Mapping the Maze: A Study of Internet shutdowns across the world

Ritik Malik
2018406

BTP report submitted in partial fulfillment of the requirements
for the Degree of B.Tech. in Computer Science & Biosciences

on June 1, 2021

BTP Track: Research

BTP Advisors
Dr. Sambuddho Chakravarty
Dr. Aasim Khan

Indraprastha Institute of Information Technology
New Delhi

Student's Declaration

I hereby declare that the work presented in the report entitled **"Mapping the Maze: A Study of Internet shutdowns across the world"** submitted by me for the partial fulfillment of the requirements for the degree of *Bachelor of Technology* in *Computer Science and Biosciences* at Indraprastha Institute of Information Technology, Delhi, is an authentic record of my work carried out under guidance of **Dr. Sambuddho Chakravarty** and **Dr. Aasim Khan**. Due acknowledgements have been given in the report to all material used. This work has not been submitted anywhere else for the reward of any other degree.

.... **Place & Date: IIIT Delhi, 1st June 2021**
Ritik Malik

Certificate

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

.... **Place & Date: IIIT Delhi, 1st June 2021**
Dr. Sambuddho Chakravarty

....

Dr. Aasim Khan

Abstract

The government regimes have frequently used internet shutdowns to curb the freedom of expression and devoice people worldwide. There could be other unintentional reasons, like power failure, device rupture, regional outages, etc. However, like before, we will mainly focus on *intentional internet shutdowns*.

There have been quite a few shutdowns from the time of our last report in Dec 2020. This time too, the reasons stated were to curb the spreading of fake news and better control over riots and protests. We again try to correlate those events with the BGP data, but this time around, with a new and better approach and extending the study to other countries as well.

We try to address the following questions:

- How do various governments implement these shutdowns?
- Are the same techniques implemented across all the ISPs?
- Can we correlate historical shutdowns with some publicly available datasets?
- Can we predict shutdowns in the future after analyzing the current trend?

Acknowledgments

I would like to express my sincere gratitude towards my supervisors Dr. Sambuddho Chakravarty, Dr. Aasim Khan for their guidance and encouragement throughout the project. Further, I would also like to thank Dr. Devashish Gosain for their constant support and motivation. They motivated me to put in my best and responded to my queries and questions promptly. I will forever be grateful to them for what I learned with their support.

Contents

1 Introduction **iv**

1.1 Motivation and Problem Description v

1.1.1 Problem Description v

2 Background and Relevant Work **vi**

3 Definitions and Hypothesis **vii**

3.1 Definitions vii

3.1.1 Autonomous System vii

3.1.2 Border Gateway Protocol vii

3.2 Hypothesis viii

4 Experimental Setup and Results **ix**

4.1 Previous Approach x

4.2 The problem x

4.3 The Solution x

4.4 The New Pipeline x

4.5 Case studies xi

4.5.1 Case Study 1 xi

4.5.2 Case Study 2 xii

4.5.3 Case Study 3 xiii

4.5.4 Case Study 4 xiii

5 Conclusion **xviii**

6 Limitations and future work **xix**

6.1 Limitations xix

6.2 Future work xix

Chapter 1

Introduction

Internet shutdowns are an absolute restriction placed on the use of internet services due to an order issued by a government body. It may be limited to a specific place and to specific period, time or number of days. It may be limited to mobile internet on smartphones, or the wired broadband that usually connects a desktop - or both at the same time.

The Internet shutdown is not a new phenomenon. There have been a lot of such incidents in India as well as other countries. The reason proposed by the government for a shutdown is to control the mob, spreading fake news and violence, but it has the opposite effect in reality. It creates a lack of medium, suppressing the local voices, which brings social unrest. Not to mention the economic loss caused by it, which accounts for $1.3+ billion for India alone in 2019 [7].

Many governments have adopted Internet shutdowns in the past, like Egypt [9], Libya [10], Iran [1], Sudan [2] and more recently like Belarus [4], Myanmar [11] and Uganda [12]. Almost all of them have resulted in mass protests and riots against the government, making it more tempting to analyze the situation. Many recent incidents led to internet shutdown like the military coup in Myanmar [11], farmer protests in India, and presidential elections in Uganda [12], *etc.* Despite the pandemic, these governments don't fail to amaze us with their tactics to maintain control and order.

1.1 Motivation and Problem Description

In this fast-growing digital era, where almost everyone relies on the internet directly or indirectly, internet shutdowns bring massive disruption to basic amenities in the lives of people living in the shutdown hotspots. Especially making it hard for the working class of people like office employees, doctors, drivers, students, etc. This forces us to question the digital freedom of the people. Also, in times of ongoing pandemic, when someone needs to book an ambulance or contact the doctor or their family, lack of internet due to shutdown seems such an unnecessary barrier.

Different tools have been developed in this field before like CAIDA IODA [3], whose aim is to develop an operational prototype system that monitors the internet, in near-real-time, to identify *macroscopic Internet outages* affecting the edge of the network, i.e., significantly impacting an AS or a *large fraction of a country*. Other projects like RIPE Atlas [16], which is the RIPE NCC's main Internet data collection system. It is a global network of devices, called probes and anchors, that actively measure Internet connectivity.

Also, there exists other tools like Censys [5], which is used to collects data on hosts and websites through daily port scan of the IPv4 address space with open-source ZMap [21]. More tools and papers discussed in Chapter 2.

1.1.1 Problem Description

In this project, we try to figure out how governments across the world implement these shutdowns and if there exist any publicly available datasets capable of detecting and correlating historical shutdowns with various geopolitical events on a regional scale. We will also observe how internet outages distinguish themselves from internet shutdowns.

We will be focusing on more recent shutdown events like in India, Uganda, Myanmar and Iran. We will try to evaluate if BGP data qualifies as one of the parameters that can help us to detect internet shutdowns on a large scale.

