Finding the Invisible: A Comparison of IRR Data and Routing Pathways

Cameron C. Gray[a], Dr. S. P. Mansoor[a]

[a]*Bangor University, Bangor, United Kingdom*

Abstract

The Border Gateway Protocol (BGP), the de facto standard protocol for Internet routing, only passes on the 'best' path for every Internet destination. As such, some connections between ISPs are entirely masked. If a network does not announce any routes of its own, the entire network could be masked as well. Internet Routing Registries (IRRs) exist to record connections among all ISPs in an attempt to combat this effect.

This paper examines how prevalent this masked link/network effect is and how well documented the connections between ISPs are in the European IRR. The result is that there is a relatively small number of discrepancies when comparing the data within a source, for example the IRR data. When sources are compared with each other a much larger number is uncovered.

Keywords: IRR, data quality, Internet routing, analysis
2010 MSC: 68-M10, 68-U35, 68-R10, 90-B10, 90-B18

1. Introduction

The Border Gateway Protocol (BGP), which provides the external/backbone routing for the Internet, is path-vector based. However, whilst any one ISP may receive differing views on the global routing table, the nature of the protocol means that only the best route is ever passed on. This leads to a classic hidden node problem [1, 2]. This problem is usually associated with wireless transmissions (due to signal ranges), but is equally valid when studying Internet routing [3].

The Internet requires a large amount of coordination; who uses what resources, when and how, who connects to whom and so on. Often these parties are

Email addresses: c.gray@bangor.ac.uk (Cameron C. Gray), s.mansoor@bangor.ac.uk (Dr. S. P. Mansoor)

competing commercial entities. This task is delegated to five Regional Internet Registries (RIRs), RIPE NCC, ARIN, LACNIC, AfriNIC, and APNIC. Each responsible for a different geographical area. ISPs must belong to one of these membership-based organisations to gain access to Internet resources, such as IP addresses and Autonomous System Numbers (ASNs). When issuing ASNs, these organisations require the recipient to register their routing connections (and any associated policies) in their IRR - which is usually part of the registrations database.

The second side of the routing coin is operational. As configuring and maintaining routers is fundamentally disconnected from the act of registering those connections in the IRR, a data quality issue can develop. Outsiders, such as other network administrators, cannot be certain of which source is correct or provides intended operation.

We hypothesise that neither the IRR data, nor operationally observed connections from routing beacons provide all the necessary data to fully reconstruct the structure of the Internet.

The remainder of this paper is structured thus: Section 2 surveys similar existing and related works. Section 3 presents experimental methods and limits. Results are shown in Section 4 with analysis in Section 5.

2. Previous Works

A similar study, conducted by Zhang et. al., has been completed in the past [4]. Their objective was to derive a 'more complete' topology of the Internet by amalgamating several data sources into a single graph. Our emphasis is based on whether a single source can be a suitable analogue for this effort or indeed if that data source contains all the information necessary to recreate the complete graph. In the Oliveira study [5], the researchers found that operational observations alone cannot be used to form a complete topological map of the Internet. The Di Battista et. al. paper [6], concerning the use of routing registries to extrapolate potential relationships among ISPs, focused on whether it was possible at all to extract the requisite information from IRRs. This work does not answer our question, making a determination on whether the data is valid or correct.

As has been found previously [7], much of the data to research this issue comes from publicly accessible route servers. The nature of BGP means that these route servers can only provide one viewpoint on the global routing table out of the many that co-exist. Alternative systems [8, 9] and algorithms [10, 11] have been proposed to combat the limits of current techniques. Paxson [12] produced a quantitative analysis of routing metrics and stability. This work was based entirely on operational observations. The choice to use only the empirically gathered data limits the possibilities as the monitoring stations were still fixed.

The Réseaux IP Européens Network Coordination Centre (RIPE NCC) has now closed their own database consistency project. This project was concerned with improving the quality of their assignment and routing registry databases. This was in response to the claim that the IRR was outdated and incomplete to the point of being useless [13]. Their work improved the quality and quantity of routing data and policies registered by ISPs. The other RIRs have completed similar projects over the past decade as well. These projects, however, do not fully address the problem.

