An Analysis of BGP Routing Table Evolution

Xiaoqiao Meng, Zhiguo Xu, Lixia Zhang and Songwu Lu
UCLA Computer Science Department, Los Angeles, CA 90095
E-mails: {xqmeng, zhiguo, lixia, slu}@cs.ucla.edu
Technical Report TR030046

Abstract

In addition to the ever growing host population, multiple other factors have contributed to the rapid growth of the global Internet routing table, such as policy routing, multi-homing, and traffic steering. In this paper we first sort the routing table entries into two broad classes, covering prefixes which represent IP address blocks that do not overlap, either partially or completely, with the address block represented by any other entry in the routing table, and covered prefixes, commonly referred to as "holes", which represent sub-blocks of those address blocks that are already represented by some shorter prefixes in the routing table. We then develop a classification methodology by identifying the different ways each covered prefix is announced to the global routing system. We inferred the causes of each covered prefix class and identified possible motives for the fragmentation of covering prefixes. Based on our analysis, we provide an empirical model of the global routing table growth by taking into account all the major factors that have been identified in this study.

1 Introduction

As the Internet continues to grow in user population, its global routing table size also experiences a rapid growth. This growth has attracted widespread attention in the network research community (see [7, 22, 25, 11] and the references therein for a sampling of the literature). Efforts have been made to characterize the global routing table growth from various different aspects, such as the trend of the growth in table size, the distribution of the prefix lengths, and the amount of IP address space covered by the routing table.

To better understand the future growth trend of the global routing table, in this paper we conduct a detailed examination of the *changes* in the BGP routing table content over the last four years. We first divide the total route entries into two classes, *covering prefixes* which represent IP address blocks that do not overlap, either partially or completely, with the address block represented by any other entry in the routing table, and *holes*, which represent sub-blocks of those address blocks that are already represented by some shorter prefixes in the routing table. For example, a routing table entry 10.0.0.0/22 is classified as a hole if the entry 10.0.0.0/19[1] is also present in the routing table. Our measurements show that the holes make up about half of the routing table entries. Furthermore, these "hole" entries change over time more rapidly than the covering prefixes, playing a more active role in the evolution of the routing table.

In this paper we analyze the growth of the numbers of holes and covering prefixes separately. First, we classify the holes into four categories based on their routing advertisement modes. For each category, we describe the effect of the holes on the data traffic flow and hypothesize the operational practice that leads to the creation of the holes in the category. Secondly, we analyze the covering prefixes and show that 40%-60% of the existing covering prefixes are fragments of previously allocated address blocks. We then classify these fragments into three categories based on the routing advertisement modes. For each category, we infer the possible causes for the fragmentation and justify the inference by case studies. Based on our findings we build an empirical model to approximate the actual routing table growth. This model takes into account multiple factors, including address allocation, fragmentation of covering prefixes and the observed trend of holes. The model would allow us to predict the routing table grows linearly in the near future.

The rest of the paper is organized as follows. Section 2 provides a brief introduction to BGP and IP address allocation. Section 3 describes the data sets and methodology used in our study. Section 4 presents our study and measurement on the entire set of route entries. We then present an analysis of holes in Section 5, and an analysis of covering prefixes in Section 6. In Section 7 we describe an empir-

[1] Without explicit mention, prefixes used in examples are purely for illustration purposes. They do not represent the reality

ical model of the BGP table size growth. We discuss the implications of our results in Section 8. Finally, we present related work and conclusions in Sections 9 and 10 respectively.

2 Background

BGP is the *de facto* standard for inter-domain routing in the global Internet. It uses *prefix*, represented by a 32-bit address and a mask length, to identify the destination network. For example, 10.0.0.0/8 specifies a block of contiguous IP addresses ranging from 10.0.0.0 to 10.255.255.255. To obtain routing information for individual prefixes, the edge routers within two neighboring ASes need to establish BGP sessions with each other to exchange routing information. Once an edge router acquires new routing information on how to reach an individual destination prefix, it announces an update message to inform its peer in the other AS. The update message for a prefix is represented by an AS path. Each AS path includes a list of ASes along which the AS originating the prefix can be reached. All the routing information collected by an edge router constitutes its BGP routing table. In reality, the two neighboring ASes might be physically scattered over large geographic regions and they might have more than one edge router pair and multiple BGP sessions.

BGP routing is greatly affected by the local routing polices of the AS. An edge router uses import policies to filter unwanted routes or alter attributes associated with the route. It also uses import policies to influence the best-route-selction process. The router selects a single *best* route for each individual destination prefix among all the routes it has received. Following this, the router employs export policies to determine whether it announces the route to its neighboring ASes and to what extent the route should be propagated. The router may also alter route attributes to enforce other local routing policies.

Internet customers apply for IP addresses from either the four Regional Internet Registries (RIRs) or ISPs. The IP address allocation also proceeds in prefix-based blocks. Classless Inter-Domain Routing (CIDR) (around 1993-1994) allowed a flexible boundary between the network-prefix and the host-number field. It provides flexible address allocation and enables more efficient utilization of the address space. CIDR also supports and encourages route aggregation, which may limit the growth of the BGP routing table. For example, if an ISP is allocated a prefix 10.0.0.0/19, it can split it into 10.0.0.0/20, 10.0.16.0/20 and assign them to two customers separately. The ISP can advertise 10.0.0.0/19 instead of two individual /20s. Such route aggregation can effectively decrease the number of route announcements to the global routing table.

3 Data sets and methodology

This section describes the data sets and research methodology used in our study.

3.1 Data sets

We use two data sets in this work: BGP routing table traces and IPv4 address allocation records. The routing table records came from the RouteView project [5]. In the most recent data set from RouteView we have used (December 1, 2002), RouteView recorded 25 peering sessions. We choose to use the routing table collected at peer 204.42.253.253 (through AS267 and it belongs to IAGNET) since this vantage point has been constantly present in the RouteView data during the past four years. To avoid potential biases due to the partial view of a single routing table trace, we verify our results by using routing tables collected at other vantage points. The IPv4 address allocation records used in our analysis are from the four Regional Internet Registries (RIRs)[3]: ARIN, RIPE, APNIC and LACNIC.

3.2 Methodology

Our analysis in this work proceeds by following four guidelines listed below:

Analyzing covering prefixes and holes differently BGP routing table growth consists of two components: holes and those covering prefixes fragmented from the allocation. We analyze the holes and covering prefixes differently for two reasons. (1) The basic function of BGP table is to provide reachability to individual IP address. If reachability is the main concern, the covering prefixes should be sufficient through hierarchical routing and multiple levels of prefix aggregation of CIDR [20]. However, in reality the BGP table includes a large number of holes that do not bring in any "new" addresses. Therefore, the motives behind such holes should be different from the covering prefixes. We have to treat them differently. (2) Covering prefixes typically evolve from the allocated blocks while holes are more closely related to their covering prefixes. The information used to infer the motives behind covering prefixes and holes is quite different.

