On the Internet Connectivity in Africa

Assane Gueye[1(✉)], Peter Mell[2], Desire Banse[2], and Faical Y. Congo[2]

[1] University of Maryland, College Park, USA
agueye@umd.edu
[2] National Institute of Standards and Technology, Gaithersburg, USA
{peter.mell,desire.banse,faical.congo}@nist.gov

Abstract. This study measures growth of Internet connectivity in Africa from 2010 to 2014 with a focus on inter-country relationships. An initial analysis reveals a modest increase in the number of participating countries but an explosive increase in the number of routers and network links. We then form the first country level topology maps of the African Internet and evaluate the robustness of the network. We study raw connectivity, pairwise shortest paths, and betweeness centrality, suggesting how improvements can be made to the inter-country African connectivity to enhance its robustness without reliance on paths traversing multiple continents.

Keywords: Africa · Internet · Connectivity · Measurement

1 Introduction

As recently as 2007, more than 70 % of internal African Internet traffic was routed to other continents (generally Europe) before reaching its final African destination [1]. This statistic suggests that internal African Internet connectivity was composed of non-communicating isolated clusters, despite the existence of fiber optic submarine cables circling the entire continent [2]. In this study we measure and document the growth of the African Internet with respect to connectivity from 2010 to 2014. We show how the African Internet is losing its fractured nature and is strengthening in its robustness to connectivity disruptions. We first focus our measurements on router to router connectivity and observe a consistent imbalance in the density of Internet infrastructures between different countries. Supporting this is a 2013 observation that 80 % of the hosts in Africa were in the country of South Africa [3]. To avoid biasing our analysis toward countries with this high density, we create a novel country-to-country connectivity map of Africa. With this approach, we evaluate the connectivity of individual countries to each other and thereby measure more uniform growth.

Our study is based on traceroute data provided by the Cooperative Association for Internet Data Analysis (CAIDA) [4]. The data contains router level topological maps of the Internet with embedded geolocation information obtained by recoding IP addresses of routers between sources chosen from a set of worldwide distributed 94 monitors and destinations chosen randomly within each/24 subnet in the IPv4 address space. With this approach, over time, each subnet is accessed from many different parts of the world,

R. Glitho et al. (Eds.): AFRICOMM 2015, LNICST 171, pp. 70–77, 2016.
DOI: 10.1007/978-3-319-43696-8_8

revealing the primary pathways through the Internet. Thus, we have confidence that we are discovering the major pathways through Africa. However, this approach has known limitations (only preferred paths can be discovered) and, because of that, our resulting connectivity analyses should be considered as worst case bounds. Using the geo-location data files provided by CAIDA, we label each router with its country. Unfortunately, this data is incomplete and a large fraction of routers miss the geolocation information. To circumvent this limitation, we use a majority vote heuristic to complete the geo-location information. For each router for which a country label is not available, we label it with the country label that is most common among its neighbors (using a random assignment whenever there is a tie). With this procedure, almost all nodes (except a tiny fraction ~0.0001 %) were assigned a geo-location. We lastly form a country level topology map from the router level topology map by merging nodes with identical country labels. Lastly, we represent the non-African continents as single nodes by merging all non-African country nodes into nodes representing their respective continents. The end result is communication interconnectivity graphs for Africa from 2010 to 2014 showing each country as a separate node and each non-African continent as a node (all multi-edges are removed). We then study country connectivity within Africa by evaluating raw connectivity, pairwise shortest paths, and betweenness centrality.

2 Data Analysis

Figure 1 shows the number of routers observed within Africa during the measurement period. The ‘A-Nodes’ line represents the number of routing nodes in Africa, the ‘AA-Links’ line the number of intra-Africa links, and the ‘AW-Links’ line the number of links between African routers and the rest of the world. The plots show a steady growth in the number of routers and links, indicating the growth in the overall infrastructure of the African Internet. From 2011 to 2014 there is growth factor of 35 in the number of routers observed in Africa. The number of observed links to other continents has also increased, but less significantly. The number of observed countries rises from 54 in 2010 to 57 in 2013 (all countries in the mainland). For the inter-country African links, we see a growth factor of 5 from 2010 to 2014, showing significant growth. However in 2014, these links accounted for only 0.3 % of the links where both routers reside within Africa. Thus, the number of these critical inter-country links is relatively small but growing rapidly. In 2014, the African routers represent about 1.6 % of the world’s routers (and 1.2 % of the world’s links). Africa’s share of worldwide routers and links modestly increases over the period of investigation while the fraction of world links that connect Africa to the other continents has stayed steady (not shown). In summary, the data indicates that most of the effort to improve Africa’s connectivity has been spent to connect nodes inside Africa. However, the countries with the greatest share of routing infrastructure have seen the most growth while the countries with the smallest share have experienced much smaller growth. Table 1 shows this disparity. The ‘top 3 countries’ (South Africa, Egypt, and Morocco) are those with the greatest number of nodes/links in 2014 while the ‘bottom 24 countries’ are those with the fewest nodes/links in 2014.