Alternative solutions have been proposed [14, 15] to handle to coordination of routing information both within and without centralisation. Whilst this research is valuable in its own right, it does little to further the situation the Internet faces now and will only face greater challenges with the eventual widespread roll-out of IPv6.

Researchers have been working to accurately model the structure of the Internet [16, 17, 18, 19]. Maps of the Internet have already been produced as a result [20, 21]. Having a correct model of the structure will enable further research into the dynamics of the system. Some investigations have already been made [22, 23] into Internet dynamics. Of particular interest is the impact of misconfiguration, whether deliberate or accidental, leading to either suboptimal routing or security incidents [24, 25]. Fully understanding these dynamics could inform further research in disparate areas; such as sociology [26], artificial intelligence [27], linguistics [28] and biology [29].

The hidden node/path problem in the Internet has been studied in an indirect way previously [30, 31, 32]. This research is more closely aimed at discovery of the dynamics of the system rather than investigations into the specific phenomena. Those studies also rely on ICMP (the Internet Control Message Protocol), making them inherently fragile. ICMP is necessarily a low priority traffic type, this means that if the targeted router is busy, the ICMP traffic will not be responded to [33].

3. Scope and Methods

In order to test the hypothesis that the routing registries can detail the full structure of the Internet on their own; we are aiming to answer the following questions:

1. How many of the Internet's inter-ISP connections can be observed?,
2. Can researchers rely on the data recorded in Internet Routing Registries as a suitable analogue for the 'real' connection data?
3. How large is the hidden node problem in the Internet routing graph?

Question 1 is important to determine whether all of the Internet can actually be observed. If we are unable to observe the connection, due to policy or contractual conditions, empirical data will be incomplete. By substituting the

registered data for empirical collections we may see patterns that cannot be observed. In order to do this we must measure the agreement between the registered and empirical data, hence question 2. From there, through data gathered to answer 1 and 2, it is possible to put a numerical limit on the hidden node problem. The experiment will also test the hypothesis that the number of discrepancies will fall into two narrow bands, with the routing registry in one and operational observations in the other.

3.1. Data Retrieval Method

The Internet is a large place; the RIPE NCC track 34,114 active ASNs [34] of 50,387 visible in the global routing table [35]. As a result, we have limited our investigation to the European Internet due to the wealth of available data and industrial access to this community. All processes discussed should work equally well for other regions, subject to the data availability. This study uses the data collected from all sources on April 14$^{\text{th}}$ 2015 at either 00:00 BST or 06:00 EDT. As the Internet is in a constant state of change, a fixed point in time is required instead of taking 'live' or current readings.

The first data set used comes from the RIPE NCC Database. It lists all ASNs, their registered routing connections and associated policy. There are 170,165 statements of routing connections/policies in this data set. The source data is in Routing Policy Specification Language (RPSL). We constructed a tool to convert between this flat, text-based structure into a more easily manipulated sparse adjacency matrix. The second data set is comprised of the full Routing Information Base (RIB) exports from the Oregon Route Views (ORV) Project [36], RIPE NCC's Routing Information Service (RIS) [37] project and Internet2 [38]. These projects have deployed a set of 'routing beacons' to strategic locations around the Internet [39]. For ORV and RIS these locations are large Internet eXchanges (IX), provider neutral networks that ISPs have joined for the express purpose of exchanging traffic. Internet2, being a closed academic and research network are able to supply complete dumps of the entire structure of their network directly. ORV have deployed 17 of these beacons, RIS - 16. There are 3 common locations between the sets; LINX - the London InterNet eXchange, PTTMetro-SP - Sao Paulo's exchange and PAIX - the Palo Alto Internet eXchange. The variety in these two sets of beacons provides a wider range of potential routes and therefore adjacencies. The Internet2 data was included, even though it is a US-based educational network, as European networks will interconnect with US ones. This provides a fuller intercontinental picture of the connections present in Europe.