Chapter 2

Background and Relevant Work

A plethora of research has been done on Internet outage detection systems, like the Trinocular [22], which uses adaptive probing to detect internet outages at the network edge level. They used ICMP probes according to the Bayesian inference. Their model of the internet was an outage centric model which is populated from long-term observations.

RiskRoute [19], a framework for mitigating Network Outages Threats, it evaluates risk via the concept of bit-risk miles, the geographically scaled outage risk of traffic in a network. The results of their analyses high-light current risks of network infrastructures and how those risks can, in some cases, be significantly mitigated using RiskRoute recommendations.

Detecting Peering Infrastructure Outages in the Wild [20], their methodology relies on the observation that BGP communities, announced with routing updates, are an excellent and yet unexplored source of information, allowing them to pinpoint outage locations with high accuracy.

Analysis of Country-wide Internet Outages Caused by Censorship [18], their primary source of data were BGP interdomain routing control plane data, unsolicited data plane traffic to unassigned address space; active macroscopic traceroute measurements, RIR delegation files, and MaxMind's geolocation database.

So we aim to find some correlation between the publicly available datasets and the historical shutdown events across the world.

Chapter 3

Definitions and Hypothesis

3.1 Definitions

3.1.1 Autonomous System

An autonomous system (AS) is a collection of connected Internet Protocol (IP) routing prefixes (like /24, /18, /16, etc.) under the control of one or more network operators on behalf of a single administrative entity or domain that presents a common, clearly defined routing policy to the internet. AS numbers are assigned in blocks by Internet Assigned Numbers Authority (IANA) [8] to regional Internet registries (RIRs) [15]. The appropriate RIR then assigns ASNs to entities within its designated area from the block assigned by IANA. *E.g.*, BHARTI AIRTEL is AS9498, VODAFONE is AS38266, *etc.*

3.1.2 Border Gateway Protocol

Border Gateway Protocol (BGP) is a standardized exterior gateway protocol designed to exchange routing and reachability information among autonomous systems (AS) on the internet. BGP is classified as a path-vector routing protocol, and it makes routing decisions based on paths, network policies, or rule-sets configured by a network administrator.

BGP used for routing within an autonomous system is called Interior Border Gateway Protocol, Internal BGP (iBGP). In contrast, the Internet application of the protocol is called Exterior Border Gateway Protocol, External BGP (eBGP).

3.2 Hypothesis

Since the government claims the mechanism of internet shutdowns as confidential, we propose the hypothesis that might explain the technique implemented -

- The ASes can stop advertising the BGP paths in the radius of internet showdown affected areas. In simple words, they can just pull the plug so that no traffic can be exchanged beyond that point.

We will perform specific tests and measures on historical internet shutdown data to see if they fit this hypothesis.

It is equally likely that the above hypothesis is incorrect and the government uses something completely different. That would be another interesting finding in itself.

Chapter 4

Experimental Setup and Results

In the last report, we talked about various publicly available datasets and were particularly interested in the Routeviews [17] dataset as our primary study target. This time, we extended that study with a new approach and new parameters with the latest ongoing internet shutdown events. We continue to build our research around Routeviews.

A brief about routeviews:

- This dataset provides us with historical BGP information about the global routing system from the perspective of several different backbones and locations around the internet.
- The project accounts for 31 unique collectors now, across the world, but with most of them residing in the US.
- They provide access to historical BGP dumps about the global routing system from their various nodes.
- These dumps are recorded actively every day with a granularity of 2 hours.

The table 4.1 is an example of general output of a routeviews dump (after formatting).

Prefix	PATH1	PATH2	PATH3	PATH4	PATH5
67.158.52.0/24	37353	37100	6453	9498	135247

Table 4.1: Here first column represents the prefix, while other column represents the ASNs

In table 4.1 67.158.52.0/24 is the destination prefix belonging to AS135247, and there exists a unique path from source AS37353 to reach this prefix at AS135247 -

AS37353 -> AS37100 -> AS6453 -> AS9498 -> AS135247

There are millions of such entries with unique paths in a particular timestamp record from an arbitrary node.

On average, after searching and combining in the dumps from all nodes, for a particular timestamp, around 100 140 unique paths are observed for a particular prefix.

4.1 Previous Approach

We created a database using different publicly available data sources, consisting of all prefixes and netblocks for top ISPs in India like Airtel, TATA, Jio, BSNL, Vodafone, Idea, etc., to their respective geolocation data.

The prefixes lists were scrapped from ipinfo.io [13] and CIDR reports [6], and the best effort geolocation was from the Maxmind API [14]. So we got our final "prefix-to-geolocation" database.

4.2 The problem

The above approach suffered from a problem that the prefixes we got from the CIDR report/ipinfo would be outdated when used for a historical event, as these prefixes are changed regularly. Some prefixes that were affected, say last year in a shutdown event, won't be listed in the new prefixes list, thus significantly affecting the results.

4.3 The Solution

This time around, we tried a unique approach to solve this issue.

Prefix	PATH1	PATH2	PATH3	PATH4	PATH5
67.158.52.0/24	37353	37100	6453	9498	135247

Table 4.2: Here 1st column represents the prefix, while other column represents the ASNs

Looking again at the general output of BGP dumps, as in table 4.2, we know that the last column is the destination AS to which this particular prefix belongs. So, after getting the dumps for a historical shutdown period, we took a list of target ASes with a significant presence in the shutdown area and then made the AS-to-prefix dataset from the dumps itself.

It means that all the prefixes in the dumps, for which we have a typical number (AS) in the last common, belong to that particular AS for that time period.

So this eliminated the risk of losing prefixes that might not be available after the event due to any reason.

