would appear to Google as if they were hosted in the country of the public resolver. Using the results collected for the rest (106 networks), we compare the intersection of the DNS and EDNS0 datasets. Note that we checked the impact that the number of probes in a network has on the number of DNS results returned to find no correlation.

Figure 7 shows the number of unique GGCs returned per *AFRINIC prefixes*. We classify these IPs into 3 categories: (*i*) IPs returned by both EDNS0 and RIPE Atlas DNS queries; (*ii*) IPs returned by just EDNS0; and (*iii*) IPs returned by just DNS. Overall, a large range of GGCs are returned to the prefixes, with 8.5% exceeding 100. There is a significant divergence between the GGCs returned via EDNS0 and DNS though. Only 29.4% are found in both datasets. 56.9% IPs were only found by RIPE Atlas DNS queries and not by EDNS0 ones, whereas 21.3% of IPs are only found by EDNS0 queries and not by DNS. It is therefore clear that DNS redirects are not equal to EDNS0 redirects; these schemes operate on either different data or a different algorithm. For 70.7% of the subnets, RIPE Atlas DNS queries gain vantage on at least twice the number of GGC IPs when compared to the EDNS0 methodology alone.

Amongst other things, the above results show that past measurements using EDNS0 [14, 27, 23] may have missed a significant amount of Google's infrastructure, at least in Africa. In total, the above analysis discovered 47 ASes hosting GGCs; 25 ASes (53.1%) could be discovered by either EDNS0 and DNS requests. However, 19 other ASes (40.4%) could only be found using RIPE Atlas DNS queries, whereas this is just 12 (25.5%) for EDNS0 queries. It is clear that the mechanisms are dealing with very different server sets in the region. This raises questions regarding whether or not simply using the EDNS0 client-subnet extension in Africa would address the known problems with remote resolvers.

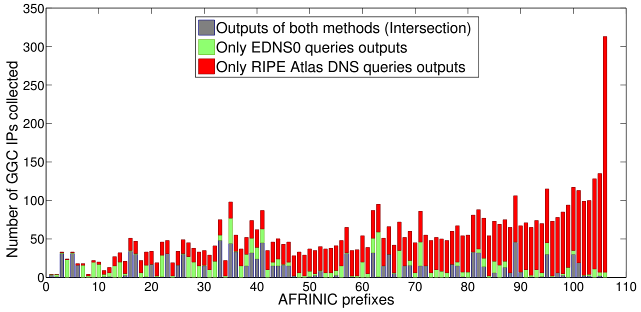


Figure 7: *Comparison of the number of unique GGCs collected via EDNS0 with those collected by RIPE Atlas DNS queries.*

6. EXPANDING TO OTHER PROVIDERS

We have so far explored Google's infrastructure. Google, however, provides only a small sample of the web operations in Africa. We therefore next examine a variety of other popular websites in the world and on the African continent.

6.1 Measuring Top Websites

To measure how other popular websites are provisioned in Africa, we inspect: (*i*) the global top 10 Alexa websites; (*ii*) the top 15 Alexa websites in Africa; (*iii*) the top 15 most popular websites in Africa listed by Afrodigit.com; and (*iv*) www.iroking.com, a well-known Nigerian video content provider on the African continent. We include websites from Afrodigit because we noted that the top Alexa websites were biased towards websites in certain countries (*e.g.*, South Africa, Nigeria, Egypt). We also added www.iroking.com to gain an understanding of video websites in Africa (because there are no local videos content websites in either the top Alexa or Afrodigit sites). Again, we utilise DNS to discover their infrastructures. We concurrently issued DNS queries from RIPE Atlas probes to each of the domains over a 4 day period on a per hour frequency (May 23–26, 2015). This allowed us to observe the location of front-end servers hosting the websites using our method from §3.5. In total, we performed 566,994 DNS queries.

