BGP Routing Table: Trends and Challenges

Alexander Afanasyev, Neil Tilley, Brent Longstaff, and Lixia Zhang

Computer Science Department, University of California, Los Angeles
{afanasev,tilleyns,blongsta,lixia}@cs.ucla.edu

Abstract. BGP is currently the most important protocol for ensuring global connectivity over the Internet. This puts a great deal of responsibility on BGP and creates a number of challenges for it. Of primary concern is the impact that various currently deployed BGP-based techniques have on the scalability of the global routing table. While these techniques provide ISPs with additional traffic management services (e.g., balancing, multi-homing, etc), they have expanded the routing table size at a pace that exceeds the allocation rate and is increasing. In our study we present a two-level analysis of BGP announcements for the period 2003–2009. First, we correlate IP allocation data with globally announced prefixes and show how efficiently ISPs announce their allocated address space. Second, we correlate BGP announcement data to itself and show various internal factors that contribute to routing table growth. Finally, we document in which regions of the world routing announcements have originated during the period of this study, and we draw conclusions about the spread of global Internet connectivity.

1 Introduction

BGP (Border Gateway Protocol) [1] is a key component enabling Internet routing worldwide. The routing protocol operates at the junction where independent networks (ASes, or autonomous systems) exchange network traffic to ensure global connectivity. Because ASes are separate networking and economic entities, BGP currently operates while essentially balancing two purposes that are for the most part orthogonal to one another. First, it interconnects all ASes in the world. Second, BGP tries to satisfy a wide variety of ISP-specific routing policies, which are governed by operating costs, a number of agreement-based and politically-based issues, network locality, multi-homing preferences, and, in some select cases, traffic connection capacity.

The routing table has expanded enormously over the past ten years due to fractionalization and finer segmentation of the IP address space. The table now maintains more entries than a hierarchical structure would have yielded that had worked with strictly consolidated blocks. Fig. 1 shows that over the past six years, the number of the global routing table entries has more than doubled. IP address allocations have also doubled in the same period, but numerically, all new allocated blocks account for less than 18% of the actual entries in the BGP routing table (≈50k new allocations from 2003 to 2009, compared to ≈300k entries in the routing table in 2009).

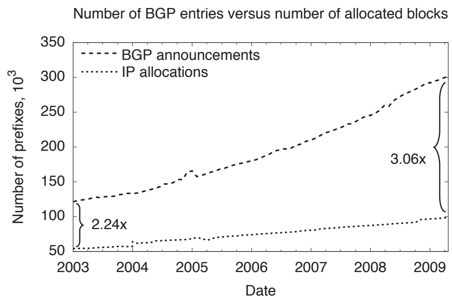


Fig. 1. Number of BGP entries compared to number of allocated IP blocks

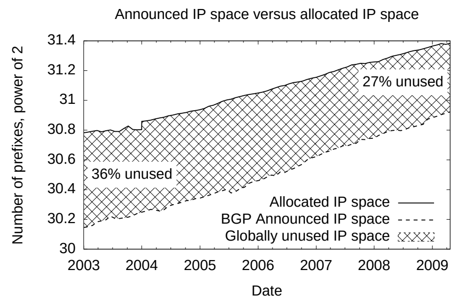


Fig. 2. Comparing the IP space that has been allocated to the amount of IP space announced in the BGP table

There is an interesting aspect to notice with the BGP table dynamics. While the routing table has experienced substantial growth in comparison to the number of new IP blocks allocated, all the IP space covered by the entries in the routing table is still only a fraction of the total measured IP space that has been allocated. This is shown in Fig. 2, which illustrates that over time, the amount of dormant IP space—allocated, but not announced in routing tables—has ranged from a little over $1/3$ to a little over $1/4$ of the total. Despite the high ratio between the number of announced prefixes in the BGP table and the number of actually allocated IP blocks (2.24x in 2003, 3.06x in 2009), a significant portion of IP space ($\approx$27%) remains unreachable globally.

In this paper we present an extensive analysis of the IP allocation and BGP announcement statistics. First, we examine the dynamics of IP address allocation and recent history regarding which prefix block sizes are most popular to allocate to ISPs (Internet Service Providers). Second, we examine the correlation between the global routing table and IP prefix allocation data. Among the common prefix block sizes allocated, we show what size blocks are most common in the BGP table. Third, we present a summary by geography of those regions around the globe that contribute to the BGP table contents each year. We also identify the regions where the most rapid development of ISP hosting has occurred. Fourth, we estimate the lifespan of BGP entries as measured between 2003 and 2009. We then discuss a number of factors that contribute to the marked growth in size of the routing table.

