Educational Administration: Theory and Practice
2024, 30(4), 6179-6184
ISSN: 2148-2403
https://kuey.net/

Research Article

Classification Of Topology For The Internet Autonomous Systems: A Study Based On The Implementation Of Internet Autonomous System

Cao YuanQing[1]* , Dr.Midhunchakkaravarthy[2]

*Research Scholar Lincoln University College Malaysia

Citation: Cao YuanQing et.al (2024) Classification Of Topology For The Internet Autonomous Systems: A Study Based On The Implementation Of Internet Autonomous System, *Educational Administration: Theory and Practice*, 30(4),6179-6184
Doi:10.53555/kuey.v30i4.2353

ARTICLE INFO

ABSTRACT

Internet routing protocols used by autonomous systems (AS) include the Border Gateway Protocol (BGP). Multiple criteria for export and import make up an AS's routing table. Many sites have suggested heuristic ways for assuming AS connections from publicly available BGP data, because of the public nature of connection unravelling. A material Delivery Network (CDN) comprised of geographically dispersed servers is a nod to the value of content consumers. It is crucial to examine the efficiency with which visitors are sent to servers that are not controlled by content providers as content delivery traffic now constitutes most of the Internet traffic. Among content delivery network (CDN) services, Akamai and Netflix rank high. The user experience is impacted by the performance of the servers in the content delivery network. This necessitates that CDNs choose the most suitable server each time a user requests content from their network. Due to the lack of authentication in BGP routes, prefixes may be taken over by ASes that do not belong to them. Approaches to address this include detecting hijacks after they have taken place and responding to them after the fact. More effective preventive measures are required to stop it in its tracks. A list of repeat hijackers that was suggested in a recent paper could make this approach possible. To make matters worse, very few individuals take over again.

Keywords *Autonomous System, Topology, Border Gateway Protocol, RIS, PCH,*

1. INTRODUCTION:

The connectedness of the Internet makes it possible to link many autonomous systems, including tens of thousands of Cases that are administered by a variety of administrative authorities. The behaviour of ASes in their interactions with one another is managed by the BGP protocol. Using BGP, every autonomous system (AS) can choose the routes that it imports and exports from its neighbours. AS relationships, which may be thought of as a kind of commercial arrangement between ASes, are the driving force behind these regulations, which are determined upon by administrators of the network. Two of the most common types of peer-to-peers (p2p) connections are known as peer-to-peer (P2P) partnerships and peer-to-consumer (P2C) partnerships. The customer is the one who is obligated to take on the financial responsibility for the function that the service provider plays in the process of transporting communications between the Internet and other networks. Even though it is possible for two ASes to freely exchange traffic with one another and with their customers, this is not permitted between the ASes and their providers or with other organisations that are regarded by peers. It is common practice for an AS to refrain from exporting its provider and peer routes to its peers or providers due to the economic paradigm that is observed. As a logical consequence of the economic paradigm that they are addressing, this is the conclusion. It is necessary to understand the commercial connections that exist between applications servers (ASes) to have a firm grip of the structure, inter-domain routing dynamics, and expansion of the Internet. Researchers on the Internet have a difficult time effectively detecting AS relationships since business links are kept as a highly guarded secret inside firms **(Akgun, 2020).**
The collecting of BGP route measurements, which is a main driving element, is the fundamental factor that is responsible for the inference of AS linkages. Because of these efforts, two projects have been formed: Route

Views, which was developed by the University of Oregon in the United States, and Routing Information Service (RIS), which was developed by RIPE in Europe. Both projects are examples of how these efforts have resulted in developments. In each of these endeavors, substantial examples of how these efforts have been used have been provided. It is the responsibility of route collectors located in various regions of the world to ensure that BGP peering sessions with other ASes are maintained. The sessions that are being discussed here are known as Vantage Points. The first appearance of VP AS numbers is the most apparent in the AS pathways that are gained by these collectors who peer with the VP in the first place. These collectors are the ones who get the AS paths. Every day, the Route Views and the RIPE route collectors listen to the BGP routing table entries of these virtual private networks (VPs) and archive each of these entries, respectively. Both processes are performed by Route Views. Peering sessions may be established between the route collector and the virtual private network (VP) in several different ways. These include directly connecting with one another or via an Internet Exchange Point (IXP). Using IXP, which is like a centralized portal, it is feasible for ASes to communicate with one another and exchange information with one another. Measures that are deemed passive are those that do not need the addition of any new traffic to the network. Passive measures are also known as passive measures (H. Haddadi, 2019).

