The Effects of Peering Agreements and Content Delivery Networks on Internet Topology Measurements

Viktor Holmgren
vikho394@student.liu.se

Yousif Touma
youto814@student.liu.se

Fredrik Wallström
frewa814@student.liu.se

George Yildiz
geoyi478@student.liu.se

ABSTRACT

As the internet's popularity grows, so does the interest for its topology. In this paper, we use public Border Gateway Protocol (BGP) data from Route Views to create a Force Directed Graph (FDG) of the internet where each node represents an Autonomous System (AS). We then use traceroutes from different geographical locations to determine the paths taken and number of AS-to-AS hops to reach the webservers of some of the most popular websites. Using this information, we provide visualizations of the paths taken on an AS level and analyze if the presence of Content Delivery Networks (CDNs) affect the paths. Our contributions are three-fold. First, we show that peering agreements have a great effect on the AS-to-AS connectivity when building a graph from publicly available BGP data. Secondly, we show how it is possible to find power law distribution among ASs. Thirdly, we present how traceroutes performed in this manner are not sufficient to provide an accurate topology of the internet and how choosing a representative set of destinations would result in a more accurate representation.

Keywords

Border Gateway Protocol, Autonomous Systems, Content Delivery Network, internet topology, traceroute

1. INTRODUCTION

Over the last decade, the internet has become more and more pervasive in our everyday lives. We use the internet in all aspects of our lives, be it sending emails at work, socializing while on the bus or using video streaming services at home. Despite our extensive usage of the internet, we know little about its topology, about what happens after our packets leave the closest router. This problem has long been of interest to many, not just academia but also industry. By using analysis and visualizations we help build an understanding of how the internet works. If we hope to improve the internet and the services that operates on it, we need to understand and measure it in efficient ways.

Although it might seem easy to get a picture of the internet topology it is actually rather difficult. There are multiple factors causing these problems in establishing a correct topology. One of the most common ways to view the internet topology is in terms of an Autonomous System (AS) graph, where each node is an AS and each edge is a connection between two ASs. An AS is a subnet, maintained by a single organization, containing a number of interconnected routers that share routing protocols. Choosing a route through the internet means choosing a path between ASs which is done using the Border Gateway Protocol (BGP). The BGP chooses routes primarily based on inter-AS business agreements, also known as peering agreements. This usually means that both organizations agree to forward each other's traffic without cost. This leads to routing based not on shortest-path algorithms but rather on lowest-cost which means some paths may be more or less used. Furthermore, we know that many services use Content Delivery Networks (CDNs) to distribute their service which means that their content is duplicated at various points across the globe to give users faster access. This also affects internet routing since the CDN may route your request to different servers depending on load or other CDN operator policies.

In this paper, we first give an introduction to some of the topics relevant to our work. We then give a thorough account of the methods used to build, measure and analyze the internet topology. In the result section we look at several aspects. First, we graphically visualize an approximation of the internet topology using BGP data from Route Views. This is done using Scalable Force Directed Placement (SFDP). We also provide statistics regarding edge distribution among the nodes. Then, we graphically visualize the results of our traceroutes in the same fashion and provide statistics to detect similar properties between the graphs.

In the subsequent sections we analyze and discuss whether peering agreements and CDNs make it harder to accurately map the internet. Finally, we discuss how our findings relate to previous findings in the area and pitfalls to be avoided by future researchers.

2. BACKGROUND

2.1 Autonomous System

An AS is a group of routers that are typically under the same administrative control, which means that they are operated by the same Internet Service Provider (ISP) or belong to the same company network. The reason that an AS is helpful is because it makes it easier for organization to run and administer its network as it wishes, while still being able to connect its network to other networks. It is also good for scale as it makes it easy to reduce the complexity of route computation in networks as large as the public internet.

ASs are connected to each other and one or more routers in an AS will have the added task of being responsible for forwarding packets to destinations outside the AS. These routers are called gateway routers. Obtaining reachability information from neighboring ASs and propagating the reachability information to all routers internal to the AS are handled by an inter-AS routing protocol. In the internet all ASs run the same inter-AS routing protocol called BGP. Each AS is also identified by a unique AS number (ASN).

