Available online at www.ijmras.com

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH AND STUDIES
ISSN: 2640 7272
Volume:07; Issue:01 (2024)

The Internet Autonomous System's implementation Categorization of Internet Autonomous Systems' Topology

CAO YUANQING[1], DR. MIDHUNCHAKKARAVARTHY[2a]
[1]PhD. Research Scholar in Information Technology, Lincoln University College Malaysia
[2]Professor in Lincoln University College Malaysia
Contact Details: [a]midhun@lincoln.edu.my

Abstract

Although the Internet has been researched since its inception, academics didn't start paying much attention to its growth and change until the mid-1990s, when it became economically viable in the United States. To begin with, it was Govindan and Reddy [16] who originally defined an AS level graph. The graph was shown as a set of nodes and linkages, with the nodes representing individual Internet domains. They found that even though the Internet has expanded greatly, the degree and route dispersion have not changed from pre-boom levels. However, in their seminal study [17], Faloutsos et al. introduced the term Internet topology. If the data from the BGP monitor was trustworthy and thorough enough, the researchers concluded that it allowed for the first full look at the AS topology. What they observed was that the distribution of AS degrees followed a simple power formula. Academic interest in Internet topology and data collection methods, graph construction, and graph approaches and analysis has increased since the publication of this influential study [18].

Keyword: RIS, PCH, AS, BGP

INTRODUCTION

Many autonomous systems are linked together through the Internet, including tens of thousands of Assess run by various administrative agencies. BGP determines how ASes communicate with each other (BGP). Using BGP, each autonomous system (AS) may choose the routes it imports and exports from its neighbours. AS relationships, a type of commercial arrangement between ASes, is what drives these policies, which are determined by administrators of the network. P2P and p2c partnerships are the two most common types of AS relationships (p2p). It is the customer's responsibility to pay for the service provider's role in transporting communications between the Internet and other networks. It is possible for two ASes to exchange traffic freely between themselves and their clients, but not between themselves and their providers or other peers. It is a logical consequence of this economic paradigm that an AS normally does not export its provider and peer routes to its providers or peers. Knowledge the Internet's structure, inter-domain routing dynamics, and evolution requires an understanding of the commercial connections between ASes. It is difficult for Internet researchers to correctly determine AS Relationships since business relationships are a well-kept secret within firms [1].

The collection of BGP path measurements is a major driving force behind the inference of AS relationships. Both of these initiatives were spawned as a result of these efforts: Route Views by the University of Oregon in the United States, and Routing Information Service (RIS) by RIPE in Europe, both of which are important examples of how these efforts have been used. BGP peering sessions with other ASes, known as Vantage Points, are maintained by route collectors located across the world (VPs). VP AS numbers appear first in the AS pathways gathered by these collectors that peer with the VP in the first place. Each day, the Route Views and RIPE route collectors listen to these VPs' BGP routing table entries and archive each of them on a daily basis. Peering sessions between the route collector and the VP can be formed either directly or via an Internet Exchange Point (IXP). An IXP is a centralised site where ASes may communicate with each other and exchange information. Passive measures are those that don't require any new traffic to be added to the network [2, 3].

Literature Review

It is the BGP routing table entries that are of main importance when attempting to determine the topology of an ASN. There are AS pathways leading from the adjacent VP's routing table towards a prefix block of IP addresses. For the sake of simplicity, we'll refer to this graph, which is termed AS Graph, for short. These nodes and edges show the links that exist between a set of autonomous systems (AS). Adding AS connections like p2c and p2p to the edges of an Annotated AS graph creates a new type of AS graph [4].

BGP routes can be obtained from other sources. One of the leading proponents of IXPs, Packet Clearing House (PCH) [3] is monitoring IXPs from all corners of the globe. Links from an Internet Routing Registry (IRR) [4], which is a collection of routing policy databases designed to give a comprehensive perspective of the Internet, are less common. Organizations like RIPE [5] and Merit Network [6] operate these databases, known as a Routing Arbiter Database (RADb). From a group of IP addresses, it is possible to extract AS linkages. For network troubleshooting, traceroute and tracert commands are commonly used. In response to the command, a list of IPv4 addresses associated with routers that have responded using the Internet Control Message Protocol is returned (ICMP). Traceroute measurements on CAIDA's Archipelago (Ark) measuring equipment are used to generate the IPv4 Routed /24 AS Connections Dataset [7] by converting IP addresses to matching ASes to build AS links. [7]. Adding traceroute measures to the network increases the amount of traffic that is monitored. To summarise, there are three methods for determining an AS link: BGP routes, Traceroute, and IRR.

