Pakistan Journal of Science, 71 (4 Suppl.): 2019, Page 69-73
ISSN: 2411-0930

COLLECTION OF AUTONOMOUS SYSTEM LEVEL TOPOLOGY USING LOOKING GLASS SERVERS

R.F. Ahmed[1,*], A. Khan[2] and M. Rizwan[3]

[1]NCBA&E, Lahore. [2]IUB, Bahawalpur. [3]KFUEIT, Rahim Yar Khan.
[*]Corresponding author's E-mail: ranafarazahmed@gmail.com

ABSTRACT: Designing Internet Topology is a task of high intensity. As, Internet is not owned by any standard body, therefore, the topology through which AS - Autonomous Systems, around the world interacts, varies on the basis of technical and socio-political circumstances. Ever increasing number of Internet users, are portraying enormous challenges to the Internet Service Providers. In this research work, we collected Border Gateway Protocol routing tables of AS from LGS - Looking Glass Servers. As an instance, we queried multiple LGS and compared the topological dataset with other available datasets. We found 378 new AS and 1740 AS links. We proved that collecting traces of Border Gateway Protocol from LGS can assist to improve the contemporary topology at AS level that is developed by BGP based projects.

Keywords: Internet Topology, Looking Glass Server, Autonomous System, Inter-domain Routing, Border Gateway Protocol.

INTRODUCTION

Internet is a heterogeneous network of inter-connected devices. It can be referred as Autonomous Systems connected with each other that use border gateway protocols for the exchange of information. In this network, each Autonomous System is a node and BGP connections among those nodes are known as links, as shown in figure 1. The importance of AS topology construction is highlighted through many studies (De-Domenico and Arenas, 2017; Khan *et al*., 2013; Tozal, 2016; Sharif and Adam, 2015). This research, intents to design an Autonomous System Level Topology using Looking Glass Server. As, Internet is not owned by any standard body, therefore, the topology through which AS around the world interacts, varies on the basis of socio-political circumstances. In this research, we collected Border Gateway Protocol routing tables of AS from LGS. As an instance, we queried multiple LGS and compared the collected topological dataset with other available datasets. We found 378 new ASN and 1740 AS links.

Due to the regularly expanding client-base of web, its ecosystem is persistently advancing to address the challenges. Extraordinary changes in the customary framework of Web have enhanced its execution, capacity and have enabled a world of new services. For Network Providers and Internet Service Providers (ISPs), encountering the rapidly emerging issues has become a technical and practical challenge.

According to the Regional Internet Registries

Statistic shows billions of Internet Protocol Addresses; version 4 and 6; have been assigned to 82,136 Autonomous Systems worldwide, as of Mon, Sep, 18, 2017, which are inter-connected by billions of intelligent and physical links, while till Mon July 08 2019, the number of AS has grown up to 91, 748. The figure 2, depicts the ASN Chronology in World zone while figure 3, portraits ASN Statistics by country in World zone. These ASN are further divided on the basis of regions i.e. AFRINIC (AFRINIC-RIRs), APNIC (APNIC-RIRs), ARIN (ARIN-RIRs), LACNIC (LACNIC-RIRs), RIPE NCC (RIPE NCC-RIRs). These statistics shows how fast the Internet is growing. Despite the continuous efforts of the research and commercial communities, to create a topological design of Internet and understanding its structure is a great challenge. Building efficient topology of the Internet at Autonomous System Level is an assignment of great significance for future protocol outlines, execution assessment, examination and reenactment.