Similar observations were made in reference [3] for the period 2004–2007 showing that this is a long term trend.

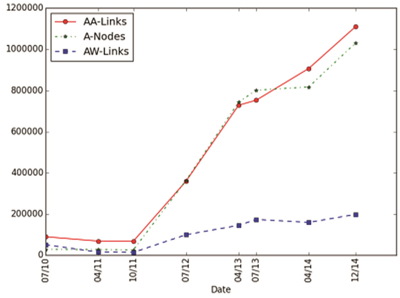

Fig. 1. Number of nodes and number links within Africa and between Africa and the rest of the world.

Table 1. Improvement in the number of links and nodes in Africa from 2010 to 2014.

	Continent	Top 3	Bottom 24	Rest of Africa
Nodes	35x	46x	6x	17x
Links	12x	15x	6x	7x

To avoid biases involving countries with more infrastructures we now evaluate the country level interconnectivity maps. Figure 2 shows the number of country-level links within Africa and between Africa and the rest of the world. Figures 3, 4, 5 and 6 show a graphic representation of the country-level connectivity for part of the investigation period. As was already seen at the router level, the number of country level links within Africa is increasing but it is always smaller than the number of country-level links to the rest of the world. This indicates that even though Africa is improving its connectivity at the country level, it largely depends on the other continents (or satellite) for Internet connectivity. In 2010 and 2011, a number of African countries (24 %) did not have any direct link to other African countries. This was mostly the case for inland countries whereas coastal countries are almost all directly connected to the Internet. In 2014, the

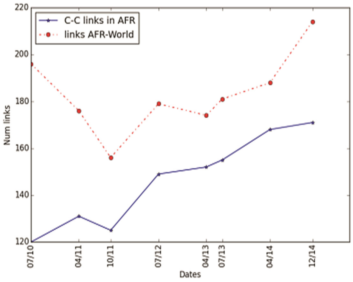

Fig. 2. Number of country-level links within Africa and between Africa and the rest of the world.

African country-level graph is connected (except for some island countries), implying that (*in principle*) any African country can now communicate with any other African country using only links within the continent.

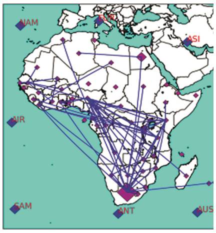

Fig. 3. 2010–07

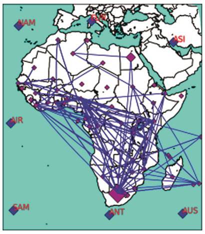

Fig. 4. 2012–07

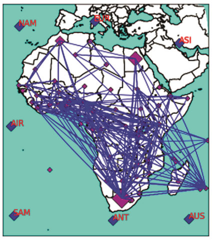

Fig. 5. 2013–07

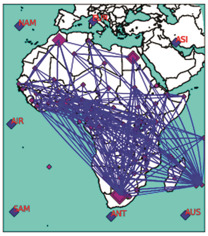

Fig. 6. 2014–12

We now study the characteristics of the country level connectivity graph. We first analyze the degree distribution of the country level map (for year 2014) and observe that most countries (always more than 25) have low degree, while a few countries (mostly South Africa) have high degree (greater than 20). The many low degree nodes and few high degree nodes is a characteristic of many engineered networks [5] and is usually referred as '*scale-free*'. Furthermore, we observed that nodes with low degrees tend to connect with nodes with high degree, and vice versa. This is another previously observed property of engineered networks and is referred as '*dissassortativity*' [6]. Another metric of interest is the '*betweenness*' centrality of individual countries, the fraction of shortest paths that go through that country (among pairs of African countries). By analyzing this metric, we observe that the majority of countries have very low betweenness centrality (carrying very little "relay" traffic). Only a small number of countries have a significant betweenness centrality. This indicates that if inter-country traffic is routed using shortest path, only a few African countries will play a role of big hubs, while most countries will carry only traffic which they have generated or which is destined to them. Also, the average number of hops between any pair of countries is always greater than 1 during the measurement period (in contrast, it is 0.5 for Europe). This means that, in average, communication between any pair of African countries has to transit through a third African country (while, in average, countries in Europe have direct link to each other).