Classifying holes To analyze the motives behind announcing holes, we classify holes into four categories based on the relations between holes and their covering prefixes. Such relations are expressed in the AS-level structure observed from the routing data. In our classification, we apply Gao's algorithm [23] to infer the commercial relationship

between AS pairs. The inferred relations can be provider-to-customer, peer-to-peer or sibling-to-sibling. Such information is used in the classification and it helps to determine the motives behind a hole in the specific scenarios.

Classifying covering prefixes fragmented from allocation The second component of the BGP table growth is those covering prefixes fragmented from the allocation. Typically an allocated block is designated to a single organization. Whenever this block is fragmented into several covering prefixes, these covering prefixes should be closely related in principle. Based on this observation and the resulting advertisement modes observed from the routing table, we classify these covering prefixes (fragmented from allocation) into three categories. We also identify practical scenarios that may lead to each of these three categories. In the categorization, we exploit the difference between "assigned" address block and "allocated" address block, which is recorded in the allocation records. According to the address allocation policy of RIRs, address blocks are distributed in two ways: (1)"Allocated" address block is distributed to ISPs for the purpose of further delegating the address space to smaller ISPs or end users. (2)"Assigned" address space is distributed to ISPs or end users without allowing for further assignment. By differentiating "assigned" and "allocated"[2], we can speculate the role that the owner of an address block is playing. This further helps us understand the possible causes behind the fragmentation.

Observing route announcements via multiple vantage points Inferring motives behind a hole (or a covering prefix fragmented from the allocation) requires capturing all the BGP announcements from the ASes involved. However, a single vantage point can only provide partial view of the global Internet [12]. Therefore, we propose to use the BGP tables collected by all the available vantage points in RouteView project. This way, the risk of losing some announced routes should be negligible given the rich redundancy provided by RouteView.

4 Evolution of route prefixes

Since the deployment of BGP4, the BGP routing table size has grown consistently faster than the number of globally routable addresses represented by the table [7][9][25][26]. Our study based on the data set collected at RouteView shows that, the global routing table size has almost doubled over the past four years but the address space increases by merely 30%.

[2]For ease of exposition, in the paper only quoted *assigned* and quoted *allocated* represent their special meanings specified by the allocation record.

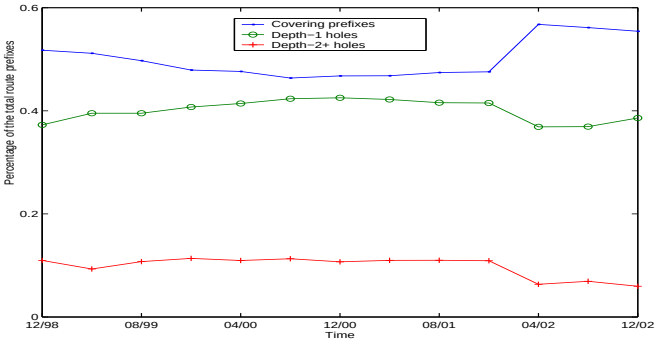

Figure 1. Growth of covering prefixes and holes

This section provides a quantitative analysis on how the routing table content has evolved during the past four years. Specifically, we show the quantitative evolution of covering prefixes and holes.

4.1 Covering prefixes and holes

We first quantify the percentage of covering prefixes and holes in the BGP table. The entire set of route prefixes can be classified into two categories: covering prefixes and holes. For illustration purposes, we use *depth-1* to denote holes that can ONLY be covered by covering prefixes. In accordance, we use *depth-2+* to denote holes that can also be covered by at least another hole. For example, given three prefixes 10.0.0.0/8, 10.0.0.0/19, 10.0.0.0/22 in the routing table and that no other prefixes can summarize or be summarized by any of these three prefixes, 10.0.0.0/8 is a covering prefix while 10.0.0.0/19 is a depth-1 hole and 10.0.0.0/22 is a depth-2+ hole.

Figure 1 depicts the percentage of covering prefixes and holes with different depth in the routing table over the four-year period of December 1998 to December 2002. The figure shows that the holes contribute about half of the routing table entries. The covering prefixes and depth-1 holes together contribute to about 95% of the entire routing table. This trend has not changed significantly over the past four years.

Since the holes constitute half of the routing table entries and their address space can been fully represented by their corresponding covering prefixes, we turn to show how these holes coexist with the covering prefixes. The study shows that the covering prefixes are taking a highly uneven distribution in terms of holes they have. As high as 90% covering prefixes have no holes while the other 10% covering prefixes bring up all the holes in the routing table.

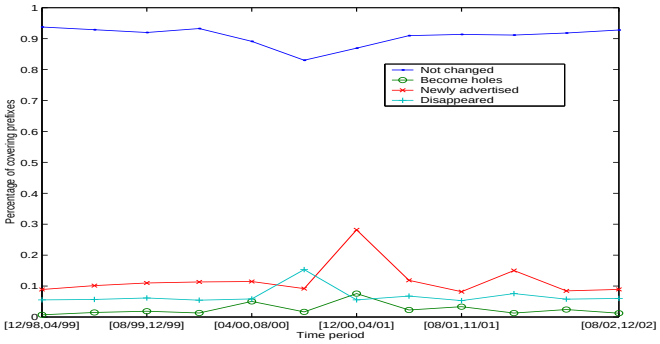

Figure 2. Evolution of covering prefixes

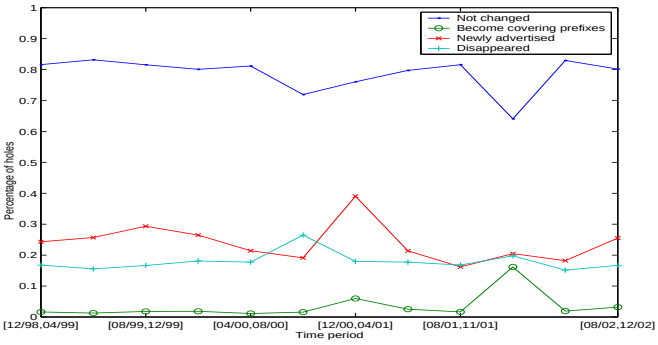

Figure 3. Evolution of holes

4.2 Evolution of covering prefixes and holes

We next study the evolution of the routing prefixes over time. Our study shows that the evolution of the global routing table includes not only the advertisement of new prefixes, but also the disappearance of historical prefixes. These historical prefixes used to be advertised but no longer exist in the current routing table. In fact, according to the statistics of [25], if no routing prefixes disappeared, the routing table would have been 5 times larger.