2.2 Border Gateway Protocol

BGP is a protocol used for determining paths for source to destination routes that go across multiple ASs. With BGP every AS can obtain subnet reachability from neighboring ASs, propagate the reachability information to all routers internal to the AS, and determine "good" routes to subnets based on the reachability information and on AS policy. With the help of BGP, each subnet can inform the rest of the internet of its existence, meaning they will not stay unknown and isolated from the rest of the internet.

The route for a given packet is determined by elimination rules in BGP if there are multiple possible paths to take. First, we consider any business agreements that the organization has that may affect which nodes we want to forward traffic through. This may be peering agreements and similarly. Second, we choose the path that results in the shortest AS-PATH which is based on how many AS hops that is needed to reach the destination. Third, we select the route with the closest NEXT-HOP router. This means that we choose the router for which the cost to a gateway router is the lowest possible and thus that the traffic exits the AS as soon as possible.

2.3 Traceroutes

A widely used mechanism for computing the internet topology is to use traceroutes. Traceroute is a utility that records the route between two hosts in a network, for example between your computer and a specified destination host, like Google.com. Traceroute will display the path taken in the internet by displaying each hop it makes between the nodes on the way to the specified destination host. It also calculates and displays the round trip time (RTT) for each node on the path.

Traceroute works by sending three packets from the source host with gradually increasing time-to-live (TTL) value. By starting with sending a packet with TTL value of one the first node that receives the packet will decrement the TTL and drop the packet because then the TTL value is zero. The node will then send a message back to the source of the traceroute with its own address as source address. Doing so, the source will know the address of the first node in the path. It can also calculate and display the RTT for each of the three packets sent to each node on the path to give an indication of consistency in the route being traced. If any of these packets are lost or timed out, the RTT is replaced by an asterisk. This will continue, with packets being sent with gradually increasing TTL value until a packet reaches the destination host, after which we have learned the address of each node in the path.

2.4 The Route Views Project

Route Views is one of the largest publicly available BGP data archives. We use Route Views archives to access information about BGP data, i.e. the BGP updates. The Route Views project is a so called route collector that allows internet users to view global BGP routing information from the perspective of other locations around the internet. What Route Views does is to set up an AS which then asks a regional ISP for any BGP announcements they receive. This Route Views AS will then save the BGP announcements it gets and release updates multiple times a day.

2.5 Tools

2.5.1 *Pyasn*

The Pyasn Python library provides us with two services. Using one of its utilities we can parse the RIB data files from Route Views and extract the AS-PATH attribute from every BGP announcement in the dataset and build a list of paths. However, its main function is to build an IP prefix lookup table and then provide an API to which we can query with a specific IP address and get the ASN of the AS in which it is located.

2.5.2 *NetworkX*

NetworkX is a graph building library that provides us with the means to generate graphs over the data that we have collected. We are also able to query the graph and generate statistics by using the NetworkX library.

2.5.3 *GraphViz*

GraphViz is a graph visualization library with which we can visualize a graph, using performant graph drawing algorithms such as SFDP.

2.5.4 *LocaPing*

LocaPing is a web based tool allowing the user to do pings to check server reachability. It also allows us to run traceroutes from their servers located in different countries spread geographically over the globe.

3. METHOD

3.1 Building an Internet Topology

High-level internet routing, i.e. inter-AS routing, is done by BGP. To build an internet topology on this level we needed access to large amounts of BGP routing tables or BGP announcements. For this purpose we used the Route Views project. We have built a Python program that, with the help of the Pyasn library, downloads the latest RIB-file containing BGP announcements from Route Views. The program then extracts the AS-PATH attribute from every BGP announcement in the dataset, building a list of paths. From this we are not only able to generate a set of all ASs but also infer all advertised connections and thus generate a complete graph using NetworkX. Here we also need to be careful as many BGP announcements contain duplicate data from different sources and as such we need to filter out duplicates before proceeding. We then ran this graph through GraphViz to generate a visualization of the said graph. The graph drawing algorithm we used was SFDP since it has good scaling properties and time complexity.

Furthermore, using NetworkX we were able to query the graph and generate a histogram over edge distribution. This gives us opportunity to analyze whether our network topology follows power law.