Table 4 summarises the results, comparing the sizes, the server geolocation and the networks hosting the websites. It can be seen that only 5 websites from the 18 regional websites in Africa actually have servers available in Africa. This is likely largely driven by the cheaper cost of hosting abroad [20]. It can also be explained by the significant inter-AS delays, due to which it is often actually more efficient (in terms of delay/QoS but not in terms of cost) to contact North America or Europe. The 5 sites hosted in Africa are in ZA, within 4 ASes. In-line with our observations about Google, the remainder are hosted in either the US or Europe. This is dominated by prominent networks, including CloudFlare and Amazon.

We next inspect the types of websites that are hosted within Africa. We discover three large commercial banks offering a full range of services, hosting their content in ZA.

Top 15 sites in Africa (by Alexa & Afrodigit)	**Type**	**#IPs caches**	**CCs host caches**	**ASes**	**Top 10 global websites** (by Alexa)	**Type**	**#IPs caches**	**CCs host caches**	**#ASes**
jumia.com.ng	E-commerce	1	DE	20546	**amazon.com**	E-commerce	4	US	2
konga.com	E-commerce	1	US	15169	**tabao.com**	E-commerce			
bidorbuy.co.za	E-commerce	1	ZA	3741					
fnb.co.za	Financial services	1	ZA	17148	**Qq.com**	Internet services	2	CN	2
gtbank.com	Financial services	1	US	26496					
absa.co.za	Financial services	1	ZA	3741					
standardbank.co.za	Financial services	1	ZA	10798					
almasryalyoum.com	News/media	5	NL, CR	13335	**google.com**	Search engine	924	18 (§ 4.1)	26
elkhabar.com	News/media	2	US	13335	**yahoo.com**	Search engine	4	US, UK	2
vanguardngr.com	News/media	1	US	14618	**baidu.com**	Search engine	1	HK	1
news24.com	News/media	1	ZA	10474					
punchng.com	News/media	1	IE	16509	**wikipedia.com**	encyclopedia	2	NL, US	2
iol.co.za	News/media	2	IE	16509					
ghanaweb.com	News/media	1	US	7859					
nairaland.com	Online community	5	US	13335	**facebook.com**	Social network	5	US, DE, NL	1
supersport.com	Sports	1	ZA	10474	**twitter.com**	Social network	7	US	2
alwafd.org	Politics	2	NL	13335					
iroking.com	Videos	2	IE	16509	**youtube.com**	Videos	41	SN, MU, US	3

Table 4: *The sizes and locations of the infrastructures of the top 18 websites in Africa (Alexa & Afrodigit), and top 10 global sites (Alexa). We classify websites by their content type.*

These are standardbank.co.za in SBISCA, absa.co.za in Internet Solutions, and fnb.co.za in First National. Amongst the considered categories, Internet banking appears to be the only one in which most of the websites are hosted locally. At first, we thought this was driven by regulation. However, it is worth noting that we also find gtbank.com (a website offering financial services) hosted in the US. In terms of hosting practices, all of the top African websites choose a single AS to host their content. Another common characteristic is that they all use a single continent except for almasryalyoum.com, a news and media website, which is hosted by Cloudflare in Europe and South America. Most importantly, those African websites use at most 4 front-end IPs to serve the 146 *AFRINIC Prefixes* hosting our probes.

In contrast, the top global Alexa websites seen from our probes have a more distributed infrastructure. Of course, we have already explored Google's reach. Alongside this, other global Alexa websites generally host their content in multiple countries and ASes. That said, none are even approaching the scale of Google's provisions for Africa. For example, for facebook.com, we discover only 5 front-end IP addresses serving content. These are all hosted in AS32934 (Facebook), but we find no presence in Africa with all servers geolocated to the US, DE, and NL. Unlike Google, Facebook does not operate within African networks, instead basing their infrastructure at their own points of presence [18]. We find similar results across all global Alexa websites, all of which have made little inroads into Africa. For instance, yahoo.com serves Africa from the GB and US (both hosted in Yahoo's AS); and amazon.com serves Africa from the US (via AS16509 Amazon and AS46475 LimestoneNetworks). That is, Google's deployment in Africa is *not* the norm. Most content providers remain external to the continent.