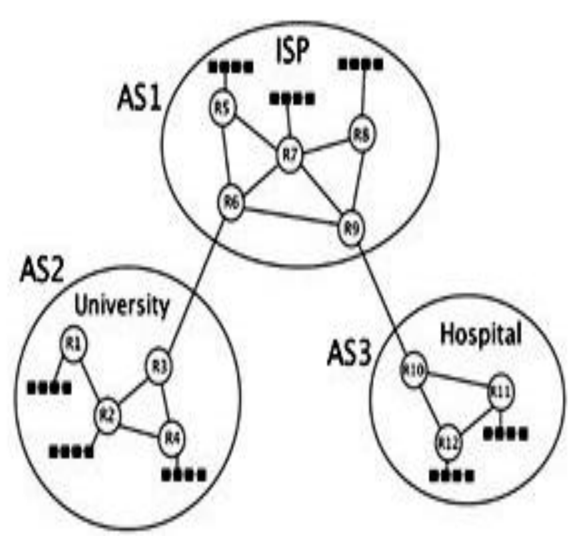

Figure 1: AS Level Topology

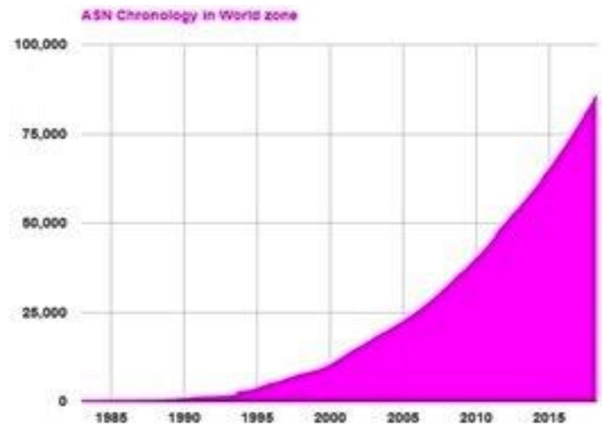

Figure 2: ASN Chronology in World zone.

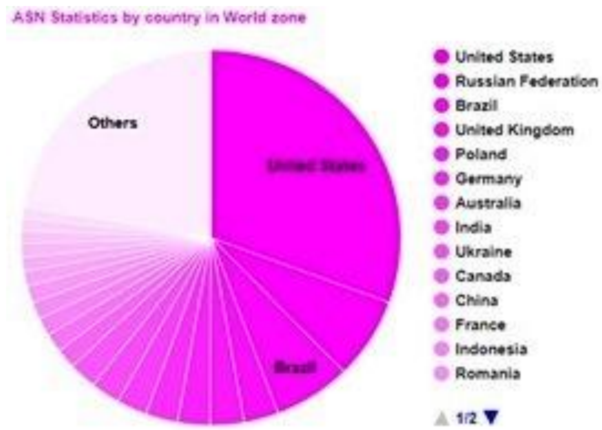

Figure 3: ASN Statistics by country in World zone.

In this research, construction of the Autonomous System level topology is focused with the help of LGS. LGS are web applications, being managed by (NOCs) Network Operation Centers or (ISPs) Internet Service Providers that facilitates to analyze the network status i.e. path, connectivity and information of routing (Khan *et al*., 2013). 378 distinct AS Numbers and 1740 AS links were found from multiple LGS across hundreds of countries. We wrote a script that automates the process of querying to LGS. Then, this LG based topology was compared with other available datasets and as a result we found many new ASN and links between them. These links and ASN were missing in the available datasets i.e. IRR (Internet Routing Registry), BGP (Internet AS-level Topology Archive., 2019), Ark (Center for Applied Internet Data Analysis) and iPlane (Madhyastha *et al.,* 2006).

MATERIALS AND METHODS

A Looking Glass is a web-based portal running on a web server that allows external users to get a look at routing and network behavior as it originates from the remote network. The importance of LG servers for AS topology construction has been highlighted in As-level topology collection through looking glass servers (Khan, *et al.,* 2013). In this research work, an application has been developed to automate the process of requesting/querying the LG servers, sequentially. Many LG servers do not entertain automated scripts/applications unless it is authorized by the relevant authorities like Sprint (Sprint.net IPv4/IPv6 Looking Glass., 2019).

In this research, Looking Glass Servers are used for the development of Autonomous System Level Topology. Links between Autonomous Systems have been gathered using ethical web scraping techniques, from Looking Glass Servers that are running in various nations, world-wide. Then, AS level topological-base is designed using LGS. A web application is developed to mechanize the process of querying to multiple LG servers, sequentially. The collected datasets are analyzed in proportion of other datasets to enlighten the uniqueness and contributions of this research work.

After scraping the LGS, we extracted ASN from the returned string and there is much more that can be explored. For instance, we have presented two most famous LGS i.e. Hurricane Electrics (Tozal, 2017) and PCH - LGS (Packet Clearing House., 2019). As shown in the Figure 4, ASN of the neighboring nodes are available and we can extract them for topological designing. Once the results are stored into the DB - database in RAW format, further analysis can be done. When we scrape data from LGS, many autonomous system numbers were inserted into DB repeatedly. To know the actual number of AS, we applied SQL Query to extract the Distinct number of AS. If we talk about PCH-LGS significantly, we extracted 2804 distinct ASN (Table 1) by applying distinct query of SQL on the RAW data set. In the end, we compared dataset collected from a couple of LGS with available datasets, IRL and Ciada Ark. We found new 378 ASN and 1740 AS links that were missing in other datasets. As shown in the Figure 4, we can see that there is a lot more information coming as a response against our query made to Looking Glass servers. Our concern is with Autonomous System Numbers to the extent of this research work. By extracting the results, we can know the total live ASN in the world and an Internet topology can be made by linking each neighboring node to local ASN from which the request was generated. The IP addresses of Neighbors can assist in building the router level Internet topology, as well.