The RIB exports contain the AS Paths for all routes received by each beacon in MRT format (defined in RFC 6396 [40]). Our tool similarly converts every pair of ASNs in these paths to a sparse adjacency matrix. This data set contains 8,465,757 unique paths and 350,553 unique AS pairs. When encoding the adjacencies, directionality was preserved.Thus A is an adjacency matrix in which there may exist $A_{i,j} \neq A_{j,i}$ as well as $A_{i,j} = A_{j,i}$.

3.2. Comparison Method

In an ideal situation, both the adjacency matrices would be symmetrical and equal. This implies that both networks have correctly configured and registered their routing relationship. As configuring a relationship and registering that relationship are disparate processes, the registration step can be overlooked. When this occurs, an asymmetric adjacency matrix results. To determine the accuracy of the data sets, two types of comparisons are required. Tests for internal agreement, to confirm both parties have registered or demonstrated the adjacency, as well as tests for external agreement where registered connections are compared to observed connections.

The internal tests are conducted by extracting both the left and right triangles from the adjacency matrix. The full symmetric matrix is then formed by;

$$AS = Tri(A) + Tri(A)^T \tag{1}$$

The addition of any matrix to the transposition of the same matrix results in a symmetric matrix. This method results in four versions of the adjacency matrix. Where these matrices do not equal each other it is deemed an error. As different numbers of networks are seen in each data source, we can only perform external agreement testing where the network is seen in both compared sources. Once the intersection is selected, the same triangulation method can then be applied. However, the test must be performed twice - to match both the upper and lower triangles from each source.

As a mismatched adjacency involves two partners, the number of errors must be treated as pairwise. Therefore the rate of errors in any adjacency matrix is calculated as:

$$Rate(A) = \frac{Error(A)}{size(A) \times (size(A) - 1)} \tag{2}$$

In Equation (2), the size of the matrix is a scalar as an adjacency matrix is necessarily square.

3.3. Connectivity Types

Our methodology also attempts to take into account the differing commercial relationships that could potentially exist. Routing registry data (the RPSL) encodes at least part of this information. Based on examining and comparing the statement of policies included in the AS Set (a formal definition of a specific non-repeating unordered AS Numbers), we classify the policy as either upstream (customer to provider), downstream (provider to customer), peer (equal footing, usually settlement free) and inconclusive (where the policy is either ambiguous or missing). Each of these connection types will affect which adjacencies will be visible operationally. Only upstream and downstream connections are guaranteed to be visible, as one network is providing access to the other. Peering connections are most susceptible to the hidden path problem. We have tried

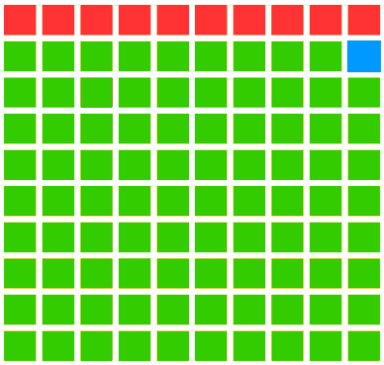

Figure 1: Waffle Chart showing Status of the 34,114 Registered European Networks (based on relative percentage). Green - Observed, Blue - Inactive, Red - Undetermined.

to counteract this potential issue by selecting data sets that include routing beacons co-located at major peering points. This strategy will not necessarily collect *all* peering connections, as so-called 'private peering' connections use dedicated physical cables.

4. Results

4.1. Coverage

Combining all data sources, a total of 59,672 unique AS numbers/networks were operationally observed in BGP paths. Of the 34,114 registered European networks, 30,486 are observed in an operational AS path. This equates to 89.37%. There are 331 AS numbers that do not have any registered connections. This data is presented graphically in Figure 1.

