Project 2: The Internet's AS Topology

ECE 578: Fundamentals of Computer Networks

Lena Voytek and Mitchell Dzurick

11/26/2019

1 Introduction:

This project describes the top level topology of the Internet by compiling and extrapolating on data about the Autonomous Systems that control it. Autonomous Systems (ASes) are the enclosed networks of large organizations connected to the wider Internet through some set of routers. These ASes are classified as either Enterprise, Transfer, or Content. Autonomous Systems make up the backbone of the Internet as they host content, move data, and provide data to individual devices. Using several recent data sets from the database of the Center for Applied Internet Data Analysis (CAIDA), general information about the interactions between these ASes and what each is used for was extrapolated. The data was parsed and manipulated using several python classes. These can be found at the following GitHub repository https://github.com/lvoytek/ECE578/tree/master/project2_Internet_Topology By extrapolating information on AS classifications, topology, tier 1 connections, and customer cones, a more rigid understanding of the core Internet structure can be determined.

2 AS Classification:

The first thing determined about the ASes was the percentage of them belonging to one of the three major classifications. These classifications, as stated by caida.org, are each defined by one of the following.

- Enterprise: Organizations, universities and companies that are mostly users
- Transit/Access: ASes inferred to be transit and/or access providers.
- Content: ASes providing content hosting and distribution systems.

Using our Python program and the provided November 2019 AS Classifications file from caida.org, the following statistics were determined:

- There are 20,509 Enterprise ASes making up 31.0% of the total
- There are 41,728 Transit ASes making up 63.1% of the total
- There are 3,852 Content ASes making up 5.8% of the total

A graphical depiction of this is shown in Figure 1 using a pie chart. With this chart, it becomes obvious that the majority of the Internet's backbone comes from Transit ASes. There are about half as many Enterprise ASes, who are consumers of the Internet's content. Meanwhile Content ASes are about one fifth of that amount, even though they are responsible for hosting the data of the Internet.

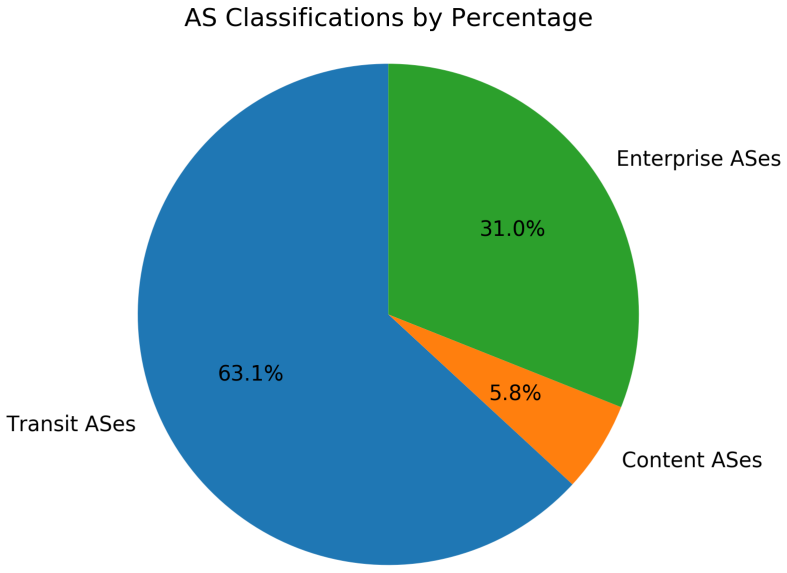

Figure 1: AS Classifications by Percentage

3 Topology Inferences Using AS Links:

In order to infer the topology of AS links, several analyses were run. Starting off, node degrees, or the number of distinct links to each AS, and the set of customers to each AS were extracted using the provided November 2019 AS relationship file. The set of IPv4 prefixes belonging to each AS were then found using the provided November 22nd, 2019 IPv4 space file. Likewise, the set of IPv6 prefixes belonging to each AS using the November 22nd, 2019 IPv6 space file. With these data, several things were determined. This includes the node degree distribution, IP space distribution, and the percentage of ASes with certain qualifications for each classification. The results of this are shown below.

3.1 AS Node Degree Distribution

The node degree distribution is the volume of ASes who are connected to a certain magnitude of other ASes. In this analysis, the magnitudes of connections that the ASes were split into were 1, 2-5, 6-100, 101-200, 201-1,000, and over 1000. The result of this is shown in the histogram in Figure 2.

