

On the potential of Google AMP to promote local content in developing regions

Amreesh Phokeer^{*†}, Josiah Chavula[†], David Johnson[†], Melissa Densmore[†], Gareth Tyson[‡],
Arjuna Sathiaselan[§], Nick Feamster[¶],

^{*}AFRICAN [†]University of Cape Town [‡]Queen Mary University of London
[§]University of Cambridge [¶]Princeton University

Abstract—The Google Accelerated Mobile Pages (AMP) project has gained lot of interest from content providers as a means to improve user experience. It introduces a number of innovations to streamline the downloading, parsing and rendering of pages. Google recently announced the hosting of more than 2B+ web pages, covering more than 900K domains. Due to the constrained nature of Internet connectivity in developing regions, AMP offers particularly exciting potential for improving web user experience in these countries. This paper provides a first look at Google’s AMP performance in Africa. We start by outlining the current web infrastructure provisioning in Africa, using local news websites as a case study. Discovering a sparse and low performance environment, we then evaluate the benefits that AMP can introduce in terms of accessibility to local content in developing regions. This study reveals that in Africa, AMP is able to reduce page load time and page size by a factor of 3 and 8 respectively. However, AMP is not a neutral technology as the search engine favours content that is using Google AMP. This raises an important question of search neutrality.

I. INTRODUCTION

As reported by the Internet Society in 2015, the vast majority of content accessed by local users, in many developing countries and emerging regions, is hosted overseas. In many cases, the content must be accessed using international links and in sometimes under-provisioned networks [1]. Previous research has shown a strong link between local content, infrastructure development and access prices [2] and while it is known that local content can be a driver for increased global internet connectivity [3], performance and quality of experience (QoE) to access those content from developing regions still remain major issues [4]–[6].

In recent years, there have been many attempts by technology companies to deal with the problem of slow mobile web. For example, Facebook Instant Articles was introduced to make news stories load faster [7], while some browsers introduced proxy-based compression mechanisms (for e.g. Opera Mini, Flywheel) to improve the experience of users on slow networks [8], [9]. A particularly prominent technology that has recently emerged, is that of Google’s Accelerated Mobile Project (AMP). AMP strives to reduce Page Load Times, by decreasing page size and complexity, whilst using its own Content Delivery Network (CDN) to distribute third-party content. Anecdotal evidence suggests that AMP has the potential to significantly boost web traffic and improve QoE [10]. We argue that these benefits will be particularly felt in

developing regions, such as Africa, where connectivity remains challenging.

This paper performs a first study of AMP-enabled web pages in Africa, by evaluating web performance when accessing local content, particularly local news websites. To achieve this, we first try to understand the local content hosting situation in Africa. In order to shed some light on the distribution of local and remote hosting, we provide a detailed analysis of where Africa’s local content is physically and topologically located. We believe such a study will provide a deeper understanding of the content hosting challenges in developing regions, including the bottlenecks faced by local and global players in terms of infrastructure and quality of service (QoS). Furthermore, we investigate the prevailing benefits of using the AMP platform for content delivery in Africa. In particular, we seek to evaluate the extent to which AMP provides cheaper and faster access to information.

II. ACCELERATED MOBILE PROJECT

The Accelerated Mobile Project (AMP) is an open-source web publishing platform, with the aim to improve web content delivery to end-users, specifically on mobile phone devices. It leverages several nascent web publishing [11] and compression technologies [12] to reduce page sizes and therefore allows considerable amount of bandwidth savings, especially for users on limited data plans. It also benefits from Google’s very large networks of content caches around the world, therefore allowing AMP pages to be distributed on the edge, at faster download speeds.

AMP pages are essentially built around three main components:

- 1) **AMP HTML**: is an extended version of HTML to support AMP properties and it has some extra html tags such as *amp-img* or *amp-video*. It also has the *link* HTML tag which allow search engines to automatically detect AMP pages.
- 2) **AMP JS**: is the AMP JS library responsible for the fast rendering of AMP pages. It is also responsible for the optimization and handling of external objects.
- 3) **AMP Cache**: The Google AMP Cache is a proxy-based content delivery network for delivering all valid AMP documents. It fetches AMP HTML pages, caches them, and improves page performance automatically. When

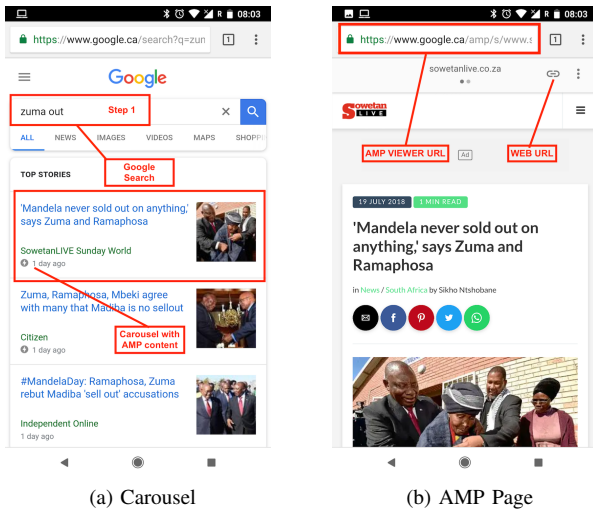


Fig. 1: AMP interface on mobile phone

using the Google AMP Cache, the document, all JS files and all images load from the same origin that is using HTTP 2.0 for maximum efficiency.