We first measure the evolution of covering prefixes in the past four years. To this end, we choose the 13 routing tables collected at the IAGNET peer and this naturally divides the past four years into 12 time periods each of which lasts for roughly four months. By comparing the routing tables at the beginning and the end of the time interval, we compute the percentage of covering prefixes that have: (1) kept as covering prefix, (2) switched from covering prefix to hole, (3) been newly announced, (4) disappeared. The result is plotted in Figure 2. It shows that 90% covering prefixes are stable on average. In a similar way and based on the same data sets, we plot the evolution of holes in Figure 3. A noticeable phenomenon reflected from the figure is that a non-trivial part of holes (24%) keep dropping out of the BGP routing table while another slightly smaller part of new holes (18%) show up. Moreover, the percentage of those dropped holes is consistently larger than the emerging ones except for period [Aug.2000, Dec.2000]. This discrepancy turns out to be the major boost to the BGP table size.

In a summary, we have demonstrated that the covering prefixes and holes have not changed much in terms of their percentage in the BGP table. However, the much more active behavior of the holes may reflect their different functionality from the covering prefixes, as we will show next.

5 Holes

As seen in Figure 1, holes contribute to about 45% of the total BGP table size. In addition, they appear to be playing a more active role in the BGP table evolution. To get a better understanding of the holes, we first classify holes into four categories, and for each category we discuss practical operations that can create holes accordingly. We also apply the classification to the real BGP data and provide some insights. The depth-2+ holes are ignored in the study since they only contribute less than 5% of the total BGP table size (see Figure 1).

5.1 A classification of holes

We classify holes into four categories: same origin AS and same AS path (SOSP), same origin AS and different AS paths (SODP), different origin ASes and same AS path (DOSP), and different origin ASes and different AS paths (DODP).

The above classification is based on the advertisement modes for both the hole and the covering prefix. Such an advertisement modes take into account metrics such as AS path of the hole, AS path of the covering prefix, origin AS of the hole (the AS that announces the hole), origin AS of the covering prefix (the AS that announces the covering prefix). Commercial AS relationships inferred by Gao's algorithm [23] are also incorporated in the classification. An implicit assumption made here is that it rarely happens that the same prefix is announced by more than one AS in the global routing table[3]. The following gives more details on each category.

[3]We do observe such prefixes in real BGP data. However, their percentage among the whole BGP entries is less than 0.1%.

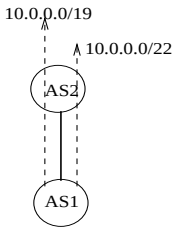

Figure 4. Same origin AS and same AS path

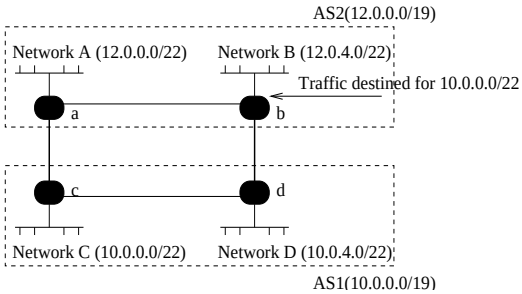

Figure 5. Hole punching between neighbor ASes to circumvent hot potato routing

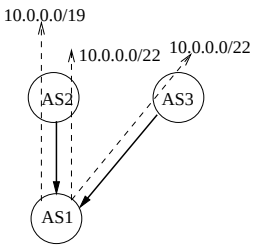

Figure 6. Same origin AS and different AS paths (type 1)

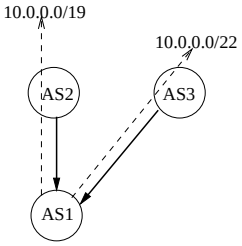

Figure 7. Same origin AS and different AS paths (type 2)

5.1.1 Same origin AS and same AS path (SOSP)

This category of holes meets three requirements: (1) The origin AS for both the hole and the covering prefix is the same. (2) Both the hole and the covering prefix share the same AS path attribute. (3) The hole is not announced via any other AS paths or by any other ASes. The third property can be checked by examining routing tables from different vantage points.

The three requirements can be illustrated by Figure 4[4]. In the example, the dashed arrow represents an announcement of a prefix. A thick line without arrow between two ASes denotes that the two ASes have any commercial relationship, while a thick line with arrow represents that the two ASes have a provider-to-customer relationship (AS at the arrowhead is the customer). In the example, 10.0.0.0/22 is announced simultaneously with the shorter prefix 10.0.0.0/19 and both share the same origin and the same AS path. Therefore, 10.0.0.0/22 should be categorized as SOSP.

We speculate that holes in this category can be created by *hot potato* routing, a common practice employed by ASes. The example of Figure 5 illustrates this cause. In the example, AS1 and AS2 are neighbor ASes and they have blocks 10.0.0.0/19 and 12.0.0.0/19, respectively. Initially, AS1 advertises 10.0.0.0/19 via its edge router *c* and *d* to AS2. Similarly, AS2 advertises 12.0.0.0/19 via edge router *a* and *b* to AS1. Using hot potato routing, once AS2 sees that traffic arriving at *b* has the eventual destination 10.0.0.0/22, which is the network *C* inside AS1, AS2 will immediately send the traffic out via the edge router closest to where the traffic enters AS2 (*b* in the example). This way, AS2 saves its backbone bandwidth by increasing AS1's burden. To circumvent AS2's hot potato routing strategy, AS1 could announce the network *C*'s block 10.0.0.0/22 via *c* in practice. Since 10.0.0.0/22 is more specific than the announced covering prefix 10.0.0.0/19, AS2 is forced to carry the aforementioned traffic to *a* through its own backbone, and then forwards to the edge router *c* inside AS1.

In the example of Figure 4, the function of announcing a hole is to redistribute incoming traffic among multiple physical links between two neighbor ASes. This is to both circumvent the hot potato policy and achieve load balancing.

5.1.2 Same origin AS and different AS paths (SODP)

The second category has to meet two requirements: (1) The hole shares the same origin AS with the covering prefix. (2) There exists at least one announcement of the hole going along an AS path different from that of the covering prefix. Compared with the first category, the hole here is not necessarily uniquely announced. In fact, we can further classify it into two types based on the number of announcement for the hole in the BGP routing table:

- Type 1: This type is exemplified by Figure 6. The hole 10.0.0.0/22 is originated by AS1 through two different AS paths. One path is via AS3, while the other is the

[4]Without explicit mention, prefixes used in the example do not represent the reality. They are purely for illustration purposes

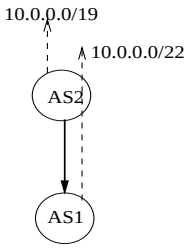

Figure 8. Different origin ASes and same AS path

same as the announced covering prefix 10.0.0.0/19. In such a scenario, traffic destined for 10.0.0.0/22 may traverse either AS2 or AS3.

- Type 2: This type is illustrated by Figure 7. The hole 10.0.0.0/22 is observed to be announced by AS1 via a single AS path. All the traffic destined for 10.0.0.0/22 will traverse the link between AS1 and AS3.

An important advantage for announcing holes in this category is that the origin AS is capable of steering incoming traffic along an AS path different from that of the covering prefix. In reality, this feature can be utilized by the AS to achieve load balancing or multihoming.