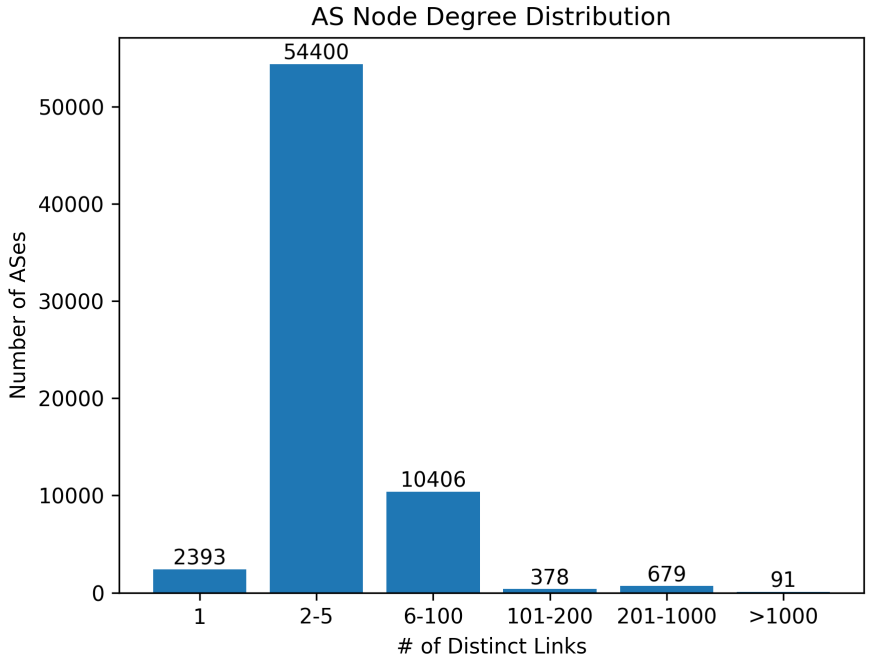

Figure 2: Histogram of AS Node Degree Distribution

The representation of the data in this manner clearly shows how the backbone of the Internet is connected together. The vast majority of AS nodes have very few connections to others. However, there is still a notable number of them that connect over 1000 nodes. This implies that most nodes focus solely on connecting a couple other nodes or consume data from only one or a couple sources. All of this describes the decentralized form of the internet. The small portion of nodes connected to an excessively large amount of nodes theoretically

describe the larger content providers and internet infrastructure companies such as Google or Facebook.

3.2 AS IP Space Distribution

The IP space distribution is the volume of ASes who own a certain magnitude of IP addresses contained in one of their IP prefixes. For this analysis, the results of IPv4 and IPv6 addresses were split up due to the many orders of magnitude more IP addresses contained in an IPv6 prefix. For IPv4, a reasonable set of bins for useful elaboration was found to be below 1,000, 1,000-10,000, 10,001-100,000, 100,001-1,000,000, 1,000,001 to 10,000,000 and over 10,000,000. The results of this split are shown in Figure 3.

