Internet Development in Africa: A Content Use, Hosting and Distribution Perspective

Enrico Calandro[1(✉)], Josiah Chavula[2], and Amreesh Phokeer[3]

[1] Research ICT Africa, Cape Town, South Africa
ecalandro@researchictafrica.net
[2] University of Cape Town, Cape Town, South Africa
jchavula@cs.uct.ac.za
[3] AFRINIC, Ebene, Mauritius
amreesh@afrinic.net

Abstract. Although a considerable investment in broadband infrastructure has improved broadband speeds across many African countries, the reliability and performance that users ultimately receive is determined also by the interconnection between networks and Internet Service Providers (ISPs) and by where the content, services and applications are hosted. Often, high latencies to remote destinations introduce significant performance bottlenecks, suggesting that, in addition to investments in higher throughput links, effort should be devoted to improving interconnection between ISPs and locating content closer to the users. By untangling the complexity of content access, use, hosting and distribution in Africa, this study offers three main contributions. First, it discusses challenges related to usage, hosting, distribution of local content and services in Africa, by developing a case on African local news websites. Second, it makes publicly available measurement data and indicators for local content use, hosting, and distribution across all African countries. And third, it provides points of policy recommendations on how to improve internet access and use, and infrastructure performance from a content perspective.

Keywords: Local content · Web hosting · Latency · Peering · Content infrastructure

1 Introduction

Between 2013 and 2017, Africa experienced the most rapid growth of international internet bandwidth in comparison to other regions, growing at a compound annual rate of 44% [1]. However, the majority of the African population continues to be offline due to high data costs [2–5] lack of local content [6], and poor network performance [7,8], despite a number of investments and projects

G. Mendy et al. (Eds.): AFRICOMM 2018, LNICST 275, pp. 131–141, 2019.
https://doi.org/10.1007/978-3-030-16042-5_13

to expand and upgrade undersea cables, and the new investments in terrestrial fibre network capacity. The average Round-Trip Times (RTT or latency) is still high, due to poor Internet peering infrastructure [9] and topological inefficiencies [10]. Not only is the peering fabric of the continent uneven, but also content infrastructure in Africa requires significant development. Studies suggest that content is a dominant component of network traffic, but local content is a major bottleneck to African connectivity [11].

While this significant investment in broadband infrastructure [12] and data centres [13,14] in Africa has improved throughput across the continent, most content in the continent, even local websites, are hosted and are delivered from overseas [15]. A number of factors have been postulated to be the reason for remote hosting of African websites. Foremost, hosting services and infrastructure have not been pervasive in Africa, rendering the provision and management of web content within the continent unattractive [16]. Although it is generally cheaper to host with remote/foreign companies than with local hosting companies, content hosted abroad must be routed back to the country of origin over international Internet transit links that, in spite of significant infrastructure investments in recent years, are still expensive. The resulting high costs to access content hosted abroad are generally borne by Internet Service Providers (ISPs), and ultimately by Internet users. The result is a negative externality, where the economic decisions of content providers to host abroad have a negative impact on ISPs' costs, which in turn increases the cost of Internet usage and limits Internet content demand [16].

1.1 Research Questions

This paper seeks to untangle the complexity of content use, hosting and distribution in Africa. More specifically, it poses the following key questions:

1. What type of content do African people consume?
2. Where is local African content hosted? Taking into account local news website, how much of the content is hosted locally vs globally?
3. How is content hosted in Africa?
4. What routes are used to access locally hosted content?
5. What is the latency for content hosted in various regions?

1.2 Contribution of the Study

In answering these key research questions, this paper makes three contributions: First, it offers a discussion on challenges related to usage, hosting, distribution and accessing of local content in Africa. Secondly, this study makes publicly available measurement data the web content infrastructure in Africa, and at the same time it illustrates what the factors affecting performance when accessing Africa's digital content are. The third contribution of this study is the provision of specific policy recommendation points on how to improve Internet adoption and infrastructure performance from a content perspective. To achieve this, the

study undertakes an active Internet measurements campaign to characterize the latencies and to geolocate web servers and routes used for Africa's online content, focussing on local news websites in each country. This paper explores the hosting patterns and performance associated with a large sample of about 1,100 local news websites. To test assumptions on content use from a user's perspective, it draws on nationally representative ICT access and use surveys conducted by Research ICT Africa (RIA) in 2017 in seven African countries.

2 Research Methods and Data Sources

The study makes use of Internet measurements, household surveys data and pricing data as the three main data sources.