Examining the coverage of the operational data, which may include networks from outside of the RIPE NCC operational area, shows: 99.76% of all adjacencies were observed in the RIPE RIS data set. 85.24% in the Oregon Route Views data set and 48.68% in the Internet2 set. 29,032 networks (48.65%) are seen in all three operational data sets.

4.2. Types of Connectivity

Using the RIPE registered connections database, there are 236,105 total adjacencies to be classified into the three connectivity types (see Section 3.3). Our method has been able to classify all but 810 (0.0034307%) adjacencies into the three main types. There are 119,090 confirmed peer connections (50.439%), 81,234 upstream connections (34.406%) and 34,971 downstream connections (14.811%). A visual representation of these relative amounts is shown in Figure 2.

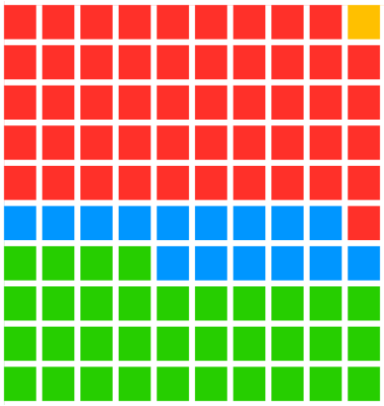

Figure 2: Waffle Chart showing Types of Registered Adjacencies (based on relative percentage). Green - Upstream, Blue - Downstream, Red - Peer, Yellow - Inconclusive.

4.3. Internal Agreement

In this experiment, 'internal agreement' is defined as a measure of the match between both ends of an adjacency within a single data source, i.e. where $A_{i,j} \neq A_{j,i}$ an error is recorded. The normalised internal agreement error rates for each data source are: (shown graphically in Figure 3)

- RIPE NCC Database/Registered Connections: 0.00021205 (123387 Errors)
- RIPE NCC RIS Observed Connections: 0.00011545 (204561 Errors)
- Oregon Route Views Observed Connections: 0.000084362 (109135 Errors)
- Internet2 Observed Connections: 0.000045803 (19325 Errors)

4.4. External Agreement

External agreement is a measure of how much a data source matches another. As there is two triangles in the adjacency matrix, essentially which half is regarded as true, the test must be completed twice. The order (i.e. which source is the upper or lower triangle) is reported as 'R' (rows) for the lower triangle and 'C' (columns) for the upper triangle. These results are shown graphically in Figure 4.

Registered Networks v. RIS (30,486 common networks)
R: Registered Networks, C: RIS Observed Paths; 0.00044130 (204,931 Errors)
R: RIS Observed Paths, C: Registered Networks; 0.00020702 (96,135 Errors)

Registered Networks v. ORV (26,884 common networks)
R: Registered Networks, C: ORV Observed Paths; 0.00039739 (143,603 Errors)
R: ORV Observed Paths, C: Registered Networks; 0.00019940 (72,252 Errors)

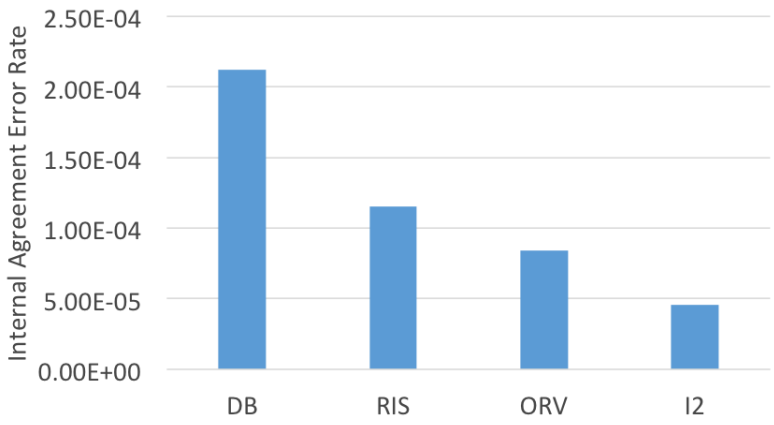

Figure 3: Internal Agreement Error Rate by Data Source.

