AS-level Topology Collection through Looking Glass Servers

Akmal Khan, Ted "Taekyoung" Kwon, *Hyun-chul Kim, Yanghee Choi
Seoul National University,*Sangmyung University
akmalshabaz,hyunchulk@gmail.com, tkkwon,yhchoi@snu.ac.kr

ABSTRACT

While accurate and complete modeling of the Internet topology at the Autonomous System (AS) level is critical for future protocol design, performance evaluation, simulation and analysis, still it remains a challenge to construct its accurate representation. In this paper, we collect BGP route announcements of ASes from Looking glass (LG) servers. By querying LG servers, we build an AS topology estimate of around 116 K AS links, from which we discover 11 K new AS links and 686 new ASes. We conclude that collecting BGP traces from LG servers can help enhance the current view of the AS topology from the BGP collector projects (e.g., RouteViews).

Categories and Subject Descriptors

C.2.2 [**Network Protocols: Routing Protocols**]:

Keywords

Inter-domain Routing, Border Gateway Protocol (BGP), Looking Glass servers

1. INTRODUCTION

The Internet consists of Autonomous Systems (ASes) that exchange inter-domain routing information using Border Gateway Protocol (BGP) [1]. The entire Internet can be viewed as an AS-level topology graph where each AS is a node, and a BGP connection between two ASes is a link. The importance of the AS topology has been highlighted through many studies, such as analyzing Internet topological properties [3,5,9], inferring AS relationships [12], building network topology generators for simulations [13], and evaluating the effectiveness of new protocols and improvements [11]. Considering the importance of the AS topology in many areas of networking research, significant efforts [2,7–10,14,19,20] have been made to discover and construct it. However, it still remains as a challenge to develop a complete and accurate view of the AS-level topology [4,7–9,22].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
IMC'13, October 23–25, 2013, Barcelona, Spain.
Copyright 2013 ACM 978-1-4503-1953-9/13/10 ...$15.00.
http://dx.doi.org/10.1145/2504730.2504758.

There have been three main approaches to construct the AS topology, each of which has its own limitations: (i) passive measurements by collecting BGP routing tables and updates [18] suffer from routing policy filters and best path selection decisions made by the neighboring ASes of BGP collectors [22], (ii) active measurements using traceroute [14,19] are error-prone and generate potentially false AS links due to non-responsive hosts or errors in converting IP addresses to AS numbers (IP-to-AS mapping) [15,16], and (iii) Internet routing registry (IRR) [21] is known to have incomplete and/or outdated information and possibly biased towards RIPE region [6,17].

In this paper, we focus on the construction of the AS topology by using Looking glass (LG) servers, which are web-based portals run by Internet service providers (ISPs) or network operation centers (NOCs), that help understand network status like connectivity, path, and routing information. We collect around 116 K AS links from 245 LG servers across 110 countries. Then, we analyze the LG server-based AS topology against BGP (IRL [18]), traceroute (Ark [19] and iPlane [20]), and IRR based [21] AS topologies. The main findings of our study are as follows.

We observe 620 neighboring ASes of the LG servers that are not sharing their BGP traces with any of RouteViews [31], RIPE-RIS [32], and PCH [33]. We find 11 K new AS links in the AS topology from the LG servers. By comparing the AS topology from the LG servers against the BGP-based AS topology (i.e., IRL [18]), we observe the increase in the connectivity of some popular content providers, such as 22% more for Microsoft. We discover 686 new ASes in the AS topology from the LG servers that are hidden from other AS topologies. Moreover, 98% of the newly discovered ASes have only one provider AS. Overall, we conclude that collecting BGP traces from the LG servers help increase the narrow view of BGP observed from current BGP collectors [22].

2. LOOKING GLASS SERVERS

Looking glass (LG) servers are web based portals operated by network operators to provide a look into the BGP routing tables of the ASes in which the server resides. For example, from a response of a query to an LG server, a network problem can be traced back to its reasons like misconfigured BGP route advertisement, wrong route aggregation, or misconfigured AS path prepending. By an LG server, we mean a web site that allows running commands (e.g., traceroute) from one or more BGP routers that are under the control of the

Table 1: A sample result of the `show ip bgp summary` command.

Router: cr1-eqx3-pa3 Local AS Number: 29075					
Command: show ip bgp summary					
Neighbor	**AS#**	**State**	**Time**	**Received**	**Sent**
195.42.144.104	6939	ESTAB	61d	36,464	153

LG server[1]. For instance, the LG server of Hurricane Electric[2] provides facilities to run LG commands on its BGP routers that are distributed across 92 locations worldwide.