3.2 Traceroutes

To get a picture of the internet that reflects how it is actually used, we performed traceroutes from fifteen different locations in the world to the webservers of the ten most popular websites. The intent of using fifteen different locations to trace from is to see how the traffic behaves. It is interesting if traffic from different locations to the same website converges at different locations depending on from where the trace is performed. It is also interesting if the traces contain very few hops until it reaches the destination. Both cases could suggest the presence of a CDN that has duplicated the content of these popular websites close to the source of the traceroute. The ten most popular websites that we accessed were chosen using Alexa Traffic Ranks [2]. Alexa ranks websites based on the browsing behavior of people in their global data panel which is a sample of all internet users. The ranking we based our traces on is from March 23rd, 2016. We chose to use the ten most popular websites because we believe it fairly accurately shows where the majority of the daily network traffic is sent. We also tried a set of ten randomly[1] selected websites, ranked between 100 and 500, to see if there are any differences. This time we used the ranking from May 10th, 2016.

To accomplish these traceroutes we used the LocaPing traceroute service from fifteen different locations around the world. The console outputs of the traceroutes were written to a text file which we parsed using a Python script.

We then use the Pyasn Python library to map each IP address in our traceroute paths to an ASN. This is done in several steps. First, we use the same RIB-file we mentioned earlier which contains all the raw BGP data from Route Views. This is converted by a Pyasn utility to a file containing IP-prefix-to-ASN mappings which can be extracted from the RIB file. Second, by giving this resulting "lookup table" file to the Pyasn library we can do a lookup of a specific IP address which will be mapped to an ASN using a longest-prefix-matching policy. The resulting path then also needed to be reduced since a number of consecutive IP addresses could be mapped to the same ASN, i.e. we needed to remove duplicates.

We then generate a graph and query it in the same way as with the BGP graph, looking at the same metrics, trying to find similarities. Here we also check whether there are any edges present which do not exist in the BGP graph. Finding such missing edges would show incompleteness of the BGP data and also hint to the existance of peering agreements.

3.3 Limitations

Gradually through our methodology we encountered some problems. We encountered the first problem when we did the traceroutes using LocaPing. In some traceroutes, all three packets to a particular node on the path were lost or experienced timeout which means we could not learn the IP address for that node on the path. In these cases, we simply omitted these nodes in that particular route because these timeouts are nothing we can handle explicitly. This will, in some cases, make it hard for us to plot the results from the traces accurately, which could lead to some parts of the graph being unrepresentative of the actual paths. In most cases though, we could successfully get the IP addresses of the nodes on the path to the destination host.

Another problem we encountered, that could make the topology unrepresentative, is the fact that we did the tracerouting using LocaPing. The service will set a maximum TTL value of 15, meaning it will not perform more than 15 hops for each host that we trace to. This means that we may not reach the destination host if the actual path requires more hops than this set value. This problem together with the timeout problem mentioned earlier might cause the traceroute to be incomplete. To decide whether these limitations affect our results, we quantify them.

The third and final problem we encountered was when we tried to map the actual IP addresses gathered from the traceroutes to an AS. In this step, we occasionally encountered an IP address that we could not map to an AS existing in our database. We have limited ourselves here as we chose to only use one route collector for our BGP data. This means that we only collect BGP updates gathered from a single source which is not representative of the internet at large. We worked around this by adding a single separate node called "None" which represents "unknown AS" in the graph. The reason for doing this is that consecutive hops in the traceroute to "unknown" ASs could either be within the same AS or not. Therefore we cannot map each hop to a unique unknown AS because it might not be the case that we have hopped out of the current AS. Much like with our traceroutes, we quantify these limitations by measuring the number of hops to this "None" node. The full impacts will be discussed in our conclusions.

4. RESULTS

From the BGP data gathered from Route Views we were able to construct a graph with roughly 54 000 ASs and 80 000 edges. A visualization of the graph is shown in Figure 1. From this graph we extracted the node degree distribution, generating the chart in Figure 2. We can observe that there are very few ASs that have a high number of connections and that most ASs are only connected to one or two other ASs. This shows that the topology follows a power law construction in terms of the node degree.