We next study the robustness of the country-level Africa graph to node and link failures. There exist several graph theoretic metrics to quantify the importance of a node (or link) in a network: (a) the degree of a country (*Deg(c)*), (b) the betweenness of a country (*Bet(c)*) which quantifies how often a country lies in the shortest path between other countries that represent the source and destination of a communication, and (c) the eigenvector centrality of the country (*Eig(c)*) which measures how connected the country is to well-connected countries (the higher it is, the more connected the country is to well-connected countries). Since all these features are important in the connectivity of a given country, we combine them with the formula $Conn(c) = (Deg(c) + Bet(c))\mathrm{e}^{Eig(\mathrm{c})}$ to define the connectivity of a country. We define the connectivity of the continent as the sum of the connectivity of its countries. We then use this metric to study the robustness of the Africa network to node and link failures by asking the following question: what is the maximum drop in connectivity when 1 (2, 3, 4, …) node (resp. link) of the network fails? Figure 7 shows the evolution of the continent connectivity when we allow up to 5 countries to fail. For each of the number of nodes (allowed) to fail, we sequentially

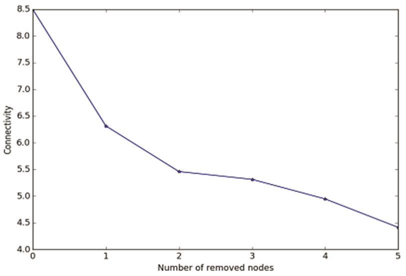

Fig. 7. Robustness to node failures.

remove the node with largest drop in connectivity. We observe a very sharp decline for the first two countries and then the connectivity decreases at a slower rate. This suggests that the first two countries are very important for the connectivity of the continent as a whole. Similar analysis for link failures shows an almost linear drop in connectivity as more links fail (not shown).

Finally, we propose an improvement of Africa's connectivity by adding new links to the country-level graph. For each additional link, we ask the question: where in the continent (i.e., between which pairs of countries) to put the link to obtain the maximum increase in connectivity. We assume that there are 19 links to be added sequentially. For each new link, we use the connectivity metric defined above to compute its best placement (i.e., which has a maximum increase in connectivity). Figure 8 shows the improvement in the continent connectivity after the addition of each optimally placed link. We can see that the connectivity improves as more links are being added. However, the curve has a few plateaus (between 5 and 8 additional links, between 10 and 12, and after 15) between which its increases. One interpretation is that after having added 5 (resp. 10) links in the continent, one does not gain much by adding more links except if one can add at least 3 (2) additional nodes. This pattern of flat region-increasing region repeats even beyond 20 links (not shown). We also see that connecting nodes that already have many connections results in a smaller payoff while connecting nodes with small number of connections results in large payoff. In other words, in order to improve Africa's connectivity, we need to build links between countries with less Internet infrastructure.

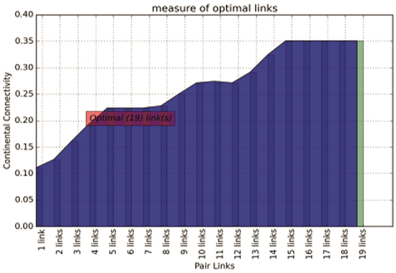

Fig. 8. Improving by adding links.