The rest of the paper is organized as follows. Section 2 presents an analysis for IP address allocation statistics. In Section 3 we analyze the composition of the BGP table and its changes over time, as well as the stability of routing table entries. Finally, we present related work and conclusions in Sections 4 and 5, respectively.

2 IP address allocation dynamics

2.1 Allocated IP block sizes

The collected data highlight a number of aspects about allocated block sizes over the past six years. Fig. 3 illustrates several notable changes. It shows the number of allocations increasing for every prefix length in every year, though at different rates. In all years, clearly /24 prefix allocations are the most prevalent. At the same time, there has been almost no increase in /16 blocks. Overall, this shows the growing trend toward smaller blocks as the IPv4 address space approaches saturation.

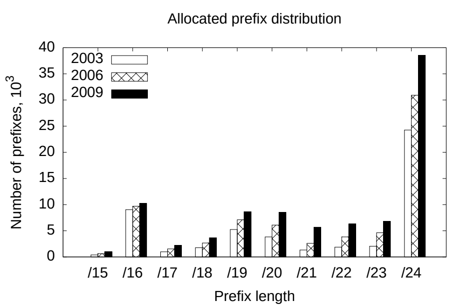


Fig. 3. Allocated prefix distribution

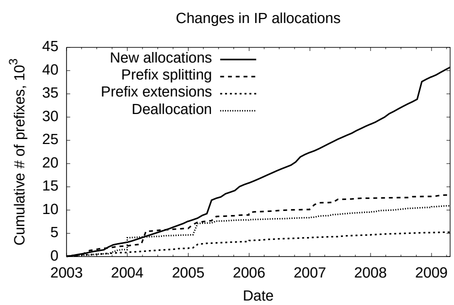


Fig. 4. Changes in IP allocations

2.2 Yearly distribution of IP allocations

Fig. 4 charts the occurrence of other events that factor into the net change of the allocation count. These include prefix splitting, prefix extension, and deallocation. Prefix splitting is the dividing of one prefix into multiple smaller prefixes (e.g., a /16 prefix can be split into two /17). Prefix extension is the aggregating of an existing prefix with its adjacent, previously unallocated address space (e.g. a /17 prefix might become a /16 prefix by including the adjacent /17 address space). Deallocation is the withdrawing of a prefix allocation, essentially releasing IP space for later use.

The number of new allocations over the studied time period exceeds 40k. A smaller amount (13k) is due to prefix splitting. A small decrease in the allocations count (5k) has resulted from prefix extension. Finally, a larger decrease is due to deallocation (10k), but the decrease is nevertheless greatly outpaced by the number of new allocations. To summarize, growth in the allocation count is due primarily to new allocations. Steady, quasi-linear growth in recent years suggests that the increase in allocations will continue at a similar, moderate pace after IPv6 deployment.

3 BGP routing table

3.1 Analysis of BGP table growth factors

The BGP routing table is growing at a rate significantly higher than the pace that RIRs (Regional Internet Registries) are allocating IP blocks. Across all observed BGP monitors [2, 3], the average number of entries in the global routing table is more than 3 times the number of IP blocks that RIRs have allocated (refer Fig. 1). This multiplication in size reflects two primary practices. First, ISPs tend to subdivide allocated IP blocks into several individual prefixes and announce them separately. Such behavior is typical among transnational providers as well as among ISP customers that have been lent parts of their service providers' address space and, in turn, independently announce subdivided IP address blocks. Second, various traffic engineering techniques (traffic balancing, multihoming, etc.) give rise to situations where the same address block is covered by several announced prefixes.

IP block fragmentation. The contents of the BGP routing table consist of IP prefixes that either match, fragment, or aggregate various IP allocation blocks. Fig. 5 shows the correlation between allocated IP blocks and announced IP prefixes and the relative proportions of these three categories over time. The *matched* curve in the figure represents IP blocks that are announced in the routing table in the exact form that they were issued by RIRs. As evident in the figure, the number of matched prefixes accounts for 1/6 of the total BGP entries presently, with the trend that this fraction is growing smaller over time.