4.4 The New Pipeline

The old pipeline for this approach was completely redesigned to accommodate the new changes. A rough overview of the same would look like this:

- The input taken by pipeline:
  - List of 30 days and 4 timestamps (approx a month) for which we want the BGP dumps
  - A list of target ASes for which we want to find the prefixes in the dataset

  - A limit variable after which the graphs should be made. *E.g.,* LIMIT=20 means that those prefixes whose advertisements in the ribs fall below 20% of the maximum in the dataset should be plotted

- Download the routeviews datasets for the defined time period
- Trim it include only the first and last columns (prefix and its corresponding AS, respectively), significantly reducing the storage space
- Now from the given target ASes, make a list of all their prefixes for each timestamp.
- Make CSV files for each of these ASes and their respective prefixes with the frequencies over the given time period.
- If they cross the LIMIT, report and make the subsequent graph

The pipeline was also optimized as compared to the previous one:

- Support removed for MongoDB, replaced by python dictionaries which is much faster, because of serialization
- Execution time for 1 test case is around 7 hours, compared to 1.5 days previously
- Efficient storage: Since only 1st and last columns of dumps are used
- Less memory usage: less than 6 GB RAM as compared to 24 GB previously

4.5 Case studies

Along with historical events, we also studied many live internet shutdown/outage events as recent as Jan, Feb 2021. The following case studies were done:

4.5.1 Case Study 1

- **Country:** Iran
- **Shutdown duration:** 16 Nov - 21 Nov 2019
- **Reason for shutdown:** The 2019 Bloody November [1], which witnessed massive protest due to new government policies
- **Approach:** We analyzed the BGP dumps for November 2019 and targeted the top 10 Iranian ISPs, and we saw a strong correlation between them. The prefixes, regardless of the AS to which they belong, showed an expected dip in their advertisements during the timeframe of the protest.

Figure 4.1 shows a graph for a randomly selected prefix from the ISP IranCell. The metadata for the graph reads as: IRAN_CELL_AS4424 -> 5.113.64.0/20 , [max = 117, min = 0]. Here, we see a *U shape* during the week of protest.

A close up view of this figure can be observed in figure 4.2. Here, we can clearly see the dip beginning at 20191116.1400

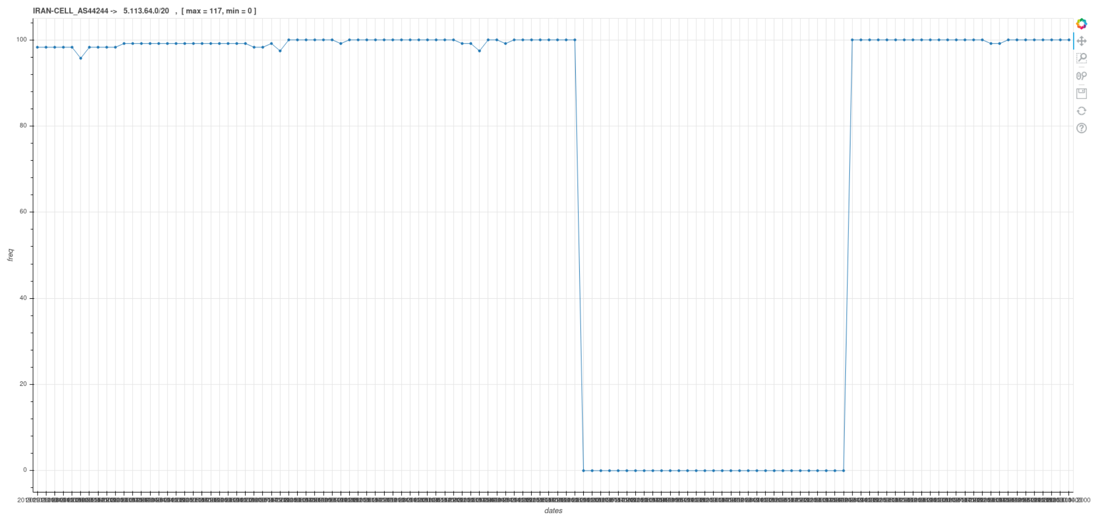

Figure 4.1: Iran #1: X-axis represents the dates in Nov 2019 While the Y-axis represents the number of unique paths to a particular prefix belonging to IranCell

Figure 4.3 is another randomly selected ISP with some random prefix. It shows the same correlation as the previous graph. The metadata of this graph read as : RIGHTEL_AS5718 -> 37.44.56.0/22 , [max = 121 , min = 0]

This proves our hypotheses that during an internet shutdown event, the BGP paths for the affected prefixes will decrease, in this case, going straight to 0.

4.5.2 Case Study 2

- **Country:** Uganda
- **Shutdown duration:** 13 Jan - 18 Jan 2021
- **Reason for shutdown:** Presidential election in Uganda [12]
- **Approach:** We analyzed the BGP dumps for January 2021 and targeted the top 10 Ugandan ISPs, and we saw a strong correlation between them. The prefixes, regardless of the AS to which they belong, showed an expected dip in their advertisements during the week of the protest.

Figure 4.4 shows a graph for a selected prefix from the ISP Airtel Uganda. The metadata for the graph reads as: AIRTEL-UGANDA-2_AS36977 -> 197.221.152.0/23 , [max = 165, min = 0]. Here also, we see a *U shape* during the week of protest.

In figure 4.5, a zoomed up version of this graph shows the prefix frequency falling to 0. The date in Y-axis is 20210114.2000 which is the starting of election week.

Figure 4.6 is another randomly selected ISP with some random prefix. It shows the same correlation as the previous graphs. The metadata of this graph read as : MTN-UGANDA_AS20294 -> 41.210.156.0/24 , [max = 145, min = 0]

But, we also saw something unusual; while these ISPs showed dips during this timeframe, some new prefixes were also added only for that particular time as if they were to counter the effect. It's hard to figure how and why such events were happening.