To be able to access an AMP page, a user typically needs to start with a Google search. The search engine will then provide a carousel (Figure 1) with a list of AMP-enabled websites - which can be identified by a *thunderbolt* sign. The AMP has essentially three types of URL which help to identify the AMP building block:

- **Web URL:** The canonical (original) web page that is published and hosted by the content provider. This is a normal HTML web page which has a special reference to an AMP version of the same page. If the page contains an *amphtml* tag, it means there is a corresponding AMP page. URL: <https://www.example.com/xyz.html>
- **AMP URL:** The AMP version of the original web page, published and hosted by the content provider. This page has special AMP HTML tag and also maintains a link to the canonical source. URL: <https://www.example.com/amp/xyz.html>
- **AMP Viewer URL:** This is a copy of the AMP URL web content which has been crawled and cached by Google. They will usually be served from the different Google caches scattered around the globe. This is the URL which the end-user will be redirected to after doing a google search (for e.g. see figure 1.b). URL: <https://www.google.com/amp/s/www.example.com/amp/xyz.html>
- **AMP CDN URL:** This URL is usually hidden from the end-user. After the crawlers have retrieved the AMP URL, objects from the AMP page are stored on the AMP project CDN. URL: <https://www-example-com.cdn.ampproject.org/c/s/www.example.com/amp/xyz.html>

In this study, we compare the Page Load time between AMP

URL, AMP Viewer URL and normal web pages.

III. METHODOLOGY

The performance of websites is defined by a mix of the underlying infrastructure (e.g. network and servers) and the structure of the web content itself. Before moving onto exploring AMP (which modifies the web content), we first inspect the available content hosting infrastructure and location. We look at both quality of service (QoS) by measuring the RTT and the number of intermediate hops (traceroute) to the news websites and we measure quality of experience (QoE) by running measurement on the page load time on both AMP and non-AMP web pages.

A. Selection of websites

To measure web performance, it is first necessary to select a set of websites. To this end, we compile a list of 1413 news websites from 54 different countries obtained directly from <http://abyznewslinks.com/>. We focus on news sites, as it is possible to gain ground-truth on their locations; we avoid using Alexa rankings, as these mostly contain international, rather than local websites (e.g., Google, Facebook, Twitter etc).

We therefore restrict our analysis to news websites that we obtained from the ABYZ News Links¹ directory. The directory does not rank the sites, but rather attempts to list all the prominent media sites for each country. We chose to study news websites only as they represent content of local interest with regards to the country where the reader base is located, but also because AMP is very popular amongst news publishers.

B. Measurement platforms

Once the list is compiled, it is then necessary to launch performance measurements towards their respective host or domain. Specifically, we are interested in understanding which paths packets are taking (*traceroute*) and how long it takes them to move from $A \rightarrow B$ (latency). For this, we use the RIPE Atlas platform. Additionally, to evaluate the impact of the hosting locations on performance from an end-user's perspective, we perform Page Load tests using the Speedchecker platform. It is important to note that the tests are run from probes in the countries where the news website is local to, such that we get the closest results possible to the end-users' experience. Below is the three platforms used:

- **RIPE ATLAS²:** is a global, open and distributed hardware-based platform made up of more than 10,000 probes around the world. At the time of the experiment, 196 RIPE Atlas probes were used from 37 countries.
- **SPEEDCHECKER³:** is a commercial software-based platform running on PC, Android and DD-WRT routers. It provides an API for Page Load tests, in which we can gather the following metrics:

¹<http://www.abyznewslinks.com/>

²<https://atlas.ripe.net/>

³<http://www.speedchecker.xyz/>

- 1) **Page Load Time (PLT)**: is time it takes between the first initiate request and when the page is fully loaded.
- 2) **Time to First Byte (TTFB)**: is the time taken for a browser to receive the first response byte from a server.
- 3) **DNS Lookup Time (DNSLT)**: is the time it takes for the probe to complete a DNS resolution for the domain.
- 4) **SSL Negotiation Time (SSLNT)**: is the time it takes to initiate and complete a full SSL handshake
- 5) **Initial Connection Time (ICT)**: is the time it takes to establish a TCP connection (i.e. a full handshake)

Page Load time (PLT) consists of: (1) *network time* (ICT, DNSLT, SSLNT) (2) *browser time*, which is the time it takes to download web objects and process the Document Object Model (DOM) and render the page. PLT is therefore the overall QoE metric we are evaluating in this experiment.

- **MAXMIND**⁴: is a geo-location database which provide an IP to Location mapping. With this, we were able to determine in which network or country the server was located.

C. "traceroute" dataset

We start with a list of 1413 websites in Africa as extracted from <http://abyznewslinks.com>. We managed to retrieve RTT (Round Trip Time) results on 1191 websites from 37 countries, the remaining websites being non-responsive. The *traceroute* measurements were repeated over a five-day period, resulting in about 19,299 successful measurements between the probes and the websites. Each *traceroute* measurement returns three final hop RTTs, meaning that in total, there were 57,897 end-to-end RTTs. However, the number of domains probed in the Page Load test, as explained below, is much lower than in the *traceroute* dataset, as we measure only AMP-enabled domains.