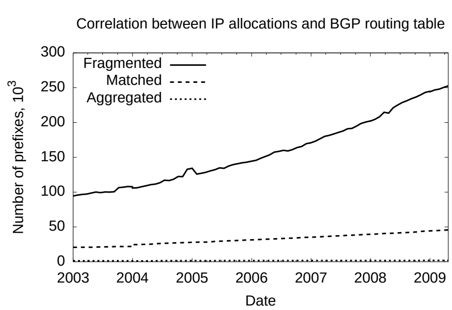


Fig. 5. Dynamics of matched, fragmented, and aggregated IP prefixes in BGP announcements

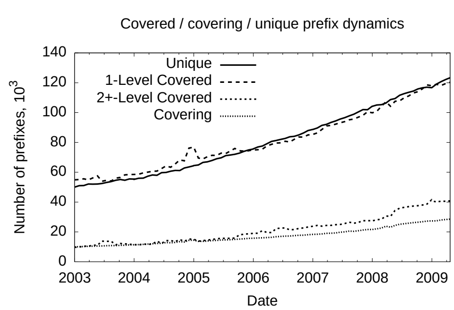


Fig. 6. Relative numbers of *covered*, *covering*, and *unique* IP prefixes among BGP announcements

ISPs are evidently not inclined to use address space in the form in which it has been allocated. For various possible reasons (e.g., geographical dispersion), ISPs split up allocated blocks into a number of sub-blocks and announce each of these independently. The *fragmented* curve in Fig. 5 represents these subdivided blocks, which account for more than 83% of all entries in the global routing table.

IP block fragmentation poses one of the primary concerns for future scalability of the BGP routing table.

The lowest curve, *aggregated*, represents IP prefix announcements that cover several allocated IP blocks. For these cases—in contrast to fragmentation—ISPs that have several adjacent IP block allocations simply announce them as a single IP prefix. As the figure illustrates, this aggregation technique is rarely employed in any measure. Its primary intent, to reduce the number of entries in the routing table, is outweighed far more by other policy routing choices. The number of aggregated prefixes in 2003 was 1,400. By 2009 this number increased only marginally, to just under 2,000 prefixes, accounting for under 1% of the entire routing table and thus is considered negligible.

This observed behavior has a measure of relevance to future IPv6 deployment. ISPs tend not to announce their allocated IP spaces in their original form. This behavior occurs regardless of the size of IP block that an ISP has been allocated. According to current RIR policy, the minimum allocation for an IPv6 block is /32 [4]. In terms of cost, the price for an IPv6 /32 block is the same as for a /19 or /20 IPv4 address block [5], meaning for an equivalent amount to obtain fewer than 10,000 IPv4 addresses, ISPs can be assigned an IPv6 block several orders of magnitude larger than the entire IPv4 space. If allocations of large blocks continue, it is likely to mitigate the problem of multiple, non-adjacent IP block allocations per customer. However, without a major change in the BGP protocol aimed at lessening incentives to announce fragmented IP prefixes, increasing the allocated IP block size will not significantly assist in reducing the size of global routing table. Table size reduction stems only from aggregatable and matching IP prefixes. In other words, only ISPs that currently use all allocated IP space as a single IP block (i.e., matched or aggregated) have any likelihood of using a bigger space provided in IPv6 also as a single block. We conclude that the upper bound of IP space announcement optimization is limited by the number of matched prefixes, which currently stands at less than 17% of all prefixes.

Duplicate announcements of IP blocks. The BGP routing table has a large number of prefix ranges that overlap each other. Such IP address coverage duplication assists in calculating an actual route by matching the destination address with the longest available prefix. Address duplication in a routing table is, in theory, an effective way to reduce the size of the routing table itself. Our observations indicate that ISPs use the fundamental IP routing feature of longest-prefix matching extensively. As Fig. 6 shows by the *1-level* and *unique* curves, the number of IP prefixes in the BGP table which are covered by exactly one bigger prefix is nearly the same as the number of unique prefixes (i.e., base prefixes). There is moreover a substantial number of prefixes that have several layers of coverage (several duplication levels—refer to *2+-level* curve). We also have found that about half of the autonomous systems (≈58% in 2003 and ≈44% in 2009) employ prefix duplication. This finding proves that prefix duplication is common practice on the Internet.