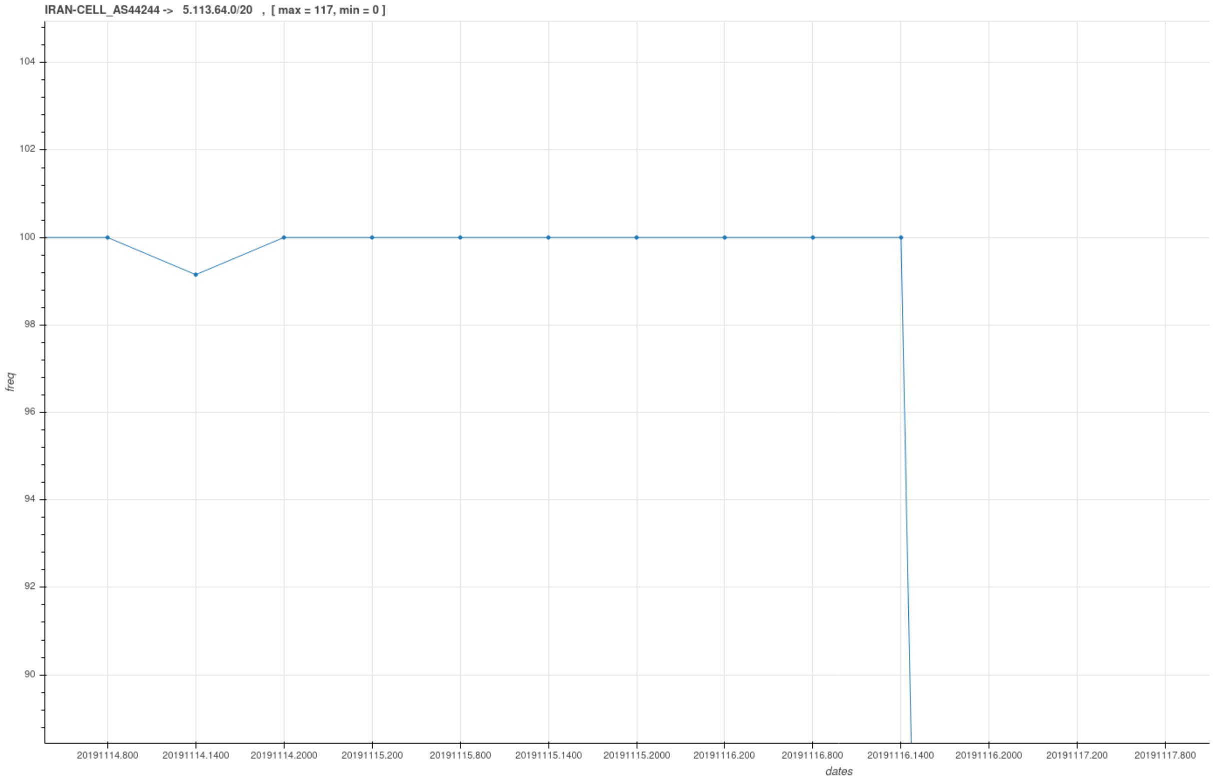

Figure 4.2: Iran #1 zoom: Zoomed in version of figure 4.1

An example of such prefix would be figure 4.7 with metadata: NIT-UGANDA_AS32774 -> 154.72.192.0/24 , [max = 143, min = 0]

The frequency of such graphs was low, but they were an exciting find.

4.5.3 Case Study 3

- **Country:** India
- **Shutdown duration:** 26 Jan - 2 Feb 2021
- **Reason for shutdown:** Farmer protest on Republic day (26th Jan), leading to internet shutdowns in Delhi borders and Haryana.
- **Approach:** We got the dumps from mid-Jan to mid-Feb 2021 and used MaxMind and IPinfo to geolocate prefixes that belonged to Haryana and surrounding areas. No significant pattern was observed in this case. The graphs were too random, as in the last report.

Some examples of graphs would be as shown in figure 4.8 and figure 4.9. These graph shows a random pattern which were not related to shutdown activity in anyway.

A majority of graphs showed poor if no correlation.

4.5.4 Case Study 4

- **Country:** Myanmar

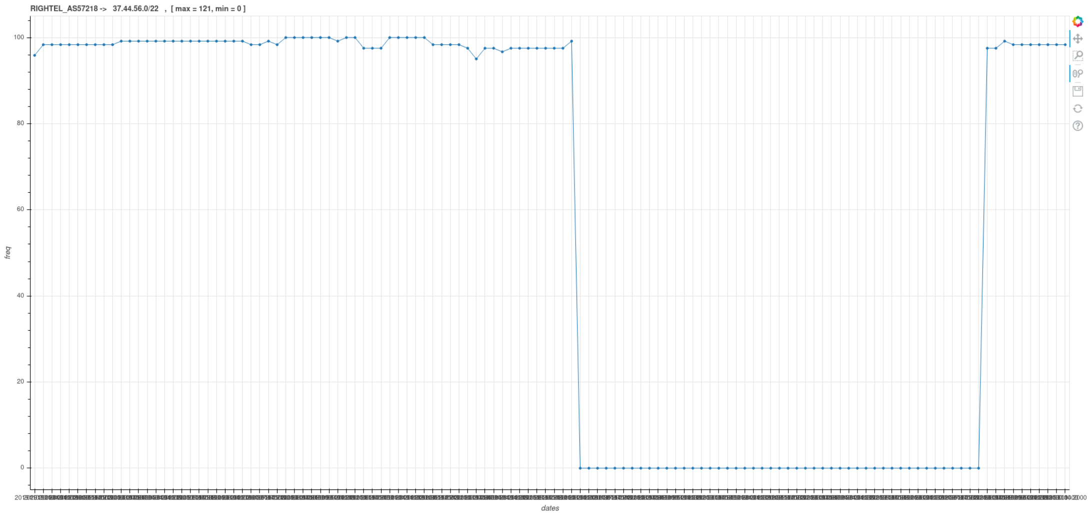

Figure 4.3: Iran #2: X-axis represents the dates in Nov 2019 While the Y-axis represents the number of unique paths to a particular prefix belonging to RIGHTEL