Registered Networks v. I2 (14,573 common networks)
R: Registered Networks, C: I2 Observed Paths; 0.00040861 (83,026 Errors)
R: I2 Observed Paths, C: Registered Networks; 0.00078194 (43,386 Errors)

RIS v. ORV (50,725 common networks)
R: RIS Observed Paths, C: ORV Observed Paths; 0.000095908 (123,385 Errors)
R: ORV Observed Paths, C: RIS Observed Paths; 0.00013399 (172,379 Errors)

RIS v. I2 (29,040 common networks)
R: RIS Observed Paths, C: I2 Observed Paths; 0.000125350 (52,853 Errors)
R: I2 Observed Paths, C: RIS Observed Paths; 0.00027381 (115,451 Errors)

ORV v. I2 (29,039 common networks)
R: ORV Observed Paths, C: I2 Observed Paths; 0.000093443 (39,397 Errors)
R: I2 Observed Paths, C: ORV Observed Paths; 0.00016098 (67,873 Errors)

5. Analysis

Initial inspection throws up what appear to be count errors; this is not the case. The total number of unique networks observed is above the 34,114 registered with RIPE NCC. This total includes networks registered under other RIRs where a European network has registered a connection. The total is also above the number visible in the global routing table as the data includes research and private ASNs that would not ordinarily be seen on the public Internet.

Similarly, the number of upstream connections should equal the number of registered downstream connections. The number of detected downstream connections is most likely underestimated as larger providers employ more advanced methods of grouping customers into AS Sets, involving multiple

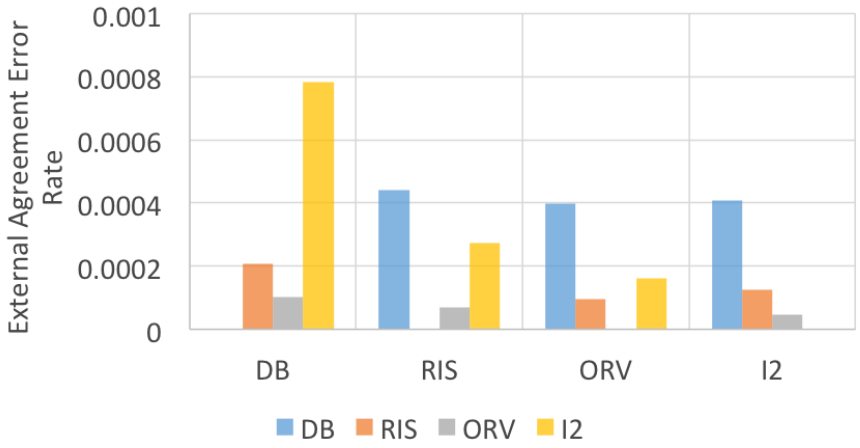

Figure 4: External Agreement Error Rate Comparison by Data Source

levels of indirection. Our RPSL parsing tool does not currently take the other levels into account when classifying a connection. However, this does not remove the potential that the mismatched count is due to human error. The upstream/downstream mismatch can also be attributed to lax policies in registering the connections in the IRR.

95% confidence values of the calculated error rate for internal agreement in each of data sources establishes that each is statistically significant from each other. This disproves one of our original hypotheses that the error rate in operationally observed data would be found in a narrow band with only the registered data set being an outlier. This finding lends support to the hidden node problem being more than a purely local (in terms of networks) phenomena.

As registering routing connections is a separate process, it is understandable how the database would be significantly different from all operational observations due to human error. The error rate from the operational data sources are between 21.6% and 54.4% of those found in the database. These differences could be wholly accounted for by the hidden path effect. It may also be a mixture of hidden paths and 'one-way' peering where one partner does not announce any routes. From a distance (i.e. outside of that particular network), there is no technical way to determine the difference and it is impractical and/or impossible to make these observations due to the vast number and varying levels of access to target networks. This is counter-intuitive to most people, as a connection either exists or not. In this case a mismatched adjacency is not stipulating that a connection doesn't exist, but it does state that a routing policy exists in one direction - but not in the other.