The importance of LG servers in constructing the Internet AS-level topology has been highlighted in many studies [2, 7, 9]. For instance, Augustin *et al.* [2] uses the `show ip bgp summary`[3] and `traceroute` commands to map IXPs, their members, and their peering matrices. While prior studies have shown the usefulness of LG servers, it is not clear what other information (apart from those available with the `show ip bgp summary` command) is available with LG servers for the purpose of collecting the AS topology. Thus, we conducted a comprehensive investigation to find out how many LG servers are operational and what functionalities are provided by individual LG servers. We first build a list of LG servers from the following sources: peeringDB [23], Traceroute.org [24], Traceroute.net.ru [25], BGP4.as [26], BGP4.net [27], and Virusnet [28]. After removing the overlapping LG servers from the above sources, we find 1.2 K LG servers, only 420 of which were in operation at the time of this study, in the month of March 2013. Our scripts can query 388 LG servers since the web sites of the other 20 LG servers are not parsable and 12 LG servers limit automated queries.

We queried 388 LG servers (running on 410 ASes) to learn their supported functionalities. We find that as many as two dozen commands[4] are supported by different LG servers, while a few of them are more widely supported than others. For example, all the 388 LG servers support `traceroute` and `ping` commands from 4.4 K (in total) locations in the Internet. Another widely supported command is `show ip bgp summary`, which is supported by 245 LG servers from 1.9 K locations. The regional Internet registries (RIR) wise distribution of 245 LG servers are as follows: RIPE (175), ARIN (40), APNIC (15), LACNIC (13), and AfriNIC (2).

Table 1 illustrates a sample result of querying a router (cr1-eqx3-pa3 operating at Paris Equinix) with the `show ip bgp summary` command through the LG server provided by Ielo (AS29075). It shows that Ielo has a BGP session with Hurricane Electric (AS6939) at Paris Equinix. It also shows other important information, such as (i) how long the BGP session has been alive (61 days), (ii) 36,464 routes received from the BGP neighbor, and (iii) 153 routes advertised to the BGP neighbor over this link.

[1]A BGP router under the control of an LG server is called an LG router.

[2]Hurricane Electric LG. http://lg.he.net

[3]`show ip bgp summary` lists the BGP sessions established with an LG router, and details the ASN and IP address of its peering BGP router, for each BGP session.

[4]The full list of LG commands along with all the datasets collected in this study can be found at http://mmlab.snu.ac.kr/traces/lg.

Table 2: A sample result of the `BGP neighbor ip advertised routes` command.

Router: cr1-eqx3-pa3 Local AS Number: 29075		
Command: BGP neighbor 195.42.144.104 advertised routes		
Prefix	**Next Hop**	**AS PATH**
149.154.80.0/21	195.42.144.71	29075 50618 57141
91.227.48.0/24	195.42.144.71	29075 50618 25091 56728

We also find that 59 LG servers (distributed over 250 locations) allow us to run the `BGP neighbor ip advertised routes` command, which helps observe IP prefix announcement(s) advertised by an LG router to its peering BGP routers. Table 2 shows a sample result of the `BGP neighbor ip advertised routes` command on the BGP router cr1-eqx3-pa3. Each row shows an IP prefix, its next hop address and AS path information.

3. AS-LEVEL TOPOLOGIES

We first describe our methodology to build an AS topology from LG servers. We also briefly describe AS topology snapshots derived from other data sources, such as BGP traces.

3.1 AS Topology derived from LG servers

We design a tool to automate a querying process to the 388 LG servers. Our tool issues 30 queries in parallel to the LG servers and waits for 15s between successive queries to the same LG server to avoid overloading them. Collecting data from an LG server is a multi-step process. First, for each LG server our tool learns, by parsing LG server websites, the supported LG commands and its LG routers to which our tool sends queries to collect the data. Second, to each LG router, our tool sends the `show ip bgp summary` command to the LG server. Third, from the returned response of `show ip bgp summary`, our tool extracts IP address(es) of the neighboring router(s) of the LG router. Fourth, by using the IP addresses of the neighboring routers, our tool sends a query of `BGP neighbor ip advertised routes` to collect the BGP routes advertised by the LG router to its neighboring routers. Finally, all the responses of the `show ip bgp summary` and `BGP neighbor ip advertised routes` commands from the LG server are stored in text files for constructing the AS topology.

We queried 245 LG servers that provide the option of running `show ip bgp summary` command from around 1.9 K locations (distributed across 110 countries), twice a week in the month of March 2013. Total 8 snapshots are combined to create an AS link dataset, which consists of around 70 K AS links. We find 77% of the AS links are intra-AS links, i.e., the source and destination ASes of a link are the same. As we are only interested in inter-AS links in this study, we filter out these intra-AS links and selected only 16 K inter-AS links. Throughout this paper, AS links refer to those inter-AS links.