CAO YUANQING：PhD Research Scholar in Information Technology, Lincoln University College, Malaysia

Understanding the relevance of different transit providers in different regions is essential to understanding the structure of the Internet and the peering agreements between ASes. A logical approach to measure the number of clients served by a transportation provider AS is to count the number of passengers it transports. The AS refers to this as the Customer Cone (CC). Those clients who are on a downward trajectory, which is a road of AS-AS interactions [8]. This knowledge is valuable for a variety of reasons. IXP operators can choose which networks to peer with according on the size of the customer cones they have at their disposal. Additionally, they may examine changes in the customer cones of peering networks and uncover connections between the customer cone size and the AS's peering activity. CAIDA has ranked an AS based on the size of its client cone. It's possible that the consumer cone size isn't an accurate depiction of the AS's customers because it differs from area to region. For example, in the United States, Level 3 is ranked first, while it doesn't play a significant role in Europe. Based on the position of AS, it is essential to differentiate the consumer cone [9].

Statement of the Problem

More than half the previous studies have examined Internet topology in terms of node degree distribution, betweenness, and average hop count at the Autonomous System (AS) level [10]. Macroscopic measures, on the other hand, fail to reflect the local characteristics of AS connections. Most earlier models, on the other hand, are basic abstractions that don't take into account the varied sorts of nodes and links. As a result, these measurements do not reflect regional differences in the Internet's evolution depending on factors such as economic, political, and business. What is the reason for this increase in size? How internationally interconnected is the Internet? How will it change regionally? Finally, the current methodologies do not provide any practical recommendations on how to allocate resources. Even for administrators of CDNs, the effort of caching material in numerous locations, creating new peering links and anticipating their performance/cost trade-offs may be difficult. [11], for example In order to assess the consequences of their actions, an in-depth understanding of the Internet architecture and its dynamics is required.

Objective of the Study

- To design the efficient internet infrastructure to satisfy the ever-increasing traffic needs

Research Questions

- Over time, and in different places, how has the Internet changed?

Research Methodology

The UCLA Internet AS-level topology archive repository [19] and the CAIDA AS relationships dataset [20] will be used to build the Internet topology graphs. In addition to their enormous BGP monitoring network, the two repositories will be chosen due to the fact that they are the only public

CAO YUANQING: PhD Research Scholar in Information Technology, Lincoln University College, Malaysia

sources that keep historical information dating back as far as 1998. There have been several previous Internet topology studies that relied on these same sources.

This will be our major source for determining AS nodes and AS linkages. BGP data collected by several BGP data collectors, including RouteViews (96), RIPE (88), PCH (85), and Internet2 (65). Every BGP path advertising broadcast or received by the routers is recorded by the collectors. UCLA obtains the routing tables for all 133 collectors and uses the routes in the tables to construct two topologies. Another difference is that one topology utilises only IPv4 addresses, while the other only IPv6 addresses. In order to illustrate the evolution of the Internet, we utilised IPv4 topologies.

A category for each AS link is assigned using the CAIDA AS relationship dataset [21]. The AS linkages in this dataset are divided into two categories: c2p and p2p. The technique provided in [22] is used to deduce the connection type from raw BGP routes ads. It will be selected to utilise the CAIDA dataset since the link inference technique will be found to have an accuracy of 99.6% for c2p connections and 98.7% for p2p links.

RESEARCH DESIGN

The UCLA data repository will be used to create the basic AS graph, which only contains AS nodes and AS linkages. A monthly breakdown of the daily data will be necessary because the graph topologies are updated every day. The UCLA dataset has certain drawbacks, such as incorrectly advertised links owing to routing table problems, path poisoning, or router failures. Only a few hours are allotted for these ephemeral occurrences. Because of this, we deleted AS connections that appeared just once in a month in order to remove any potential false pathways.