5.1.3 Different origin ASes and same AS path (DOSP)

The third category meets four requirements: (1) Origin AS of the hole is different from the origin AS of the covering prefix . (2) The AS path of the hole is exactly a subset of the covering prefix's AS path, i.e., after the hole is propagated from its origin AS, it goes along the same AS path as the announcement of the covering prefix. (3) The hole is not announced by more than one AS or along more than one AS path. (4) AS1 is a customer of AS2 and AS1 has no other providers. These four features are exemplified by Figure 8.

We speculate that such a scenario is typically brought up by the customer AS1's requirements for more fine-grained local routing policies. In the example described by Figure 8, AS1 is single-homed. Ideally AS2 can safely aggregate 10.0.0.0/22 and make a single announcement 10.0.0.0/19. This would not lose AS1's connectivity. However, AS1 may have routing policies different from AS2. For example, due to business concerns, AS1 may want to control the propagation range of its BGP announcement. He then declare such a policy through tagging the BGP community attribute with appropriate value. Accordingly, to satisfy such requirements from AS1, AS2 can not do the aggregation. Instead, it will announce the hole 10.0.0.0/22 appropriately[5].

[5]In BGP community, a provider typically does not perform *proxy aggregation*, i.e., the provider would not aggregate more specific routes originating from the customers. See [2] for the detailed explanation

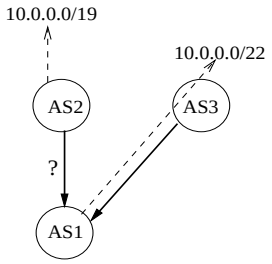

Figure 9. Different origin ASes and different AS paths (type 1)

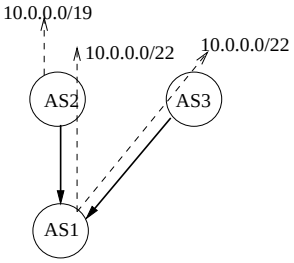

Figure 10. Different origin ASes and different AS paths (type 2)

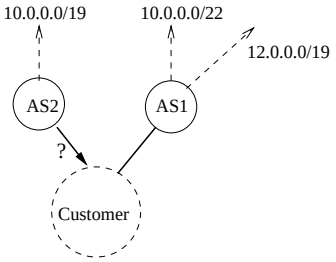

Figure 11. Different origin ASes and different AS paths (type 3)

5.1.4 Different origin ASes and different AS paths (DODP)

The last category is characterized by two features: (1) The covering prefix and the hole have the same origin AS. (2) There exists at least one announcement of the hole that traverses an AS path different from the covering prefix. This category can be further divided into three types:

- Type 1: Type 1 imposes two additional requirements: (1) The AS path for all the announcements of the hole is different from the AS path of the covering prefix. (2) The origin AS does not announce any prefixes other than the hole.

  Figure 9 illustrates this type. In the example, AS1's announcement of 10.0.0.0/22 steers all the incoming traffic through AS3. We speculate that such a scenario can arise by two activities: (1)The end user AS1 switches to a new provider AS3 while still using its old addresses assigned by AS2. (2) The end user AS1 is multihoming to AS2 and AS3. To differentiate these two cases, additional information about the real traffic between AS1 and AS2 is required. As a side note, we infer AS1 to be an end user, simply because AS1 does not announce any other prefixes and it fits the normal behavior of an end user.

- Type 2: In addition to the two requirements of Type 1, Type 2 adds another requirement that the hole be also announced by its origin AS via the same AS path as the covering prefix, exemplified by Figure 10. A common practice leading to such scenario is that the origin AS of the hole is doing multihoming and it seeks to balance the incoming traffic among connections to its multiple providers. Note that this scenario is similar to SODP Type-1 (see Figure 9) in terms of impact on the incoming traffic. However, in SODP Type-1, the origin AS of the hole is the same as the origin AS of the covering prefix; therefore, it typically reflects activities of large ISPs. The current scenario is more likely to represent activities of end users.

- Type 3: Type 3 differs from Types 1 and 2 by requiring that the origin AS of the hole also announce prefixes other than the hole (illustrated in Figure 11, where AS1 announces both the hole 10.0.0.0/22 and another prefix 12.0.0.0/19). Such a scenario is likely to reflect the multihoming activities of customers that are not ASes themselves or customers that have private AS numbers[6]. Typically such customers are non-transit users, i.e., they only send or receive traffic. A large number of organizations fall into this situation. Except for the original AS, this scenario actually shares the same traffic characteristics as Type 1 (Figure 9).

[6]Private AS numbers are typically not shown in AS path

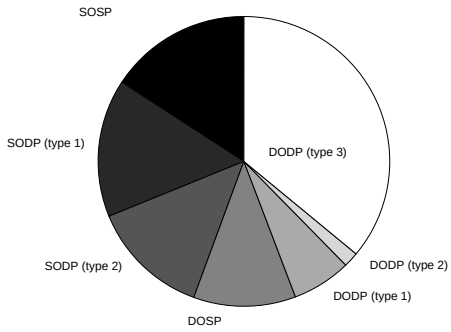

Figure 12. Fraction of holes in the four categories (based on the routing table collected by IAGNET peer 204.42.253.253 on Dec.1, 2002)

5.1.5 Discussion

In summary, we identify the most likely motives for each of the above four categories:

- SOSP: traffic steering between neighbor ASes.
- SODP: load balancing among multiple AS paths.
- DOSP: requirements for more fine-grained local routing policies.
- DODP: multihoming of either ISPs or end users.

The above speculations can indeed be disrupted by misconfigurations and many other unexpected operations. [14] reported that origin misconfigurations are not highly unlikely in the Internet. During origin misconfigurations, some BGP routers may inject excessive internal prefixes, typically specific ones, into the global BGP table. Holes raised by this reason can still be classified into any of the above four categories, but obviously our reasoning does not apply. Another unexpected operation could be that some people may mistakenly announce another ISP's block, instead of waiting for the problem to be fixed by the error makers. The ISP announces a more specific prefix which turns out to be a hole.

5.2 Classification results

We now study the distribution of the above four categories for holes based on the real BGP data traces. To this end, we first analyze the BGP table collected at peer 204.42.253.253 on Dec.1, 2002. The results are given in Figure 12. The figure shows that the fourth category (DODP) is the most popular one (44%). This indicates that among all the causes we enumerated in this section,

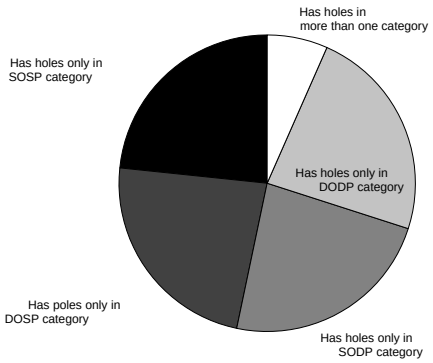

Figure 13. Fraction of covering prefixes having holes falling into the four categories

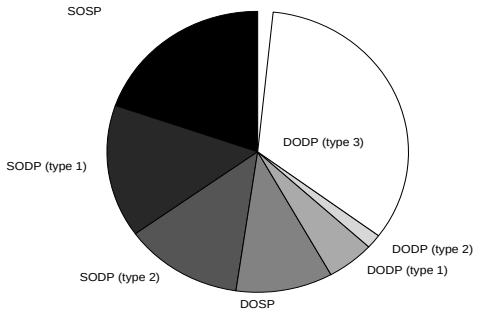

Figure 14. Fraction of holes in the four categories (based on the routing table collected by ATT peer 192.205.31.33 on Dec.1, 2002)

multihoming has the most important impact on the BGP table size. Among the three DODP types, Type-3 generates the largest number of holes. This result is reminiscent of the fact that most of the Internet customers are not service provider and do not perform their own routing policies. The second and third largest categories of holes are DOSP and SOSP, contributing 28% and 16%, respectively. Both reflect the impact of traffic engineering practice.