In Table 1, we can see topological design of ASN extracted from Hurricane Electric LGS (Electric). Our script scrapes the data and stores the ASN into the database after applying REGEX filters on the string. This technique can be used with all LGS to develop real-time Internet topology.

Table 1: Hurricane Electric - AS Level Topology.

Local ASN	Neighboring Nodes
6939	1200
6939	26496
6939	29561
6939	8302
6939	75641

```
1) Router Name: MyIX, Kuala Lumpur, Malaysia:
Result
BGP router identifier 74.80.94.4, local AS number 3856
RIB entries 1229390, using 75 MiB of memory
Peers 60, using 148 KiB of memory
Peer groups 8, using 128 bytes of memory

Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
74.80.94.1      4    42    2534    2304        0    0    0 1d14h20m        2
206.220.231.55  4  3856    2297  307257        0    0    0 1d14h20m        0
218.100.44.67   4 24218    5094    4604        0    0    0 1d14h20m      692
218.100.44.76   4 24028    2611    2304        0    0    0 1d14h20m       81
218.100.44.80   4  9534    2343    2305        0    0    0 1d14h20m      155
218.100.44.84   4 10204    2568    2304        0    0    0 1d14h20m       41
218.100.44.89   4 38199       0       0        0    0    0 never    Active
218.100.44.92   4 15169    5179    4605        0    0    0 1d14h20m      445
218.100.44.101  4 55799       0       0        0    0    0 never    Active
218.100.44.102  4 45352    2563    2304        0    0    0 1d14h20m       67
218.100.44.104  4 45960    2556    2305        0    0    0 1d14h20m       99
218.100.44.109  4 55720    2308    2304        0    0    0 1d14h20m       41
218.100.44.111  4 45352    2558    2304        0    0    0 1d14h20m       81
218.100.44.112  4  9930    2382    2305        0    0    0 1d14h20m      202
218.100.44.121  4 131310       0       0        0    0    0 never    Active
218.100.44.126  4 55420    4844    4605        0    0    0 1d14h20m        1
218.100.44.132  4  8674    2304    2304        0    0    0 1d14h20m        5
218.100.44.138  4 55822    6863    2305        0    0    0 1d14h20m     2155
218.100.44.139  4 55822    6862    2305        0    0    0 1d14h20m     2154
218.100.44.144  4 132111    2536    2304        0    0    0 1d14h20m       47
218.100.44.146  4 32934    4776    4605        0    0    0 1d14h20m       30
218.100.44.157  4 56286    2665    2305        0    0    0 1d14h20m        4
218.100.44.160  4 24514    2560    2305        0    0    0 1d14h20m       47
218.100.44.163  4 132435  364988    2304        0    0    0 1d14h20m   671499
218.100.44.164  4 132514       0       0        0    0    0 never    Active
218.100.44.173  4 58910    4496    2304        0    0    0 1d14h20m        8
218.100.44.177  4 56109    5255    2304        0    0    0 1d14h20m     3479
218.100.44.185  4 13335    4785    4605        0    0    0 1d14h20m      219
218.100.44.188  4 133421    2536    2305        0    0    0 1d14h20m       37
218.100.44.195  4  9930    2386    2305        0    0    0 1d14h20m      215
218.100.44.198  4 134809    4778    4605        0    0    0 1d14h20m       14
218.100.44.213  4 15932    4666    4605        0    0    0 1d14h20m      224
218.100.44.220  4 45102    7623    6902        0    0    0 1d14h20m        1
218.100.44.221  4 45102    7630    6902        0    0    0 1d14h20m        1
218.100.44.223  4 135553    2541    2304        0    0    0 1d14h20m       12

Total number of neighbors 35
```

Figure 4: Response from PCH LGS.