We also queried 59 LG servers (out of the 245 ones) that provide the option of running `BGP neighbor ip advertised routes` command, once a week in the month of Mar. 2013. Their LG routers are located in 250 locations distributed across 40 countries. Moreover, these LG servers advertise routes to 5 K routers of their neighboring ASes. From the

BGP traces collected from the 59 LG servers, we extracted around 2 million AS paths and broke down these AS paths into around 103 K AS links.

Overall, by running the `show ip bgp summary` and `BGP neighbor ip advertised routes` commands on the LG servers, we have collected 116 K unique AS links (103 K+16 K=119 K-3 K overlapping AS links). To the best of our knowledge, this is the first study that investigates not only `show ip bgp summary` but also `BGP neighbor ip advertised routes` commands to construct the AS topology.

3.2 Other AS-level Topologies

In this section, we describe other AS topologies: BGP based (IRL [18]), traceroute based (Ark [19], iPlane [20]), and IRR [21] based, all of which are used in this paper. We have decided to use the recently (and regularly) published AS topology datasets only, not including ones such as Ono [10], which had been collected using BitTorrent P2P clients in 2007-2008. Since we cannot quantify how much of this AS topology dataset is outdated, we decided not to use it. Similarly, we exclude DIMES [14] as it has not been updated since Apr. 2012. Note that all the AS topologies are collected in March 2013.

IRL: UCLA IRL [18] regularly publishes the AS topology extracted from the BGP traces shared by RouteViews [31], RIPE-RIS [32], Packet Clearing House (PCH) [33], and Internet2 [34]. We find 179 K AS links in the AS topology published by IRL in March 2013.

Ark: CAIDA Archipelago (Ark) [19] provides the AS topology derived from the traceroute based measurements. As of March 2013, there are 71 Ark monitors distributed across the Internet. We used Ark's IPv4 Routed/24 AS link dataset, which includes direct and indirect AS links. A direct AS link means that a pair of connected ASes have a pair of adjacent hops in the traceroute path, while an indirect AS link means that two connected ASes are separated by one or more unmapped or non-responsive hops. We discarded indirect AS links and used direct AS links. We find 116 K direct AS links from the Ark dataset in March 2013.

iPlane: Madhyastha *et al.* proposed the iPlane [20] service that performs traceroute from around 300 PlanetLab locations daily to map the Internet topology. We find 81 K AS links from the iPlane AS topology in March 2013.

Internet Routing Registry (IRR): The Internet Routing Registry (IRR) [21] is a set of globally distributed databases with which ASes can register their routing and address related information. The IRR has been reported to contain stale records [6]. Nevertheless, the IRR contains AS links that are not observed in BGP and traceroute based AS topologies [9]. we downloaded the IRR data from the RIPE Routing Registry (RR) [29] and RADB [30] in March 2013. To derive AS link information from the IRR, we investigate *aut-num* objects' import, export and default attributes. We extract 206 K AS links from the IRR, most of which (75%) come from the RIPE RR.

To remove possibly outdated AS links from the IRR dataset, earlier studies like [7] have relied on `changed` attribute of an IRR record, which shows the last updated date of the IRR record. However, we find that RIPE no longer publishes `changed` attribute in the IRR records. To remove outdated AS links from the IRR, we collected the AS topology snapshots published by IRL [18] dating back till 2004. We want to check whether some of the AS links in the IRR were observed in the historical BGP based AS topologies. Such comparison confirms that some of the AS links found in the IRR were operational in BGP but are not currently visible from BGP. Using the above method, we find as many as 46 K (out of 206 K) outdated IRR AS links. We removed these AS links and have used around 160 K AS links from the IRR.

4. COMPARISON OF BGP FEEDERS

There are three popular BGP collector projects: RouteViews [31], RIPE-RIS [32], and PCH [33]. The ASes sharing their BGP traces to the BGP collector projects are known as BGP feeders [22]. In this section, we are interested in finding out whether, by querying LG servers, we can collect BGP traces from ASes that are not BGP feeders of RouteViews, RIPE-RIS, and PCH. Such analysis indicates whether BGP traces collected from the LG servers help discover new AS links that are not found in the other AS topology datasets (e.g., IRL [18]). Moreover, collecting BGP traces from new BGP feeders help widen our limited view of BGP observed from current BGP collectors [22].

We have collected information regarding the BGP feeders (i.e., ASNs and IP addresses of routers) of RouteViews [31], RIPE-RIS [32], and PCH [33] from their websites in the month of March 2013. The comparison between the BGP feeders of different projects are based on the ASN and IP address of BGP routers sharing the BGP traces. That is, if the ASN and/or IP address of a BGP router matches between the LG servers and RouteViews BGP feeders, then it is considered that the same BGP router (of an AS) is sharing its BGP traces with both RouteViews and the LG servers.