For determining the AS relationship type, we used CAIDA data from the Topology data set. As part of the inference technique in [23], we did not filter the CAIDA dataset because erroneous information is filtered out of it. Although the CAIDA dataset did not have information on the connection type for 10% of the linkages in the basic AS graph, we had to infer the link type for those links.

We used the following inference methods for connections of unknown type:1. After determining the number of peers and the node degree of the incident ASes, we looked at each link's number of peers and node degree. Peer-to-peer (p2p) links will be established when two or more incident ASes had comparable node degrees2 and two or more peers in common. When Zhou discovered that the Internet topology displays a rich-club connection, this categorization will be applied in Gao's technique [24].

CAO YUANQING：PhD Research Scholar in Information Technology, Lincoln University College, Malaysia

Data Analysis

We try out several models apart from SVM (146), K-Nearest Neighbor (147), and Random Forest (148). Among these models, the Random Forest classifier performs the best. The Receiver-Operator Curve shows the compromise between sensitivity and specificity (ROC). As can be seen, both the False Positive and True Positive rates are maintained at less than 10% and more than 80%, respectively. An area under the ROC curve of 0.98 was obtained for this classification job. This model's Precision-recall curve will illustrate the compromise between the two metrics. For the purpose of gauging the Random Forest classifier's precision and recall, we plan to split the samples 70% to 30% for training and testing purposes. Exactly one thousand times, we'll go through this procedure again.

Conclusion

We presented a machine learning approach to inferring edge types in AS graphs generated from open-source data. With the help of the Gentle AdaBoost machine learning approach and the five node properties extracted from the AS graph, a classifier for p2p and p2c edges was learned. We apply our method to the categorization of three AS graphs: a BGP network, a traceroute graph, and an IRR graph. There are two datasets used to evaluate each classifier. The BGP dataset serves as the basis for the first test set, while the AS connection inference dataset developed at CAIDA serves as the basis for the second. Combining the three individual AS graphs allows for the computation of edge types in an AS graph. We analyse three different graphs and one composite graph to determine their unique characteristics. All three graphs feature a very high number of distinct p2p and c2c edges. Each. Integrating the three graphs gives us a much fuller view of the p2p and p2c ecosystems on the Internet.

References

[1] D. Achlioptas, A. Clauset, D. Kempe, and C. Moore. On the bias of tracer- oute sampling. In STOC, ACM, volume 1581139608, page 0005. Citeseer, 2005.

[2] M. B. Akgun and M. H. Gunes. Bipartite internet topology at the subnet- level. In Network Science Workshop (NSW), 2013 IEEE 2nd, pages 94–97. IEEE, 2013.

[3] R. Albert and A.-L. Barab´asi. Topology of evolving networks: local events and universality. Physical review letters, 85(24):5234, 2000.

[4] J. I. Alvarez-Hamelin, L. Dall'Asta, A. Barrat, and A. Vespignani. k-core de- composition of internet graphs: hierarchies, self-similarity and measurement biases. arXiv preprint cs/0511007, 2005.

[5] B. Augustin, B. Krishnamurthy, and W. Willinger. Ixps: mapped? In Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference, pages 336–349. ACM, 2009.

[6] A.-L. Barab´asi and R. Albert. Emergence of scaling in random networks.

science, 286(5439):509–512, 1999.

[7] A. Baumann and B. Fabian. Who runs the internet?-classifying autonomous systems into industries. In WEBIST (1), pages 361–368, 2014.

[8] M. Baur, U. Brandes, M. Gaertler, and D. Wagner. Drawing the as graph in 2.5 dimensions. In Graph Drawing, pages 43–48. Springer, 2005.

[9] K. Boitmanis, U. Brandes, and C. Pich. Visualizing internet evolution on the autonomous systems level. In Graph Drawing, pages 365–376. Springer, 2008.

[10] J. Brodkin. Why YouTube buffers: The se- cret deals that make-and break-online video. http://arstechnica.com/information-technology/2013/07/why-youtube-buffers-the-secret-deals-that-make-and-break-online-video/, July 2013.

[11] T. Bu, N. Duffield, F. L. Presti, and D. Towsley. Network tomography on general topologies. In ACM SIGMETRICS Performance Evaluation Review, 2002.

[12] T. Bu and D. Towsley. On distinguishing between internet power law topol- ogy generators. In INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, volume 2, pages 638–647. IEEE, 2002.

