This was a pleasure to read (even as a "drafty draft")! Looking forward to final paper

-BJ

The MIRA Project: A Study of Internet Resiliency among Countries in Africa

Blaise Viateur Niyigena, Esther Kamau, Isaac Manzi, Pasqua Ruggiero

May 11th, 2021

1 Abstract

The heightened importance of the Internet has indicated the need for a solution to determine which countries are in higher need of more secure Internet infrastructure to prevent significant damages from power outages. The nonprofit organization AFRINIC has created the MIRA project to construct a methodology to measure Internet resiliency for countries in Africa and to thus be able to provide recommendations to achieve higher resiliency. This framework was constructed by taking weighted sums of selected metrics (pertinent to measuring Internet resiliency) and using this model as the country's Internet resiliency score. These results are displayed in a dashboard to facilitate the comparison of scores between countries. Calculating the Internet Resiliency scores for each country revealed that (insert results). From the results, it can be inferred that (insert discussion points)

good way of pointing out what needs to be done. thanks!

2 Introduction

The security and reliability of Internet connectivity is of utmost priority to many today, especially since the COVID-19 pandemic has revealed the importance of stayed connected to each other during a time of limited physical social interactions (Deloitte, 2020). Though there has been a large emphasis on the strengthening of Internet during these times, the distribution of these measures has not been equal across all countries.

In particular, many low-income countries do not get to benefit from as stable of Internet connectivity as many others in the world due to facing various issues such as under-provisioned networks, lack of proper cable infrastructure, or even redundant interconnection systems. These areas are more prone to widespread and high impacting internet outages when their infrastructure is compromised, in the case of a cable break or a power failure. The results from the power outages are incredibly detrimental and impact the entire Internet ecosystem, leading to revenue loss to their digital economy(AFRINIC, 2021). The impact of the Internet outages could be drastically decreased if these countries had the

(insert space)

ability to thoroughly audit their Internet Infrastructure and were to implement best practices for building resilient Internet infrastructures.

The prevalence of discrepancy between Internet resiliency among African countries and the desire to provide Internet support to low-income countries is the driving factor that lead to AFRINIC, a nonprofit Internet registry and research organization, to start the MIRA (Measuring Internet Resilience in Africa) project. The MIRA project is the result of a cross-collaborative effort between AFRINIC Carnegie Mellon University's Pittsburgh and Africa campuses.

The aim of the MIRA project is to create a framework that can evaluate a country's capability to provide reliable means of Internet connectivity and to be able to do so in times of crises(AFRINIC, 2021). The framework is to be created by creating an index composed of various network resiliency metrics. This framework, which will be referred to as the Internet Resiliency Index, will be used to assign a score to each country that reflects the resilience of their physical and logical Internet infrastructure. By doing so, recommendations can be provided to help networks and countries achieve higher resilience(AFRINIC, 2021).

3 Data

The primary types of data to be used for this analysis is recent data collected for the various resilience metrics that have been determined to be most important for creating the resilience index (this is explained further in the Methods section). Most data was is obtained through open Internet data sources, though data that was not readily available was acquired through either requests to private organizations for their internal data, or by utilizing AFRINIC's internet probes as a primary source for gathering this information. Table 1 dictates the different metrics that were selected to be included in the Internet resiliency index. As indicated below, there were 15 metrics that were selected to be included in this framework.

Category	Metric	Measurement
QoS	Throughput	Fluctuation in throughput
	Latency	Latency to local services (ms)
	IPv6 capability	IPv6 capability of the ISP network (count)
Security	Routing	% of prefixes covered by IRR object
	AS hegemony	Compute the AS dependency of network
	MANRS Score	The consolidated MANRS score
	DDos Potential	Level of risks posed to other countries
	Spam Infection	%
Infrastructure	IXPs per 10M	Number of IXPs per 10M inhabitants
	IXP efficiency	% of ASes present at the IXP
	Upstream	Number of upstream providers
	Cable landing stations	Number of cable landing stations per capita/km2
	reach	% of population within 10-Km reach
	degree distribution	Degree distribution of cable entering/leaving a country/city
Affordability	Affordability	How affordable is Internet services in this country ($)

this could stand to be a little bigger (maybe 150-200%)

Table 1: Table of metrics selected to be incorporated for calculating a country's Internet Resiliency Score.

leave some space here so table caption doesn't blend into text.