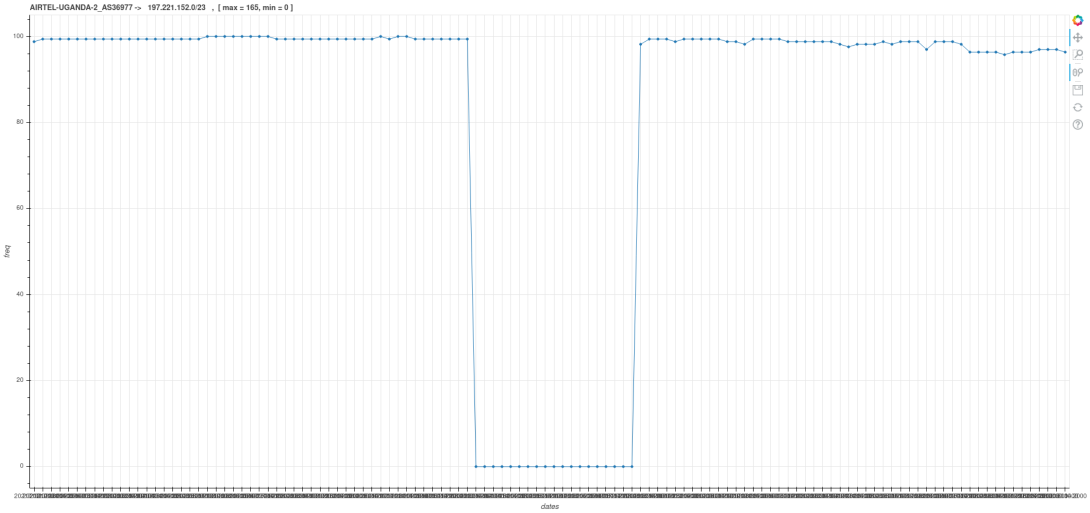

Figure 4.4: Uganda #1: X-axis represents the dates in Jan 2021, While the Y-axis represents the number of unique paths to a particular prefix belonging to Airtel Uganda

- **Shutdown duration:** 2 Feb & 6-8 Feb 2021
- **Reason for shutdown:** A military coup followed by a nationwide internet shutdown. This shutdown took place in Myanmar in 2 phases. The first was on 2nd Feb, and the second was on 6th to 8th Feb 2021. [11]
- **Approach:** We analyzed the BGP dumps for Feb 2021 and targeted the top 15 Myanmar ISPs, and we saw a strong correlation between them. Regardless of the AS to which they belong, the prefixes showed an expected dip in their advertisements during these two-time frames of the protest.

We had about 316 prefixes, out of which 279 showed a good correlation, which is around 80%.

Sample graphs for correlation would be figure 4.10 and figure 4.11 with metadata MYAN-POSTS-TELE_AS45558 -> 43.224.43.0/24 , [max = 101, min = 0] and POST-TELECOM_AS9988 -> 203.81.64.0/21 , [max = 170, min = 5] respectively.