[13] Caida ark project. http://www.caida.org/projects/ark/.

[14] Caida datasets. http://data.caida.org/datasets/.

[15] Caida. Ipv4 and ipv6 as core: Visualizing ipv4 and ipv6 internet topology at a macroscopic scale in 2014. http://www.caida.org/research/topology/ as_core_network/2014/, 2014.

[16] H. Chang, R. Govindan, S. Jamin, S. J. Shenker, and W. Willinger. Towards capturing representative as-level internet topologies. Computer Networks, 44(6):737–755, 2004.

[17] H. Chang, S. Jamin, and W. Willinger. Internet connectivity at the as- level: an optimization-driven modeling approach. In Proceedings of the ACM SIGCOMM workshop on Models, methods and tools for reproducible network research, pages 33–46. ACM, 2003.

[18] H. Chang, S. Jamin, and W. Willinger. To peer or not to peer: Modeling the evolution of the internets as-level topology. Ann Arbor, 1001:48109–2122, 2006.

[19] H. Chang and W. Willinger. Difficulties measuring the internet’s as-level ecosystem. In Information Sciences and Systems, 2006 40th Annual Confer- ence on, pages 1479–1483. IEEE, 2006.

[20] K. Chen, D. R. Choffnes, R. Potharaju, Y. Chen, F. E. Bustamante, D. Pei, and Y. Zhao. Where the sidewalk ends: Extending the internet as graph using traceroutes from p2p users. In Proceedings of the 5th international conference on Emerging networking experiments and technologies, pages 217–228. ACM, 2009.

[21] Q. Chen, H. Chang, R. Govindan, and S. Jamin. The origin of power laws in internet topologies revisited. In INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceed- ings. IEEE, volume 2, pages 608–617. IEEE, 2002.

CAO YUANQING：PhD Research Scholar in Information Technology, Lincoln University College, Malaysia

[22] R. Cohen and D. Raz. The internet dark matter-on the missing links in the as connectivity map. In INFOCOM, 2006.

[23] A. Dhamdhere and C. Dovrolis. Ten years in the evolution of the internet ecosystem. In Proceedings of the 8th ACM SIGCOMM conference on Internet measurement, pages 183–196. ACM, 2008.

[24] A. Dhamdhere and C. Dovrolis. The internet is flat: modeling the transi- tion from a transit hierarchy to a peering mesh. In Proceedings of the 6th International COnference, page 21. ACM, 2010.

[25] A. Dhamdhere and C. Dovrolis. Twelve years in the evolution of the inter- net ecosystem. IEEE/ACM Transactions on Networking (ToN), 19(5):1420– 1433, 2011.

[26] G. Di Battista, M. Patrignani, and M. Pizzonia. Computing the types of the relationships between autonomous systems. In INFOCOM 2003. Twenty- Second Annual Joint Conference of the IEEE Computer and Communica- tions. IEEE Societies, volume 1, pages 156–165. IEEE, 2003.

[27] X. Dimitropoulos, D. Krioukov, M. Fomenkov, B. Huffaker, Y. Hyun, G. Ri- ley, et al. As relationships: Inference and validation. ACM SIGCOMM Computer Communication Review, 37(1):29–40, 2007.

[28] X. Dimitropoulos, D. Krioukov, G. Riley, and K. Claffy. Classifying the types of autonomous systems in the internet. SIGCOMM Poster, 151, 2005.

[29] B. Donnet and T. Friedman. Internet topology discovery: a survey. Commu- nications Surveys & Tutorials, IEEE, 9(4):56–69, 2007.

[30] S. N. Dorogovtsev and J. F. Mendes. Evolution of networks: From biological nets to the Internet and WWW. Oxford University Press, 2013.

[31] N. Economides. The economics of the internet backbone. NYU, Law and Economics Research Paper, (04-033):04–23, 2005.

[32] B. Edwards, S. Hofmeyr, G. Stelle, and S. Forrest. Internet topology over time. arXiv preprint arXiv:1202.3993, 2012.

[33] Eu telecom’s rules. https://ec.europa.eu/digital-agenda/en/ telecoms-rules/.

CAO YUANQING：PhD Research Scholar in Information Technology, Lincoln University College, Malaysia