CN3530 - Computer Networks

K Vivek Kumar
CS21BTECH11026

Adarsh Suresh Bhende
CS21BTECH11008

Jarupula Sai Kumar
CS21BTECH11023

August 25, 2023

Abstract

The visualization of the assignment is being hosted on Vercel and the link to the created webpage can be accessed below: https://cn3530-assign1.vercel.app/

1 Assignment Statement

The Internet works as large-scale and complex interconnections among many routers belonging to various ISPs. If you want to know how those routers and ISPs are interconnected between the source and destination hosts, how would you do so?

Perform traceroute from many source hosts to many destination hosts with showing AS numbers.

2 Assignment Procedure

The entire assignment was completed through a systematic approach, following the steps outlined below:

2.1 Step I: Presentation Approach for Understanding the Internet

We chose to convey an understanding of the Internet's topology by visually plotting the network paths onto a global map. To achieve this, we utilized the 'folium' package in Python, which allowed us to create an animated visualization of network routes. The resulting visualization was converted into HTML pages, capturing both points and lines on the map. These HTML pages were categorized based on the sources from which the traceroutes were conducted. The corresponding Python code and generated HTML pages can be accessed in our GitHub repository:

Repository Link:
https://github.com/v-house/assignment-1-cn3530.git

2.2 Step II: Tracerouting to Various Sources and Destinations

We initiated traceroutes from multiple sources to various destinations to explore the paths taken across the globe. Below, we outline the sources and destinations used in this analysis:

2.2.1 Sources

We collected traceroute data from various sources using the following methods:

- **Mumbai, Maharashtra:** We utilized the 'dotcom-tools' looking glass website, available at https://www.dotcom-tools.com/network-trace-test.

- **IITH Network:** Traceroutes were initiated from our laptop's terminal connected to IITH-LAN, using the following command:

  ```
  $ traceroute <destination domain or ip>
  ```

- **Amsterdam, Netherlands:** We performed traceroutes using the Proton VPN on our personal laptop, employing a similar command as mentioned above.

- **Sydney, Australia:** We performed traceroutes using the following website which can be accessed from the link below: Sydney Tracerouting:
  `https://perfops.net/traceroute-from-sydney`

- **Japan:** We performed traceroutes using the Proton VPN with the Japan server on our personal laptop, employing a similar command as mentioned above.

2.2.2 Destinations

We conducted traceroutes on various website domains to explore diverse routes. Below are some of the domains we analyzed:

- `google.com`
- `gmail.com`
- `yahoo.co.jp`
- `lookup.icann.org`
- `japantimes.com`
- `cisco.com`
- `indiatoday.in`
- `netflix.com`
- `spain.info`
- `discord.com`
- `paypal.com`
- `openai.com`
- `duckduckgo.com`

- `stackoverflow.com`

The raw data collected for all the routes we performed is available in the shared Google Sheet we used throughout the assignment project to store plot details. You can access the 'RAW_DATA' section in the following sheet. The URL is provided below:

Google Sheet Data:
`https://docs.google.com/spreadsheets/d/1ZuNJxUlX7y1exhHOBRBQ1ZtEih9MzCVHy_NRCmONXDY/edit?usp=sharing`

2.3   Step III: Generating Visualization Plots

In order to create visualizations using the folium package in Python, we needed latitude and longitude coordinates for mapping the data points. To gather this information, we utilized an IP-to-address conversion website. You can access the website via the following link:

IP-to-Address Conversion Website:
`https://whatismyipaddress.com/ip-lookup`

This platform provided us with essential details such as latitude, longitude, the organization associated with the IP address, and the AS (Autonomous System) number of the IP address. For each example, this data is stored in the 'understanding_¡number¿' section of the same Google Sheet. To examine the collected information, refer to the 'Latitudes, Longitudes, and Organization Names' under the sheet sections: 'Understanding_1', 'Understanding_2', 'Understanding_3', 'Understanding_4', and 'Understanding_5'.