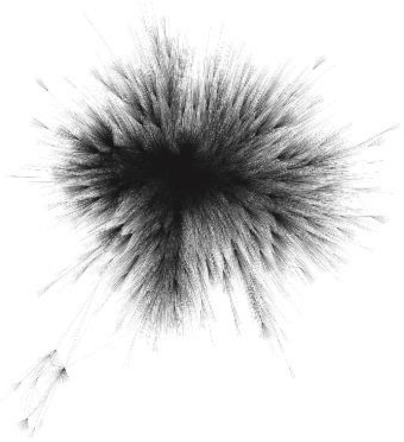

Figure 1. BGP graph

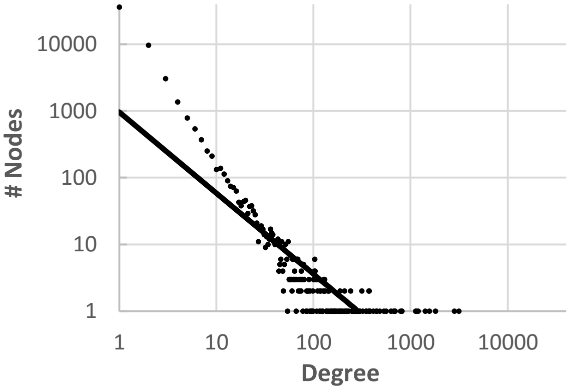

Figure 2. Node degree distribution in the BGP graph. Plotted with a best fit line.

[1] Random numbers generated from Random.org

From our traceroutes we have several results. First, we measured the number of successful traces in our 15 attempts from the different hosts. We define a successful trace as reaching the destination host IP given by the Domain Name System (DNS) resolution done from the traceroute server at LocaPing. The data gathered from this can be seen in Figure 3. Here we see that either the share of successful traces to the host was very high, as is the case with Google.com or Qq.com, or there were no successful traces. If we look more closely at individual traces, we have measured the average number of hops and the average number of successful hops, which is shown in Figure 4.

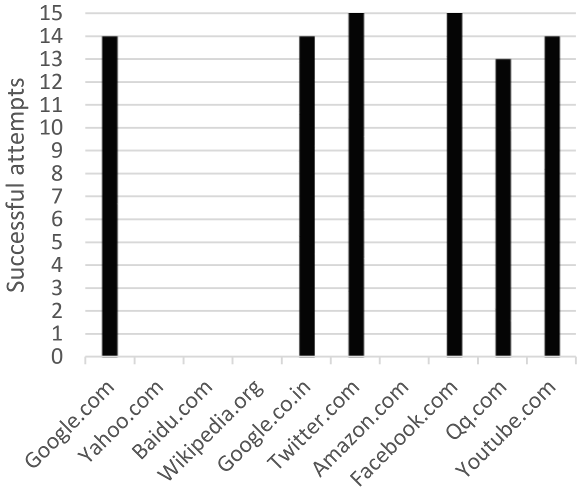

Figure 3. Attempted and successful hops in traces.

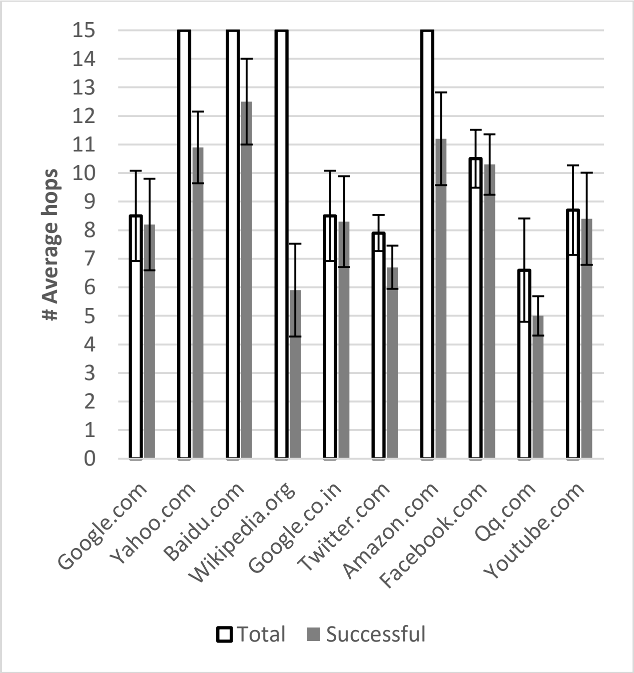

Figure 4. Average number of router hops with 95% confidence interval.

Furthermore, during the mapping of IP addresses to ASNs we had some addresses which the Pyasn library was not able to map to an AS in our database, which resulted in hops to the aforementioned "None" AS. From all the paths we had a total of 68 hops to "None" with 361 AS hops in total. From the traces we also extracted the average number of successful hops for each host, which is shown in Figure 5.