We also examine the covering prefixes to see which category of holes they are generating most. For this study, we only consider covering prefixes that have holes and plot the percentage in Figure 13. The figure shows that among all the covering prefixes that have holes, 78% of them have holes falling into a single category. We therefore conjecture that in reality most ASes only undertake one practice among traffic engineering, different routing policies and multihoming.

All our measurements presented so far are based on the routing table collected by a single vantage point. To reduce biases in the study, we apply our categorization to the announced holes in the routing table collected by ATT peer 192.205.31.33 at the same time point. Comparing the result (Figure 14) with Figure 12, we can see that the difference is minor.

6 Covering prefixes

This section examines the covering prefixes, which contribute to over a half of the BGP table size. Ideally, covering prefixes should match the allocated address blocks accordingly. However, our measurements show that, more than half of the covering prefixes are fragmented from their original allocated blocks and this ratio is still increasing.

In the following, we first analyze the impact of address allocation on covering prefixes. Our study reveals that fragmentation of the allocated blocks significantly increases the

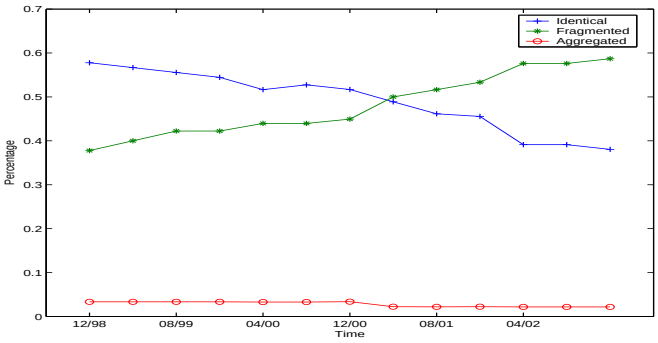

Figure 15. Change of the composition of covering prefixes in terms of their relationships with allocations

number of covering prefixes. Consequently, we carefully examine the fragmentation by classifying the involved covering prefixes into three categories. For each category, we explain the possible underlying motives and also provide some real examples.

6.1 Impact of allocation on covering prefixes

The impact of address allocation on the number of covering prefixes is manifested in two ways. First, a large number of class C prefixes have been allocated before the deployment of CIDR (around 1993-1994). Most of these class C prefixes cannot be aggregated and still get advertised in the global routing table. Second, more than half of the covering prefixes are fragmented from a small number of allocated blocks.

In general, three types of relationships exist between advertised covering prefixes and their original allocated blocks, i.e., identical, fragmented and aggregated.

- *Identical* The covering prefix is identical to an allocated block. For example, for a covering prefix 3.0.0.0/8 in the routing table, we find an identical block 3.0.0.0/8 allocated on Feb.23, 1998.

- *Fragmented* The covering prefix can be summarized by an allocated block. For example, for a covering prefix 9.2.0.0/16, we find that a larger block 9.0.0.0/8 was allocated on Dec.16, 1988.

- *Aggregated* The covering prefix is found to contain the address space of several allocated blocks. For example, a covering prefix 24.48.0.0/13 contains two blocks, 24.48.0.0/14 and 24.52.0.0/14, which were allocated on Jun.14, 1996 and Apr.16, 2001 respectively.

We now plot the fraction of covering prefixes that fall into each of the above three categories in Figure 15. It shows that the covering prefixes that are fragmented from address allocations become more and more popular, i.e., rising from 40% to 60% during the time period (Dec.1998-Dec.2002). For brevity, we call this type of covering prefixes as *fragments*. The rest of this section categorizes these fragments and analyzes the possible motives behind them through both reasoning and case study.

6.2 A classification of fragmentation and underlying motive speculation

We now categorize how an allocated address block is chopped into several fragments. The rationale for this categorization results from the inherent relationships between fragments and the corresponding allocation. Similar to our previous categorization of holes, a handful of metrics including origin AS, AS path are used in the classification. We still employ Gao's algorithm [23] to infer AS relationships. More importantly, we take advantage of the difference between "allocated block" and "assigned block" to infer whether the block owner is an ISP or an non-transit customer. Such information helps to determine causes behind the fragmentation.

- *Different origin ASes and different AS paths* Fragments chopped from the same allocated block are advertised by different ASes. In the example of Figure 16, 10.1.0.0/16 is the allocated block, two smaller blocks, 10.1.0.0/17 and 10.1.128.0/17, are advertised separately by AS1 and AS2.

- *Same origin AS and different AS paths* Fragments chopped from the same allocated block are advertised by the same AS. However, these advertisements are propagated along different AS paths. In the example of Figure 17, 10.1.0.0/16 is the allocated block, and two covering prefixes, say 10.1.0.0/17 and 10.1.128.0/17, chopped from 10.1.0.0/16, are both announced by AS1. However, the two announcements are through different AS paths via AS2 and AS3, respectively.

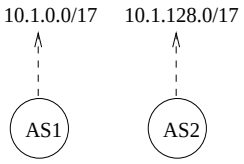

Figure 16. Covering prefixes (fragments of a single allocated block) have different origin ASes and AS paths

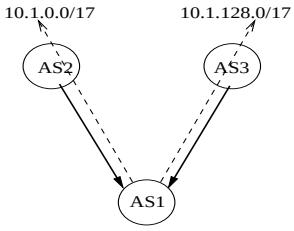

Figure 17. Covering prefixes (fragments of a single allocated block) have the same origin AS and different AS paths

- *Same origin AS and same AS path* This is better exemplified by Figure 18 where 10.1.0.0/16 is the allocated block while 10.1.0.0/17 and 10.1.128.0/17 are announced.

6.2.1 Different origin ASes

In this category we differentiate two scenarios based on whether the address block is recorded to be *assigned* or *allocated* in the allocation record [3]. These two types of address delegation have quite different purposes. *Assigned*

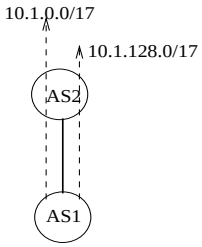

Figure 18. Covering prefixes (fragments of a single allocated block) have the same origin AS and AS path

blocks are typically for non-transit customers (stub ISP or end users) for their usage (not allowing further distribution), while *allocated blocks* are usually given to large ISPs (or regional Internet registries) for subsequent distribution. We discuss two cases separately and speculate that the causes for fragmentation in these two scenarios are quite different.