(are you using the \begin{table} \end{table} environment?)

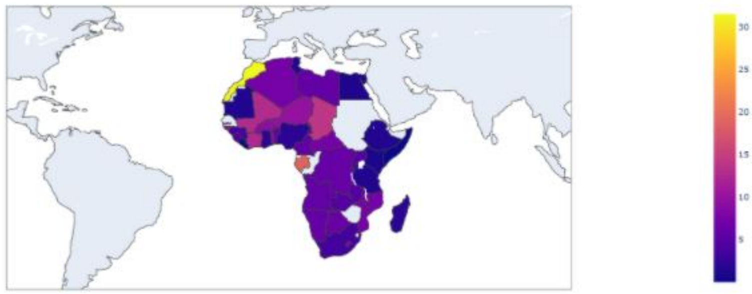
Before these datasets could be incorporated into our framework, various levels of preprocessing were carried out to ensure that the data were of high quality, tidy, cleaned, standardized, and sufficiently representative of countries in Africa. The ideal cleaned datasets would consist of 57 rows - one row per country of interest as well as the subsequent Internet measures. A more detailed explanation of the preprocessing steps are included in Technical Appendix 1.

great

Another step that was taken before creating the aggregate index was to perform an initial exploratory data analysis to determine if there were any potential patterns regarding country representation for all datasets, as well as visualize the metrics' measurements across Africa. This was done by calculating the correlations between metric measurements. Additionally, geographic maps were created, which served to be an important and useful tool for comparing individual metric measurements between African countries.

For the data to be aggregated, the raw values for each indicator were converted to an equivalent scale and unit using a min/max normalization for quantitative measurements.

Figure 2 indicates one example of the geographic maps. This particular map represents the measurements for cheapest prepaid mobile plan (Affordability). In this map, there are various trends prevalent, such as the most expensive plans being in Western and North-Western Africa, while the cheapest plans mostly lie along the Eastern coastline . More details on the implementation of these analyses, as well as descriptions and results of further exploratory analyses can be found in Technical Appendix 2. great



could stand to be a bit larger

use {table} environment in LaTeX

Figure 2: Choropleth map for Affordability created using Python's Plotly library.

4 Methods

The best method to measure the Internet resiliency of African countries was determined to be the creation an index composed of various aggregated network resiliency metrics. Such aggregation will enable us to derived summary values that can quickly and intuitively give indications of network resilience.

The first step towards creating this tool consisted of extensive research on various Internet security, quality of service, infrastructure and affordability metrics that would be most plausible to incorporate into the Internet Resiliency Index. This procedure required not only formal research on various Internet metrics, but also included research into other Internet security indexes that have been created and previous suggestions from AFRINIC personnel(AFRINIC). (space) Additionally, it was crucial to ask for the opinion of subject matter experts in the selection of these various metrics. This primarily took the form of structured discussions between the researchers and the experienced personnel at AFRINIC.

There were various criteria that were considered when choosing the most relevant metrics of measurement to include in the Internet Resiliency Index. There are many types of Internet measurement metrics that could be useful, but considering certain criteria before selecting the metrics helped in selecting only the highest quality forms of measurements were used. At the end, there were fifteen metrics selected to be included. The qualities of consideration were as follows. A comprehensive list of the various metrics considered can be found in Technical Appendix 3.

this sentence would go better after the list of six qualities

1. Measurements that are easily attainable, whether it be through open source data or through attainment of internal data. Though the theoretical foundation of the Internet Resilience Index is very important, there is a need for real data to be able to test and display the scores produced by this framework.

2. Given time restraints, the data must include the corresponding countries of measurements, or at the very least be able to be easily joined with other datasets of country descriptions to map one metric measure per country. As the goal of this framework is to eventually prescribe recommendations for improving

Internet infrastructure and security for low-income countries, it is important to have a basis for how the countries score on each metric.

3. The metrics must have data available that is recent (within the last two years). This way, the most accurate scores can be calculated per country. This also provides incentive to recollect/reattain more updated data in future years when the metric scores are updated.