Table 3 shows the number of common BGP feeders (ASes and routers) sharing their BGP traces with the RouteViews, RIPE-RIS, PCH, or LG server datasets. The diagonal (in bold) is the number of BGP feeders available only in one dataset; either in RV, RIPE, PCH, or LG servers. We observe differences in the number of ASes and router IPs overlapping between different datasets. For example, 63 neighboring ASes of LG servers are sharing their traces with RV. However, only 36 router IPs are matched between LG servers and RV. Further investigation leads us to find the following two reasons for such mismatches: (i) An AS can be peering on an IPv4 connection with RV while on an IPv6 connection with LG servers. In that case, when two datasets are compared to check for the overlapping ASNs and router IPs, the observed router IPs can be different in both datasets though they are with the same ASN. (ii) An AS can be peering with RV at a different location in the Internet from where an LG server is located, thus the observed router IPs between the two datasets can be different as well, while they have the same ASNs.

Moreover, we find that 545 (out of 1.1 K) neighboring ASes of the LG servers overlap with RouteViews, RIPE-RIS, or PCH. More importantly, we observe that 620 neighboring ASes of the LG routers are not sharing their BGP traces with RouteViews, RIPE-RIS, nor PCH. We further inspect the number of routes announced by each neighboring ASes of LG servers to find that 70% (of the 1.1 K) neighboring ASes of the LG servers announce a small number (1 to 100) of BGP routes, since most of these ASes are stub ASes. The remaining (30%) neighbors of LG servers announce BGP routes in the range of 100 to 450 K. Overall, we were able

Table 3: The number of overlapping and unique (in bold) ASes and peering routers between various BGP feeders (RouteViews (RV), RIPE-RIS, PCH, and LG servers).

Collector (Total # of ASes and Routers)	RV ASes (Routers)	RIPE	PCH	LG servers
RV (179 and 368)	**72 (276)**	46 (27)	76 (44)	63 (36)
RIPE (343 and 599)	46 (27)	**51 (314)**	235 (215)	191 (133)
PCH (1.2 K and 2.7 K)	76 (44)	235 (215)	**719 (2 K)**	428 (615)
LG servers (1.1 K and 3.3 K)	63 (36)	191 (133)	428 (615)	**620 (2.6 K)**

Table 4: The number of overlapping and unique (in bold) AS links between various AS topology datasets.

Source (Links)	**LGs**	**IRL**	**IRR**	**Ark**	**iPlane**
LGs (116 K)	**11 K**	99 K	46 K	67 K	45 K
IRL (179 K)	99 K	**51 K**	62 K	75 K	48 K
IRR (160 K)	46 K	62 K	**93 K**	36 K	24 K
Ark (116 K)	67 K	75 K	36 K	**30 K**	51 K
iPlane (81 K)	45 K	48 K	24 K	51 K	**25 K**

to collect 128 BGP routing tables of around 450 K prefixes from the LG servers, which is approximately equal to the size of full BGP routing table in the current Internet [35].

The analysis presented so far in this section suggests that there are many ASes who are willing to publicly share their BGP traces by operating LG servers, which in turn begs the question that why such ASes have yet to offer feeds to route collectors. We suggest two possible reasons: (i) In the past, network operators were motivated to share their BGP feeds to the route collectors in order to advertise their rich connectivity and dominance (especially Tier-1's) in the Internet [22]. However, they may not need to do that any more as maintaining an LG server serves that purpose too. Besides, maintaining an LG server by an AS is helpful for operational reasons such as troubleshooting routing issues. (ii) BGP collector projects such as RouteViews have presence at a limited number of locations in the Internet (e.g., large IXPs) and mostly collect traces from ASes present at those locations (e.g., members of large IXPs). Thus, RouteViews can not collect traces from the ASes which are not located at these locations but are sharing their feeds to the LG servers.

5. ANALYSIS OF AS-LEVEL TOPOLOGIES

In this section, we present the analysis of the AS topology generated by the LG server datasets in comparison to the ones from the other datasets.

5.1 Overlapping/Unique AS Links

We compare AS topologies generated from different datasets, to find overlapping and unique AS links among them. Such analysis are important in quantifying how many AS links are newly discovered from the LG servers. Table 4 indicates the number of common AS links for each pair of datasets. The diagonal zone from upper-left to bottom-right (in bold) reports the count of AS links appearing in only one dataset, i.e., either in LGs, IRL, IRR, Ark or iPlane only. We observe 11 K unique AS links in the AS topology from the LG servers. We find 51 K unique AS links in IRL, which suggests that there are a significant number of AS links not observed

Table 5: Top 10 ASes in terms of the number of newly discovered AS links through LG servers. The number of AS links found in the IRL (2nd column), the number of newly discovered AS links from the LG servers (3rd column), and the corresponding increase in the AS connectivity in percentage (4th column).