D. AMP-enabled news article links

Before performing any page load tests, we scrape the 1191 active websites, and looked for a special `<amphtml>` tags, i.e. we select only AMP-enabled web pages. Figure 2 shows the distribution of AMP and non-AMP news websites in Africa. We end up with a list of 194 AMP-enabled news websites in 22 countries, meaning 16% of active news websites are using AMP in Africa.

As the number of article links vary by website varies considerably, we select a random list of 10 AMP-enabled web articles (per news website) that we could find during the scraping process. This sampling produces a dataset of 1477⁵ individual URLs, on which we run Page Load tests both on the normal web pages and their equivalent AMP page. So the total number of individual pages probes were 4431 for the three types of URL.

⁴<https://www.maxmind.com>

⁵Some websites had less than 10 article links when scrapped

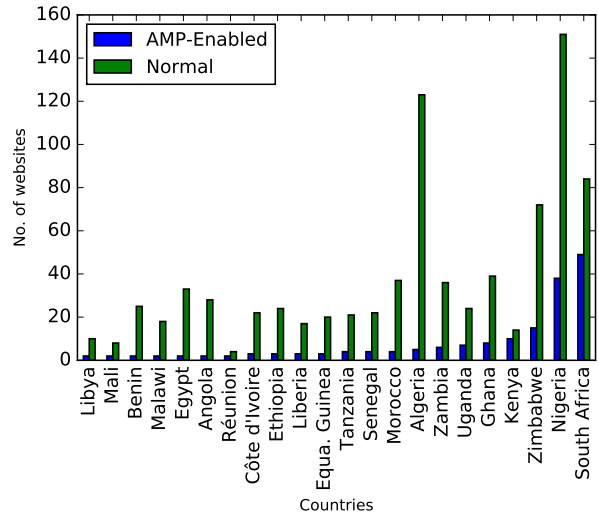


Fig. 2: Distribution of AMP-enabled vs non-AMP websites

TABLE I: Summary of measurements obtained

Data collected		Traceroute	Page Load
# of domains		1191	194
# of countries		37	22
# of ASNs		223	51
# of unique web pages		-	1477
# of URLs probed			
Original/AMP Viewer/AMP CDN		-	4431
RIPE Atlas	Probes	196	-
	Measurements	57,897	-
Speedchecker	Probes	-	225
	Measurements	-	9224

E. "Page Load" tests

The SpeedChecker platform provides an API which allows you to select a destination (URL) and a set of probes from which the measurements are launched. For this experiment, we select exclusively probes running on Android devices. It is important to note that SpeedChecker only provide access to Android probes running on WiFi only. Each API call provides results from a maximum of 10 measurement points, selected randomly from the available pool. Additionally, we repeat the whole set of Page Load tests (on the three types of URLs) at five different times, during a week, before aggregating the results. From the 4431 different URLs measured, we retrieve a total of 9224 measurements data points i.e. we measure from an average of two probes for each URL.

IV. MEASUREMENTS AND RESULTS

This section describes both the *traceroute* and *page load time* measurements to the local news websites, which helped to reveal information such as latency and geolocation of content servers. We first compare URL types i.e. AMP vs non-AMP pages and then we regions of access by running the page load measurements from EU, US and Africa to understand the impact of remote hosting on QoE, on accessing the news websites. The full set of page load measurements

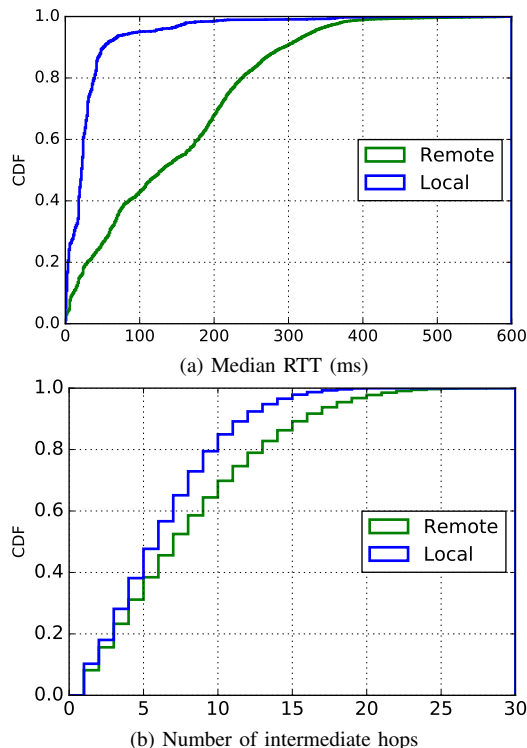


Fig. 3: Median latency to remotely hosted content is 7 times longer than locally hosted content

were repeated five times and each API call produced 10 page loads data points.