First, in order to measure what type of content people in Africa consume, the study draws on the Research ICT Africa (RIA) #AfterAccess survey[1], which delivers nationally representative results for households and individuals. The survey is based on enumerator areas (EA) of national census sample frames as primary sampling units. Through the survey, we could establish what type of content African people consume.

Second, Internet measurements were conducted to gather information regarding where Africa's web content is hosted, as well as to assess the associated performance and cost implications. The first task was therefore to identify websites that would be considered representative of Africa's local web content. In this study, Africa's web content is defined as content that is primarily generated and consumed within each African country. It was decided therefore to study local news and media websites, which by definition constitute a significant body of local content in Africa. A list of local news websites made for every African country was compiled from ABYZ News Links[2], an online directory of links to online news sources from around the world organized on a geographical basis.

Third, to answer to the question of whether local news websites are locally hosted within their countries or not, Traceroute data was analysed to determine the networks that host each of the measured websites, as well as the networks through which traffic flows between the websites and the measurement vantage points. Subsequently, the geographical location of each web-hosting server was determined. MaxMind[3] geolocation database was used to obtain the network information, which includes the networks' Autonomous System Numbers (ASNs) and network names. The geolocation database was also used to identify the countries related to each IP in the dataset, both the websites' web-servers and routers along the paths to the websites. The country-level geolocation was preferred as it has been shown to have relatively higher accuracy compared to city-level geolocation [17,18]. The Traceroute measurements were repeated over a five-day period, resulting in about 19,299 successfully measurements between

[1] https://researchictafrica.net/2017/08/04/beyond-access-surveys-questionnaires-methodologyand-timeframe/.

[2] Available at the following link: http://www.abyznewslinks.com/.

[3] https://www.maxmind.com/en/geoip2-country-database.

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. A Traceroute measurement is considered successful if an IP route can be determined from the source to the web-hosting network, thereby also being able to reveal a delay estimate to the website. Each successful measurement contains the IP address of a website's hosting server, a series of IP hops from the vantage point up the server, as well as the delays (round-trip time, RTT) at each hop (router). Also, each Traceroute result is made up of multiple records, one record for each of the multiple hops on the path. Consequently, the final dataset was made up of 256,654 records, with each record comprising source and destination addresses, as well as the IP hop and RTT from the source to that hop.

3 Data Analysis

3.1 Internet Users' Perspective

Table 1 below shows that mobile phone penetration in African countries has not reached the 100% as reported by ITU for some African countries. The mobile phone technology continues nevertheless to scale rapidly with more than 50% of the African population owning a mobile phone. Migration to higher speed networks and smartphones continues apace, with mobile broadband connections set to reduce the historical digital divide. In four of the seven countries surveyed, more than 20% of respondents have used the Internet, in contrast to the 3 poorest of the countries surveyed, Mozambique, Rwanda, and Tanzania, which recorded 9,70%, 8,21%, and 13,53% respectively.

As the vast majority of people in all seven countries access the Internet through their mobile phone, the low Internet penetration in these countries can be attributed to low smartphone penetration which, except for Tanzania, is lower than 20% compared to South Africa (55,53%), Ghana (34,27%), Kenya (27,57%) and Nigeria (23,83%). Surprisingly, Tanzania's smartphone penetration is above 20% but Internet penetration remains lower. This could be attributed to supply-side issues such as data prices or the dearth of skills to enable Internet use.

In terms of barriers to internet use (Table 2 below), as expected, data cost has been reported as the main obstacle. For a number of respondents, internet is considered a time-consuming activity, as it seems that lack of time is another relevant barrier to internet use. In Tanzania (28.36%), South Africa (24.22%), and Mozambique (36.5%), instead, the users perceive that internet speed is not sufficient for a seamless internet access. Lack of content in local language, on the other hand, and in contrast to previous studies on local content in Africa, is not considered one of the main obstacles to internet use, except in Rwanda, where 8,49% of the respondents expressed some concern related to lack of content in local languages.