AS Name (ASN)	**In IRL**	**New in LGs**	**Percent Incr.**
Level3 GBLX (AS3549)	3,290	112	3.40%
Abovenet(AS6461)	1,119	109	9.74%
Google (AS15169)	164	31	18.90%
Globalnet (AS31500)	115	29	25.22%
GlobalSol. (AS12713)	86	27	31.40%
Microsoft (AS8075)	122	27	22.13%
Yahoo (AS10310)	150	23	15.33%
Amazon (AS16509)	132	22	16.67%
EdgeCast (AS15133)	112	19	16.96%
Facebook (AS32934)	99	19	19.19%

from the BGP feeders of LG servers. The IRR contains 93 K unique AS links, some or many of which might have been outdated. However, as noted earlier in Section 3.2, there is no clear way of filtering out the outdated information from the IRR. Finally, Ark and iPlane have their contribution of unique AS links, while many of these AS links may not be accurate either, due to the traceroute measurement issues such as IP-to-AS mapping [15, 16].

To better understand the geographic location of newly discovered AS links from the LG servers, we use the Team Cymru WHOIS service [37] to map the ASes in the newly discovered 11 K AS links to the regional Internet registries (RIRs). We find 6.5 K AS links whose source AS (in an AS link) is from RIPE, 3.4 K from ARIN, 0.8 K from APNIC, 295 from LACNIC, and 88 from AfriNIC. Considering that the IRR registration is popular in RIPE [6], it is surprising to observe that the largest portion (6.5 K out of 11 K) of newly discovered AS links are from RIPE. That is, there are still many ASes in the RIPE region who do not correctly maintain their routing policies in the IRR. Further inspection reveals that most of these 11 K AS links do not go beyond their RIR regions. For instance, 80% of the newly discovered AS links from the the RIPE region have both of the source and destination ASes in that region.

Finally, we analyze the differences in the AS connectivity of all the 4 K source ASes (i.e., every source AS of the 11 K newly discovered AS links) with respect to the AS connectivity of the ASes observed in IRL [18], which is known as the most complete BGP-based AS topology dataset [7]. Table 5 shows, for each of top 10 ASes in terms of the number of new AS links found from our LG dataset, the number of AS links found in the IRL (2nd column), the number of newly discovered AS links from the LG servers (3rd column), and the corresponding increase in AS connectivity in percentage (4th column). We found 112 and 109 new AS links for the Tier-1 ASes Level 3 and Abovenet, respectively, through the LG servers. We also observe the increase in the AS connectivity of some other large ASes, e.g., 18.9% increase in the AS connectivity of Google (AS15169).

Why do the LG servers miss AS links observed in other datasets? While the AS topology collected from LG servers reports new AS links (11 K), it is missing many AS links visible in other datasets, e.g., 51 K in the IRL. This can be due to the following reasons. First, BGP feeders may provide a full feed to the router collector projects such as RouteViews while they may share only partial feed to LG servers due to economic relationships (such as Peer-to-Peer) with ASes operating the LG servers. Second, LG servers also suffer from vantage point bias. More specifically, depending on the view of a BGP feeder and its location in the Internet, only a specific part of AS topology can be discovered by an LG server. Third, it is not clear whether all the AS links published by the traceroute and IRR datasets are correct.

Why do the other datasets not see AS links discovered using LG servers? The reasons are different depending on different AS topology datasets. First, the incompleteness of BGP-based datasets has been reported in a few studies [7, 22], which is due to the limited number of ASes sharing their BGP feeds to RouteViews and RIPE-RIS. Moreover, AS topology view from BGP-based datasets is also biased, as current route collector projects have better views of the core rather than the other parts of the Internet; Tier-1s more actively share their BGP traces than other ASes [22]. Second, traceroute-based datasets suffer from limited vantage points, selectively probing prefixes, IP-to-AS mapping issues [15, 16], and can not discover backup links [20]. Finally, many of the ASes, except in the RIPE region, do not actively use the IRR to register their routing policy related information [17,36]. Thus, topology collection from LG servers result in discovering otherwise unobserved part of the Internet as LG servers provide new BGP feeders from geographically diverse locations in the Internet. Note that new AS links found in the AS topology from the LG servers can be hidden from other router collectors due to traffic engineering practices (such as route aggregation) of ASes in BGP [7, 10].