A. Routes to reach news websites

With the 57,897 trace paths obtained (including RTT values), we compute the median latency for locally hosted websites versus remotely hosted websites (see Figure 3.a). As expected, the cost in terms of latency to reach a remotely hosted website is much higher (by seven times). This is also confirmed by the number of intermediate hops as in Figure 3.b, which is slightly higher for remotely hosted websites, as the use of international connectivity brings in added delay [4]. We argue that usage of AMP can potentially change this situation, as AMP pages can be served from Google AMP caches (GGC), which are found in most major locations in Africa⁶ or from Cloudflare AMP caches⁷, the current two AMP cache providers.

B. Location of hosting providers

From Figure 4.a below we observe that the US takes the lion's share in hosting African content, with about 56% of all the websites being hosted by American companies. Within Africa, South Africa leads in the content hosting business, hosting about 7% of all of Africa's remotely hosted news websites (i.e., minus those that belong to South Africa). The rest of the websites, about 25%, are hosted in various countries

⁶<https://peering.google.com/#/infrastructure>

⁷<https://amp.cloudflare.com/>

TABLE II: Number of web objects collected on main pages

Object type	Amount	Size (KB)	Percentage
image	46200	1699477	78.9
script	13072	416825	19.35
video	17	409	0.01
audio	5	36970	1.71

in Europe. As previously mentioned, having content hosted remotely adds to the overall page load time and therefore can impact negatively the overall quality of experience of a user. This is what AMP aims to tackle by reducing web site complexity and load times. In section IV-D we study the difference in page load time when African content is accessed from Europe or US as opposed from within Africa itself. We later repeat the same study on AMP-enabled web pages to compare and contrast (see figure 6).

C. Distribution of content networks

Similar to the geolocation analysis, network-level analysis shows that most of the websites are hosted by foreign companies. We scraped all of the 1191 news websites and retrieved the images, javascript, video and audio files from their main page. In total, we retrieved 57,444 objects as shown in Table II amounting to ≈ 2 GB of data. As expected, images represented a larger share (78%). To understand where those objects were hosted, we analysed the URLs of the objects and retrieved their domains (e.g., *fbcdn* is for Facebook, *wp.com* is for Wordpress). Without much surprise, Wordpress (US) takes the biggest share of the market, hosting about 72% of the websites, followed by NetDNA (8.3%), StackPath (8.1%) and Cloudflare (4.1%). We removed data from Google and Facebook as they were mostly pointing to ad-related objects and therefore skewed the results. Otherwise Google represented 23% and Facebook 9.7%, which were mostly consisted of javascript documents and Facebook platform images. This correlates with our finding on the geolocation of the domains, which is mainly dominated by the US, where Wordpress is predominantly hosted.

What is important to note is that the leading content hosting providers for Africa's remotely hosted news websites are largely based on Cloud infrastructure and make use of CDNs. For e.g. Cloudflare, Akamai, Microsoft Azure, Amazon have their PoPs in multiple places in Africa. Wordpress is currently the most used CMS in the world (60%) and this might explain why Google has strategically partnered with Wordpress to disseminate AMP pages using the CMS plugin.

From a performance perspective, Figure 4.c shows that there is some benefit of using CDN to host local content as, in doing so, it reduces the latency to the content by a factor of two. This performance gain, while it does contribute to achieving better QoE, comes with a cost, which might not be affordable for small publishers in developing countries [13]. Domains using CDN were identified by the CDN domain in their URL, the network name of their ASN (Autonomous System Number) also reveal which CDN is currently hosting

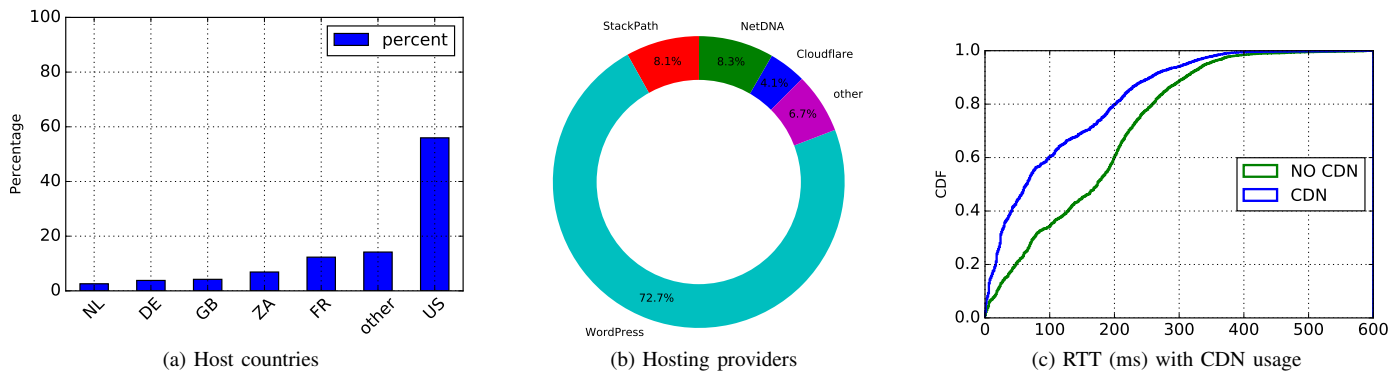


Fig. 4: Content hosting and distribution

the website. Having content closer to the edge is therefore what content providers should be aiming for. Google AMP currently uses two AMP Cache provider namely Google AMP Cache and Cloudflare AMP Cache. The platform automatically chooses from which cache provider the content should be served. Our study therefore evaluates tries to understand the benefit of using Google AMP project is leveraging their global CDN infrastructure to provide low-delay access to content.