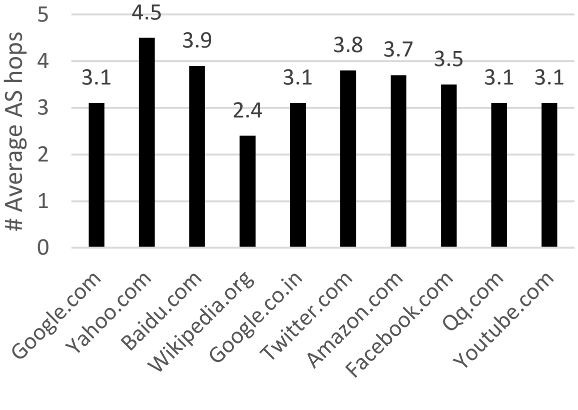

Figure 5. Average number of AS-to-AS hops.

Figure 6 shows the routes that are taken when we do traceroutes to the webservers of the ten most popular websites. As mentioned earlier, this is done from 15 different locations. The graph is visually similar to the BGP graph, with many nodes only having one connection and a few nodes having many connections. Just like with the BGP graph, we extracted some information regarding node degree distribution from this graph which can be seen in Figure 7. We can see power law distribution, just like in Figure 2.

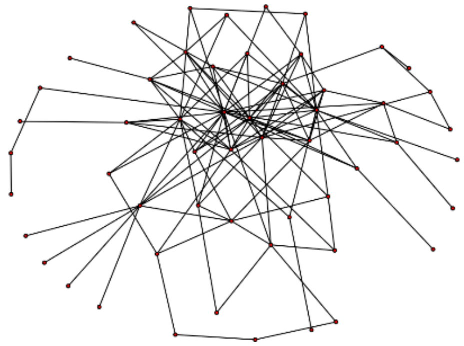

Figure 6. Traceroute graph.

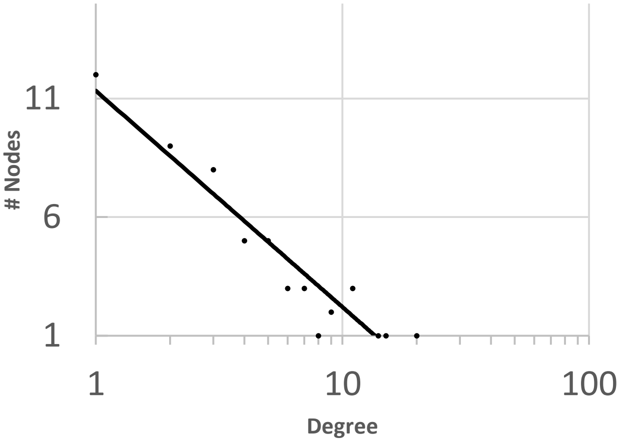

Figure 7. Node degree distribution in the traceroute graph. Plotted with a best fit line.

When we analyzed the random set, the resulting graph had similar characteristics to when we used top websites, primarily when it comes to power law. One difference was that the average number of AS-to-AS hops was noticeably larger for the random set compared to the top websites. Finally, we measured whether there exists any edges in the traceroute graph for the top websites, which do not exist in the BGP graph. Here we found 52 new unique edges.

5. CONCLUSIONS

Route collectors such as Route Views have long been essential to researchers when it comes to creating internet topologies. The BGP data they provide gives us an insight in the routing for the internet on an AS-level. However, it is not without flaws. First, only using a single route collector can result in a skewed result. To get a more accurate picture we would need to gather data from a large amount of different route collectors distributed geographically all over the world. Gregori et al. [4] conclude that the BGP data is in fact highly incomplete even when gathered from multiple collectors because of the types of ASs from which the BGP data is usually gathered. This is due to smaller AS-to-AS connections not being announced to these collectors and are therefore not included in the topologies that are inferred from the data. Looking at the node distribution in our topology we see this, as most ASs have very few connections. Although it follows a power law construction you could have expected more ASs having a medium number of connections (perhaps in the range 11-20). A possible cause for this distribution is peering agreements. Apart from the paid customer-provider links, many ASs also have "free" connections with peers in an attempt to lower costs. We can see the presence of such agreements due to there being edges in the traceroute graph which do not exist in the AS-level topology inferred by the BGP data. Furthermore, several studies have shown that even the most carefully inferred AS maps in use today still miss a substantial portion of AS connectivity of the peering type, with estimates varying from 35% to 95% [3][5].

