ECE 478/578: Fundamentals of Computer Networks

Project # 2. Studying The Internet Topology

1 Preliminaries

- Read the project description carefully
- You must form a group of **up** two people.
- You are free to use any programming environment or data analysis tool.
- Files needed for the project (available on D2L): 20210401.as2types.txt, 20150801.as2types.txt, 20241101.as-rel2.txt, routeviews-rv2-20241106-1400.pfx2as.txt, 20241001.as-org2info.txt

2 Project Description

Using real topology data, you are to study the Internet topology at the Autonomous System (AS) level. We will use the data sources available at the Center for Applied Internet Data Analysis (CAIDA) at `http://www.caida.org/`.

2.1 AS Classification – 20 points

Explore the classification of ASes to (a) transit/access ASes, (b) content ASes, and (c) enterprise ASes. Relevant information for the classification techniques can be found at `https://www.caida.org/catalog/datasets/as-classification/`. Use the following files (available on D2L), containing the AS number and class of all the ASes detected in BGP path advertisements and peering points.

- The AS classification published on April 2021: *20210401.as2types.txt*
- The AS classification published on August 2015: *20150801.as2types.txt*

Generally, enterprise ASes are expected to be stubs, whereas content ASes are expected to be multi-homed.

Graph 1: Create a pie chart plot that shows the (%) distribution of ASes to the three classes for 2015 and 2021. Comment on your findings. What is the majority of ASes on the Internet topology? How does it compare to other types of ASes? How did the distribution change over the years? (**20 points**).

2.2 Topology Inference Through AS links - 60 points

Explore the AS relationships by studying the Internet topology through AS links. Use the *20241101.as-rel2.txt* dataset which contains every link between ASes as it is inferred from BGP path advertisements. The file format is explained at at `https://www.caida.org/catalog/datasets/as-relationships/` and is as follows:

p2c link: <provider-AS>|<customer-AS>| -1 |<source>
p2p link: <peer-AS>|<peer-AS>| 0 |<source>

Ignore the last field that indicates the method for collecting AS link information. For each AS_i compute

1. The global node degree, defined as the number of distinct links (all types) incident to the AS_i.
2. The customer degree, defined as the number of direct customers of AS_i.
3. The peer degree, defined as the number of peers of AS_i.
4. The provider degree, defined as the number of providers for AS_i.

Graph 2: Create histograms of the AS node degree distribution (one histogram for each type of degree) using the following bins: 0, 1, 2-5, 6-100, 101-500, 501-1,000, >1,000. Explain what you observe concerning the degree distribution on the Internet for each type of degree (**20 points**).

Use the *routeviews-rv2-20241106-1400.pfx2as.txt* dataset which contains the IPv4/IPv6 Prefix-to-Autonomous System (AS) mappings derived from RouteViews data. This dataset contains data about the IP prefixes advertised by each AS, with one prefix-AS mapping per line. The tab-separated fields are IP prefix, prefix length, and AS number. More details can be found at Visit `https://www.caida.org/catalog/datasets/routeviews-prefix2as/`.

Graph 3: Create a histogram of the IP space size assigned to each AS. Use a binning method of your choice depending on the range of results that you obtain. Explain how you selected the bins. Explain what you observe concerning the IP space distribution across ASes (**20 points**).

Graph 4: Recreate the pie chart chart (Graph 1) that shows the (%) distribution of ASes to the three classes, by using the following classification (Do so only for the 2024 dataset) (**20 points**):

- Enterprise ASes: any AS without customers or peers.
- Content AS: Any AS with no customers and at least one peer.
- Transit AS: Any AS with at least one customer.

What do you observe regarding the data consistency when comparing to Graph 1?

2.3 Inference of Tier-1 ASes - 20 points

Infer the list of Tier 1 ASes by computing the largest clique (complete graph) in the AS topology graph. To compute the largest clique follow this simple greedy heuristic.

- Rank all the ASes according to their global degree and organize them to a descending order set $R = \{AS_1, AS_2, ...\}$, where $degree(AS_i) > degree(AS_{i+1})$.
- Initialize the clique $S = \{AS_1\}$.
- If AS_2 is connected to AS_1, via any type of link add it to S. That is $S = \{AS_1, AS_2\}$,
- If AS_3 is connected to AS_1 and AS_2, add it to S. That is $S = \{AS_1, AS_2, AS_3\}$.
- Terminate when you find the *first* AS that is not connected to all ASes in S.

Table 1: Report the size of the T1 list that you were able to discover and also report the first 10 ASes that were added to S. You can use the AS-to-organization mapping database *20241001.as-org2info.txt* (`http://www.caida.org/data/as-organizations/`) to map AS numbers to the organizations that own them (**20 points**).

The table should contain the AS number, the AS organization, and the number of incident links.

Note. It is possible that you do not find 10 ASes that form a clique and the algorithm terminates within the first 2-3 iterations. In this case, eliminate the AS that is NOT connected to all previous ASes in the clique and see if the next AS on your ranked list fits the clique criteria. Attempt to find a clique of 10 ASes by looking down the ranked list of ASes, by considering up to the top 50 ASes. Example. Say AS_1, AS_2, and AS_3 are added to the clique but AS_4 is not connected to all previous ASes. Inspect if AS_5 or AS_6 fits the clique criteria, i.e., they are connected to AS_1, AS_2, and AS_3.

2.4 Extra Credit: Customer Cones and AS Rank - 20 points

Study the definition of the customer cone at `https://www.caida.org/catalog/datasets/as-relationships/`. Briefly, the AS customer cone for AS_i is defined as AS_i itself plus all the ASes that can be reached from AS_i following only p2c links. In other words, AS_i's customer cone contains AS_i, plus AS_i's customers, plus its customers' customers, and so on.

Compute the customer cone size of every AS in:

- Number of ASes.
- Number of advertised IP prefixes.
- Number of unique IP addresses.
- Percentage of ASes, advertised prefixes, and IP addresses.

Table 2: Present the top 15 ASes ranked by the customer cone *in the number of reachable ASes* (**20 points**).

Use the following table format to present your results

AS rank	AS #	AS name	AS degree	customer cone					
				number of			percentage of		
				ASes	IP Prefix	IPs	ASes	IP Prefix	IPs
1									
...									
15									

The AS name is inferred from the AS-to-organization mapping (`http://www.caida.org/data/`

`as-organizations/`). The AS degree refers to the global node degree as computed in 2.2. The number of IP prefixes refers to the unique prefixes that are reachable in the AS customer cone. The number of IPs refers to the number of unique IP reachable in the AS customer cone (found by looking at the length of each IP prefix and summing the number of hosts per prefix). The percentages are normalized over the total number of ASes, IP prefixes, and IPs respectively.

[1] Luckie, Matthew *et al.*, "AS Relationships, Customer Cones, and Validation" in *Proceedings of the 2013 Conference on Internet Measurement Conference* (ICM), pp. 243–256, 2013.