In the first scenario, we claim that a common practical operation that can lead to the fragmentation is the address block owner's multihoming. A real example is: 9.0.0.0/8 was allocated to IBM Corporation on Dec.16, 1988 for its own usage. Since IBM's own networks are widely located in the world, IBM breaks 9.0.0.0/8 into multiple sub-blocks and assigns to its subnetworks in different places. These subnetworks will connect to ISPs that are geographically close. Consequently, in the routing table on Dec.1, 2002, we observe that one sub-block 9.2.0.0/16 was advertised by UUNet (AS701, USA) while two other sub-blocks 9.184.112.0/20 and 9.186.144.0/20 were advertised by AS3786 which belonged to a Korean ISP.

In the second scenario, address blocks are allocated to ISPs for subsequent distribution. A common practice that can lead to the fragmentation is that the ISP further distributes sub-blocks to its customers which have their own AS number. Similar to the arguments for proxy aggregation, each customer will announce its sub-block separately for requiring finer-granularity routing policies. A real example is as follows. On June 09, 2000, 24.24.0.0/14 was allocated to Road Runner (an ISP). From the routing table on Dec. 1, 2002, we notice that: three sub-blocks, 24.24.0.0/19, 24.24.32.0/19 and 24.24.64.0/19, are advertised by AS11351. Another sub-block 24.24.96.0/19 is advertised by AS11707. Two sub-blocks, 24.24.192.0/20 and 24.24.208.0/20, are advertised by AS1668. According to the AS relationships and the AS registration database maintained by WHOIS [4], we find out that AS1668 belongs to GNN Hosting Service (an ISP) while AS11351 and AS11707 are two customer ASes managed by ServiceCo LLC (Road Runner) and Time Warner Cable (Road Runner) respectively. Although AS1668 is the single provider for both AS11351 and AS11707, it does not aggregate fragments originated from its two customers.

Based on whether the origin AS is the same or not, we classify fragments into different groups, and each group is as a *fragmentation group, (type I)*. In theory, since they're owned by the same organization and fall in the same administration domain, they should have been aggregated. The conjecture deems reasonable by the two separate advertisements of 24.24.0.0/19 and 24.24.32.0/19 in December 2002 being replaced by 24.24.0.0/18 in January 2003. So we claim that the difference between the number of fragment group (I) and the total number of covering prefixes that they have included would be the maximum saving in the routing table size if each fragment group (I) were replaced by a single aggregate prefix.

Seen in Figure 19, the number of fragmentation group (I) represented by line + is always above two times that of the allocated blocks that were allocated to end users and actually grows a little faster than the latter. It indicates that this routing practice may be growing more and more popular.

Similar trend is also observed for the address block that was allocated to an ISP in Figure 20. And coincidentally, the connectivity between the customer AS and the provider ASes in the Internet is growing. To our conjecture, the real users of these fragmented sub-blocks are more and more willing to defy the address aggregation at the ISP. So the benefit from the hierarchical address distribution mechanism has been seriously questioned.

In Figures 19 and 20, the total number of fragmentation group (I) is much smaller than that of the fragments. It indicates that usually one fragmentation group (I) includes multiple covering prefixes, especially in the scenario where the allocation is made for the the purpose of further distribution. According to our above analysis, these fragments chopped from the same allocated block and advertised by the same AS have better chances to be aggregated than those originated from different ASes or those belonging to different allocated address blocks. However, the big gap between the line + and the line * indicates that the "failure of aggregation" happens quite commonly. Generally, it can be further classified into the following two categories: same origin AS with different AS paths, and same origin AS with same AS path, which will be presented in the following two sub-sections.

6.2.2 Same origin AS and different AS paths

The AS who originates several covering prefixes in the same fragmentation group (I) have two or more upper-layer providers/peers and is desiring to optimize the use of all available links between itself and the provider/peer ASes. So a common practice is to force different covering prefixes being propagated through different upper layer providers and consequently to steer the in-bound traffic.

To quantify the popularity of this practice, we classify the covering prefixes in the same fragmentation group (I) that share the same origin AS into different groups in terms of its AS path. Each group is termed as *fragmentation groups, (type II)*. Seen in Figure 19 and Figure 20, the line of fragmentation group (II) is always above the line of fragmentation group (I), which exactly shows the existing of this routing practice. However, the negligible difference indicates that to use multiple upward AS links to distribute inbound traffic cannot account for the advertisement of most fragments.

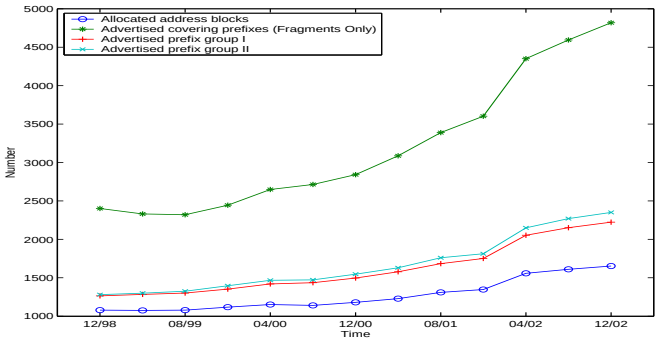

Figure 19. Relationships between covering prefixes (only for fragments of allocations) and allocations made to end users for their own use

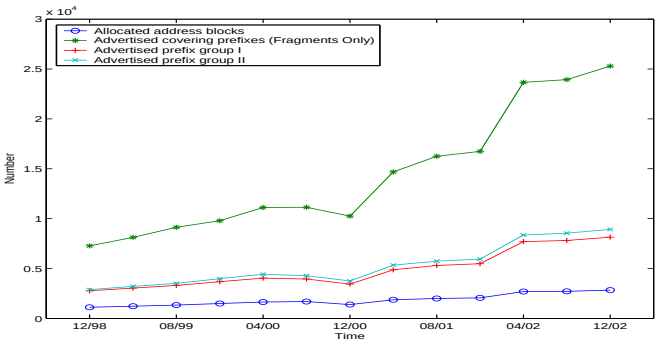

Figure 20. Relationships between covering prefixes (only for fragment of allocations) and allocations made for the purpose of subsequent distribution, usually to ISPs

6.2.3 Same origin AS and same AS path

It is well known that the AS link has only logical meaning. Physically, an AS link may include many connections between the routers at the two ends of the AS link, likely at different locations. For example, with the ISP map at router level built by Rocketfuel [28], we find the number of links at router level between AS3356, (Level 3), and AS1668 is 72 at many different places around the whole country. So it is sensible for a customer AS to make better use of all these physical links to the provider AS. The advertisement of fragments well indicates that some practice has been done to steer the in-bound traffic. A real case presented in the following paragraph serve as a detailed explanation.