2. BACKGROUND OF THE STUDY:

Regrettably, the phenomenon of domain prefix theft is quite widespread in the realm of the internet. It is possible for prefix hijacking to occur for several reasons, including unintentional BGP route misconfigurations and malicious prefix hijacking, the latter of which may result in significant financial losses, disruptions to service, and even breaches of privacy. It is possible that these hijackings affect every AS on the planet, or they may merely affect a small fraction of ASes in a certain region. The ASC 4391" The BGP upgrade that was implemented in response to PCCW Global has now propagated over the whole of the internet (AS3491). Because of this, Pakistan Communications Corporation was able to deceive YouTube users located all over the different countries. Because of this, the siege continued for a total of two hours. The number of unassigned prefixes that were produced by China Net (AS23724) in the year 2010 was about equivalent to the size of the whole global routing table. China Net is a China Telecom network that normally creates around forty prefixes that are assigned individually. These prefixes are only available to a small number of networks that are located outside of China, which is a grateful development. A previous event occurred in April 2010, when a China Telecom AS hijacked more than 50,000 prefixes, which resulted in a deviation of fifteen percent of the Internet traffic for a period of fifteen minutes. During the approximately five minutes when CDTDBC (Compamia de Telecommunicators do Brasil Central AS16735) exposed its whole prefix database to its upstream providers, similar attacks took place. The 2018 3ve takeover was a planned criminal takeover that aired advertisements for actual companies and the United States Air Force, resulting in the theft of $29 million in phony ad money. It is possible that these hijackings were inadvertent; however, it is more likely that they were purposeful. The fact that this illegal enterprise was allowed to continue for a whole year is a very unfortunate development. The theft of more than 1.5 million IP addresses made this possible. Because there is no means to track the routes, most of these issues over traffic diversion are triggered by unintentional misconfigurations of the BGP traffic routing protocol. These threats have been mitigated by cryptographic methods like Secure-BGP, which restrict route broadcasting to just those routes that have been verified as legitimate. One of the most significant challenges that has been encountered is the high computational and storage expenses that are involved with the installation of public key infrastructure (PKI). One novel approach to reducing the number of anomalies in routing is to ignore routes that seem to be suspicious. When it comes to dealing with outliers, it could be challenging to have strategies in place **(A. Baumann, 2020).**

how the link between several fields evolves over time and how it functions. When operators are trying to determine the origin of BGP misconfigurations, they may look at the topology of the association. Using Argus, the instances of prefix hijacking may be quickly discovered and categorised into a wide variety of distinct groupings. This programming also provides a possible reason for the hijacking of the aircraft. Since the hijacked AS routers were unable to interact with the hosts in the victim prefix, a solution was established. Using ARTEMIS, administrators of networks can manage anomaly detection systems and attach their own conditions to attacks. The results of a recent research successfully classified cases of BGP hijacking into four categories: typos, prepending difficulties, origin alterations, and counterfeit AS routes. The accuracy of this classification was 95.71 percent. These strategies for detecting hijacking of the control plane BGP connection make use of BGP metrics at the autonomous system (AS) level. To carry out testing, it is necessary to use tools such as traceroutes to measure the data plane. According to the findings of a recent survey, most users (78.6%) make use of hijacking detection software that is provided by a third party. BGPmon should be considered one of the top monitoring tools. The use of phishing by around 17.3 percent of networks is said to have resulted in the disclosure of sensitive information by individuals who were members of email lists such as NANONG. A novel approach to dealing with the problem of BGP hijacking has been proposed by researchers in a study that was only recently released. To identify repeat offenders, this approach involves the examination of BGP data that was collected at intervals of five minutes between the years 2014 and 2018. Because of this, they can comprehend the patterns that these ASes progressively leave behind. Using the NANOG mailing list and

MANRS, they painstakingly assemble a total of twenty-three serial hijackers. Additionally, they make use of a machine learning classifier that was constructed with the use of data that was taken from the routing parameters of ASes. Using this technology, network operators were able to identify potentially malicious patterns within routing data and proactively prevent BGP hijacks that are produced by repeat hijackers.