5.2 Overlapping/Unique ASes

Table 6 shows the number of common ASes for each pair of the datasets. The diagonal zone from upper-left to bottom-right (in bold) reports the count of ASes appearing in only one dataset. We observe 686 unique ASes from LG servers while IRL has only 181 unique ASes. There are around 6.2 K unique ASes in the IRR; mostly from RIPE and APNIC regions suggesting the popularity of the IRR usage in these regions [17]. We do not find any unique ASes in Ark, which suggests their accurate use of the BGP traces since Ark relies on BGP traces from RouteViews to convert IP-to-AS paths collected from traceroute measurements. Finally, iPlane shows 281 unique ASes. On further investigation, we find that these ASes were not operational in BGP in the month of March 2013. For example, (i) AS1448 (DPEC Partners) had been visible until February 18, 2013, (ii) AS40027 (Netflix) had been visible until February 23, 2012, and (iii) AS42411 (Vereya) had been visible until March 16, 2012, but not any more. The reason why we observed 281 unique ASes in iPlane is because it uses BGP traces that contain older data collected from July 2006 to March 2013 during the IP to AS mapping process, constructing its AS topology.

Table 6: The number of overlapping and unique (in bold) ASes found in various AS topology datasets.

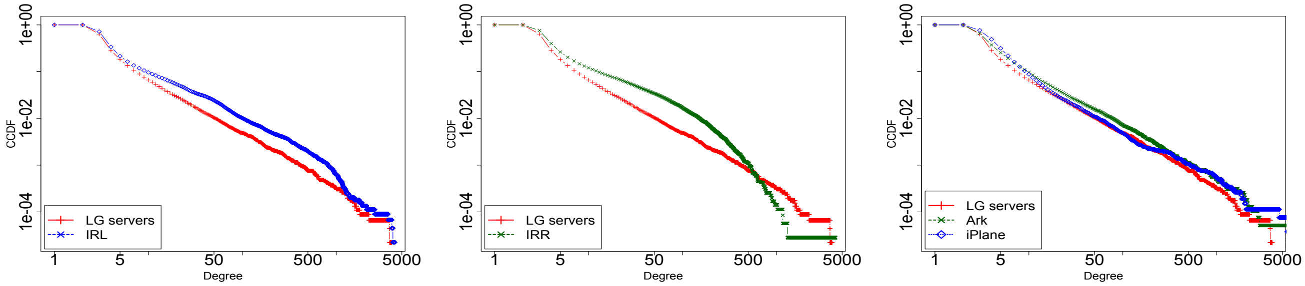
Source (ASes)	**LGs**	**IRL**	**IRR**	**Ark**	**iPlane**
LGs (45.4 K)	**686**	44.1 K	28.7 K	36.9 K	25.4 K
IRL (44.9 K)	44.1 K	**181**	28.8 K	37.1 K	25.5 K
IRR (35.7 K)	28.7 K	28.8 K	**6.2 K**	24.5 K	17.4 K
Ark (37.1 K)	36.9 K	37.1 K	24.5 K	**-**	23.6 K
iPlane (26.1 K)	25.4 K	25.5 K	17.4 K	23.6 K	**281**

We check the RIR wise distribution of these 686 new ASes found from the LG servers, and find that 548 ASes are operational in ARIN, 71 in RIPE, 31 in APNIC, 28 in LACNIC, and 8 in AfriNIC. Most of the ASes are from the ARIN region, possibly due to the lower popularity of the IRR registration practice in that region, compared with the RIPE and APNIC ones [17]. We also check the length of IP prefixes announced by the new 686 ASes to find that 90% of the ASes announce IP prefixes of length between /19 and /24, which suggests that these IP prefixes are possibly aggregated by their provider ASes. Thus, the practice of IP prefix aggregation in BGP is one of the possible reasons that many of the new ASes observed from the LG servers are hidden from RouteViews, RIPE-RIS, and PCH BGP collectors. Yet, we observe that the practice of IP prefix aggregation in BGP may not be the only reason for these 686 hidden ASes since we also observe a few IP prefixes of length between /16 and /18, which needs further study.

Finally, we check the provider degree (i.e., the number of provider ASes of an AS) of each of the 686 ASes, and find that 98% of the ASes have only one provider AS. That is, an AS with a small provider degree is more likely to be hidden behind their provider ASes [7, 9]. Still, there are some ASes with their provider degrees ranging between 2 and 10. For instance, AS410 (754th Electronic Systems Group) has 10 provider ASes. To further investigate 686 hidden ASes, we queried two popular WHOIS services: Team Cymru and Hurricane Electric's WHOIS service [38]. These WHOIS services reportedly have access to the BGP feeds of ASes that are not the BGP feeders of RouteViews and RIPE-RIS [10]. Thus, they can provide some more insights related to the operation of these hidden ASes. While the WHOIS records confirm the operation of 150 (out of 686) ASes found from the LG servers, most of the ASes do not have their details in the WHOIS services either.

5.3 AS Degree Distribution

In this section, we investigate whether the different methods of collecting AS topologies result in different AS degree distributions. The **AS degree distribution** is the probability that a randomly selected AS is k-degree: $P(k) = n(k)/n$; where n is the number of ASes and $n(k)$ is the number of ASes with degree k. The degree distribution is the most frequently used topology characteristic [5].