Our results from our traceroutes show that some of the most popular websites in the world cannot be reached within 15 router hops from any of the 15 locations we traced from. The reason for not reaching the destination host probably depends on at least one of two factors. The first factor is that the maximum TTL value of 15, set by LocaPing, is insufficient for reaching the host. This is probably the case when tracing to Baidu.com, which we can conclude since there are rather few timeouts to Baidu.com, as seen in Figure 4, but still zero successful traces, as seen in Figure 3. The other factor is that we get several timeouts in some traces. This can be seen in the case of Wikipedia.org, which has a very small share of successful router hops in its traces, zero of which are successful traces. A possible reason for this could be that traceroute uses User Datagram Protocol (UDP) packets for sending its probes. These packets might intentionally be dropped by the servers of for example Wikipedia.org. An indicator of this is that tracerouting using a different tracing tool that does not send UDP packets, but rather Internet Control Message Protocol echo requests, is successful when sent from the same location where the UDP trace fails.

We conclude that traceroutes are not adequate for the purpose of measuring the internet and inferring its topology. The traceroutes need to be carefully selected if this method would work. The destination host is particularly important for tracerouting. As we have seen in our results, we failed reaching 40 percent of our chosen destination hosts. Clearly our measurement of the internet is not as representative as it could be. Acharya et al. [1] have found that a specific kind of network is necessary to solve this problem. They show that special classes of networks are needed, paired with specific traces between extremely many nodes. The general conclusion is that the problem is not solvable for large classes of networks, including the internet.

From looking at both the BGP graph, the traceroute graph and the data we have extracted from them we can see that they are similar, even if they are of completely different order of magnitude. As mentioned earlier, the traceroute graph uses 52 new edges which implies that they are both of value when building a topology. We see that it is necessary to use both BGP data and traceroutes to be able to infer a more correct topology.

From our results we can also see that the average number of AS-to-AS hops taken to the servers of the ten most popular websites is between three and four hops if we do not include the ones where we never reach the destination. A reason for the low number of AS-to-AS hops is that many of the top ten websites have both the interest and often the money to spend on CDNs to increase user experience. This, together with few router hops indicates that we reach the destination server quickly. One exception to this is Wikipedia.org which even though being very popular is a nonprofit organization and therefore might not have the resources of similarly popular sites. In this case, the few number of AS-to-AS hops is the result of many router hops experiencing timeout which leads to the data being incomplete.

By changing the set of websites to the random set, we achieved similar graph characteristics, although the number of AS-to-AS hops increased. This is likely due to these websites having less resources to spend on CDN infrastructure and thus the path becomes longer. Besides changing to a random set we also tried removing all top hosts to which we did not have any successful traces to. This only resulted in a dataset too small to draw any conclusions from and thus we did not choose to include or analyze that data any further.

To conclude, we have shown that building correct internet topologies is difficult and requires the extensive usage of various tools, primarily BGP data and traceroutes. These datasets also needs to be gathered from multiple sources and compiled into a single large dataset of ASs and their connections.

6. REFERENCES

[1] Acharya, H. B., Gouda, M. G., 2009. Brief announcement: the theory of network tracing. *PODC 09 Proc. 28th ACM symp. Princ. Distr. Comp.,* 318-319.

[2] Alexa Internet Inc. 2016. http://www.alexa.com/topsites.

[3] He, Y., Siganos, G., Faloutsos, M. and Krishnamurthy, S. 2009. Lord of the links: a framework for discovering missing links in the internet topology. *IEEE/ACM Trans. Netw.* 17, 2 (Apr. 2009), 391-404.

[4] Gregori, E., Improta, A., Lenzini, L., Rossi, L. and Sani, L. 2012. On the incompleteness of the AS-level graph: a novel methodology for BGP route collector placement. *IMC '12 proc. 2012 ACM conf. Internet meas. conf.,* 253-264.

[5] Oliveira, R. V., Pei, D., Willinger, W., Zhang, B., and Zhang, L. 2008. In search of the elusive ground truth: the internet's as-level connectivity structure. *SIGMETRICS 08 Proc. 2008 ACM SIGMETRICS int. conf. meas. model. comp. Sys.,* 217-228.