3 Related Work

Several studies are available showing African Internet accessibility (see [2, 7]). The African Economic Outlook [7] provides and analyzes data on telecommunication investments, access to information technology, technology penetration, and connectivity in Africa. They have reported 16 undersea cables connecting Africa to the Americas, Europe and Asia (e.g., SAT3/WASC, EASSy, Seacom, TEAMS, RCIP, GL01, MaIN, ACE). With the exception of Eritrea and Western Sahara, all coastline countries have a cable landing on their shore. This has helped triple the Africa Internet access in the last decade. On the other hand, landlocked countries were (up to 2010) mainly connected via satellite (VSAT). However, our data shows that this is changing with inland countries connecting to the undersea cable via fiber optic cables traversing neighboring coastline countries. Most of the observations in [7] are also made in this paper, although it is based on (different) topological data set. Our paper, however, goes one step further by introducing a novel country-level connectivity graph of Africa and studies its properties. We also investigate improvement to the current country-level connectivity. [8] provides an analysis of the distribution of Internet infrastructure in Africa for the years prior to our study (with some of the same trends being found with respect to a few countries dominating African network growth). The ping end-to-end reporting (PingER) project is, like CAIDA, an Internet End-to-end Performance Measurement (IEPM) project that monitors end-to-end performance of Internet links [9] using the simple and common "ping" test. [10] uses PingER to shown the low presence of Africa in the world Internet, a steady improvement of Africa connectivity since late 2010, and the disparity in improvement among the African countries. Our study, although based on a different data set, confirms such observation. Their paper, however, does not study country-level connectivity nor does it consider the robustness and improvement analyses carried in our study.

4 Conclusion

This study has shown that Africa's Internet connectivity has significantly improved during the period from 2010 to 2014. Both the fraction of worldwide Internet backbone routers attributable to Africa and the number of intra-Africa links have risen substantially. This is important given that much of the African inter-country connectivity had been previously routed through other continents. On a downside, we note that most of the router growth occurred in African countries that already had a robust infrastructure. That said, the countries with less infrastructures also generally experienced Internet infrastructure growth. The analysis of a novel country to country connectivity map has shown an increasing participation of African countries over the time period of study. We also see a significant increase in the number of direct links between African countries. Even with these added links, however, the connectivity is still not as robust as Europe where the average hop length is much less. With respect to the rest of the world, we see an increase in links from African to other continents. At a deeper level, our connectivity metrics also reveal a highly *scale-free* nature in the country to country Africa connectivity graph. Most countries have low degree and a few have high degree. There is a *negative assortativity* whereby low degree nodes tend to connect to high degree nodes and not to each other. The problem with this current architecture is that it is susceptible to node failure. The majority of African countries are dependent upon just a few other African countries for their intra-continental Internet access. However, judicious placement of additional links (i.e., links among the low degree nodes) can reduce the fragility induced by the scale free nature. This translates in the need for direct Internet links between countries with less Internet infrastructure in order to make that African Internet stronger as a whole.

Acknowledgements. This research was sponsored by the National Institute of Standards and Technology (NIST) and partially accomplished under NIST Cooperative Agreement No. 70NANB13H012 with the University of Maryland. The views and conclusions contained in this document should not be interpreted as representing the official policies, either expressed or implied of NIST or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes, notwithstanding any copyright notation hereon.

References

1. Africa Waiting for Net Revolution: BBC News. http://news.bbc.co.uk/2/hi/technology/7063682.stm. Accessed 13 July 2015
2. ITU: Telecommunication/ICT Markets and Trends in Africa (2007). http://www.itu.int/ITU-D/ict/statistics/material/af_report07.pdf. Accessed 13 July 2015
3. Livraghi, G.: Data on Internet Activity in Africa (Hostcount). http://www.gandalf.it/data/africeng.htm. Accessed 13 July 2015
4. CAIDA: Macroscopic Internet Topology Data Kit. http://www.caida.org/data/internet-topology-data-kit/. Accessed 13 July 2015
5. Barabási, A.-L., Albert, R.: Emergence of scaling in random networks. Science **286**, 509–512 (1999)
6. Newman, M.E.J.: Mixing patterns in networks. Phy. Rev. **67**(2), 026126 (2003)
7. African Economic Outlook: Technology Infrastructure and Services in Africa. http://www.africaneconomicoutlook.org/en/theme/ict-africa/technology-infrastructure-and-services-in-africa/
8. Data on Internet Activity in Africa (Hostcount). http://www.gandalf.it/data/africeng.htm. Accessed 13 July 2015
9. Ping End-to-End Reporting (PingER). Project http://www-iepm.slac.stanford.edu/pinger/
10. Les Cottrell, R.: How bad is Africa’s Internet? IEEE Spectrum. http://spectrum.ieee.org/telecom/internet/how-bad-is-africas-internet