4. The metric must have data available that has sufficient coverage of African countries. Few metrics that were initially considered had less than fifty percent coverage of African countries in the available data. Incorporating this much missing data would not be ideal and would make the process of comparing metrics more sparse and thus more difficult.

5. The metrics must not be duplicates of each other. This is to avoid double-penalizing countries that may rank low on certain metrics(that have similar measures incorporated into the framework) and to give space for the inclusion of more insightful metrics to be incorporated instead.

6. Most importantly, the metrics are indicative of a certain aspect of a country's Internet resiliency.

After selecting the most relevant, available and representative metrics to create the framework, the metrics were to be grouped into various categories based on their type of measurement. This was used to determine the levels of weights that are necessary to consider in creating the Internet Resiliency Index. This selection of categories was primarily influenced by the introductory MIRA paper(AFRINIC, 2021), which highlighted important categories of consideration. The four final categories that were selected were Quality of Service, Security, Infrastructure and Affordability. (space)

After structuring the metrics into categories, one of the final steps in the creation of the theoretical framework itself was to assign weights to each category and metric. There were various methodologies and considerations that were taken into account when selecting the weight for each metric. The main methodology that was used for weight assignment was the ad-hoc weighting scheme explained in "The Inclusive Internet Index 2020 Methodology Report"(The Economist Intelligence Unit, 2021). Their "Inclusive Internet" aggregate index had weights assigned based on the Internet life cycle, which is defined as a ranking of the most important aspects of Internet. Their ranking followed a 40-30-20-10 rule, with the categories being Internet Availability, Internet Affordability, Internet Accessibility and Internet Readiness (The Economist Intelligence Unit, 2021). Though these were not the categories that were included in the Internet Resiliency framework, the idea of the availability of Internet services being more important than it's affordability (after all, the question of affordability can not be even considered if services are not available) was used to assign weights to each category. Assigning weights to the individual metrics in each category was more difficult, but it followed a similar process to that of selecting the metrics to be included in the Internet Resiliency Index. These metrics were ranked among others in their category based on data availability and coverage, as well as it's importance to describing each category.

Upon assignment and validation of weights, the aggregated metric is mapped to numerical scores with the formula indicated in Figure 3. This would be calculated for each African country and would be its Internet resiliency score. Additionally, each score was translated into a qualitative representation (such as low, medium, high, and critical). The qualitative representation is to be used to help organizations properly assess and prioritize their decision processes.

use the \begin{figure} \end{figure} environment

$$Y = w_{c_1} * (w_{m_1} * m_1 + \ldots) + w_{c_2} * (w_{m_1} * m_1 + \ldots) + ...$$

Figure 3: Formula derived to calculate Internet Resiliency Scores. The variable c represent category, while m represents the individual metric.

you don't need to do this, but a more precise indexing would be m_{c_1 1}, m_{c_1 2}, etc. for the first category, and m_{c_2 1}, m_{c_2 2} etc. for the second category, and so forth, rather than the generic m_1, m_2, etc.

To facilitate the comparison of scores between countries, the team decided that the best way to visualize the Internet resiliency score was through the creation of a dashboard. Various frameworks and packages were considered such as RShiny, D3, Python Dash and Highcharts, while factors such as setup requirements, data loading and handling, interactivity, and customization helped determine which visualization tool was the final selection. In the end, after thoroughly analyzing the benefits and drawbacks of each software, Python Dash was selected as the best platform to visualize these results, as it was the platform that best could be incorporated with our analyses, which were also implemented in Python.

The final aspect of the creation of the Internet Resiliency framework to be considered was the potential recalibration and updating of model parameters and data. Thus, a database was created using Apache Superset to facilitate the process of acquiring new data and updating the model (include more details about database once completed)

5 Results

After thorough consideration of each metric, category and necessary weights, the metrics were grouped and assigned various levels of weights to be used to calculate the Internet Resiliency score. Table 2 below shows the final metrics that were selected, the category they were assigned to, and the weights per category and individual metrics.