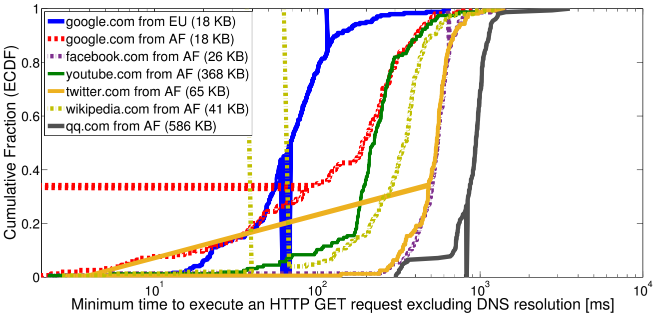
6.2 Website Performance

Our previous performance measurements (§4.5) have solely focused on end-to-end packet delay. We next explore the HTTP performance characteristics of all websites studied using the RIPE Atlas probes, including 242 extra probes randomly spread across Europe for comparison. We launched HTTP requests every 12 hours during the period June 2–5, 2015 from every probe to every website's homepage. To reduce the impact of differences in page size and third party objects, we only fetch the homepage HTML; we do *not* request images, adverts, javascript, etc. This results in a mean page size of 169 KB, with a standard deviation of just 166 KB (we include website size in the figures). Figure 8(a) shows the minimum time to fetch the global Alexa websites from each probe (measured by the length of the TCP connection). Again, we take the minimum to observe the best case scenario for each probe.

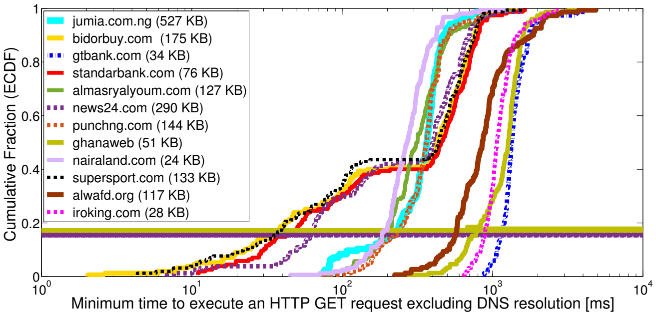
First, it can be seen that performance levels are extremely heterogeneous, with different probes getting wildly different performance. Interestingly, this heterogeneity varies amongst websites. Most noteworthy is Google in Africa, which has samples ranging from 2 ms to 1,250 ms: The average is 200.9 ms, while the interquartile range is 224.4 ms. We find the greater heterogeneity of Google is caused by the presence of GGCs in some of the probes' networks. The median load time in networks hosting a cache is just 148 ms compared to an overall median of 190.2 ms. Moreover, 60.7% of probes in ASes hosting GGCs have a delay that is below the average for the continent. That said, this still leaves many users who obtain slower than the average; in fact, only 26.2% have a delay that is below that of the median of the HTTP performance from Europe (67.6 ms) and only 32% have an HTTP performance below its mean (84.6 ms). This is also not simply caused by the previously reported highly DNS resolution times. Even when ignoring the DNS resolution times, we see that only 35% of probes in Africa fetch google.com in under 100 ms; in contrast, this is 78% in Europe. Furthermore, the average of the HTTP performance from Europe to Google is more than twice the one experienced from Africa. For medians, it is thrice.

In contrast to google.com, the other websites seen from Africa on Figure 8(a) have greater density around the mean (indicated by a sharp upturn in their CDF). This is because their infrastructures are not as well distributed in Africa as Google's. Consequently, *most* African users have similar (worse) performance. More precisely, the median of the HTTP requests performed by the RIPE Atlas probes hosted in African networks is 223.8 ms towards youtube.com, 339.8 ms towards wikipedia.com, 540 ms towards twitter.com, 549.1 ms towards facebook.com, and 943.41 ms to qq.com.