D. Measuring web quality of experience

In this section, we describe the results obtained from measuring the *Page Load* dataset. Firstly, the aim of the experiment is to understand the difference in Page Load time between the different URL types (AMP URL, AMP CDN URL, AMP Viewer URL and WEB URL) and secondly, we want to understand the difference in QoE for a user located in Africa as opposed to being located in the US or EU, where most of the African news content are actually hosted. For the latter, the Page Load tests were run for the **WEB URL** and their equivalent **AMP Viewer URL**⁸, using probes in EU, US and Africa.

COMPARING URL TYPES:

Page Size: Figure 5.a describes the differences between the three URL types in terms of page size. As expected, we see a clear difference in page size between WEB URL and AMP URL, confirming that AMP strips down or compresses many objects before rendering. In 80% of the pages we have explored, AMP pages is 8x smaller than normal web pages. As we shall see below, this definitely has a positive impact on the overall Page Load time.

Time to First Byte: As it can be seen from figure 5.b, the TTFB (i.e. latency) for AMP CDN URL is much smaller than AMP URL, AMP Viewer and WEB URL. As AMP CDN URL is served from an edge location (for e.g. Google or Cloudflare), this explains why latency to reach the content is almost three times smaller. We can also observe AMP URL and Web URL have more or less the same TTFB, as both web pages are generally hosted on the same server. In section

⁸We choose the AMP Viewer URL as point of comparison as it is the URL that the user will navigate to after doing a search on Google.

Page Load Time: Figure 5.c is a CDF of Page Load Time where it can be observed that the fastest is AMP Viewer URL. One main challenge we have with measuring Page Load time of the AMP Viewer URL, is that Android Speedchecker probes in Africa are scanty, leading to some skewness in the data. On the other hand, we see a better distribution in the EU/US measurements. Additionally, we observe that on average, an AMP URL will take around 10s to load while it takes 17s for a WEB URL to load.

COMPARING REGIONS OF ACCESS:

Time to First Byte: Figure 6.a is a CDF of TTFB that have been collected using probes from Africa, Europe and the US. As it can be seen on average, it is faster to access African content from EU or US than from Africa itself by a factor of 1.5. This is a well known situation which can be due to many factors as explained earlier such as circuitous routing or remotely hosted content. However, with regards to AMP, access to the first byte is much faster in Africa than from EU or US. This perhaps may have to do with the fact that AMP pages are hosted on Google caches, located within an ISP in the country or close by.

Page Load Time (PLT): PLT is the most important metric in this analysis as it is a proper indication of QoE. It is a function of the DNS Lookup Time and SSL Negotiation Time, the latency to the server (\approx TTFB) and the amount of data that need to be download. Besides, Page Load time can also be influenced by the link quality (throughput, congestion, jitter, etc). In figure 6.b, it is very clear that loading AMP is faster than loading normal pages, across the three regions. The plot also shows that normal pages are loaded faster in the EU/US than in Africa (at least by 1.5x). With regards to AMP, it seems to be faster to load an AMP page from Africa than from EU/US.

DNS Lookup Time: Figure 6.c shows that it is generally faster to resolve DNS queries on AMP, but it is interesting to note that it is faster to resolve the domain names from Africa than from the US. An explanation could be that DNS resolvers in Africa already cache the domain (as they are from Africa) as opposed to resolver in the US.