Category	Metric	Proposed Individual Metric Weight According to Internet Lifecycle	Proposed Individual Metric Weight	Measurement
QoS - 25%	Throughput	Availability - Quality (very important)	33.30%	Fluctuation in throughput
	Latency	Availability - Quality (very important)	33.30%	Latency to local services (ms)
	IPv6 capability	Availability - Quality (very important)	33.30%	IPv6 capability of the ISP network (count)
Security - 25%	Routing	Availability - Quality (very important)	20%	% of prefixes covered by IRR object
	AS hegemony	Availability - Quality (very important)	20%	Compute the AS dependency of network
	MANRS Score	Availability - Quality (very important)	20%	The consolidated MANRS score
	DDos Potential	Availability - Quality (very important)	20%	Level of risks posed to other countries
	Spam Infection	Availability - Quality (very important)	20%	%
Infrastructure - 35%	IXPs per 10M	Availability - Infrastructure (very important)	12.50%	Number of IXPs per 10M inhabitants
	IXP efficiency	Availability - Infrastructure (very important)	12.50%	% of ASes present at the IXP
	Upstream	Availability - Quality (very important)	25%	Number of upstream providers
	Cable landing stations	Availability - Infrastructure (very important)	12.50%	Number of cable landing stations per capita/km2
	reach	Availability - Infrastructure (very important)	25%	% of population within 10-Km reach
	degree distribution	Availability - Infrastructure (very important)	12.50%	Degree distribution of cable entering/leaving a country/
Affordability - 15%	Affordability	Affordability - Price	100%	How affordable is Internet services in this country ($)

again, this table could stand to be 150-200% bigger.

use {table} environment

Table 2: Final weights and metrics selected to be used for calculating the Internet Resiliency Score.

(Include description of final dashboard functionalities - user selection of weights per category and metric, incorporation and display of Internet resiliency scores, qualitative assignment of scores to display and deployment to AFRINIC site)

(Include screenshot of scores, both quantitative and qualitative, as displayed on the dashboard)

(Explain final format of database and functionality)

cool. looking forward to it!

6 Discussion

(Discuss implications of various weights and what effects these can have on considering Internet Resiliency scores per country)

(Discuss trends in Internet resiliency scores from the dashboard and what they tell us about trends in Africa)

(Discuss functionality of dashboard and implications of having this tool, as well as shortcomings)

(Include implications of database creation and any shortcomings that need to be addressed in the future)

Current future steps to consider (to be formalized and extended once deliverables are completed):

1. Utilization of other frameworks to create metrics.

good start; looking forward to it!

If possible it would be nice if final paper also contains link to the dashboard.

2. Updating Subject Expert Harnessing tool to receive feedback from Subject Matter Experts and to help validate final model.

3. Incorporation of other metrics and considering metrics not necessarily directly related to the Internet like were used in the The Economist's Intelligence Unit report(The Economist Intelligence Unit, 2020) - such as literacy rate, gender politics, etc.

4. Collection of unavailable data per selected metric using Internet Probes.

5. Expanding the usage of this tool to countries in other continents.

6. Discussion of further study of Internet Measurement Probes and benefits, other measurements that can be incorporated.

7 References

these are sensibel references. please adjust format to ASA format.

AFRINIC (2021), *Measuring Internet Resilience in Africa (MIRA)*. Africa.

AFRINIC, *State of Internet Measurement in Africa - A Survey*. Africa.

Deloitte (2020), *Connectivity Resilience Amidst COVID-19*. Africa.

The Economist Intelligence Unit (2020), *The Inclusive Internet Index 2020 Methodology Report*.

Gotesman, Ryan (2018), *An Introduction to Geographical Data Visualization*. towardsdatascience.com.

(Include all sources of data)