3. PURPOSE OF THE RESEARCH:

Since IP addresses are dynamic and must be able to accurately reflect the current location of a device inside the Internet, it is not possible to issue IP addresses in the factory way MAC addresses are. Using Dynamic Host Configuration Protocol (DHCP), a MAC module can acquire an IP address for itself. This allows the module to accurately reflect its current location in respect to the topology of the Internet. "I would appreciate it if they could provide my MAC module with an IP address," DHCP asks. When a device's MAC layer module wishes to connect to a new media, the device invoke the Dynamic Host Configuration Protocol (DHCP). The DHCP module in question is responsible for sending out a request that contains the MAC address of the MAC layer module. The MAC layer module then takes that address and broadcasts it to all the other devices that share that physical media segment. In a unicast DHCP response from a DHCP server, a digital media access control (MAC) address is assigned to an IP address. After that, the DHCP client module sends the newly assigned IP address to the network layer so that it may be entered into the translation table. The answer contains several different network setup settings, such as the maximum datagram size, the addresses of other servers (such as DNS servers) that transform human-readable names into IP addresses, and the binding of an IP address to the MAC module that is used by DHCP. In addition to this, it contains the address of one or more routers, which enables the user to access any location they want.

4. LITERATURE REVIEW:

The entries in the BGP routing table are the most essential considerations when trying to ascertain the topology of an ASN. The presence of AS paths connecting the two sites may be determined using routing information from a nearby virtual private network (VP) and an IP address prefix block. This graph, known as the AS Graph, was used by them to simplify matters. Each of the AS in this collection has its own unique series of connections, represented here as nodes and edges. Annotated AS graphs may have new kinds of AS graphs created by linking AS nodes along their edges. One example of this is the p2c and p2p linkages **(E. Gregori, 2020).**
Multiple sources make it feasible to get BGP routes. Packet Clearing House (PCH) monitors IXPs worldwide and is one of the most outspoken supporters of IXPs. The goal of the several databases that make up Internet Routing Registries (IRRs) is to provide a complete picture of the network. Compared to links from this registry, there are more typical types of links. The RADb databases are managed by both individuals and organizations such as RIPE and Merit Network. Extracting AS connections is achievable given a set of IP addresses. The traceroute and tracer commands are often used for network troubleshooting. When a router responds to an ICMP command, the protocol's responsibility is to give a list of IPv4 addresses linked with those routers. Datasets for IPv4 Routed /24 AS Connections were generated using traceroute data taken by CAIDA's Archipelago (Ark) monitoring system. This is accomplished by converting IP addresses to their respective ASes and then building AS connections. Implementing traceroute methods into the network significantly increases the quantity of traffic that is monitored. They may check for the presence of an AS connection primarily using BGP routes, Traceroute, or IRR. An in-depth familiarity with the Internet's design and the peering agreements between ASes is essential for comprehending the many transit providers from surrounding areas. Simply counting the number of passengers carried by a transport provider AS should give them a good idea of the number of people it serves. customers whose trajectory is dropping are known as the Customer Cone (CC) according to the AS. This is because it is a route of since-AS interactions; customers in this stage are declining. This information is valuable for several reasons. When deciding which networks to peer with, operators of internet exchange points may take client cone size into account. Peering network customer cone size changes may also be investigated, along with the relationship between the AS's peering activity and the size of the customer cone. An AS's CAIDA rating is proportional to the size of its client cone. Due to size differences across locations, the consumer cone cannot be an exact representation of the AS's real customer. Level 3, for instance, has widespread acclaim in the United States but almost little traction in Europe. Critical to understanding AS's stance is dissecting the consumer cone. These two couldn't be more different from one another **(V. Giotsas, 2019).**

5. RESEARCH QUESTIONS:

- Just how does the Internet function in reality?
- What relationships do the main players have with one another?
- How has the Internet evolved over time and in different parts of the world?

- Is it possible to predict how the Internet was evolve in the future based on what they know about its current state?
- How can they improve the Internet's design to accommodate the increasing volume of users?

6. RESEARCH METHODOLOGY:

For the purpose of constructing the Internet topology graphs, the UCLA Internet AS-level topology archive repository and the CAIDA AS relationships dataset was used. The two repositories were selected not only because of their vast BGP monitoring network, but also because they are the only public sources that maintain historical information going back to 1998. This is the reason why they were picked. These same sources have been used in a number of earlier studies that have been conducted on the topology of the Internet.
The majority of their information for identifying AS nodes and AS links were come from this source. BGP data gathered by a number of different BGP data collectors, such as RouteViews (96), RIPE (88), PCH (85), and Internet. Collectors are responsible for recording each and every BGP path advertisement that is either broadcast or received by the routers. The routing tables for all 133 collectors are obtained by UCLA, and the routes included inside the tables are used in the construction of two topologies. A further distinction lies in the fact that one topology makes use of only IPv4 addresses, whilst the other topology makes use of only IPv6 addresses. They made use of IPv4 topologies in order to demonstrate how the Internet has developed over time.
Through the use of the CAIDA AS relationship dataset, a category is allocated to each and every AS connection. This particular dataset contains AS links that are separated into two distinct categories: c2p and p2p. For the purpose of determining the connection type based on raw BGP route advertisements, the method described in is used. Due to the fact that the link inference approach was discovered to have an accuracy of 99.6% for c2p connections and 98.7% for p2p connections, it was chosen to make use of the CAIDA dataset.