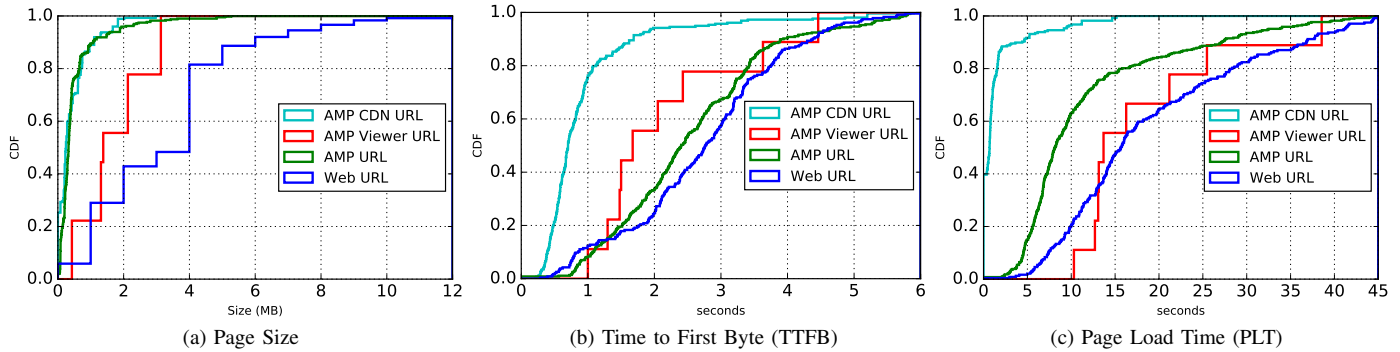


Fig. 5: Larger page size, TTFB and PLT for WEB URL

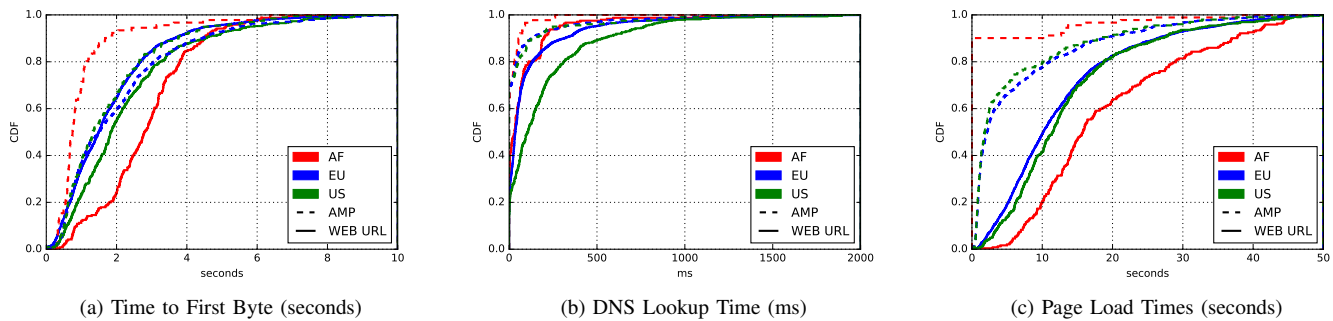


Fig. 6: Page Load Times much higher from Africa (50% larger than 30sec), but lower using AMP (80% lower than 10sec)

E. Location of AMP Caches

To supplement our analysis in the previous section, we use traceroute and reverse DNS lookup to determine the location of the Google AMP caches as they would be accessed from the 22 countries in the AMP study. Google AMP caches share the same domain name `cdn.ampproject.org` and the caches are hosted by anycast servers using the same range of IP addresses. To determine those locations, we run a traceroute measurement campaign⁹ using RIPE Atlas from 19 countries (no probes were available from Liberia, Equatorial Guinea and Mali at the time of the experiment). Each AMP cache has a reverse DNS domain for e.g. the Mombassa cache is `mba01s08-in-f1.1e100.net`, where *mba* is the airport code.

We observe that AMP pages from Eastern and Southern Africa countries are predominantly fetched from Mombassa, Northern Africa from Milan, Frankfurt and Marseilles and Western Africa from Frankfurt and London. We found two interesting cases Ethiopia fetching content from Stockholm and Angola from Washington DC, with relatively higher latencies. Formoso *et al.* also found the same exact two countries which did not share the same latency characteristics as other neighbouring countries [5]. The reasons are that Angolan ISPs

usually peer with networks where Portuguese content is hosted (Brazil or Portugal) and the fact that Angola is well connected to the US through the undersea cable¹⁰ via Brazil. Ethiopia has only one incumbent operator and international connectivity are sometimes carried out using satellite links.

Another notable observation is that Nigeria is hosting a Google AMP cache and latency from within the country is very low (1.434ms). But none of the other Western African countries in the study is fetching content from Lagos. This again proves a lack of coordination and peering in the Western African region, as explained by Formoso *et al.* [5].

F. Discussion

Our study clearly shows that PLT of WEB URLs are higher, when we try to access the news websites from Africa as opposed to from US and EU. This corroborates with our findings on the location of the content itself which confirms that hosting in the US and in the EU, has indeed an incidence on the overall PLT. The median PLT in Africa is 15s vs 10s in Europe or US. However, it can also be seen that their AMP equivalent pages have a much smaller PLT. We see a 80% reduction in PLT for webpages on AMP. The benefits are particularly pronounced in our Africa measurements, which

⁹We run a traceroute measurement, collect 2437 traces using 58 probes from 19 countries

¹⁰<https://www.submarinecablemap.com/>

TABLE III: Location of AMP Caches when accessed from selected African countries

Source Country	AMP Cache Location	Average Latency (ms)
Nigeria	Lagos	1.022
Tanzania	Mombassa	6.262
Kenya	Mombassa	6.581
Uganda	Mombassa	15.344
Algeria	Milan	33.638
South Africa	Mombassa	46.457
Zimbabwe	Mombassa	68.419
Libya	Frankfurt	80.021
Egypt	Marseilles	80.141
Morocco	Marseilles	86.705
Malawi	Mombassa	91.684
Senegal	Frankfurt	104.025
Cote D'Ivoire	Frankfurt	139.215
Angola	Washington DC	192.245
Benin	London	195.214
Ethiopia	Stockholm	207.943
Reunion	Amsterdam	211.792
Ghana	London	236.224

on average save 12s vs. 8s in Europe, between AMP and non-AMP pages.

Through the above, we want to give an overview of the current content hosting situation in Africa. Because of a lack of infrastructure and proper incentives to host local content locally, most of Africa's local content today is hosted in remote locations. Even if there is more and more deployment of cloud service providers in Africa, we see from the plots that it is still slower to access content from a remote location, than from Africa itself. Obviously, we hope to see the situation improve over time as more CDNs deploy their PoPs in Africa.