To explore the comprehensive dataset of latitudes, longitudes, and organization names, visit the following Google Sheet:

Google Sheet Data:
`https://docs.google.com/spreadsheets/d/1ZuNJxUlX7y1exhHOBRBQ1ZtEih9MzCVHy_NRCmONXDY/edit?usp=sharing`

2.4   Step IV: Writing Python Code for HTML Page Generation

To create the visualizations, we developed Python code that leveraged the 'Folium' package from the Python library. Writing the code was facilitated by consulting the official Folium package documentation, which proved to be a valuable resource in the process. You can access the official documentation through the following link:

Folium Python Package Documentation:
`https://python-visualization.github.io/folium/`

Additionally, we referred to the following article for guidance in crafting the code:

Real Python Article on Using Folium for Web Maps
`https://realpython.com/python-folium-web-maps-from-data/`

The complete code for this stage is available in the following GitHub repository:

Repository Link:
`https://github.com/v-house/assignment-1-cn3530.git`

When executing this Python code, it generates an HTML page with the filename specified within the code. Prior to running the code, ensure that the 'folium' package is installed by executing the following command:

```
$ pip install folium
```

The HTML page is generated in the same directory as your python folder. And also we made a slight edit in the HTML page by adding a Navbar head to show the color notations used in the page. You can see some glimpse of the visualization from the image below:
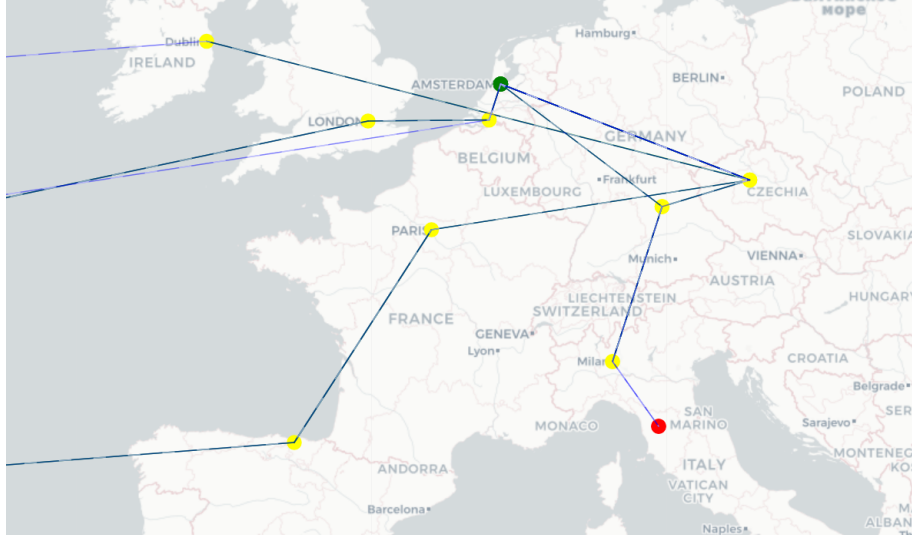

Figure 1: Visualization of Internet Topology

2.5 Step V: Hosting the Assignment (Enhanced Presentation)

With our HTML pages prepared, we decided to enhance the accessibility by hosting the assignment online. This idea led us to streamline the code and structure it conventionally for optimal presentation, making the repository ready for hosting. After exploring various free hosting services, we discovered Vercel, a platform that significantly simplified the hosting process. By connecting our GitHub repository to our Vercel account and following their hosting instructions, we successfully hosted our webpage on the Vercel platform. You can access our assignment webpage through the following link:

Hosted Assignment Website:
   https://cn3530-assign1.vercel.app/

We extend special thanks to Vercel for their assistance.

2.6 Step VI: Drawing Conclusions and Gained Insights

The final phase of our project involved summarizing our learning, insights, and conclusions drawn from this assignment. The accompanying PDF provides detailed information regarding this aspect.