Incorporating the registered connections we can estimate the numeric bounds for this situation. Of the 29,032 networks found in all operational data sets, we find that 16,911 of 165,720 (10.2055%) errors are present in every permutation of the adjacencies. Matching these networks to registered connections 14,569 networks exist in all four data sources. Only 859 errors (out of a total of 166,974) exist in every operational data set and do not have a registration either. This represents 0.0051445% of errors being legitimately absent relations. Using these two figures, we can estimate that between 0.0051446% and 9.6911% of adjacency agreement errors are due to the hidden path problem.

Assuming the upper bound for the hidden path quantity, 99.0611% of European autonomous systems are accounted for. The remaining 0.9389% equates to 320 networks. As this is less than the 331 inactive/non-connected networks we are able to narrow the bound for hidden paths further and account for all registered networks. $\frac{0.9389}{320} \times 331 = 0.97117\%$, a difference of 0.03227% making the final hidden path quantity as 9.65883%.

Even though the quantity of hidden nodes and paths can now be calculated, we are still unable to state what the missing connections are. This means that the absolute structure of the Internet cannot be determined without intricate placement of more routing beacons. Our findings do support the hypothesis that IRR data can be used as a suitable analogue for objective measurements in general situations. With over 90% of the structure and connections that can be observed directly found in the registration data. However, researchers would need to account for the extra error inherent in the registered data. This error is still (in absolute terms) small, so would not affect any macro-scale observations.

There is a caveat for this data and analysis. RIPE place an emphasis on maintaining valid and correct information in their database. It is incumbent on the Local Internet Registries (LIRs) that are members of RIPE to enforce this policy. LIRs run the risk of losing the ability to obtain or modify the entries relating to their Internet resources if they fail in this duty. Other RIRs have different data quality and consistency policies. For example, ARIN (the North American registry) do not require registration of routing policies at all. North America relies on a 3$^{\text{rd}}$ party database owned and operated by MERIT Network Inc. for this purpose. Access to the 'RADb' (or Routing Assets Database) [41] is on a subscription basis, incurring additional costs for providers and LIRs. As such there will be a different level of error and missing data when basing a study with their data set.

6. Conclusions

In conclusion, the results show that a majority of the structure of the observable European Internet was found in the registered data. This means that researchers can accept the RIPE NCC database as a suitable analogue for empirically collected data on the structure of the European Internet. As at the 14$^{\text{th}}$ April

2015, researchers can trust that 90.33928% of the structure is represented in the IRR data. Note; whilst the data used in this paper is available in the repository archives, re-running the experiment with current data will change the figures slightly. The Internet is a dynamic system and whilst the impact of changes will be locally significant, the system as a whole does not really change. The collected data shows that the remaining 9.65883% are hidden paths/nodes. These could never be observed without costly deployments of routing beacons to smaller and smaller networks. Counter-intuitively the number of downstream customers (those who are charged to receive traffic from a larger network), is under reported. This is due to the method that larger ISPs use to manage their routing policies. The selection of routing beacons is crucial for this task, with some areas of massive overlap and others with almost no representation. Combining different sets of beacons can overcome the lack of coverage for some areas, but care must be taken with oversampling.

We have also found an upper bound for the hidden node problem in the European Internet routing graph. This number can only ever be a bound as the dynamics of the Internet and the type of problem will cause natural fluctuations as well as timings of updates and provisioning of new connections - or indeed de-commissioning of old connections. This is a critical finding for developing accurate maps or models of the Internet. This quantifies how much of the map needs to be 'filled in'. We can therefore extrapolate the remaining structure - this will be an estimate, but a more accurate estimate than previous efforts. Combining finding the number of hidden nodes/paths, with classifying the connections observed, allows a more accurate model to be produced. Not just with the correct number but also the correct type of connections represented.

Acknowledgements

The authors would like to thank the following for providing access to their data sets, making this paper possible.