Figure 8(a) can also be compared to Figure 8(b), which presents the same data for the top African websites (from Alexa and Afrodigit). We find that the top African websites get approximately equivalent performance to the global top websites, suggesting that these regional services have made little effort to optimise their local distribution on the continent. Curiously, the regional websites on Figure 8(b) can also be separated into roughly three groups of varying load times. We note that the ones gaining highest performance are predominantly hosted in Africa, *e.g.*, supersport.com



(a) *Distribution of minimum time to execute an HTTP GET request per probe (ms) from Europe (EU) and Africa (AF) to top global Alexa websites.*



(b) *Distribution of minimum time to execute an HTTP GET request per probe from Africa to selected top local Alexa & Afrodigit websites.*

Figure 8: *HTTP fetch time for websites from RIPE Atlas probes. Website sizes are in parentheses.*

and standardbank.co.za, confirming the benefits that could be gained by services located themselves in Africa. In all cases, these are based in ZA, where infrastructure is well developed and affordable. Unfortunately, the worst performing local websites get even lower performance than the globally popular equivalents, indicating that these are not well provisioned. Unsurprisingly, these are those that are based in either the US or Europe. Thus, for African websites that do not have the scale of Google, we argue that they should take steps to host locally, hopefully encouraging global websites to follow suit too. In the future, as inter-AS connectivity improves, the introduction of shared caches in Africa could strongly incentivise this.

7. CONCLUSION AND DISCUSSION

This paper has explored the deployment of web infrastructure in Africa. Whilst recent studies have begun to measure the topology of the African Internet, we argue that this only addresses a subset of the challenges. By characterising the African web setup, we shed light on a topic that impacts greatly the quality of service received by users in Africa. The provision of web infrastructure will, thus, impact end-user experience greatly, even in the face of improving mobile and wireline connectivity.

We have shown that Africa is far from being self sufficient in terms of its hosting infrastructure. We began by studying Google's deployment. Although we discovered over a thousand caches across half of the African countries, we found that most users are still reliant on US infrastructure. We next showed that, unlike Google's global deployment, most GGCs in Africa are hosted in third party networks. These third party networks nearly always exclusively serve their own customers, with the exclusion of those connected to local IXPs (*e.g.*, JINX, CINX, TIX or NAPAfrica). Due to poor peering, we find that, in many cases, reaching a geographically nearby African cache actually has a higher delay than contacting the US. As such, sharing cache capacity across networks can only work with improved operator cooperation [26, 7].

That said, we find that Google is considerably more developed in Africa than other providers. We analysed both global and regional websites to find that even local websites are hosted outside of the continent. In fact, only 5 out of the 18 regional website front-ends surveyed were hosted locally (all in ZA). The cheaper cost of hosting abroad and the significant inter-AS delays amongst African ASes are two possible reasons for this. In all cases, we find clear trends showing that these hosting decisions have severe ramifications for performance. We consistently observed higher HTTP load times for non-Google websites hosted outside of the continent. For those hosted within the continent, we see roughly consistent performance, although it is not yet equivalent to the performance seen in Europe.

There are a number of key implications from our work. We have clearly shown that superior connectivity in Africa is only one part of the equation — it is also necessary to ensure that services are appropriately provisioned. Thus, content providers should begin to improve their presence there. Intuitively, popular regional providers should be the front runners in this effort. Although perhaps not immediately financially beneficial, this could act as a powerful catalyst for Internet uptake, which will result in revenues in the future.

Combining the above, we can therefore propose some steps that should be taken by both network operators and web providers: (*i*) operators must improve peering between networks to enable cache capacity to be shared cheaply and with low delay; (*ii*) content providers must concurrently be encouraged to host caches at existing IXPs; (*iii*) network operators must correct their DNS configuration settings to rely on local DNS for resolution; and (*iv*) public DNS resolvers should be placed in Africa (*e.g.*, at some of the 36 African IXPs [28]) to reduce the overheads for clients that continue to use them. These steps are complementary, with the ability of all stakeholders to encourage each other. For instance, if Google were to redirect more clients to GGCs hosted in Africa, network operators would be encouraged to increase peering to reduce the cost of these redirections.