These high proportions of *1-level* and *2+-level* covered prefixes indicate there are other incentives and benefits for prefix duplication, in addition to the theoretical routing table optimization. One factor we consider is a multi-provider connection for end-networks (so-called multi-homing of *stub networks*). As Oliveira et al. [6] have stated, more than 70% of all BGP announcements belong to multi-homed stub networks. In other words, the global routing table is employed to serve local or semi-local routing interests for most customers. Since these routing interests have importance primarily on a local scale, it is unlikely that the outside world follows widely divergent routing paths to reach various providers' connections to a multi-homed customer. While the need for covered prefixes is evident to accommodate various routing preferences, they need not be shared universally in routing tables. Accordingly, with counter-incentives to IP prefix fragmentation, a significant reduction in the size of the BGP routing table is conceivable. For example, deployment of geography-based techniques (e.g., GIRO protocol [7]) can assist in reducing the global routing table to $1/4$ of the current size. As an area of further research, separate means for multi-homing and traffic engineering tasks might be provided.

3.2 Analysis of the BGP table contents

BGP announcements by geographical region. In examining the global routing table content, we have conducted a country-based analysis of the distribution of globally announced IP prefixes. Fixed-time snapshots point out the major contributors to the global routing table and give an understanding of the Internet's penetration globally. Country assignments for IP prefixes are taken from RIR data, which provide enough precision for a global-scale analysis.

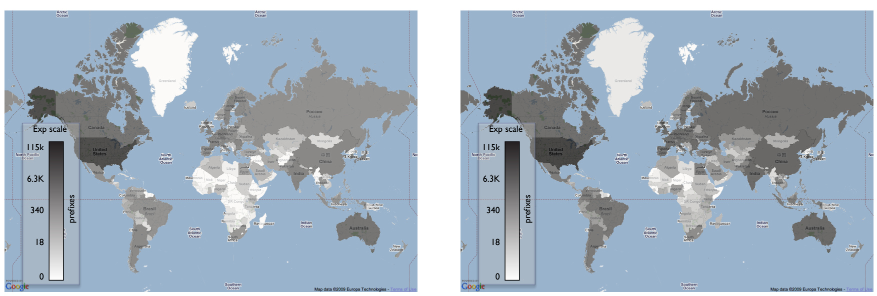
By way of summarizing results, we present collected information for the six countries worldwide that have the most impact on the global routing table in 2009 (Table 1). Matching their 2003 and 2009 numbers side-by-side, they provide easy comparison of changes over time. All countries contribute in greater numbers to the BGP table. The United States retains its leading position. Meanwhile the ordering of the rest of the contributors significantly changes. For example, South Korea becomes the second major contributor to the size of the global routing table, and China ranks third.

Comparing IP space usage to the number of announcements attributed to a given country, China, Japan, the European Union, Germany, and South Korea are responsible for the most amount of announced address space, behind the United States. These five geographic areas are a different ordering from that given in Table 1 (South Korea, China, Australia, India, Russia). This fact highlights that some countries announce a large number of relatively small prefixes (e.g., in South Korea one prefix covers on average 5,900 addresses), whereas some announce a small number of large prefixes (e.g., in Japan one prefix covers on average 37,200 addresses). If this difference in usage efficiency occurs because of additional government regulations, then for future IPv6 deployment, a similar host of regulations should be considered globally.

Table 1. Statistics for announced IP prefixes and corresponding IP space

Country	**Prefixes**		**IP addresses**	
	2003	2009	2003	2009
US	65k	116k	760M	1,170M
South Korea	2k	14k	27M	84M
China	2k	13k	28M	224M
Australia	6k	11k	47M	38M
India	2k	11k	3M	18M
Russia	1k	9k	6M	27M

Fig. 7 shows a comparison of Internet penetration globally between years 2003 and 2009. Note that in both illustrations, shades of gray assigned to countries correspond to an exponential scale. Comparing 2009 with 2003, all regions show various degrees of greater Internet hosting. While the United States retains the lead position throughout the years measured, several regions with emerging economies exhibit marked change (toward darker shades) within the six-year span, such as Russia, China, and India.