- RIPE NCC: RIPE DB, RIS BGP Dumps.
- Oregon Route Views Project: BGP RIB Dumps.
- GRNOC/Internet2: BGP RIB Dumps.

References

[1] A. Tsertou, D. I. Laurenson, Insights into the hidden node problem, in: Proceedings of the 2006 international conference on Wireless communications and mobile computing, ACM, 2006, pp. 767–772.

[2] A. Rahman, P. Gburzynski, Hidden problems with the hidden node problem, in: Proceedings of 23rd Biennial Symposium on Communications, 2006, pp. 270–273.

[3] H. Chang, R. Govindan, S. Jamin, S. Shenker, W. Willinger, On inferring as-level connectivity from bgp routing tables, Tech. rep., Citeseer (2002).

[4] B. Zhang, R. Liu, D. Massey, L. Zhang, Collecting the internet as-level topology, ACM SIGCOMM Computer Communication Review 35 (1) (2005) 53–61.

[5] R. Oliveira, D. Pei, W. Willinger, B. Zhang, L. Zhang, The (in) completeness of the observed internet as-level structure, IEEE/ACM Transactions on Networking (ToN) 18 (1) (2010) 109–122.

[6] G. D. Battista, T. Refice, M. Rimondini, How to extract bgp peering information from the internet routing registry, in: Proceedings of the 2006 SIGCOMM workshop on Mining network data, ACM, 2006, pp. 317–322.

[7] H. Kong, The consistency verification of zebra bgp data collection, Jupiter 146 (2003) 240–68.

[8] Y. Shavitt, E. Shir, Dimes: Let the internet measure itself, ACM SIGCOMM Computer Communication Review 35 (5) (2005) 71–74.

[9] N. Spring, R. Mahajan, D. Wetherall, T. Anderson, Measuring isp topologies with rocketfuel, Networking, IEEE/ACM Transactions on 12 (1) (2004) 2–16. doi:10.1109/TNET.2003.822655.

[10] Y. Zhang, Z. Zhang, Z. M. Mao, C. Hu, B. MacDowell Maggs, On the impact of route monitor selection, in: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement, IMC '07, ACM, 2007, pp. 215–220. doi:10.1145/1298306.1298336.

[11] L. Gao, On inferring autonomous system relationships in the internet, IEEE/ACM Transactions on Networking (ToN) 9 (6) (2001) 733–745.

[12] V. Paxson, End-to-end routing behavior in the Internet, in: ACM SIGCOMM Computer Communication Review, Vol. 26, ACM, 1996, pp. 25–38.

[13] M. Kühne, R. Wilhelmm, E. Aben, P. Palse, Interesting Graph - How Complete is the RIPE Routing Registry, Tech. rep., RIPE NCC, Amsterdam (2010).

[14] G. Goodell, W. Aiello, T. Griffin, J. Ioannidis, P. D. McDaniel, A. D. Rubin, Working around bgp: An incremental approach to improving security and accuracy in interdomain routing., in: NDSS, 2003.

[15] L. Gao, J. Rexford, Stable internet routing without global coordination, IEEE/ACM Transactions on Networking (TON) 9 (6) (2001) 681–692.

[16] J. Winick, S. Jamin, Inet-3.0: Internet topology generator, Tech. rep., Technical Report CSE-TR-456-02, University of Michigan (2002).

[17] S. Zhou, R. J. Mondragón, Accurately modeling the internet topology, Physical Review E 70 (6).

[18] S. Carmi, S. Havlin, S. Kirkpatrick, Y. Shavitt, E. Shir, A model of internet topology using k-shell decomposition, in: Proceedings of the National Academy of Sciences, Vol. 104, 2007, pp. 11150–11154.

[19] A. Broido, et al., Internet topology: Connectivity of ip graphs, in: ITCom 2001: International Symposium on the Convergence of IT and Communications, International Society for Optics and Photonics, 2001, pp. 172–187.