Topology derived from LG servers: We wrote a python script to automate the querying process. Our tool issues query to multiple LG servers e.g. Hurricane Electric (Electric) and PCH (Packet Clearing House). Our script is time barred and starts every loop after an interval of 1 minute to the same LG server to avoid overloading them. It's a multi-step process to collect data from LGS. By querying LG servers, we became capable of running **show ip bgp summary** command from 150 locations distributed among 90 countries, once a week in the month of April 2019. We collected about 7 snapshots to develop an AS link dataset, consisted on about 50,000 AS links. We explored that about 90% of the links are intra-AS links that means the source and destination AS of a link are the same. Because this research work focuses on inter-AS links, we filter out intra-AS links and selected 9,000 inter-AS links. Overall, we collected total 60,000 ASN and 23,000 distinct ASN among them.

Level topological datasets: There are many other Autonomous System Level Topologies. Some of those are:

- Traceroute based Ark - CAIDA Archipelago (Center for Applied Internet Data Analysis) portraits the topology of AS that is derived from traceroute based measurements.
- iPlane (Madhyastha *et al.,* 2006) service executes traceroutes from hundreds of PlanetLab locations daily to construct the topology of Internet.
- BGP based UCLA IRL (Internet AS-level Topology Archive), extracts Autonomous System topology from the Border Gateway Protocol traces and regularly publishes it. These traces are shared by Internet2 (Internet2), PCH (Packet Clearing House), RouteViews (University of Oregon Route Views Project), RIPE-RIS (RIPE).
- IRR (Internet Routing Registry) based, it is a system of distributed datasets and they contain IP Addresses and routing related information of Autonomous Systems.

All the aforementioned topologies collect their own datasets and construct topology on the basis of those. None of them uses the technique that we have implemented in this research work.

RESULTS AND DISCUSSION

Many service providers are collecting autonomous system numbers and links between them. The intention is to study the AS level topology. But the approach to collect them is quite different from the proposed one. Datasets collected from looking glass servers are real-time and updated. While the content of other datasets is said to be outdated and have been proved by many studies (Tozal, 2016; Tozal, 2017). Specifically, if some ASN is terminated from the network, there is no mechanism to eradicate it from the dataset other than collecting real-time ASN from looking glass servers. Therefore, to justify the authenticity of our work, we compared different datasets to analyze the differences and similarities among them. As an instance, we have collected ASN from multiple Looking Glass servers, the comparison results may not be quite surprising but to elaborate the process we have portrayed the comparison results of two most famous available topological datasets provided by Caida Ark (Center for Applied Internet Data Analysis) and IRL (Internet AS-level Topology Archive) collected from September 2017 to July 2018. We found 5790 unique ASN among them. After extracting the results, we compared our collected topological dataset with it and found 712 new ASN that were missing in Caida Ark and IRL datasets along with 2396 new links.

Else than collection, pre-processing of the dataset was another challenge. To compare the both available datasets with our collected topology, we pre-processed them and by using Python script, Regex and SQL queries, stored the ASN in a relational database so that different SQL queries can be applied on them to extract the desired results.

After applying DISTINCT and UNIQUE SQL queries we found 378 new autonomous system numbers that were missing in the datasets of Caida Ark and IRL. These new ASN were found in the topology collected by Looking Glass servers.

Overlapping and unique as links: After collection of AS level topology from LG servers, we compared AS topologies generated from different datasets, to find unique and overlapping AS links among them. Such analysis is essential to justify our work and to sort out how many AS links are newly discovered from the LG servers. Table 2, indicates the number of common AS links for each pair of datasets. The diagonal zone from upper-left to bottom-right (in bold) reports the count of AS links appearing in only one dataset, i.e. LGS, IRL, Ark. We collected 9000 unique AS links in the AS topology from the LG servers. We find 25000 unique AS links in IRL, it depicts that there are vast number of links that may not be observed from the BGP feeders of LG servers. The Ark has unique 32000 AS links, while many of these AS links are doubted not to be accurate either due to the traceroute measurement issues such as IP-to-AS mapping.

Finally, we analyze the differences in the AS connectivity of all the sources with respect to the AS connectivity of the AS observed IRL, that is known as the most accurate BGP-based AS level topological dataset. Table 3, shows most famous organizations in terms of the number of new AS links found from our dataset, column 2 contains the number of AS found in the dataset of IRL (Internet AS-level Topology Archive) and 3rd column have the newly discovered AS links from LG servers.