Table 1. Mobile phone, smartphone, and internet use across 7 African countries. (Source: #AfterAccess RIA surveys, 2017)

Country	Mobile phone (%)	Smartphone (%)	Internet use (%)
Ghana	73,87	34,27	26,00
Kenya	86,94	27,57	25,59
Mozambique	39,73	17,01	9,70
Nigeria	64,42	23,83	30,22
Rwanda	48,16	9,02	8,21
South Africa	83,84	55,53	49,72
Tanzania	58,52	22,12	13,53

Table 2. What does limit you to from using the internet? (Source: #AfterAccess RIA surveys, 2017)

	Ghana	Kenya	Mozambique	Nigeria	Rwanda	South Africa	Tanzania
No limitation	11,96	16,95		19,92	21,58		
Lack of time	21,78	20,16	11,59	15,65	18,13	10	25,62
Data cost	51,51	45,42	43,28	32,25	48,7	47,15	40,64
Lack of content in my language	**3,59**	**1,96**	**6,43**	**0,26**	**8,49**	**3,32**	**3,68**
Speed of internet	7,53	11,63	36,5	18,11	1,01	24,22	28,36
Privacy concern	0,47	0,51	2,44	2,98	2,08	3,18	0,89
Worried about getting virus/malware	0,6	0,74	8,85		9,97	3,77	0,85
Not allowed to use it (by family/spouse)	0,69	0,07	5,45	1,02	2,95	2,88	0,57
Find it difficult to use	1,87	0,95	5,36	1,65	5,99	2,23	4,14

3.2 Geolocation of African News Content Hosting

The hosting and geolocation analysis indicates that about 85% of the news websites are hosted outside the countries in which they belong, i.e. the website is owned and it is local to one country, but is hosted in another country. This is, hereafter, referred to as remote hosting. Analysis of remote hosted websites reveals that most of them are hosted in Europe and the US.

Figure 1 below shows a country-level distribution of locally hosted websites versus remotely hosted. Almost all the countries in the sample have less than 30% of their websites hosted locally, and about half of all the countries have less than 10% local hosting. South Africa (ZA) appears to have a high percentage of local hosting with 46%.

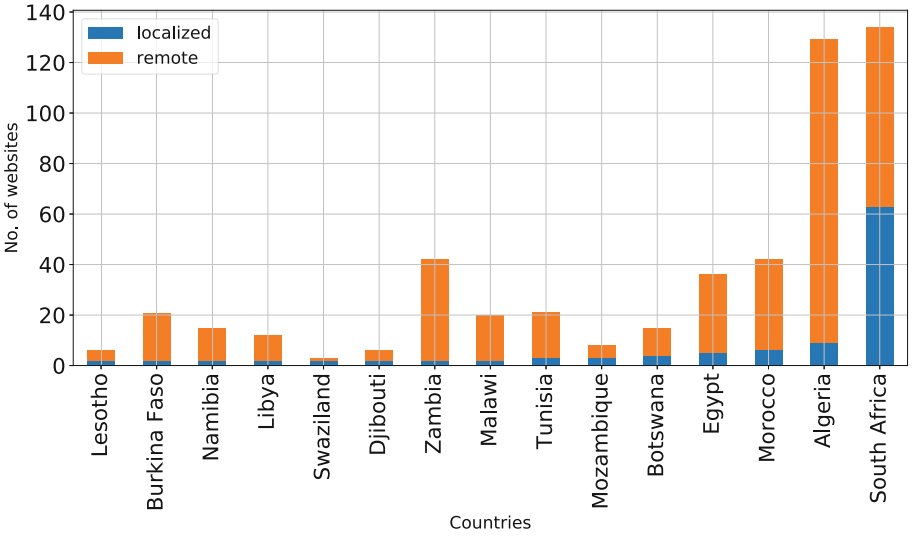

Fig. 1. Number of websites per country local vs remote

We observe that the US takes the lion's share in hosting African content, with about 58% of all the websites being hosted by American companies. Within Africa, South Africa leads in the content hosting business, hosting about 14% of all of Africa's remotely hosted news websites (i.e. minus those that belong to South Africa). The rest of the websites, about 20% are hosted in various countries in Europe (notably, 9% in France, 4% in Germany, and 3% in Great Britain). This signals low participation of the continent's companies in content hosting. Most of the websites that were observed to be hosted within Africa were based in South Africa, while the majority of the rest are hosted in either the US or Europe.

We also found that about 45% of all the IP hops (i.e. Internet path) for accessing African websites from African countries traversed outside African clients' home countries. Internet packets travel mostly through US networks, and about 23% pass through European networks. South Africa takes about 8% of all IP hops for traffic traversing to other African countries.