7. CONCEPTUAL FRAMEWORK:

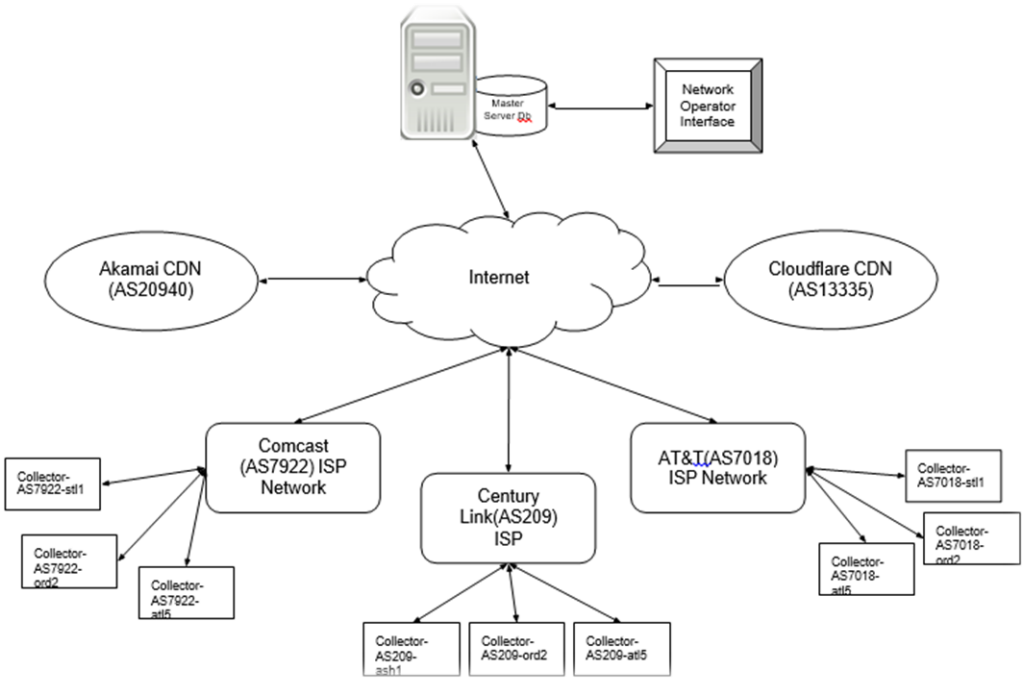

Table 1: The Ten Best Networks Relative to their Cones

8. RESULT:

They forecast 5,753 client cones for 3,290 PVPs worldwide (LDCCs) using a mix of population density and average journey times. Not only may PVPs produce LDCCs, but they can also aggregate all of the LDCCs produced by that PVP to form a CCC.

ASN	Our Cone Size	CAIDA Cone Size
AS3257 (1)	19,256	18,886
AS6762 (2)	12,379	14,319
AS6939 (3)	10,771	9,501
AS3491 (5)	7,825	4,561
AS1273 (4)	5,322	5,805
AS6461 (6)	4,330	4,415
AS9002 (7)	3,471	3,656
AS20485 (8)	2,883	3,153
AS12389 (9)	2,589	2,815
AS4323 (10)	2,265	2,288

Inferred AS relations are used to generate the CAIDA customer cone datasets. They compare the CCCs that their technique produces to the March 1, 2016 customer cone dataset that CAIDA released. They share 3265 common ASes with the CAIDA database. Their cone and the CAIDA cone had identical 2008 AS case percentages of 61.5%. Their smaller cone has 201 ASes, or 6.2% of all ASes, whereas CAIDA's bigger cone contains 1056 ASes, or 32.3% of all ASes. Since their identities are made public, the ten networks that were combined in may be considered full NSPs.

As a result of peering with PCH at several IXPs, 1158 ASes have multiple LDCCs. Six hundred eighty-one LDCCs (58.8%) have varying prefix counts, whereas 512 (44%) have varying sizes. So, different peering networks was often use different prefix sets and client cones. Network service providers, cable and digital service providers, and content providers make up the bulk of the 956 (or 92.5%) ASes that have their firm categories mentioned in PeeringDB out of 1158. Displays the number of ASes, the percentage of ASes with non-matching cone diameters, and the percentage of ASes with non-matching prefix counts at different locations for each of the three company kinds. It has been shown that NSPs have a vast array of cones and prefixes at various places. In most cases, the cones used by access providers and content providers are same; however, the prefixes might differ substantially. There are more than twice as many ASes using location-dependent prefixes as there are using location-dependent cones across all three categories of enterprises.