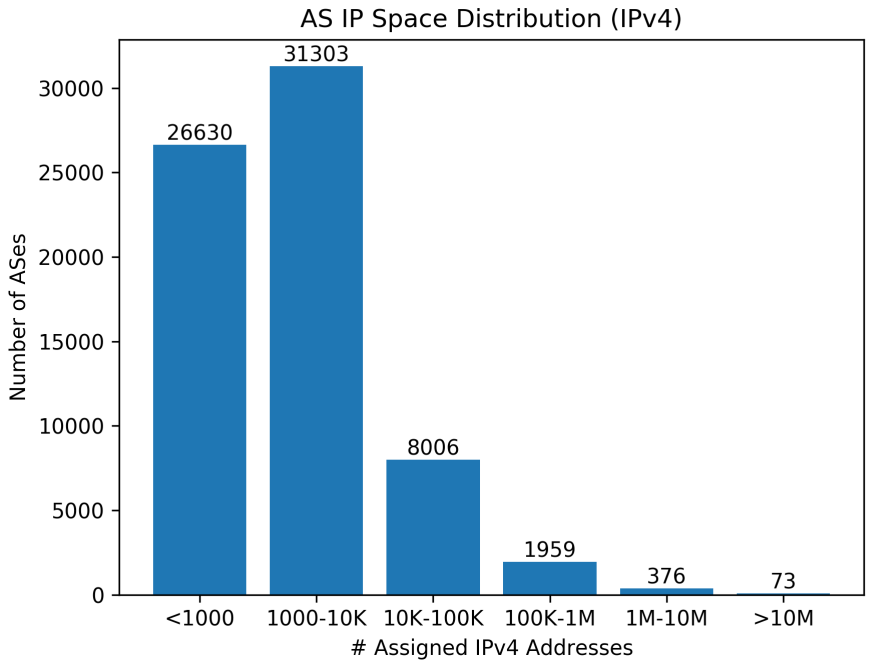

Figure 3: Histogram of AS IPv4 Space

Once again, the IPv4 distribution helps confirm the notion of a distributed Internet combined with a small set of very far reaching ASes. The majority of nodes with under 10,000 IPs are probably made up mostly of transfer ASes which focus mostly on connecting larger entities. Meanwhile Enterprise ASes for groups such as Internet Service Providers require a far greater set of addresses to give out to the general population. Although slightly unclear, the distribution of IPv6 addresses, shown below in Figure 4, describes the same thing.

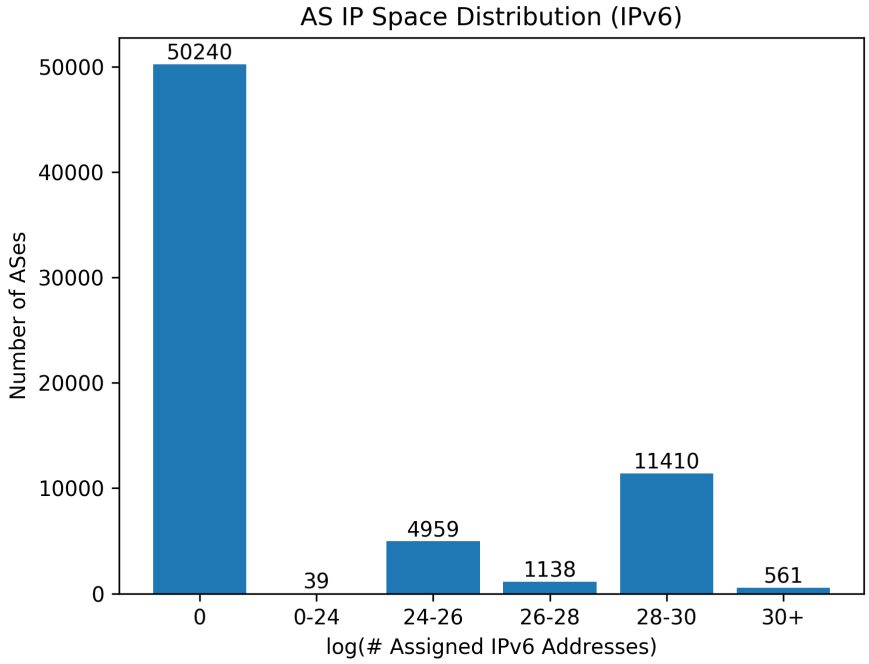

Figure 4: Histogram of AS IPv6 Space

A striking difference between IPv6 and IPv4 distribution is the number of nodes with zero allocated addresses, and the logarithmic scale required to show the others. The 50,240 ASes without any allocatable IPv6 addresses can be blamed on how recent the protocol was implemented and what it is for. IPv6 is mostly used to describe devices belonging to individual people, as they were the main contribution to the shortage of IPv4 addresses. This means the main consumers of these addresses are Enterprise ASes rather than Transit. As such, Transit ASes and most Content ASes have no use for IPv6 addresses. The ridiculous number of addresses given to the remaining minority is due to the fact that IPv6 addresses are 128-bits long, which results in exponentially larger address spaces compared to the 32 bits of IPv4.

3.3 A More In-Depth Classification

The pie chart from Figure 1 on page 2 gives insight into the distribution of classification of AS nodes. However, it fails to state anything about each classification's role in the topology of the Internet. In order to do this, the classifications were split up once more in terms of their roles with connections and customers. Enterprise nodes were split into those that have customers and/or peers, and those that do not. Transit ASes were split by those without customers, and those with at least one. Finally, Content ASes were split between those with no customers and at least one peer, and others. The results for this are shown below in Figure 5.