Seen in the routing table in December 2002 at vantage point 204.42.253.254 (IAGnet), two fragments, say 24.25.32.0/19 and 24.27.128.0/19 that are chopped from the allocated address block 24.24.0.0/14, are advertised by AS1668 and through the provider AS3356. To our conjecture, AS1668 may advertise the specific prefixes only to the BGP peers at AS3356 through their respective closest egress routers. In this way, AS1668 can ensure that traffic destined him flows into through the routers closest to the respective networks represented by the two prefixes. So this trick can circumvent the ”hot potato routing” (see the description in section 5.1.1) played by AS3356. By using some geographic mapping techniques [27], we’re confirmed that 24.25.32.0/19 and 24.27.128.0/19 are located geographically apart. And further, with the tool of traceroute [30], we observe the traffic destined to 24.25.32.0/19 and 24.27.128.0/19 flows into AS3356 (Level 3) from the same ingress router but flows out to AS1668 via different egress routers.

On the other hand, if in-bound traffic toward different fragmented covering prefixes cannot be distributed over multiple links, ASes, with the consideration of scalability, should aggregate them together. For example two fragments chopped from the previous allocated address block, say 24.24.192.0/20, 24.24.208.0/20, originated from AS1668 and through the provider AS3356 in December 2002. They are actually located quite close geographically and may have been connected to AS3356 through the same physical link. The latest routing table in January, 2003 shows that both of them have been withdrawn and replaced by 24.24.192.0/19.

7 An empirical model for the BGP table size growth

Sections 5 and 6 classify both the holes and the fragmentation of the allocation. We now seek to build a simple model to empirically approximate the BGP table size

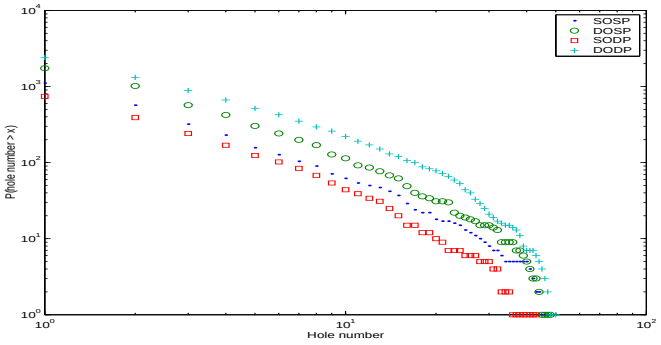

Figure 21. Histogram of holes contained by covering prefixes (log vs. log)

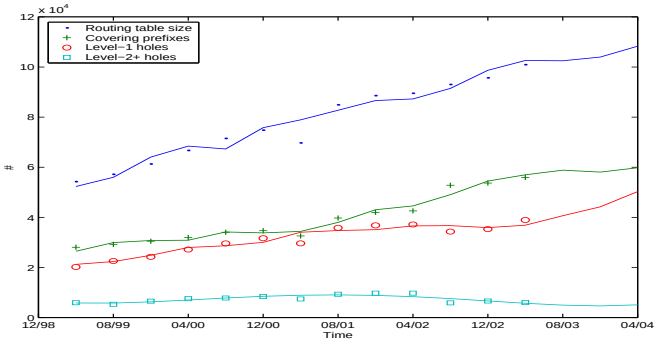

Figure 22. Prediction of the BGP table size

growth. This helps us to predict the routing table size in the near future and answer questions such as whether the increase is linear or exponential.

We first review the evolution of route prefixes using previous analysis and measurement. Based on the data we have used, about 55% to 65% of new allocated blocks are identically announced in the BGP table, less than 10% of them are aggregated. The rest of the allocated blocks are fragmented into four types of covering prefixes. All these new covering prefixes, together with a large number of existing prefixes without holes, constitute the covering prefixes. These covering prefixes actually contain all the theoretically reachable addresses in the global routing table. However, besides reachability, the BGP routing table has been utilized to achieve various goals such as load balancing, connection backup and different local routing policies. All these goals stimulate the creation of large number of holes.

Based on the above description, the model incorporates the following modules.

- *Modeling address allocation* As can be seen from Figures 19 and 20, the rate of newly address allocation remains constant over time and it can be well approximated by polynomial fitting. We perform the polynomial fitting for every allocation size ranging from 8 to 24 and we also treat "assigned block" and "allocated block" separately.

- *Modeling evolution of allocated blocks* The fraction of newly allocated blocks being identically announced (or aggregated) varies over time. Its trend (see Figure 17) is approximated by polynomial fitting. The rest of the newly allocated blocks are fragmented into covering prefixes in the way we have described in Section 6. We model this fragmentation process by polynomially approximating the curves in Figures 19 and 20. Note that we still differentiate prefix length. We also distinguish "assigned block" and "allocated block".

- *Modeling the formation of holes* The previous module gives the estimated number of total covering prefixes. To estimate the number of holes that should coexist with these covering prefixes, we first use polynomial fitting to approximate the percentage of covering prefixes that have holes. Since the number of depth-2+ holes is comparatively small, we approximate its evolution by polynomial fitting. The rest are depth-1 holes, and can be classified into four categories as introduced in Section 5. We first plot the histogram of the number of such depth-1 holes contained by a covering prefix in Figure 21. The figure is based on the routing table on Dec.1, 2002 and is plotted in log-log scale. From the figure, the number of holes (in any category) contained by the covering prefix exhibits strong indication of heavy-tail distribution. We therefore use heavy tail-distribution to approximate the number of holes given the number of covering prefixes.

In essence, by combining our previous analysis about holes, allocation and covering prefixes, our model is capable of approximating the BGP table size in a more realistic manner. We use the past four years' BGP data collected by peer 204.42.253.253 as input and use the model to approximate and predict the BGP table. The results are plotted in Figure 22. It shows that the BGP table size estimated by the model fits the input well, and we also observe that in the near future, the trend of the BGP table size increase is more like linearly than exponentially. This conclusion tallies with [25].

8 Implication

Our results have several implications.

First, our categorization and inferring the underlying motives for route prefixes help to understand why the BGP

table size has become so large and what kinds of ongoing network operations are more likely to drive the increase. Therefore, the results can be used to motivate and evaluate approaches that are aimed to tackle the scalability problem of BGP table. We enumerate some of them in the following.

- Our results indicate a requirement for modifying the current allocation policy. Since the fragmentation of allocated blocks occurs so frequently and generates many additional route entries, address allocation should be conducted in a more strict way and address renumbering can be enforced.
- Some researchers have proposed to decouple the traffic engineering functionality from the BGP table. Such a proposal should be evaluated for covering prefixes and holes differently. It also needs to treat traffic engineering between adjacent ASes and traffic engineering among multiple AS paths differently.
- An ongoing task in the Internet community is to achieve multihoming in IPv6 [29]. Since now we have seen the popularity of multihoming in IPv4 and its impact on the routing table size, a different solution to achieve it in IPV6 is preferred.