Overlapping and unique as: Table 4, shows the number of common AS for each pair of datasets. From upper-left to bottom-right in diagonal zone reports the count of AS appearing in only one dataset. We observe 378 unique AS from LG servers while only 258 in IRL. There are about 9 K unique AS in IRR. While no unique AS was found in Ark. It shows their absolute use of BGP traces, since Ark relies on BGP traces from RouteViews to convert IP-to-AS paths collected from traceroute measurements. In the end, we observed iPlane, which have 346 unique AS. We also find that these AS were not operational in the month of Nov 2018. For example, AS136274 (Cloud Servers Pvt Ltd) had been visible until March 01, 2017 and AS137258 (Shirazi Investments Pvt Ltd.) had been visible until April 17, 2016, and AS138590 (Prime Networks) had been visible until Dec 29, 2016 but not anymore. The reason why we observed 346 unique AS in iPlane is because it uses BGP traces that contains older data collected from Dec 2010 to Jan 2018 during the IP to AS mapping process, constructing its AS level topology.

Table 2: The number of similar and distinct Autonomous System links between different topological datasets.

Source (AS)	LGS	IRL	Ark
LGs (30 K)	9 K	35 K	50 K
IRL (65 K)	35 K	25 K	31 K
Ark (51 K)	50 K	31 K	32 K

Table 3: Most famous organizations of Pakistan and their newly discovered Autonomous System level connections located through LGS.

Source (AS)	LGS	IRL	Ark
LGs (30 K)	9 K	35 K	50 K
IRL (65 K)	35 K	25 K	31 K
Ark (51 K)	50 K	31 K	32 K

Table 4: Overlapping and unique (in bold) AS found in various AS topology datasets.

Sources (AS)	LGs	IRL	IRR	Ark	iPlane
LGs (18.5 K)	378	57 K	34.3 K	39.3 K	54.3 K
IRL (50.2 K)	55.3 K	258	33.9 K	36.4 K	44.1 K
IRR (41.9 K)	25.3 K	58.7 K	9 K	26.2 K	88.2 K
Ark (41.1 K)	39.3 K	36.4 K	29.1 K	-	52.6 K
iPlane (37.1 K)	19.7 K	44.4 K	21.2 K	30.0 K	346

Conclusion: In this research work, the less-known abilities of Looking Glass servers are depicted with the help of which we can model the topology of the Internet. By running the show ip bgp summary command from multiple Looking Glass servers, we collected thousands of AS links and numbers. By collecting this dataset, we constructed an AS level topology. It is absolute that Looking Glass servers assists in augmenting the present AS topology collection efforts reliably as BGP based techniques are less error prone in proportion to the ones that are traceroute-based. Perhaps, it can be assumed that more Looking Glass servers will be deployed on the Internet in near future as it is an important operational tool of AS. So, we believe that the community of research needs to be aware about the abilities and facilities provided by the LGS to explore and construct the Autonomous System Level Topology.

Future work directions: As future work, the script (Ahmed., 2019) that we developed to automate the querying process can be further improved, so that it can scrape data more efficiently. Data analysis can be done in contrast to other available topological datasets with a larger LGS dataset, while, the response that we receive by querying LGS, contains information other than just ASN. Further experimentation and research can be done that how the rest of the data can help in making more accurate topological design of Internet.

Acknowledgement: I, really admire the cooperation and supervision made by my honorable mentor, **Dr. Akmal Khan**. He has been an ultimate source of guidance, throughout this research work.

REFERENCES

De-Domenico, M. and A. Arenas (2017). Modeling structure and resilience of the dark network. *Physical Review E*, *95*(2), 022313.

Khan, A., H.C. Kim, T.T. Kwon and Y. Choi (2013). A comparative study on IP prefixes and their origin ases in BGP and the IRR. *ACM SIGCOMM Computer Communication Review*, *43*(3), 16-24.

Khan, A., T.T. Kwon, H. Kim and Y. Choi (2013). AS-level Topology Collection Through Looking Glass Servers. In IMC (Vol. 13, pp. 23-25).

Madhyastha, H.V., T. Isdal, M. Piatek, C. Dixon, T. Anderson, A. Krishnamurthy and A. Venkataramani (2006). iPlane: An information plane for distributed services. In *Proceedings of the 7th symposium on Operating systems design and implementation* (pp. 367-380). USENIX Association.

Sharif, S.M. and H.A. Adam (2015). Internet Autonomous Systems (ASes) Level Topology Modeling. University Of Khartoum Engineering Journal, 4(2).

Tozal, M.E. (2016). The Internet: A system of interconnected autonomous systems. In *2016 Annual IEEE Systems Conference (SysCon)* (pp. 1-8).

Tozal, M.E. (2017). Autonomous system ranking by topological characteristics: A comparative study. In 2017 Annual IEEE International Systems Conference (SysCon) (pp. 1-8).