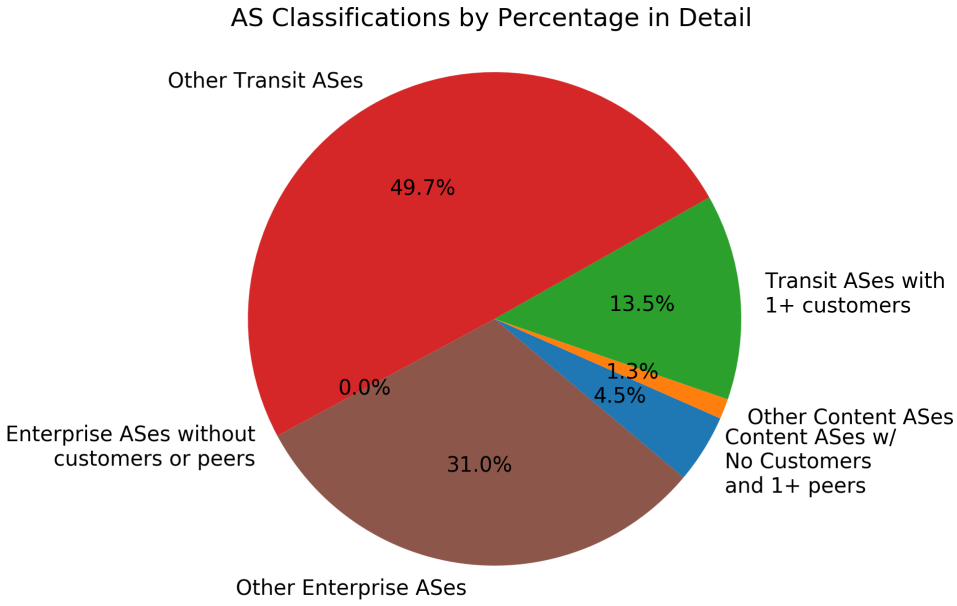

Figure 5: A More Detailed Overview of AS Classification by Percentage

The new classifications chart has several implications for the topology of the Internet. For Enterprise nodes, it shows that every single one must have at least one peer and/or customer. Transit ASes are a bit more interesting in that the majority of them, equivalent to almost half of all ASes, have no customers. This makes sense as the point of transfer nodes is not to be a provider, but rather to transfer data by connecting nodes to each other. Lastly, Content ASes are mostly made up of nodes that also only have peers and not customers. As such, Content ASes are mostly made to provide information to transfer nodes so they can send it off to Enterprise nodes and subsequently the general population.

4 Inference of T1 ASes:

The table in Figure 6 shows the top 10 Tier 1 ASes ranked on their degree. The Tier 1 ASes were determined by computing the largest clique, using a simple greedy heuristic. The heuristic is outlined below:

- Rank all the ASes by their degree, and place in a set $R = \{AS_1, AS_2, AS_3, ...\}$.
- initialize the clique, S with the AS_1.
- if AS_2 is connected to AS_2, 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.

NOTE: due to the fact that it was not possible to fill a clique of at least 10 ASes using the above greedy heuristic, the first 50 ASes that did not meet the criteria in R were ignored, before terminating the algorithm.

rank	degree	AS name
1	8052	Hurricane Electric LLC
2	5311	Level 3 Parent, LLC
3	4042	SG.GS
4	3207	Angola Cables
5	2961	Onecom Global Communications LTD
6	2885	Telekomunikasi Indonesia (PT)
7	2843	Convergenze S.p.A.
8	2498	COMMCORP COMUNICACOES LTDA
9	2435	Digital Telecommunication Services S.r.l.
10	2403	Neterra Ltd.

Figure 6: Top 10 Inferred Tier1 ASes based on degree

The table shows that Hurricane Electric LLC is the leading Tier 1 AS simply because Hurricane Electric LLC has the largest degree out of every AS, and therefore was put into the clique initially before any other AS. **The size of the clique is 35 different ASes**.