Table 2: The Three Business Types' Dependency on Customer Cone Location

Business Type	Count	Nonident. Cone	Nonident. Prefix
NSP	386	66.8%	76.2%
Cable/DSL/ISP	305	39.7%	54.7%
Content	193	29.5%	58.5%

As a result of peering with PCH at several IXPs, 1158 ASes have multiple LDCCs. Six hundred eighty-one LDCCs (58.8%) have varying prefix counts, whereas 512 (44%) have varying sizes. So, different peering networks was often use different prefix sets and client cones. The bulk of the 956 ASes (or 92.5% of the 1158 ASes they surveyed in PeeringDB) are identified as network service providers, cable and digital service providers, or content providers. To see how many ASes, what proportion of ASes have non-matching cone diameters, and what proportion of ASes have non-matching prefix counts at different places for each of the three company kinds. It has been shown that NSPs have a vast array of cones and prefixes at various places. In most cases, the cones used by access providers and content providers are same; however, the prefixes might differ substantially. There are more than twice as many ASes using location-dependent prefixes as there are using location-dependent cones across all three categories of enterprises.

9. DISCUSSION:

In networks that peer with PCH route collectors and broadcast customer routes to the collectors, they presented a technique for calculating customer cones. The truth is accurately reflected in the resultant customer cones since peering networks only broadcast their customer routes to the collectors. Using this

method, they can find the client cones connected to 112 IXPs around the nation for 3290 networks. They contrast their results with those of CAIDA, whose buyer's cone is constructed using p2c links uncovered by their AS relationship inference method. According to their findings, CAIDA's consumer demographics consistently leave out several important demographics. Half of the 1158 networks they discovered across several IXPs made use of many cones in various locations. Customer cones for NSPs were highest in the top ten ASes that they examined, with Cable following closely after. The client cone of a content provider is very tiny. There are many possible applications for their client cones. At an IXP, networks may peer with each other depending on a variety of factors, one of which is the customer cones of the other networks. Practical data. Second, monitoring the growth of customer cones in peering networks might provide light on the connection between AS peering behavior and the size of brand cones. The third possible area of investigation is the correlation between peering and deepening occurrences and the growth or decline of an AS's client base.

10. CONCLUSION:

Researchers offered a machine learning method to interpret AS graphs built from open-source data for inferring edge kinds. Together, the five node properties extracted from the AS graph and the Gentle AdaBoost machine learning method allowed for the development of a classifier for p2p and p2c edges. They apply their approach to the classification process in order to categorize three AS graphs: a BGP network, a traceroute graph, and an IRR graph. They use two datasets to evaluate each classifier. Both sets of tests are based on datasets that were created at CAIDA: one using the BGP dataset and the other using the AS connection inference dataset. The three independent AS graphs may be combined to get the edge types of an AS graph. They look at one composite graph, three separate graphs, and one combined graph to see what makes each one unique. An abundance of distinct edges linking p2p and c2c are present in all three graphs. By combining the three graphs, they can see the Internet's peer-to-peer and peer-to-consumer ecosystems much more clearly.

11. REFERENCES:

1. M. B. Akgun and M. H. Gunes. Bipartite internet topology at the subnetlevel. In Network Science Workshop (NSW), 2013 IEEE 2nd, pages 94–97. IEEE, 2020.
2. Baumann and B. Fabian. Who runs the internet? -classifying autonomous systems into industries. In WEBIST (1), pages 361–368, 2020.
3. V. Giotsas, M. Luckie, B. Huffaker, et al. Inferring complex relationships. In Proceedings of the 2014 Conference on Internet Measurement Conference, pages 23–30. ACM, 2019.
4. E. Gregori, A. Improta, L. Lenzini, L. Rossi, and L. Sani. On the incompleteness of the as-level graph: A novel methodology for bgp route collector placement. In Proceedings of the 2020 ACM Conference on Internet Measurement Conference, pages 253–264, 2020.
5. H. Haddadi, D. Fay, S. Uhlig, A. Moore, R. Mortier, and A. Jamakovic. Mixing biases: Structural changes in the as topology evolution. In Traffic Monitoring and Analysis, pages 32–45. Springer, 2019.