3.3 Network-Level Analysis of Africa's News Sites

Similar to the geolocation analysis, network-level analysis shows that most of the websites are hosted by foreign companies. Taking into consideration all the sampled African news websites, Cloudflare Inc. (US) takes the biggest share of the market, hosting about 22% of the websites. Following in the far distance is OVH SAS (France) with 8%, OPTINET (South Africa) at 6%, Google LLC and GoDaddy.com (both US) at 5% each, and Unified Layer (US) at 4%, and HETZNER (South Africa) at 3%.

With regards to the hosting market share, i.e. if only remotely hosted websites are considered, Cloudflare take an even bigger share of 26%, followed by OVH SAS (9%), Google LLC (6%), GoDaddy.com (5%) and Unified Layer (5%). What is interesting to note is that the leading providers for Africa's remote hosted news websites are largely based on Cloud infrastructure and make use of content distribution networks.

3.4 Delay Analysis (Round Trip Times) to Access Locally and Remotely Hosted Content

Analysis of round-trip times (RTTs) when accessing the websites from each of the countries shows significant RTT differences between locally hosted websites and those remotely hosted as shown in Fig. 3. The maps in Fig. 2 below highlights country-level RTT differences for local and remote hosting. The median RTTs for locally hosted websites is about 50 ms, whereas for remote hosted websites, the median RTTs range between 100 ms and 300 ms.

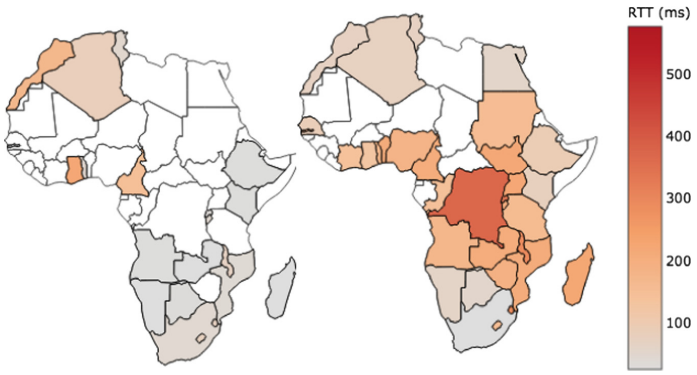

Fig. 2. National median RTT comparing locally (left) vs remotely (right) hosted content

The range of median RTTs shows significant differences between countries for websites hosted within each country. We found out that out of the 22 countries where local hosting performance was measured, 16 countries registered local median RTTs of less than 50 ms, while the other 6 countries (Angola, Malawi, Lesotho, Algeria, Cameroon, Morocco, and Ghana) had local median RTTs ranging between 50 ms and 200 ms. The analysis shows that in some countries, the median RTTs for locally hosted websites is higher than for websites that are remotely hosted. Examples include Ghana, where the median RTT for locally hosted websites was found to be 205 ms, whereas for remote websites, the median RTT was 127 ms; and Morocco, where the local median RTT was 152 ms against a remote median RTT of 68 ms. RTTs of over 100 ms for locally hosted websites

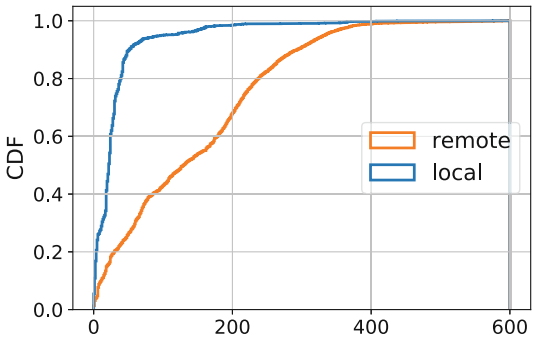

Fig. 3. RTT to access locally vs remotely hosted websites

could suggest circuitous path, where locally hosted websites are accessed through Internet paths that traverse other countries.

This further indicates a lack of peering, where interconnections between local network is done through networks in remote countries. In Ghana for example, one probe reached the locally hosted www.ghana.gov.gh by traversing remote networks, chronologically through Ghana, South Africa, UK, South Africa, and Ghana, resulting in a delay of 380 ms. Similarly in Morocco, a locally hosted website www.leseco.ma was reached by a probe in Morocco by traversing four networks, first in Morocco, then France, Ireland, Canada, and back to Morocco, experiencing a delay of 160 ms. These examples illustrate that in the absence of local peering, local hosting of websites tends to force much more circuitous routes for accessing the content than when the websites are in foreign countries, resulting higher delays and poor performance for consumers.