V. IS AMP A VIABLE ALTERNATIVE?

Our study shows that there is indeed a benefit of using AMP as a means of content delivery, especially for users in developing regions. By reducing the size of the page, users on limited data plans can also save on expensive mobile data usage [14]. AMP is therefore a promising web publishing technology that not only is beneficial for the end-users but also allows content publishers to increase their reach and maintain their reader-base.

However, AMP comes with lot of controversies. It forces end-users to remain within the Google's domain and as a consequence it diverts traffic towards websites hosted by Google [15]. At a scale of billions of users, AMP would reinforce Google's dominance in the web as it can be potentially considered as the *de facto* mechanism for fast content delivery. The fact that third party content is now within the realm of Google, the search engine can choose to prioritise which content end-users should see first, at the detriment of non-AMP enabled web pages. This situation further raises questions of centralisation of data and search neutrality [16].

The success of the World Wide Web and the Internet is due to consensus building around open standards developed by the World Wide Web Consortium (W3C) and the Internet

Engineering Task Force (IETF). Web acceleration technology, like AMP, should be no exception and should ideally become a protocol adopted and developed by the W3C. At a minimum, web acceleration technology should be made available as open source technology that can be deployed in ISP DMZs or locally owned community networks and adapted to local needs. Examples of open source web acceleration are Ziproxy¹¹ and Varnish¹². One of the challenges is that these solutions cannot currently handle HTTPS. Accelerating content that uses HTTPS would require trusted edge caching services where content providers can place their content closer to the user edge of the network. These edge cached services could then run web acceleration technology such as AMP or Ziproxy to further accelerate this content to users if necessary. With web acceleration being in control of the ISP or community network, acceleration parameters could be tuned to the users preferences or network conditions.

VI. CHALLENGES AND LIMITATIONS

This study is a preliminary study of AMP and Web Quality of Experience. Some of the challenges we faced were: some websites were not responsive and had to be removed from our initial hitlist. Scraping the websites to retrieve news articles (only) was challenging as content organisation differs from website to website. Furthermore, our study is limited by a few factors. The Speedchecker platform is a software-based platform, where probes are running on end-users' mobile phones or PCs. This makes the number of active probes very volatile. Therefore, we have to run the experiment multiple times to make sure we have enough measurement data points for analysis. Another issue is the number of active probes in Africa, which is far less than in other regions. There are also very few Android probes on the platform in Africa, which explains why our results for AMP Viewer URL were a bit skewed. Finally, Speedchecker does not (yet) provide the HAR (HTTP Archive) files associated with the Page Load time test; this would have enabled us to make a more in-depth analysis of the internals of AMP. AMP also has a pre-fetching mechanism, whereby content is loaded in the background, while the mobile phone user is browsing the search results. The aim is to pre-fetch content that the user is likely to access, therefore optimizing further the browsing experience. In this study, we did not measure the pre-fetching mechanism.

VII. RELATED WORK

In a study commissioned by Google in 2017, Forrester Consulting conducted a Total Economic Impact (TEI) study to evaluate the potential return on investment (ROI) publishers are making with the deployment of AMP [17]. They based their study on a composite organization representative of existing AMP eCommerce vendors and found out that there was an increase by 20% in profit growth from sales conversion rate. They also found a 10% year-over-year increase in AMP site traffic, which also resulted in profit growth. Additionally,

¹¹<https://en.wikipedia.org/wiki/Ziproxy>

¹²<https://varnish-cache.org/>

in a survey conducted, publishers also benefit from improved discoverability (Google search prioritization) and increased traffic. They also noted improved customer and reader engagement, whereby some interviewees experienced a 60% increase in pages per visit, 0.3% increase in returning visitors and time spent on AMP pages doubled. The study was conducted on four AMP use cases (two eCommerce vendors and two publishers). In another recent study by Charbeat [18] (a content analytics service), they evaluated mobile audience behaviors and content interaction on AMP and Facebook Instant Article (FIA). They showed that AMP roughly loads four times faster than standard mobile site. In terms of user experience, they observed that average user engagement time is 48s as opposed to 36s for a standard mobile page, i.e. 35% increase in reader's engagement time. Furthermore, Doll *et al.* presented the first formal statistical report on the relationship between AMP adoption and publisher traffic [19]. They collected data from 159 publishers which launched AMP in 2017. They found that AMP adoption resulted in a 22% average Google traffic increase. The above studies focus mostly on the economic aspects and provide statistics on the ROI from a publisher's perspective (reader's engagement and increased web traffic) but none of the reports produced a wide-scale analysis of the performance gain from an end-user's perspective. Our study, though limited to measuring page load time, provides a preliminary empirical analysis of how AMP can benefit publishers of local content in a developmental context.