8 Technical Appendix

great start

8.1 Data Extraction

Procurement of AS Hegemony data through API calls for each African country from December 1st, 2020 to April 19th, 2021. To be completed:

```
A2_Africa = AfricaISOcodes['A2']

AS_score_dict = dict()
for x in range(0,len(A2_Africa)):
    AS_score_dict[A2_Africa[x]] = x

AS_score = 0
for country in A2_Africa:
    AS_score=0
    for month in ['12', '1', '2', '3', '4']:
        for day in range(1,32):
            if day < 10:
                day = '0'+str(day)
            else:
                day=str(day)
            if month == '12':
```

```
            url = 'https://ihr.iijlab.net/ihr/api/hegemony/countries/
            ?country=' + country + '&timebin=2020-' +  month + '-' +
            day + 'T00:00:00Z'
            resp = requests.get(url)
            if (resp.ok):
                try:
                    data = resp.json()
                    for i in range(0, len(data['results'])):
                        AS_score +=
                        data['results'][i]['hege']*data['results'][i]['weight']
                except:
                    print(month, day, "failed")
        else:
            url = 'https://ihr.iijlab.net/ihr/api/hegemony/countries/
            ?country='  + country + '&timebin=2021-' +

            month + '-' + day + 'T00:00:00Z'
            resp = requests.get(url)
            if (resp.ok):
                try:
                    data = resp.json()
                    for i in range(0, len(data['results'])):
                        AS_score +=
                        data['results'][i]['hege']*data['results'][i]['weight']
                except:
                    print(month, day, "failed")
    AS_score_dict[country] = AS_score
    print(country, AS_score_dict[country])
```

8.2 Data Processing and Cleaning

Percent Population within 10-KM Reach data cleaning and exploration. To be completed:

```
#read in data and select rows and columns of relevance
capacity_data_pop = pd.read_excel("ITU Broadband Capacity Indicators
2019.xlsx",skiprows=1)
capacity_data_pop = capacity_data_pop[capacity_data_pop.columns[0:9]]
capacity_data_pop.columns = capacity_data_pop.iloc[0]
capacity_data_pop = capacity_data_pop.iloc[2:195]
capacity_data_pop = capacity_data_pop[['Country name', 'IsoCode',
'10-km Range']]

#merge with ISO3 Codes
capacity_data_pop_cleaned = pd.merge(capacity_data_pop, AfricaISOcodes,
```

```
left_on ='IsoCode',
right_on='A3', how='inner')
capacity_data_pop_cleaned = capacity_data_pop_cleaned.drop(['Country',
'A2', 'IsoCode',
'Region'], axis=1)

capacity_data_pop_cleaned['10-km Range'] =
capacity_data_pop_cleaned['10-km Range'].astype(float)

capacity_data_pop_cleaned.describe()

capacity_data_pop_cleaned.loc[capacity_data_pop_cleaned["10-km Range"]
== 0]

#create pandas report
profile = ProfileReport(capacity_data_pop,
title='capacity_data_pop Profiling Report', explorative=True)
profile.to_file("C:/Users/pas/documents/github/MIRA_AFRINIC/
Documentation/Data Exploration Reports/capacity_data_pop.html")

data = dict (
    type = 'choropleth',
    locations = capacity_data_pop_cleaned["A3"],
    locationmode='ISO-3',
    z=capacity_data_pop_cleaned["10-km Range"])
map = go.Figure(data=[data])
py.offline.plot(map)

capacity_data_pop_cleaned["standardized 10-km Range"]  =
(capacity_data_pop_cleaned['10-km Range']-capacity_data_pop_cleaned
['10-km Range'].min())/(capacity_data_pop_cleaned['10-km Range'].max()-
capacity_data_pop_cleaned['10-km Range'].min())

capacity_data_pop_cleaned.to_pickle("C:/users/pas/documents/github/
mira_afrinic/data/
pickle_files/10 KM Reach.pkl")

fig = px.line(capacity_data_pop_cleaned, x='Country name', y='10-km Range')
fig.update_xaxes(showgrid=False)
fig.update_yaxes(showgrid=False)
fig.show()

fig = px.line(capacity_data_pop_cleaned, x='Country name',
y='standardized 10-km Range')
fig.update_xaxes(showgrid=False)
fig.update_yaxes(showgrid=False)
```

```
fig.show()
```

bigger please so a person can read!

Correlation code:

```
# calculate the correlation matrix
corr = merged_1.corr()

# plot the heatmap
sns.heatmap(corr,
        xticklabels=corr.columns,
        yticklabels=corr.columns)
```

8.3 Metric Selection

(Include more information about how metrics were selected)

8.4 Metric Aggregation

(Include more information about how weights were assigned)