(a) 2003 (b) 2009

Fig. 7. Geographical distribution of the number of announced prefixes. Given on an exponential color scale, from light (least) penetration to dark (greatest).

An extended presentation illustrating the geographical distribution of IP prefixes and IP space is available online, at `http://lasr.cs.ucla.edu/afanasyev/09-routing-map/`

Lengths of announced IP prefixes. Fig. 8 presents the distribution of announced prefix lengths, classified by each year. The majority of the global routing table entries are /24 prefixes and account for more than 53% of the entries. The number of /24 prefixes has nearly doubled between 2003 and 2009. At the same time, the number of blocks allocated that are actually /24 in size is 4 times smaller (refer Fig. 3). On the one hand, this is evidence that large numbers of stub networks (i.e., relatively small customer networks) use announcements of small address blocks to implement multi-provider connectivity. On the other hand, the excessive number of /24 prefixes poses a question about the relevance of current and future algorithms for IP space assignment (e.g., sequential, bisection allocation, and GAP algorithm [8]). If the majority of ISPs tend to fragment an assigned address space into small chunks and announce them separately, attempts at minimizing the number of assigned IP blocks per organization will yield little effect and have only limited impact.

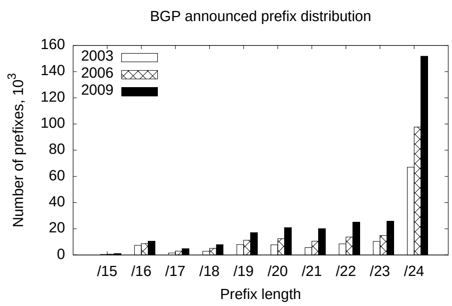


Fig. 8. Distribution of announced IP prefix lengths

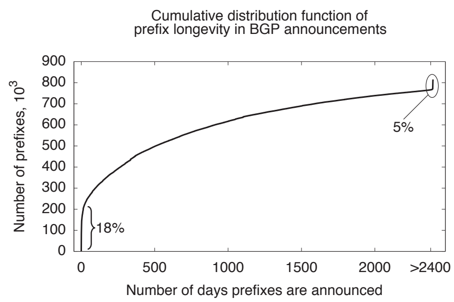


Fig. 9. Longevity of prefixes in BGP announcements

The results in this section indicate a tight relationship between global routing table growth and IP space fragmentation and duplication. If we could suppress the majority of non-globally related (i.e., locally concerned) announcements, such as bursts of /24 prefixes due to local traffic engineering, local and semi-local multihoming support, etc., the size of the global routing table could be significantly reduced.

Longevity distribution of BGP entries. Another aspect of the BGP announcement analysis is determining the stability of the global routing table. A cumulative distribution function of prefix longevity appears in Fig. 9. The step on the far right of the graph indicates that about 5% of all prefixes remain stable over the six-year period (≈40k equal to 15% of the current global routing table). On the far-left side a large number of prefixes (18% of all prefixes observed) are active for only short periods of time (i.e., 1–3 days). One portion of these short-lived prefixes is likely composed of spammers that are known to hijack somebody's (or even no one's) prefix, announce it for a brief time, and send virtually untraceable spam messages [9]. Another portion can be attributed to configuration errors. The rest can be explained by normal BGP operations, where some prefix becomes briefly visible, as for example when a primary network channel malfunctions.

With regard to prefix length, we observe that /24 is prevalent both for long- and short-lived prefixes. However, most entries in the BGP table in the range from /25 to /32 are composed almost entirely of short-lived (unstable) prefixes.

Besides a fixed number of highly stable routes, more than half of the prefixes appear in the table for less than 500 days. It is unlikely then that a given route is visible for a very long time. This observation underscores that the composition of the routing table is highly dynamic and will pose challenges for future research.

4 Related work

Past studies [10–12] have characterized growth of the Border Gateway Protocol routing table in terms of the prevalence of special announcements to suit traffic engineering purposes. They also have measured the number of appearing and disappearing announcements in the BGP routing table, the latency between allocation and prefix appearance in BGP announcements, and the level of unallocated address announcements. Since the time of those studies (2003–2005), the BGP routing table has continued its growth. Our study examines the current state of the BGP routing table and quantifies how the high-level picture has changed from earlier measurements.