In addition, the empirical model in the paper can be readily employed to predict the future routing table size. It can also be used to gauge the impact of different causes.

9 Related work

Huston is among the first to study BGP routing table growth. In [7][8], he measured the BGP table size from multiple aspects and enumerated several operations that could contribute to the table increase. Our work differs from these studies by making a more detailed classification of the BGP table growing components and their physical causes. We also provide an empirical model for the BGP table size growth. In another related work [21], the authors evaluate the redundancy of the BGP table by introducing the notion of *policy atoms* . Their focus was to devise various techniques that could be applied to reduce the BGP table size without losing much connectivity information. In [25], Alaettinoglu analyzed the BGP table growth and also identified several causes such as multihoming, engineered prefixes and punching holes. However, his work did not differentiate covering prefixes and holes, and his classification was merely based on the origin AS. In our approach, we distinguish covering prefixes and holes, and also take into account other relevant information. Bu, Gao and Towsley's work [22] is perhaps the closest in spirit to our work. In [22], they ascribed the BGP table growth to four factors of multihoming, failure to aggregate, load balancing and fragmentation. In addition, to predict the BGP table size growth, they proposed a power-law model which involves the number of ASes and the number of prefix clusters originated by each AS. We share some motivations with [22]. However, we did not seek to ascribe the BGP size growth to a handful of factors, since in many cases the underlying factors cannot be discerned by merely examining the routing data. In addition, the empirical model of our work considers more factors including address allocation, and evolution of covering prefix and holes. We noticed recently that Savola proposed a scheme to categorize route prefixes [15] based on their advertisements. Compared with our analysis, Savola's study is mainly about multihoming and is based on a limited number of routes.

10 Conclusion

The global Internet routing table continues to grow over time. To help estimate the future growth trend, in this paper we analyzed the changes of the global routing table over the last four years to identify the major factors that have contributed to the growth. We first categorized the routing table entries into two broad classes, covering prefixes and holes, each of which contributing to about half of the total routing table size. We then characterized the changes and growth of both prefix types. We observed that, over the last four years, the number of covering prefixes has been increasing much faster than the number of new address allocations, because about 40%-60% of the covering prefixes come from the fragmentation of previously allocated address blocks. During the same time period, although the total number of holes has increased at the same rate as the covering prefixes, the composition of the holes has been changing more rapidly, while about 20% of the holes are newly added into the routing table, a slightly less percentage (15% - 20%) of existing holes, are being removed.

To further identify the motivations behind the observed increase of both covering and holes, we classified both types of prefixes by their routing advertisement modes. With the map of commercial relationships between ASes, we inferred the proportion of contributions by different classes of routing practices today, including traffic steering, multihoming and different local routing policies. To further verify our inference, we conducted several case studies to make sure that real data traffic has followed the routes that the inferred routing practice could have expected. Lastly, based on the measurements, we gave an empirical model of the BGP table growth.

References

[1] Y.Rekhter and T.Li, A Border Gateway Protocol 4 (BGP-4), Internet RFC 1771.

[2] E.Chen an J.Stewart, A framework for inter-domain route aggregation. Internet RFC 2519.

[3] IPv4 Address allocation records, ftp://ftp.arin.net/pub/stats/{arin, ripencc, apnic, lacnic}

[4] WHOIS database, http://whois.{arin, apnic, ripe}.net

[5] Route views project, http://www.antc.uoregon.edu/route-views

[6] D.G.Andersen, N.Feamster, S.Bauer and H.Balakrishnan, Topology inference from BGP routing dynamics, *Internet Measurement Workshop 2002*, 2002.

[7] G.Huston, Analyzing the Internet's BGP routing table, *The Internet Protocol Journal*, 4(1), March 2001.

[8] G.Huston, BGP routing table statistics, http://www.telstra.net/ops/bgp/

[9] G.Huston, BGP Update *Presented on the routing area meeting at 54th IETF*, July 2002.

[10] G. Huston, Commentary on inter-domain routing in the Internet, Internet RFC 3221.

[11] O.Maennel and A.Feldmann, Realistic BGP traffic for test labs, *In Proceedings of SIGCOMM'02*, Pittsburgh, PA, 2002.

[12] H.Chang, R.Govindan, S.Jamin, S.J.Shenker and W.Willinger, On inferring AS-level connectivity from BGP routing tables, *proceedings of INFOCOM'02*, INFOCOM 2002.

[13] S.Bellovin, R.Bush, T.G.Griffin and J.Rexford, Slowing routing table growth by filtering based on address allocation policies, *http://citeseer.nj.nec.com/493410.html*

[14] R.Mahajan, D.Wetherall and T.Anderson, Understanding BGP misconfiguration, *proceedings of SIGCOMM'02*, SIGCOMM 2002.

[15] P.Savola, Categorizing prefix advertisements. *http://staff.csc.fi/psavola/mhome/node46.html*.

[16] E. Gerich, Guidelines for management of IP address space, Internet RFC 1466.

[17] Y.Rekhter and T.Li, Architecture for IP Address Allocation with CIDR, Internet RFC 1518.

[18] V.Fuller, T.Li, J.Yu and K.Varadhan, Classless inter-domain routing (CIDR): an address assignment and aggregation strategy, Internet RFC 1519.

[19] Internet domain survey, http://www.isc.org/ds/

[20] P.S. Ford, Y. Rekhter and H.W. Braun, Improving the Routing and Addressing of IP, *IEEE Network*, May 1993.

[21] A.Broido and k.claffy. Analysis of RouteViews BGP data: policy atoms, *Proceedings of network-related data management (NRDM) workshop Santa Barbara*, CAIDA, May 2001.

[22] T.Bu, L.Gao and D.Towsley, On routing table growth, *Proceedings of Globe Internet 2002* (http://www-net.cs.umass.edu/ tbu/).

[23] L.Gao, On inferring autonomous system relationships in the Internet, *IEEE Global Internet*, Nov 2000.

[24] L.Subramanian, S.Agarwal, J.Rexford and R.Katz, Characterizing the Internet hierarchy from multiple vantage points, *INFOCOM 2002*.

[25] C.Alaettinoglu, RIPE/RIS project BGP analysis: CIDR at work, *NANOG Oct. 2001, IETF August 2001*.

[26] Phillip Smith, http://www.apnic.net/stats/bgp

[27] V.N.Padmanabhan and L.Subramanian, An Investigation of Geographic Mapping Techniques for Internet Hosts, *In Proceedings of SIGCOMM'01*, San Diego, CA, 2001.

[28] N.Spring, R.Mahajan and D.Wetherall, Measuring ISP Topologies with Rocketfuel, *In Proceedings of SIGCOMM'02*, Pittsburgh, PA, 2002.

[29] K.Lindqvist, Multihoming in IPv6 by multiple announcements of longer prefixes, *Internet draft*, Dec.2002.

[30] T.Kernen, http://www.traceroute.org