(a) LG and BGP-based (b) LG and IRR (c) LG and Traceroute-based

Figure 1: AS degree distribution (CCDF) of the AS topology obtained from the LG server dataset, as compared to those of other datasets.

Figure 1 shows the CCDFs of the AS topology datasets plotted along with that of the LG server-derived AS topology. Figure 1(a) shows that IRL provides more connectivity information related to moderate degree ASes than the LG servers.

Figure 1(b) shows that IRR has many more AS links for the moderate degree ASes which implies that the IRR is popular in the realm of smaller ISPs. Large ISPs have shown little interest in the IRR as it is difficult to manage complex routing policies in the IRR [36]. Figure 1(c) shows that LG servers and iPlane closely match for the low degree ASes. However, iPlane reports more AS links for higher degree ASes. To find out the reasons, we analyze the traceroute IP paths collected from the iPlane in March 2013. We observe that iPlane has a selective list of IP prefixes (120 K out of approximately 450 K IP prefixes that are currently operational in BGP [35]) to probe the Internet and this list seems to concentrate more on probing the core Internet [20]. Thus, iPlane discovers more connectivity of higher degree ASes. Finally, Ark reports more connectivity of the moderate degree ASes than the LG severs and iPlane. Overall, traceroute based projects such as Ark and iPlane suffer from limited vantage points, selectively probing IP prefixes, errors in the translation from IP to AS Path. Such factors impact the AS topology view observed from traceroute measurements, which can also be error prone [15, 16].

6. RELATED WORK

There have been a number of measurement studies related to the AS topology discovery [2, 7–9, 14, 18, 19]. To quote the most recent efforts, He *et al.* [9] provide a large scale comprehensive synthesis of the available routing data sources such as BGP routing tables, IRR, and traceroute data. Augustin *et al.* [2] build on the work of He *et al.* [9], but the focus is on the IXP substrate, not on the AS topology as a whole. Active measurement platforms such as Ark [19], DIMES [14], and iPlane [20] are providing the AS topology views, but suffer from the small number of vantage points to run traceroute measurements. To overcome the limitation, Chen *et al.* [10] propose to send traceroute probes from a large number of (992,000 P2P user IPs in 3,700 ASes) P2P clients.

So far, LG servers have been considered as a secondary source of inter-domain routing data for discovering links in the Internet topology [2, 7, 9, 18]. That is, LG servers have been used to augment some AS links to the AS topology extracted from BGP traces [7], or used to help verify the AS links found in the IRR [9]. To the best of our knowledge, this paper is the first to show that LG servers are yet another non-negligible source for building Internet AS topology. Moreover, collecting BGP traces from the LG servers can help widen the narrow view of BGP observed from the current BGP collector projects, such as RouteViews, RIPE-RIS, and PCH [22].

7. CONCLUSION

In this paper, we highlight the less-known capabilities of Looking glass (LG) servers to construct Internet AS topology. By collecting **show ip bgp summary** command responses from 245 LG servers (from 1.9 K locations in 110 countries) and **BGP neighbor ip advertised routes** command responses from 59 LG servers (from 250 locations in 40 countries) in March 2013, we build an AS topology estimate of around 116 K AS links. We newly discovered 11 K AS links and 686 ASes that are not found in BGP, traceroute, and IRR based AS topologies. Clearly, LG servers help in augmenting the current AS topology collection efforts reliably as BGP based methods are less error prone as compared to traceroute-based ones. However, the AS topology view from the LG servers suffers from limited vantage points of the LG servers and BGP export policies employed by the neighboring ASes of LG servers. Overall, we envision that more LG servers are deployed in the Internet in the future as they are important operational tools deployed by ASes. Thus, the research community needs to be aware of the facilities provided by them to discover Internet AS topology.

8. ACKNOWLEDGEMENTS

We are grateful to the anonymous IMC reviewers for their insightful comments. This research was supported in part by the KCC (Korea Communications Commission), Korea, under the R&D program supervised by the KCA (Korea Communications Agency) (KCA-2013-11-911-05-002) and by the MSIP (Ministry of Science, ICT & Future Planning), Korea in the ICT R&D Program 2013, and by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2013R1A1A2010474). The corresponding authors of this paper are Ted "Taekyoung" Kwon and Hyun-chul Kim.

9. REFERENCES

[1] Y. Rekhter, T. Li, and S. Hares. A Border Gateway Protocol. *RFC 4271*, January, 2006.

[2] B. Augustin, B. Krishnamurthy, and W. Willinger. IXPs: mapped?. *ACM IMC*, November, 2009.

[3] G. Gursun, N. Ruchansky, E. Terzi, and M. Crovella. Routing state distance: a path-based metric for network analysis. *ACM IMC*, November, 2012.