Geoff Huston's Potaroo project [13] presents up-to-date measurements of the BGP routing table growth and IP allocation dynamics, dating back to 1994. However it does not analyze the impact that fragmentation and address space duplication have on the BGP table growth over time and how they affect estimates of the routing table size in the future.

The contribution of our research is to document where routing announcements are originating around the world—not necessarily a measure of where most new Internet traffic is occurring, but a way to witness the spreading of Internet infrastructure connectivity around the globe. Coupled with an analysis of how long these announcements stay in the routing table—a measure of table stability—it is possible to make some projections of how the Internet continues to diversify.

5 Conclusions

We have analyzed BGP announcement snapshots provided by the University of Oregon Route Views and RIPE NCC Routing Information Service projects. Between 2003 and 2009, a span of six years, the average size of the BGP routing table has more than doubled. Likewise the IP address allocations have also doubled in the same period. Numerically, however, all the new allocated blocks have added less than 18% of the total entries in the BGP routing table.

We have identified several primary causes of the accelerated BGP routing table growth. They are as follows: First, fragmentation of allocated IP blocks; more than 80% of announced prefixes are from allocated IP blocks that have been subdivided. Second, announced space duplication; more than 54% of the address space is covered at least twice in the global routing table. This duplication highlights an emerging problematic trend of using the global routing table to serve local interests, e.g., to implement traffic engineering and multi-provider connections.

The content analysis of BGP routing table announcements shows that the majority of globally announced prefixes (>50%) are of /24 size. This strengthens the conclusion that a substantial number of entries in the global routing table serves local, not global, interests of small customer networks. The content of the global routing table is highly dynamic. Although there is a small portion of highly stable entries (<15%), the remainder of the BGP table content fits an exponential tapering-off distribution for prefix longevity.

Our examination of the geographical distribution of IP allocation and BGP announced prefixes shows a depth Internet penetration around the globe that is wide-ranging. We have observed a number of quasi-exponential distributions for various measurements, including for the following: the geographical distribution of the number of allocated prefixes, the numbers of the corresponding address spaces, the number of announced prefixes, and the corresponding globally announced address spaces. Moreover these distributions have not significantly changed in character over the last six years.

References

1. Rekhter, Y., Li, T., Hares, S., et al.: RFC1771 – A border gateway protocol 4 (BGP-4). RFC (March 1995)
2. University of Oregon: Route Views Project. `http://routeviews.org/`
3. RIPE NCC: Routing Information Service (RIS). `http://www.ripe.net/ris`
4. APNIC, ARIN, RIPE NCC: IPv6 address allocation and assignment policy. ripe-466 (February 2009)
5. ARIN: Annual fee scedule. `https://www.arin.net/fees/fee_schedule.html` (2009)
6. Oliveira, R.V., Zhang, B., Zhang, L.: Observing the evolution of Internet as topology. In: Proc. of SIGCOMM'07 conference on Applications, technologies, architectures, and protocols for computer communications, New York, NY, USA, ACM (2007) 313–324
7. Oliveira, R., Lad, M., Zhang, B., Zhang, L.: Geographically informed inter-domain routing. In: Proc. of ICNP'2007. (2007) 103–112
8. Wang, M., Dunn, L., Mao, W., Chen, T.: Reduce IP address fragmentation through allocation. Proc. of ICCCN 2007 (August 2007) 371–376
9. Ramachandran, A., Feamster, N.: Understanding the network-level behavior of spammers. SIGCOMM Comput. Commun. Rev. **36**(4) (2006) 291–302
10. Meng, X., Zhang, B., Huston, G., Lu, S.: IPv4 address allocation and the BGP routing table evolution. SIGCOMM Comput. Commun. Rev. Special Issue on Internet Vital Statistics **35**(1) (January 2005) 71–80
11. Xu, Z., Meng, X., Lu, S., Zhang, L., Wittbrodt, C.J.: IPv4 Address Allocation and the Evolution of the BGP routing table. Technical Report TR-03009, UCLA Computer Science Department, Los Angeles, CA 90095, USA (March 2003)
12. Meng, X., Xu, Z., Zhang, L., Lu, S.: An analysis of BGP routing table evolution. Technical Report TR-30046, UCLA Computer Science Department, Los Angeles, CA 90095, USA (October 2003)
13. Huston, G.: IPv4 Address Report. `http://www.potaroo.net/tools/ipv4/` (2009)