3 Findings

While doing the assignment and obtaining various data, we came across the following observations:

- Virtually all the requests dispatched via IITH WiFi to access foreign websites traverse either through the Mumbai (Krunalshah Software Private Limited and Yotta Network Services Private Limited) or Vijaywada (Reliance Jio Infocomm Limited) ISP routers.
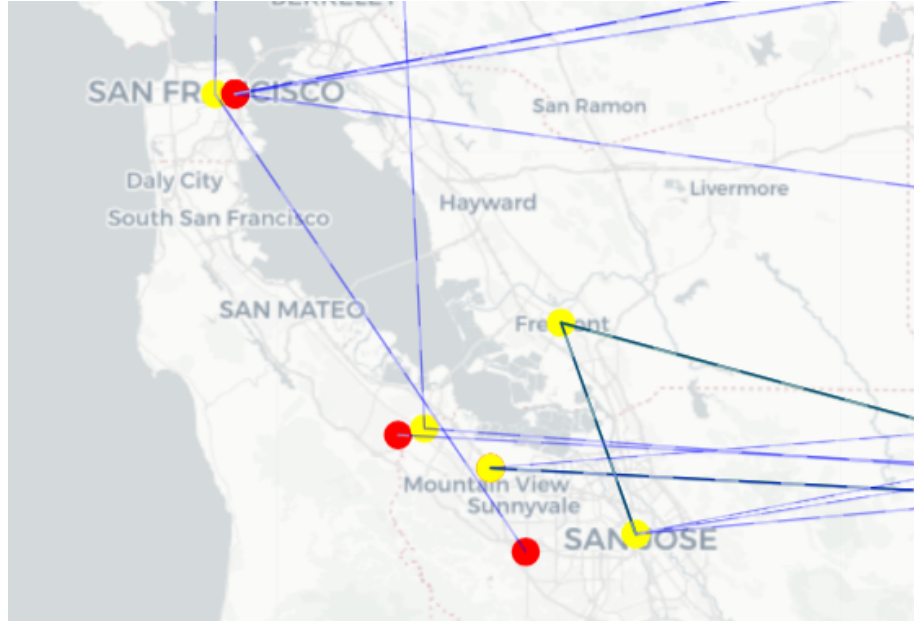

Figure 2: Higher frequency of routers near San Francisco

- The termination points of the majority of website destinations are within the ISP router located in San Francisco. This can be attributed to Amazon Cloud Services basing their operations in the San Francisco region. Look the figure 2 above.

- In our exploration of AS numbers, we have discovered that AS numbers encompass a range of IP addresses, and a specific organization manages this range. Consequently, when two different IP addresses fall within this same range, it results in both IPs being associated with the same AS number and organization.

- We have observed that a significant number of government websites, such as sbi.com and defense websites, do not provide responses when subjected to traceroute probing. Instead, they exhibit a "request timeout" behavior, suggesting a private and restricted network configuration.

- During traceroute sessions conducted via a VPN connection, we consistently encountered a private IP address as the initial hop. This occurrence is in line with VPN protocols, which conceal the true geographical location of the user, hence masking the originating IP address with a private one.

- In our exploration of AS numbers, we have discovered that AS numbers encompass a range of IP addresses, and a specific organization manages this range. Consequently, when two

different IP addresses fall within this same range, it results in both IPs being associated with the same AS number and organization.

- During a traceroute operation, we encounter IP addresses that disclose their presence but withhold location information, while others remain entirely unresponsive, neither revealing their IP address nor providing any response.

4   Conclusion

In conclusion, the analysis of Internet topology provides valuable insights into its structure and connectivity. The combination of data collection, processing, visualization, and analysis offers a comprehensive understanding of the modern Internet.
The learnings towards this assignment was really great! Understanding Internet was really a fun activity.

5   Deliverables

We have submitted the following files:

- Raw data and the furnished data as a google sheet (also downloaded and attached in the zip file).

- Link to the webpage made by us and hosted on Vercel.

- HTML files for building the webpage.

- Python files for running over all the data and obtaining visualizations.

- Report, written on Latex.

6   References

- `https://whatismyipaddress.com/`

- `https://www.whois.com/whois/`

- `https://python-visualization.github.io/folium/`

- `https://www.dotcom-tools.com/`

- `https://www.iplocation.net/`

Thank You