This is just the start of our expanding campaign of Internet measurements in Africa. Its dynamism means that constantly increasing scrutiny should be placed on how it evolves. We intend for our work to provided guidance for ISPs and content providers wishing to deploy in Africa. Our future work will therefore focus on expanding our findings to deep dive into services beyond that of Google and Alexa sites. We are also developing design solutions to mitigate some of the problems observed.

Acknowledgements

We would like to thank Massimo Candela, Lucky Green, Lia Hestina, Victor Sanchez, Anna Mandalari, Pablo Camarillo, Joel Rosental, Jose Felix Kukielka, and the RIPE Atlas team. Arjuna Sathiaseelan is funded by H2020 RIFE project (Grant agreement No.644663).

8. REFERENCES

[1] Google Africa Blog. http://google-africa.blogspot.co.uk/.

[2] African Union. African Internet eXchange System. http://pages.au.int/axis, September 2015.

[3] AFRINIC. AFRINIC Database. ftp://ftp.afrinic.net/pub/stats/afrinic/, 2015.

[4] Alexa. Alexa Websites. http://www.alexa.com/topsites/, May 2015.

[5] APNIC. APNIC Database. ftp://ftp.apnic.net/pub/stats/apnic/, 2015.

[6] ARIN. ARIN Database. ftp://ftp.arin.net/pub/stats/arin/, 2015.

[7] J. Bezzina. Interconnection Challenges in a Converging Environment. The World Bank, 2005.

[8] Z. Bischof, F. Bustamante, and R. Stanojevic. Need, Want, Can Afford - Broadband Markets and the Behavior of Users. ACM IMC, 2014.

[9] F. Cheng, R. K. Sitaraman, and M. Torres. End-user Mapping: Next Generation Request Routing for Content Delivery. In *Proceedings of the 2015 ACM Conference on SIGCOMM, SIGCOMM*, volume 15, 2015.

[10] M. Chetty, S. Sundaresan, S. Muckaden, N. Feamster, and E. Calandro. Measuring Broadband Performance in South Africa. In *Proceedings of the 4th Annual Symposium on Computing for Development*, page 1. ACM, 2013.

[11] Cisco Visual Networking Index. Forecast and Methodology, 2009–2014, 2010.

[12] P. Cristel, C. Luca, V. Stefano, and R. Bush. From Paris to Tokyo: On the suitability of ping to Measure Latency. In *IMC '13 Proceedings of the 13th IMC*, pages 427–432, 2013.

[13] Ekinops. Liquid Telecom Deploys New Optical Network in Africa Using Ekinops Long-Haul DWDM Technology. http://www.ekinops.net/en/press-releases/liquid-telecom-deploys-new-optical-network-in-africa-using-ekinops-long-haul-dwdm-technology, May 2015.

[14] X. Fan, E. Katz-Bassett, and J. Heidemann. Assessing Affinity Between Users and CDNs sites. In *TMA*, 2015.

[15] R. Fanou, P. Francois, and E. Aben. On the Diversity of Interdomain Routing in Africa. In *PAM*, March 2015.

[16] R. Fanou, G. Tyson, P. Francois, and A. Sathiaseelan. Technical Report: African Content Measurement Campaign. https://techrep _cdma:PDQ7Rjkj@fourier. networks.imdea.org/external/techrep_cdma/index, June 2015.

[17] A. Gupta, M. Calder, N. Feamster, M. Chetty, E. Calandro, and E. Katz-Bassett. Peering at the Internet's Frontier: A First Look at ISP interconnectivity in Africa. In *Passive and Active Measurement*, pages 204–213. Springer, 2014.