[4] B. Ager, N. Chatzis, A. Feldmann, N. Sarrar, S. Uhlig, and W. Willinger. Anatomy of a Large European IXP. *ACM SIGCOMM*, August, 2012.

[5] P. Mahadevan, D. Krioukov, M. Fomenkov, X. Dimitropoulos, K. Claffy, and A. Vahdat. The internet AS-level topology: three data sources and one definitive metric. *ACM SIGCOMM CCR*, vol. 36, no. 1, January, 2006.

[6] K. Butler, T. Farley, P. McDaniel, and J. Rexford. A Survey of BGP Security Issues and Solutions. *Proc. of the IEEE*, vol. 98, no. 1, January, 2010.

[7] R. Oliveira, D. Pei, W. Willinger, B. Zhang, and L. Zhang. The (In) Completeness of the Observed Internet AS-level Structure. *IEEE/ACM TON*, vol. 18, no. 1, February, 2010.

[8] H. Chang, R. Govindan, S. Jamin, S. Shenker, and W. Willinger. Towards Capturing Representative AS-Level Internet Topologies. *Computer Networks*, vol. 44, no. 6, April, 2004.

[9] Y. He, G. Siganos, and M. Faloutsos. Lord of the Links: A Framework for Discovering Missing Links in the Internet Topology. *IEEE/ACM TON*, vol. 17, no. 2, April, 2009.

[10] K. Chen, D. Choffnes, R. Potharaju, Y. Chen, F. Bustamante, D. Pei, and Y. Zhao. Where the Sidewalk Ends: Extending the Internet AS Graph Using Traceroutes From P2P. *ACM CoNEXT*, December, 2009.

[11] John S. Otto, M. Sanchez, D. Choffnes, F. Bustamante, and G. Siganos. On Blind Mice and the Elephant – Understanding the Network Impact of a Large Distributed System. *ACM SIGCOMM*, August, 2011.

[12] X. Dimitropoulos, D. Krioukov, M. Fomenkov, B. Huffaker, Y. Hyun, K. claffy, and G. Riley. AS relationships:Inference and validation. *ACM SIGCOMM CCR*, vol. 37, no. 1, January, 2007.

[13] P. Mahadevan, C. Hubble, D. Krioukov, B. Huffaker, and A. Vahdat. Orbis: rescaling degree correlations to generate annotated internet topologies. *ACM SIGCOMM*, August, 2007.

[14] Y. Shavitt, E. Shir. DIMES: Let the Internet measure itself. *ACM SIGCOMM CCR*, vol. 35, no. 5, October, 2005.

[15] Y. Zhang, R. Oliveira, Y. Wang, S. Su, B. Zhang, J. Bi, H. Zhang, and L. Zhang. A Framework to Quantify the Pitfalls of Using Traceroute in AS-Level Topology Measurement. *IEEE JSAC*, vol. 29, no. 9, October, 2011.

[16] P. Marchetta, W. Donato, A. Pescapé. Detecting Third-party Addresses in Traceroute Traces with IP Timestamp Option. *PAM*, March, 2013.

[17] A. Khan, H. Kim, T. Kwon, and Y. Choi. A comparative study on IP prefixes and their origin Ases in BGP and the IRR. *ACM SIGCOMM CCR*, vol. 43, no. 3, July, 2013.

[18] UCLA IRL, http://irl.cs.ucla.edu/topology.

[19] CAIDA ARK. http://www.caida.org/projects/ark.

[20] H. V. Madhyastha, T. Isdal, M. Piatek, C. Dixon, T. Anderson, A. Krishnamurthy, and A. Venkataramani. iPlane: An Information Plane for Distributed Services. *USENIX NSDI*, May, 2006.

[21] Internet Routing Registry. http://www.irr.net.

[22] 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. *ACM IMC*, November, 2012.

[23] PeeringDB, http://www.peeringdb.com.

[24] traceroute.org, http://www.traceroute.org.

[25] traceroute.net.ru, www.traceroute.net.ru.

[26] bgp4.as, http://www.bgp4.as/looking-glasses.

[27] bgp4.net, wiki, http://www.bgp4.net.

[28] virus-net.ru, http://www.virus-net.ru.

[29] RIPE DB, ftp://ftp.ripe.net/ripe/dbase.

[30] RADB, ftp://ftp.radb.net/radb/dbase.

[31] RouteViews, http://www.routeviews.org.

[32] RIPE-RIS, http://www.ripe.net/ris.

[33] Packet Clearing House, https://www.pch.net/resources/data.php.

[34] Internet2, http://ndb7.net.internet2.edu/bgp.

[35] BGP potaroo, http://bgp.potaroo.net.

[36] Route Filter Panel, Nanog43, June, 2008. http://www.nanog.org/meetings/nanog43.

[37] Team Cymru WHOIS service, http://asn.cymru.com.

[38] BGP Toolkit, http://bgp.he.net.