ECE 478/578: Fundamentals of Computer Networks

Project # 2. The Internet Topology

Due Tuesday Nov. 8, Electronically via D2L at 11:59 pm

1 Preliminaries

- Read the project description. When you are finished reading it, read it again.
- You must form a group of **(strictly)** two people. If you do not have a partner please post on Piazza.
- Due date: Tuesday November 8, 2012 11:59 pm via D2L.
- You are free to use a programming language of your choice.

2 Project Description

You are to study the Internet topology at the Autonomous System (AS) level using real topological data. 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, using the dataset available at `http://www.caida.org/data/as-classification/` (file name: 20150801.as2types.txt.gz). Generally, enterprise ASes are expected to be stubs, whereas content ASes are expected to be multi-homed.

Graph 1: Create a pie chart that shows the (%) distribution of ASes to the three classes (**20 points**).

2.2 Topology Inference Through AS links - 60 points

Explore the AS relationships be recreating the Internet topology through AS links. Use the dataset available at `http://www.caida.org/data/as-relationships/` (file name: 20161001.as-rel2.txt.bz2 , under the serial 2 directory, dated Oct. 21st). This dataset contains every link between ASes as it is inferred from BGP path advertisements. The file format 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 node degree, defined as the number of distinct links (all types) incident to the AS_i.
2. The set of ASes to which AS_i is a provider (direct customers of AS_i.

3. The set of IP prefixes associated with AS_i. This association can be found at `http://www.caida.org/data/routing/routeviews-prefix2as.xml`. Use the latest file available.

Graph 2: Create a histogram of the AS node degree distribution using the following bins: 1, 2-5, 5-100, 100-200, 200-1,000, <1,000. The y axis should be normalized to the total number of ASes that were enumerated in the dataset. Explain what you observe with respect to the degree distribution on the Internet (**20 points**).

Graph 3: Create a histogram of the IP space assigned to each AS. Use a binning method of your choice depending on the range of results that you obtain. For simplicity, normalize the IP space assigned to each AS to the total IP space. Explain what you observe with respect to the IP space distribution across ASes (**20 points**).

Graph 4: Recreate the pie chart that shows the (%) distribution of ASes to the three classes, by using the following classification (**20 points**):

- Enterprise ASes: any AS with degree less or equal to two and no customers or peers.
- Content AS: Any AS with degree less or equal to two, no customers and at least one peer.
- Transit AS: Any AS with at least one customer.

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

2.3 Inference of T1 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 degree and place them to 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, 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: Give the size of the T1 list and the first 10 ASes that were added to S. You can use the AS-to-organization mapping (`http://www.caida.org/data/as-organizations/`) to map AS numbers to the organizations that own them (**20 points**).

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

Study the definition of the customer cone at `http://as-rank.caida.org/?mode0=as-intro#customer-cone`. 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 number of ASes* that they can reach using p2c links (**10 points**).

Table 3: Present the top 15 ASes ranked by the customer cone *in percentage of IP addresses* they can reach using p2c links (**10 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									