[18] Q. Huang, K. Birman, R. van Renesse, W. Lloyd, S. Kumar, and H. C. Li. An Analysis of Facebook Photo Caching. In *Proceedings of the Twenty-Fourth ACM Symposium on Operating Systems Principles*, pages 167–181. ACM, 2013.

[19] Internet World Stats. Internet World Stats: Usage and Population Statistics. http://www.internetworldstats.com/stats.htm, June 2015.

[20] M. Kende and K. Rose. Promoting Local Content Hosting to Develop the Internet Ecosystem. ISOC Report, 2015.

[21] LACNIC. LACNIC Database. ftp://ftp.lacnic.net/pub/stats/lacnic/, 2015.

[22] Liquid Telecom. Liquid Telecom network map. http://liquidtelecom.com/about-us/network-map, May 2015.

[23] C. Matt, F. Xun, H. Zi, E. Katz-Basset, H. John, and G. Ramesh. Mapping the Expansion of Google's Serving Infrastructure. In *IMC*, 2013.

[24] MaxMind. GeoIP. http://www.maxmind.com/en/geolocation_landing, 2015.

[25] NSRC. Map of African Terrestrial and Undersea Fibre Networks. https://afterfibre.nsrc.org/, October 2015.

[26] T. Nyirenda-Jere and T. Biru. Internet Development and Internet Governance in Africa. ISOC Report, 2015.

[27] J. S. Otto, M. A. Sánchez, J. P. Rula, and F. E. Bustamante. Content Delivery and The Natural Evolution of DNS: Remote DNS Trends, Performance Issues and Alternative Solutions. In *Proceedings of the 2012 ACM conference on Internet measurement conference*, pages 523–536. ACM, 2012.

[28] Packet Clearing House (PCH). Internet Exchange Point Growth. https://prefix.pch.net/applications/ixpdir/summary/growth/, April 2014.

[29] Packet Clearing House (PCH). PCH IXP directory. http://prefix.pch.net/images/applications/ixpdir/ip_asn_mapping.txt, October 2015.

[30] PeeringDB. http://www.peeringdb.com/private/exchange_list.php.

[31] PlanetLab. PlanetLab. http://planet-lab.org, 2015.

[32] I. Poese, S. Uhlig, M. A. Kaafar, B. Donnet, and B. Gueye. IP Geolocation Databases: Unreliable? *ACM SIGCOMM Computer Communication Review*, 41(2):53–56, 2011.

[33] RIPE NCC. Global RIPE Atlas Network Coverage. https://atlas.ripe.net/results/maps/network-coverage/, May 2015.

[34] RIPE NCC. OpenIPMap database. https://labs.ripe.net/Members/emileaben/infrastructure-geolocation-plan-of-action, May 2015.

[35] RIPE NCC. RIPE Atlas - Raw Data Structure Documentation. https://atlas.ripe.net/docs/data_struct/, May 2015.

[36] RIPE NCC. RIPE NCC Database. ftp://ftp.ripe.net/ripe/stats/, 2015.

[37] S. Song. African Undersea Cables. https://manypossibilities.net/african-undersea-cables/, May 2015.

[38] S. Song. Mapping Terrestrial Fibre Optic Cable Projects in Africa. https://afterfibre.net/, May 2015.

[39] F. Streibelt, J. Böttger, N. Chatzis, G. Smaragdakis, and A. Feldmann. Exploring EDNS-client-subnet Adopters in Your Free Time. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 305–312. ACM, 2013.

[40] A.-J. Su, D. R. Choffnes, A. Kuzmanovic, and F. E. Bustamante. Drafting Behind Akamai (Travelocity-Based Detouring). In *ACM SIGCOMM '06 Proceedings of the 2006 conference on Applications, technologies, architectures, and protocols for computer communications*, volume 36, pages 435–446. ACM, 2006.

[41] Team Cymru. Team Cymru Services. https://www.team-cymru.com/.

[42] Y. Zaki, J. Chen, T. Pötsch, and T. A. Lakshminarayanan Subramanian. Dissecting Web Latency in Ghana. In *IMC*, November 2014.