8.5 Dashboard Creation

Code for creating dashboard - to be updated.

```
reach_data = pd.read_csv("10 KM Reach Cleaned.csv")
affordability_data = pd.read_csv("Affordability Cleaned.csv")
ddos_data = pd.read_csv("ddos Cleaned.csv")
degree_distribution_data = pd.read_csv("Degree Distribution Cleaned.csv")
ipv6_data = pd.read_csv("IPv6 Cleaned.csv")
irr_data = pd.read_csv("IRR Cleaned.csv")
```

```
landing_stations_data = pd.read_csv("landing stations Cleaned.csv")
spam_data = pd.read_csv("Spam Cleaned.csv")

os.chdir("c:/users/pas/documents/github/mira_afrinic/data/raw data")
#import country coordinates - obtained from https://datahub.io/core/
geo-countries#resource-countries
with open('countries.geojson') as f:
    data_json = json.load(f)

#merge data for now
merged = pd.merge(reach_data, affordability_data)
merged = pd.merge(merged, ddos_data)
merged = pd.merge(merged, degree_distribution_data)
merged = pd.merge(merged, ipv6_data)
merged = pd.merge(merged, irr_data)
merged = pd.merge(merged, landing_stations_data)
merged = pd.merge(merged, spam_data)

countries = merged["ISO_A3"].unique()
metrics = ['standardized 10-km Range', 'standardized affordability',
       'standardized amplified count', 'standardized links per node',
       'standardized ipv6 counts', 'standardized irr',
       'standardized landing stations', 'standardized spam']

 app = dash.Dash(__name__)

 #add individual slider labels
#add hover values
app.layout = html.Div([
    html.P("Weights:"),
    html.Label(id='reach-weight'),
    dcc.Slider(
        id='reach_weight',
        min=0.1,
        max=1.0,
        step=0.05,
        marks={i: '{}'.format(float(i)) for i in range(5)},
        value=0.1,
    ),
    html.Label(id='affordability-weight'),
    dcc.Slider(
        id='affordability_weight',
        min=0.1,
        max=1.0,
        step=0.05,
        marks={i: '{}'.format(float(i)) for i in range(5)},
```

```
    value=0.1,
),
html.Label(id='ddos-weight'),
dcc.Slider(
    id='ddos_weight',
    min=0.1,
    max=1.0,
    step=0.05,
    marks={i: '{}'.format(float(i)) for i in range(5)},
    value=0.1,
),
html.Label(id='links-weight'),
dcc.Slider(
    id='links_weight',
    min=0.1,
    max=1.0,
    step=0.05,
    marks={i: '{}'.format(float(i)) for i in range(5)},
    value=0.1,
),
html.Label(id='ipv6-weight'),
dcc.Slider(
    id='ipv6_weight',
    min=0.1,
    max=1.0,
    step=0.05,
    marks={i: '{}'.format(float(i)) for i in range(5)},
    value=0.1,
),
html.Label(id='irr-weight'),
dcc.Slider(
    id='irr_weight',
    min=0.1,
    max=1.0,
    step=0.05,
    marks={i: '{}'.format(float(i)) for i in range(5)},
    value=0.1,
),
html.Label(id='landingstations-weight'),
dcc.Slider(
    id='landingstations_weight',
    min=0.1,
    max=1.0,
    step=0.05,
    marks={i: '{}'.format(float(i)) for i in range(5)},
    value=0.1,
```

```
    ),
    html.Label(id='spam-weight'),
    dcc.Slider(
        id='spam_weight',
        min=0.1,
        max=1.0,
        step=0.05,
        marks={i: '{}'.format(float(i)) for i in range(5)},
        value=0.1,
    ),
    dcc.Graph(id="choropleth"),
])

@app.callback(
    Output("choropleth", "figure"),
    [Input("reach_weight", "value"),
    Input("affordability_weight", "value"),
    Input("ddos_weight", "value"),
    Input("links_weight", "value"),
    Input("ipv6_weight", "value"),
    Input("irr_weight", "value"),
    Input("landingstations_weight", "value"),
    Input("spam_weight", "value"),])

def display_choropleth(reach_weight, affordability_weight, ddos_weight,
                       links_weight, ipv6_weight, irr_weight, landingstations_weight,
                       spam_weight):
    data = dict (
        type = 'choropleth',
        locations = merged["ISO_A3"],
        locationmode='ISO-3',
        z=merged["standardized spam"])
    map = go.Figure(data=[data])

    return map

 app.run_server(debug=False)
```

8.6 Database Creation

(Include additional details of database creation here)