5 Customer Cones and AS Ranks:

The final method of information extrapolation was to use the concept of "customer cones" to show the extent of AS connections. A sample of ASes and their customer cones were taken and displayed in the following two tables. The first is the top 15 ASes in terms of the number of other ASes that they reach.

rank	AS	AS name	AS degree	ASes	IP Prefix	IPs	% ASes	% IP Prefix	% IPs
1	3356	Level 3 Parent, LLC	5311	58506	817698	3095690514	85.7	98.5	72.1
2	1299	Telia Company AB	1928	55842	790497	3096515651	81.7	95.2	72.1
3	174	Cogent Communications	5969	53542	746690	2913394262	78.3	89.0	67.8
4	3257	GTT Communications Inc.	1985	53205	761889	2919819429	77.8	91.8	68.0
5	6762	TELECOM ITALIA SPARKLE S.p.A.	515	49312	714088	2727472589	72.1	86.0	63.5
6	3491	PCCW Global, Inc.	687	48667	713608	2690222857	71.2	85.9	62.6
7	2914	NTT America, Inc.	1746	47136	669106	2481638038	69.0	80.6	57.8
8	196753	Fiber Telecom S.p.A.	291	46789	679339	2566712808	68.5	81.8	59.8
9	1299	Telia Company AB	71	46786	679325	2566705384	68.5	81.8	59.8
10	6939	Hurricane Electric LLC	8052	46697	678646	2563745512	68.3	81.7	59.7
11	6461	Zayo Bandwidth	1979	45781	655588	2324588913	667.0	79.0	54.1
12	6453	TATA COMMUNICATIONS (AMERICA) INC	724	45369	671179	2615130319.0	66.4	80.8	60.9
13	701	MCI Communications Services, Inc. d/b/a Verizon Business	1397	44924	681151	2589625875	65.7	82.0	60.3
14	1273	Vodafone Group PLC	311	42052	632531	2259185450	61.5	76.2	52.6
15	3549	Level 3 Parent, LLC	2199	41782	608041	2258708726	61.1	73.2	52.6

Figure 7: Top 15 ASes ranked by the customer cone in number of ASes that they can reach using p2c links.

The table shows that the top ASes reach out to the majority of other ASes. Likewise they also reach most IP prefixes and allocatable IPs. This makes sense as, being the backbone of the internet, most networks need to have some connection to them by some means. The second table, which displays the top 15 ASes by rank in reachable IP addresses, also shows this trend.

AS rank	AS	AS name	AS degree	ASes	IP Prefix	IPs	% ASes	% IP Prefix	% IPs
1	3356	Telia Company AB	1928	55842	790497	3096515651	81.7	95.2	72.1
2	1299	Level 3 Parent, LLC	5311	58506	817698	3095690514	85.6	98.5	72.1
3	174	GTT Communications Inc.	1985	53205	761889	2919819429	77.8	91.8	68.0
4	3257	Cogent Communications	5969	53542	746690	2913394262	78.3	89.9	67.8
5	6762	TELECOM ITALIA SPARKLE S.p.A.	515	49312	714088	2727472589	72.1	86.0	63.5
6	3491	PCCW Global, Inc.	687	48667	713608	2690222857	71.2	85.9	62.6
7	2914	TATA COMMUNICATIONS (AMERICA) INC	724	45369	671179	2615130319	66.4	80.8	60.9
8	196753	MCI Communications Services, Inc. d/b/a Verizon Business	1397	44924	681151	2589625875	65.7	82.0	60.3
9	1299	Fiber Telecom S.p.A.	291	46789	679339	2566712808	68.5	81.8	59.8
10	6939	Telia Company AB	71	46786	679325	2566705384	68.5	81.8	59.8
11	6461	Hurricane Electric LLC	8052	46697	678646	2563745512	68.3	81.7	59.7
12	6453	NTT America, Inc.	1746	47136	669106	2481638038	69.0	80.6	57.8
13	701	Sprint	372	39590	608540	2422883862	57.9	73.3	56.4
14	1273	Zayo Bandwidth	1979	45781	655588	2324588913	67.0	79.0	54.1
15	3549	Vodafone Group PLC	311	42052	632531	2259185450	61.5	76.2	52.6

Figure 8: Top 15 AS ranked by customer cone in percentage of IP addresses they can reach using p2c links.

ASes found in this table can reach insane numbers of single IP addresses. Multiple duplicates to the previous table appear here, as is to be expected. However, some new parties also show up. These new groups include ISPs such as Sprint which directly provide addresses to consumers.