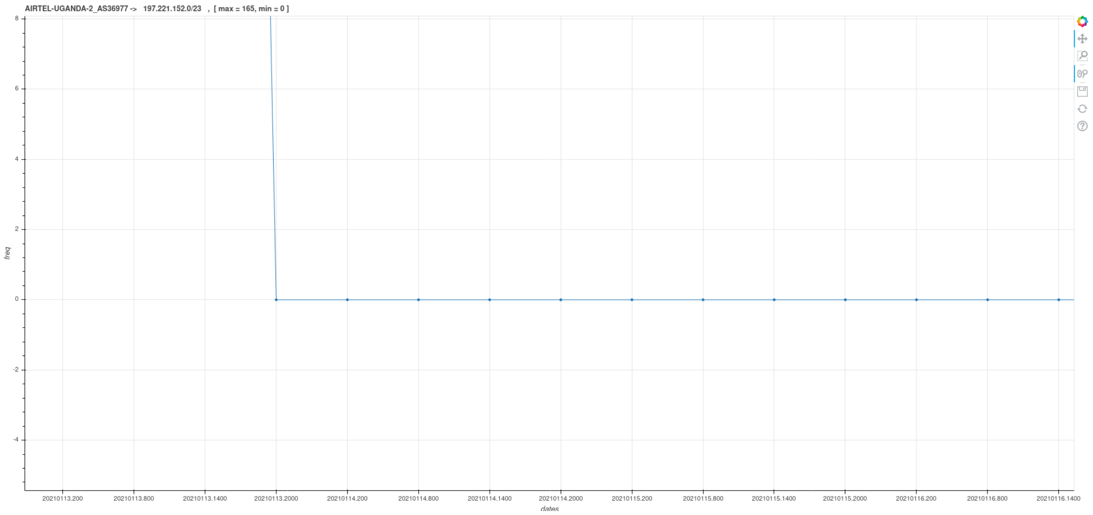

Figure 4.5: Uganda #1 zoom: Zoomed in version of figure 4.4

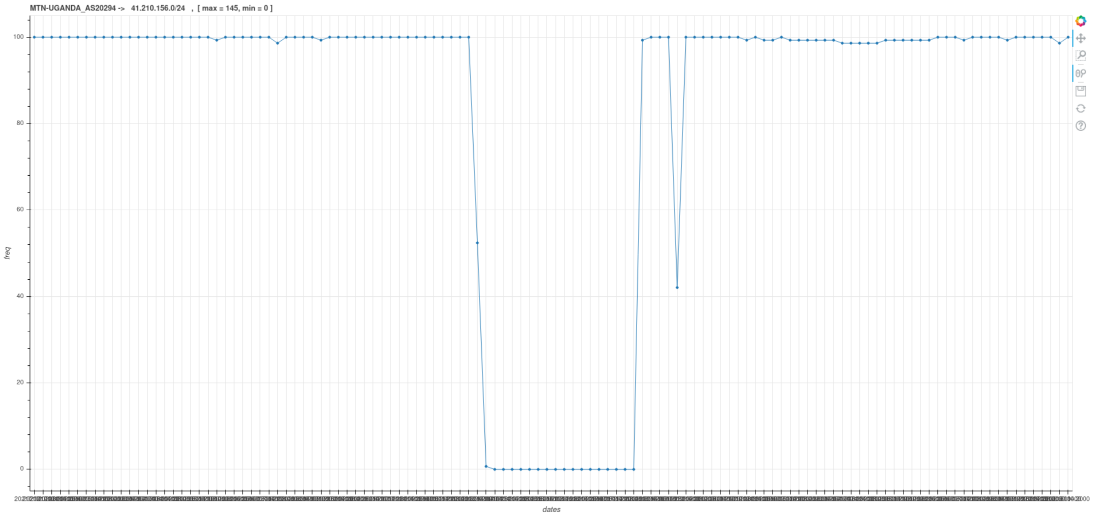

Figure 4.6: Uganda #2: X-axis represents the dates in Jan 2021, While the Y-axis represents the number of unique paths to a particular prefix belonging to MTN Uganda

In figure 4.10 we see 2 sharp dips, straight to 0. The first one is on 2nd Feb and the second on 6th - 8th Feb, which shows a good correlation with the actual event.

Similar patter can be found in figure 4.11. The frequency of this particular prefix advertised globally went from around 170, all the way down to 5, during those two time periods.

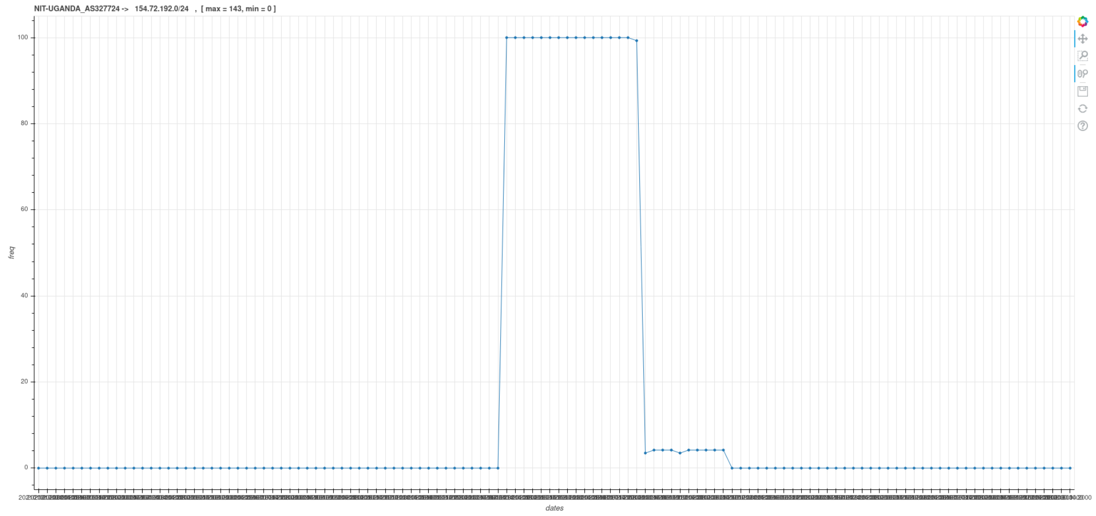

Figure 4.7: Uganda #3: These type of special graphs shows a unique characteristic, they are active only during the shutdown period

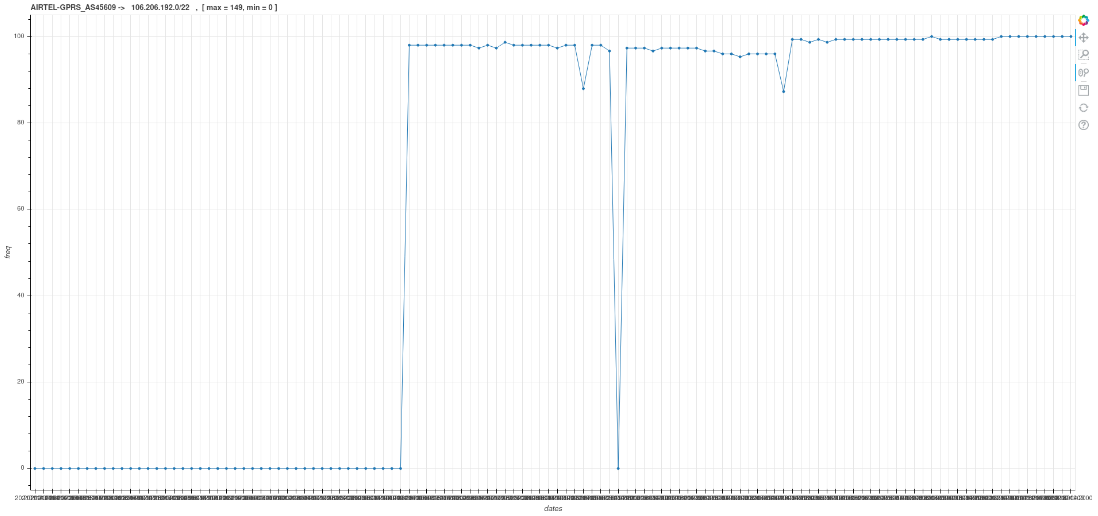

Figure 4.8: India #1: X-axis represents the dates in Jan-Feb 2021, While the Y-axis represents the number of unique paths to a particular prefix belonging to AIRTEL-GPRS