[20] B. Cheswick, H. Burch, S. Branigan, Mapping and visualizing the internet., in: USENIX Annual Technical Conference, General Track, 2000, pp. 1–12.

[21] B. Huffakec, K. Claffy, Y. Hyan, M. Luckie, H. On, CAIDA's IPv4 & IPv6 AS Core AS-level Internet Graph, Tech. rep., Cooperative Association for Internet Data Analysis (2014).
URL `http://www.caida.org/research/topology/as_core_network/pics/2013/ascore-2013-jan-ipv4v6-poster-7x4.pdf`

[22] R. Pastor-Satorras, A. Vázquez, A. Vespignani, Dynamical and correlation properties of the internet, Physical review letters 87 (25) (2001) 258701.

[23] V. Paxson, End-to-end internet packet dynamics, IEEE/ACM Transactions on Networking (TON) 7 (3) (1999) 277–292.

[24] E.-y. Kim, L. Xiao, K. Nahrstedt, K. Park, Secure interdomain routing registry, Information Forensics and Security, IEEE Transactions on 3 (2) (2008) 304–316.

[25] X. Liu, P.-D. Zhu, Y.-X. Peng, Internet registry mechanism for preventing prefix hijacks, Journal of Software 20 (3) (2009) 620–629.

[26] C. Haythornthwaite, Social network analysis: An approach and technique for the study of information exchange, Library & information science research 18 (4) (1996) 323–342.

[27] N. R. Jennings, K. Sycara, M. Wooldridge, A roadmap of agent research and development, Autonomous agents and multi-agent systems 1 (1) (1998) 7–38.

[28] A. Bavelas, Communication patterns in task-oriented groups., The Journal of the Acoustical Society of America (1950) 725–730.

[29] A. Shimbel, Structural parameters of communication networks, The bulletin of mathematical biophysics 15 (4) (1953) 501–507.

[30] H. Chang, S. Jamin, W. Willinger, Inferring AS-level Internet topology from router-level path traces, in: ITCom 2001: International Symposium on the Convergence of IT and Communications, International Society for Optics and Photonics, 2001, pp. 196–207.

[31] R. Mahajan, N. Spring, D. Wetherall, T. Anderson, User-level internet path diagnosis, in: ACM SIGOPS Operating Systems Review, Vol. 37, ACM, 2003, pp. 106–119.

[32] V. Jacobson, Pathchar: A tool to infer characteristics of internet paths (1997).

[33] G. Lu, Y. Chen, S. Birrer, F. E. Bustamante, C. Y. Cheung, X. Li, End-to-end inference of router packet forwarding priority, in: INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE, IEEE, 2007, pp. 1784–1792.

[34] RIPE NCC, RIPE Database - Split by Type - Autonomous System Numbers, `ftp://ftp.ripe.net/ripe/dbase/split/ripe.db.aut-num.gz` collected 2015-04-14 (April 2015).

[35] G. Huston, CIDR Report - AS2.0, `http://www.cidr-report.org/as2.0/` collected 2015-04-14 (April 2015).

[36] University of Oregon, Route Views Project, `http://archive.routeviews.org/` collected 2015-04-14 (April 2015).

[37] RIPE NCC, RIS Raw Data, `https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris/ris-raw-data` collected 2015-04-14 (April 2015).

[38] Global Research Network Operations Center @ Indiana University, Internet2 RIB Dumps, `http://ndb7.net.internet2.edu/bgp/` collected 2015-04-14 (April 2015).

[39] N. Chatzis, G. Smaragdakis, A. Feldmann, On the importance of Internet eXchange Points for today's Internet ecosystem, arXiv preprint arXiv:1307.5264.

[40] L. Blunk, M. Karir, C. Labovitz, Multi-Threaded Routing Toolkit (MRT) Routing Information Export Format, RFC 6396 (Proposed Standard) (Oct. 2011).
URL `http://www.ietf.org/rfc/rfc6396.txt`

[41] MERIT Network Inc., RADb, `http://www.radb.net` visited 2015-04-14 (April 2015).