From a performance perspective, it is interesting to make some comparison between remote websites that are hosted by CDN-enabled networks (see Fig. 4). In a nutshell, CDN refers to groups of servers that are distributed in various geographic locations and work together to provide fast delivery of Internet content. The CDN takes content that is otherwise hosted on a single server, and replicates it to a set of distributed servers that are deemed to be closer to the intended consumers of such content. On the other hand, content that is not supported by CDN infrastructure generally remains within its original server location. This means that, while locally hosted content will be closer to the local consumers, CDN-based content can be brought closer to consumers even if the original servers are in distant locations. The expectation, therefore, is that level of delays for CDN-based websites should be similar to locally hosted websites. Of course, this assumes that the CDNs have nodes within or close to the respective countries. In this regard, the South African case is worth of a mention. Although the country has only about 46% of the news websites hosted locally, it can be seen the median RTTs for local and remote websites are almost the same; 22 ms and 25 ms respectively. As was mentioned earlier, the leading remote hosting providers for Africa (Cloudflare, OVH SAS, Google LLC, GoDaddy.com

and Unified Layer) operate on Cloud infrastructure and distribute their content via CDN services. It is also worth noting that South Africa hosts a number of CDN nodes, including Cloudflare, which has two; one in Cape Town and another node in Johannesburg[4]. This means that although the original web hosting is in remote countries, the actual website content is generally served from within the country. CDNs are helpful primarily to bring content hosted overseas closer to its users, and the increasing in local traffic might become an incentive for local ISPs to peer locally.

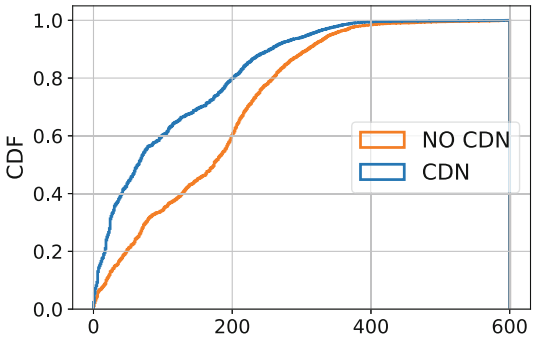

Fig. 4. RTT for websites using CDN vs NO CDN

4 Discussion

As opposed to previous studies on Internet adoption, this study does not find evidence of lack of content in local languages as one of the main obstacles to internet use, except in Rwanda. 85% of the African local news websites are hosted outside of their respective countries, mostly Europe and US, while the majority of the websites that were observed to be hosted within Africa were based in South Africa. Almost all the countries in the sample have less than 30% of their websites hosted locally, and about half of all the countries have less than 10% local hosting. 68% of all the Internet path for accessing African websites from African countries traversed outside Africa, mostly through US and European networks. A network-level analysis showed that most of the local African news websites are hosted by foreign companies.

The leading providers for Africa's remote hosted news websites are largely based on Cloud infrastructure and make use of content distribution networks to mirror the content locally. One direct consequence of remote hosting is that African network operators have to use significant levels of international bandwidth to fetch African content for their clients. The cost of this international bandwidth is in the end passed on to the Internet users in Africa.

Furthermore, geolocation of website hosting has significant implication on performance. The median RTTs for locally hosted websites is generally lower

[4] https://www.cdnplanet.com/geo/south-africa-cdn/.

than the RTTs for remote hosted websites. However, in some countries, the median RTTs for locally hosted websites is higher than for websites that are remotely hosted. High RTTs levels for locally hosted websites is an indicator of circuitous paths, which are due to lack of local ISP peering. Rather, in these cases interconnections between local networks is done through networks in remote countries. In the absence of local peering, local hosting of websites tends to resolve into circuitous routes for accessing the content than when the websites are in foreign countries, resulting in higher delays and therefore in poor performance for Internet users. By bringing remotely hosted content closer to the users, CDN-enabled networks reduce delays for CDN-based websites. Although level of delays for CDN-based websites are similar to locally hosted websites, CDNs do not improve the performance of locally hosted content in cases where there is lack of local peering. While access network infrastructure has improved in Africa, the content infrastructure has lagged behind. There is a considerable difference between access infrastructure and content infrastructure. While supply-side policy and regulatory interventions have positively affected access infrastructure (notably the roll out of mobile networks in the case of Africa), content infrastructure has not always been enabled by national policy and regulatory interventions resulting in most of the content in Africa being hosted and accessed from overseas. This configuration is not ideal, as it increases latency levels, and costs to access content. Not having a reliable hosting infrastructure affects also in-country delivery infrastructure. Without the necessary local peering, locally hosted content can exhibit equally higher access latencies.