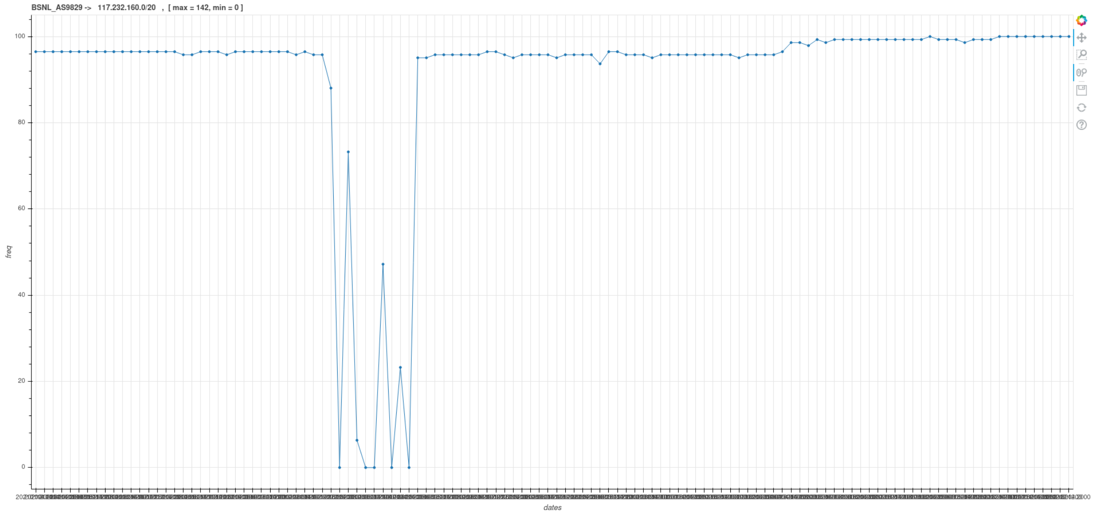

Figure 4.9: India #2: X-axis represents the dates in Jan-Feb 2021, While the Y-axis represents the number of unique paths to a particular prefix belonging to BSNL

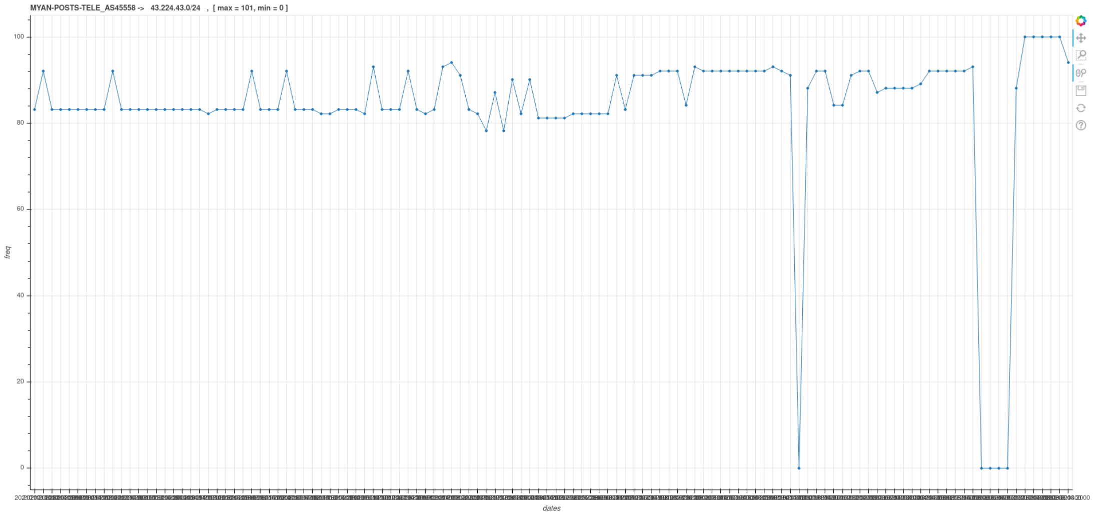

Figure 4.10: Myanmar #1: X-axis represents the dates in Jan-Feb 2021, While the Y-axis represents the number of unique paths to a particular prefix belonging to MYAN-POSTS-TELE

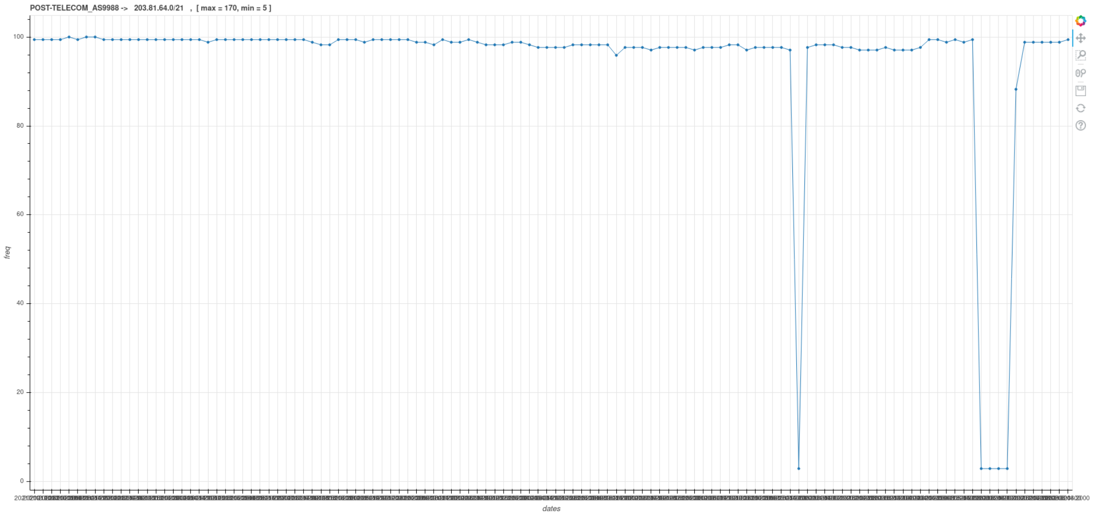

Figure 4.11: Myanmar #2: X-axis represents the dates in Jan-Feb 2021, While the Y-axis represents the number of unique paths to a particular prefix belonging to POST-TELECOM