VIII. FUTURE WORK

This study of AMP in Africa provides a good understanding of its potential as an alternative content delivery mechanism, which can significantly boost the consumption of local content in developing regions. In the future, we want understand in greater details the web optimization mechanisms AMP is using. Understanding the impact of pre-fetching, both in terms of end-user performance and cost, is also an important aspect to consider, especially in the context where mobile data plans are prohibitive. The latter would require a more controlled experiment in a lab setting. Additionally, AMP makes use of the Google AMP caches and Cloudflare AMP caches to serve content closer to the end-user, understanding the location and performance to those cache nodes would be necessary.

IX. CONCLUSION

This paper has evaluated the performance impact of Google's AMP in Africa. We have analysed the hosting situation of 1191 African news websites, focusing on such metrics as RTTs and number of hop between African vantage points and the websites, hosting countries and networks, pages sizes, and page load times. We further do an analysis on 194 of the 1191 websites that were found to be using the AMP infrastructure by comparing access via AMP and traditional web hosting. Our results indicate a significant reduction in website sizes as downloaded onto web clients by a factor of 8. This reduction can make for significant savings in data costs for Internet users in Africa, especially for mobile

Internet users. In terms of performance, our results indicate that page load times are much higher in Africa compared to EU and US (50% of website load times in Africa are over 15s, compared to EU/US where 70% of pages load under 15s), but significantly lower using AMP (80% lower than 10sec). When AMP is used, we observe a significant improvement in page load, with over 80% of the pages loading under 10s. While AMP's performance improvements in Africa are obvious, further discussion needs to be had in terms of how the platform promotes externalization of Africa's local content. We believe the findings would allow to better understand the current underpinning of local content generation and consumption in developing regions.

REFERENCES

- [1] M. Kende and K. Rose, *Promoting Local Content Hosting to Develop the Internet Ecosystem*, 2015. [Online]. Available at <https://www.internetsociety.org/resources/doc/2015/promoting-local-content-hosting-to-develop-the-internet-ecosystem/>. [Accessed on 15-Nov-2018].
- [2] Internet Society, Economic Cooperation, United Nations Educational, Scientific and Cultural Organization, Oecd, Isoc, and Unesco, "The Relationship Between Local Content , Internet Development," *OECD Digital Economy Papers*, vol. 217, no. 217, p. 141, 2013.
- [3] J. Bhargava and A. Klat, "Content democratization - How the Internet is fueling the growth of creative economies," tech. rep., 2017.
- [4] J. Chavula, A. Phokeer, A. Formoso, and N. Feamster, "Insight into Africa's country-level latencies," in *2017 IEEE AFRICON Sci. Technol. Innov. Africa, AFRICON 2017*, 2017.
- [5] A. Formoso, J. Chavula, A. Phokeer, A. Sathiaseelan, and G. Tyson, "Deep diving into africa's inter-country latencies," in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, pp. 2231–2239, April 2018.
- [6] A. Gupta, M. Calder, N. Feamster, M. Chetty, E. Calandro, and E. Katz-Bassett, "Peering at the internets frontier: A first look at isp interconnectivity in africa," in *International Conference on Passive and Active Network Measurement*, pp. 204–213, Springer, 2014.
- [7] N. Newman, "Journalism, media, and technology trends and predictions 2017," 2017.
- [8] V. Agababov, M. Buettner, V. Chudnovsky, M. Cogan, B. Greenstein, S. McDaniel, M. Piatek, C. Scott, M. Welsh, and B. Yin, "Flywheel: Google's data compression proxy for the mobile web.," in *NSDI*, vol. 15, pp. 367–380, 2015.
- [9] R. Netravali and J. Mickens, "Remote-control caching: Proxy-based url rewriting to decrease mobile browsing bandwidth," in *Proceedings of the 19th International Workshop on Mobile Computing Systems & Applications*, pp. 63–68, ACM, 2018.
- [10] CHRIS BREAU and BRADLEY DOLL, *Research Study: Only 1 in 3 publishers see a clear traffic boost from AMP*. [Online]. Available at <http://blog.chartbeat.com/2018/08/23/research-study-1-3-publishers-see-clear-traffic-boost-amp/>. [Accessed on 15-Nov-2018].
- [11] W3C, *Distributed and syndicated content*, 2017.
- [12] J. Alakuijala and Z. Szabadka, "Brotli compressed data format," vol. RFC 7932, 2016.
- [13] Nitin Rao, "Bandwidth Costs Around the World," 2016.
- [14] A. Phokeer, M. Densmore, D. Johnson, and N. Feamster, "A first look at mobile internet use in township communities in south africa," in *Proceedings of the 7th Annual Symposium on Computing for Development*, p. 15, ACM, 2016.
- [15] Multiple, "AMP letter," 2018.
- [16] P. L'Ecuyer, P. Maillé, N. E. Stier-Moses, and B. Tuffin, "Non-neutrality of search engines and its impact on innovation," *Internet Technology Letters*, vol. 1, no. 1, p. e10, 2018.
- [17] Forrester, "Cost Savings And Business Benefits From Building Fast, High Performing Mobile Experiences With AMP," tech. rep., 2017.
- [18] J. Saroff, "The New Speed of Mobile Engagement - Chartbeat Blog," *Chartbeat*, pp. 1–8, 2018.
- [19] A. B. B. Doll and C. Breaux, "Amped about AMP ? Time to look more carefully," tech. rep.