5 Conclusion and Policy Implications

This study has provided empirical evidence on the current configuration of web content hosting, access, and distribution in Africa, and has demonstrated that the status of the African content infrastructure is alarming. All African countries heavily rely on foreign services, both to host, to access, and to distribute local content in Africa. Latency levels to remotely hosted local content are high as well as costs of accessing remotely hosted local content. Most of the public policy strategies on improving local content in Africa focused on demand-side interventions, such as the creation of content in local languages, and on developing skills on web content production and consumption. While these policies are important, bodies in charge of the governance of the internet are urged to identify ways of facilitating local markets for content hosting, access and distributions by focusing on: (1) incentivising investments on data centres and web farms in Africa, to stimulate economies of scale for the local web hosting market; (2) encouraging local news websites to move the content closer to the users in Africa, by incentivising the use of CDN-enabled networks and by reducing prices for local hosting, (3) Facilitating peering relationships between ISPs and investing in local exchange points to reduce latency; and (4) incentivising ISPs to peer in local exchange points.

References

1. Mauldin, A.: International Internet Capacity Growth Just Accelerated for the First Time Since 2015 (2018). https://blog.telegeography.com/international-internet-capacity-growth-just-accelerated-for-the-first-time-since-2015. Accessed 12 Nov 2018
2. Mothobi, O.: Botswana telecommunications limp a decade after policy changes (2017)
3. Mothobi, O.: South African data prices static for two years but consumers not flocking to cheapest product offering (2017)
4. Mothobi, O.: SADC not bridging digital divide, policy brief 6 (2017)
5. A4AI: 2017 Affordability Report: Regional Snapshots & Country-Specific Highlights. https://a4ai.org/2017-affordability-report-regional/. Accessed 15 Nov 2018
6. Amos, I.: Is the Lack of Local Content Hindering Internet Adoption in Afrika? (2016). https://www.iafrikan.com/2016/09/21/is-the-lack-of-local-content-hindering-internet-adoption-in-afrika-2/. Accessed 15 Nov 2018
7. Chetty, M., Sundaresan, S., Muckaden, S., Feamster, N., Calandro, E.: Measuring broadband performance in South Africa. In: Proceedings of the 4th Annual Symposium on Computing for Development, p. 1. ACM (2013)
8. Fanou, R., Francois, P., Aben, E.: On the diversity of interdomain routing in Africa. In: Mirkovic, J., Liu, Y. (eds.) PAM 2015. LNCS, vol. 8995, pp. 41–54. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-15509-8_4
9. Chavula, J., Phokeer, A., Formoso, A., Feamster, N.: Insight into Africa's country-level latencies. In: AFRICON 2017, pp. 938–944. IEEE (2017)
10. Formoso, A., Chavula, J., Phokeer, A., Sathiaseelan, A., Tyson, G.: Deep diving into Africa's inter-country latencies. In: IEEE INFOCOM 2018 - IEEE Conference on Computer Communications, pp. 2231–2239, April 2018
11. OECD/ISOC/UNESCO: The relationship between local content, internet development and access prices (217) (2013)
12. Song, S.: The Internet is U-Shaped (2017)
13. Data Cent. Dyn. Africa Tracks the Growth of the African Colo and Cloud Markets (2018). https://www.datacenterdynamics.com/news/dcdafrica-tracks-the-growth-of-the-african-colo-and-cloud-markets/. Accessed 15 Nov 2018
14. Nelson, V.: MainOne to Build Three New Data Centres Across West Africa (2014). https://thestack.com/data-centre/2018/09/06/mainone-to-build-three-new-data-centres-across-west-africa/. Accessed 15 Nov 2018
15. Bram, U.: Why Your Internet Connection is Slow Wherever You are in Africa (2015). https://qz.com/africa/472028/why-your-internet-connection-is-slow-wherever-you-are-in-africa/. Accessed 15 Nov 2018
16. Kende, M., Rose, K.: Promoting local content hosting to develop the internet ecosystem. ISOC report (2015)
17. Poese, I., Uhlig, S., Kaafar, M.A., Donnet, B., Gueye. B.: IP geolocation databases: unreliable? ACM SIGCOMM Comput. Commun. Rev. **41**(2), 53–56 (2011)
18. Shavitt, Y., Zilberman, N.: A geolocation databases study. IEEE J. Sel. Areas Commun. **29**(10), 2044–2056 (2011)