Chapter 5

Conclusion

We analyzed four countries viz., India, Iran, Myanmar and Uganda for remotely detecting the shutdowns. We saw a good correlation of the shutdown events with the BGP data in Iran, Myanmar, and Uganda, where the shutdown was, on a national scale, imposed deliberately by the government. However, we still failed to detect poorly executed small-scale shutdowns in India.

So it would be safe to conclude that BGP data is *one of the many factors* that play an essential role while detecting internet shutdown or outages on a *macroscopic scale*. More such parameters exist and should be taken into consideration to arrive at a full-proof conclusion. We can't rely solely on BGP.

Chapter 6

Limitations and future work

6.1 Limitations

There are some limitations that we encountered while working on this project:

- Trusting BGP data is a leap of faith, as BGP itself is a very complex algorithm in which financial relationships have the upper hand against cost and efficiency.
- Since the number of Routeviews probes is limited, it is expected that not all data can be captured.
- If there is some change in a prefix, then according to BGP, it can take hours to get it registered in the global routing tables. Thus a mismatch in time and information can be an expected thing.
- There were many shutdowns/outages which went unreported and thus, it would be difficult to decide whether some dips in the graphs belong to unreported shutdowns or, in actuality, it is a false positive

6.2 Future work

In order to remove all the ambiguity about the results, we plan the following inspections:

- We plan to correlate this data with other months as well, when there was no protest, to prove these graphs' uniqueness.
- We also plan to geolocate the prefixes, which showed a significant dip in the graphs and then verify it with the news.
- In order to increase confidence in our hypothesis, we need to do more of such case studies but this time focussing more on the regional ISPs
- We can correlate for other countries as well, such as Belarus
- Lastly, we need to correlate the Routeviews data with other publicly available projects, which we mentioned at the beginning, like Censys, CAIDA ARK, IODA and RIPE Atlas.

Bibliography

[1] 2019 internet blackout in iran. https://en.wikipedia.org/wiki/2019_Internet_blackout_in_Iran.

[2] 2019 internet shutdown in sudan. https://globalvoices.org/2020/06/08/internet-shutdowns-in-sudan-the-story-behind-the-numbers-and-statistics/.

[3] Caida ioda project. https://ioda.caida.org/.

[4] Censorship in belarus. https://en.wikipedia.org/wiki/Censorship_in_Belarus.

[5] Censys project. https://www.censys.io/.

[6] Cidr reports. https://www.cidr-report.org/.

[7] Economic loss due to internet shutdown in india 2019. https://timesofindia.indiatimes.com/business/india-business/india-lost-1-3bn-due-to-4196-hours-of-no-internet/articleshow/73179287.cms.

[8] Internet assigned numbers authority. https://www.iana.org/.

[9] Internet shutdown in egypt 2011. https://en.wikipedia.org/wiki/Internet_in_Egypt#2011_Internet_shutdown.

[10] Internet shutdown in libya 2011. https://en.wikipedia.org/wiki/Internet_censorship_in_the_Arab_Spring#Libya.

[11] Internet shutdown in myanmar. https://netblocks.org/reports/internet-disrupted-in-myanmar-amid-apparent-military-uprising-JBZrmlB6.

[12] Internet shutdown in uganda. https://www.apc.org/en/news/uganda-2021-general-elections-internet-shutdown-and-its-ripple-effects.

[13] Ipinfo. https://www.ipinfo.io/.

[14] Maxmind. https://www.maxmind.com/en/home.

[15] regional internet registries. https://www.iana.org/numbers/allocations/.

[16] Ripe atlas project. https://atlas.ripe.net/.

[17] University of oregon route views archive project. http://archive.routeviews.org/.

[18] Dainotti, A., Squarcella, C., Aben, E., Claffy, K. C., Chiesa, M., Russo, M., and Pescapé, A. Analysis of country-wide internet outages caused by censorship. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference* (New York, NY, USA, 2011), IMC '11, Association for Computing Machinery, p. 1–18.

[19] Eriksson, B., Durairajan, R., and Barford, P. Riskroute: A framework for mitigating network outage threats. In *Proceedings of the Ninth ACM Conference on Emerging Networking Experiments and Technologies* (New York, NY, USA, 2013), CoNEXT '13, Association for Computing Machinery, p. 405–416.

[20] Giotsas, V., Dietzel, C., Smaragdakis, G., Feldmann, A., Berger, A., and Aben, E. Detecting peering infrastructure outages in the wild. In *Proceedings of the Conference of the ACM Special Interest Group on Data Communication* (New York, NY, USA, 2017), SIGCOMM '17, Association for Computing Machinery, p. 446–459.

[21] Lee, S., Im, S., Shin, S., Roh, B., and Lee, C. Implementation and vulnerability test of stealth port scanning attacks using zmap of censys engine. In *2016 International Conference on Information and Communication Technology Convergence (ICTC)* (2016), pp. 681–683.

[22] Quan, L., Heidemann, J., and Pradkin, Y. Trinocular: Understanding internet reliability through adaptive probing. In *Proceedings of the ACM SIGCOMM 2013 Conference on SIGCOMM* (New York, NY, USA, 2013), SIGCOMM '13, Association for Computing Machinery, p. 255–266.