routing, Internet, BGP

Internet Expansion, Refinement and Churn

ANDRE BROIDO, EVI NEMETH, KC CLAFFY
CAIDA, San Diego Supercomputer Center,
University of California, San Diego.
E-mail: {broido, kc, evi}@caida.org
1

Abstract.
We analyze the evolution of the global Internet interdomain routing system on AS, prefix and IP address level granularities, using snapshots of RouteViews BGP tables from 1997 to 2001. We introduce the notion of *semiglobally routed prefixes*, those present in the majority of backbone tables, and classify them into *standalone* – those which have no subsets, no supersets; *root* – have subsets, but no supersets; and *subset, or more specific*, which are subsets of other blocks.

Using these distinctions we find that from 1999 to 2001 many measures of routing system complexity demonstrated stability in the form of slow growth, dynamic equilibrium, and occasional contraction.

We find that many net change measures reflect contributions of opposite sign, and that true measure of variation, or churn, is the sum of their absolute magnitudes rather than the difference. Appearance and disappearance of prefixes, ASes and RouteViews peers, as well as status changes (an AS changing from transit to non-transit, or a prefix shifting from a standalone prefix to a root prefix) are instances of routing system *churn*. One advantage of using our notion of semiglobal prefixes is that they exhibit less churn than global prefixes (those prefixes common to all backbone tables) and as such allow for derivation of more robust macroscopic statistics about the routing system.

We study route prefix instability at a medium time granularity for late 2001 using 2-hour snapshots of BGP tables, and find that half of all prefix reannouncements (*flips*) are contributed by 1% of all ASes, with government networks, telecoms in developing countries and major backbone ISPs at the top of the list of instability contributors. Small ASes (those who originate only a few prefixes into the global routing system) do not contribute more than their fair share of either route entries or churn to the global routing system. We conclude that during 1999-2001 many Internet metrics were stable, and that the routing system's growth and instability are mostly caused by large and medium-sized ISPs.

1 INTRODUCTION

The aim of this paper is to characterize changes in Internet routing characteristics over the last two years – 1999-2001. We classify quantitative measures of the Internet's growth and complexity into extensive (volume and size) and intensive (relative and structural) metrics. Our observations confirm that many intensive quantities were invariant during this time period and that many extensive quantities were semi-invariant; that is, they scaled polynomially with the Internet's growth.

1.1 AUTONOMOUS SYSTEMS AND BGP

Global routing in the Internet is negotiated among independently operated sets of networks known as Autonomous Systems (AS). An AS is an entity that:

- connects one or more networks to the Internet;
- applies its own policies to the exchange of traffic;
- has a globally unique identifier (AS number)

AS policy is used to control routing of traffic between networks via specific connections. These policies are articulated in router configuration languages and implemented by the Border Gateway Protocol (BGP) [1, 2].

A BGP message announces the reachability of a specific network via a specific router. Reachability information includes an AS path which is a sequence of ASes. BGP assumes that:

- this path is traversed by the BGP message
- the advertised network can be reached via this path

BGP also assumes that all ASes in the path forwarded the message in accordance with that AS's company policies and *ipso facto* agree to accept traffic destined to the advertised network. A BGP table links a network prefix

identifier with the AS path. This table is an important component of the packet forwarding process in a BGP-enabled router.

In addition to packet forwarding, a representative set of core BGP tables can be used to monitor the evolution of the Internet architecture. BGP data reflects consumption of several vital and finite resources [3]: IP addresses, AS numbers, entries in routing table, CPU cycles in routers and bandwidth consumed by routing update traffic. It is important to know whether this consumption is spread evenly over all Internet entities or rather driven by a small subset of them whose behavior may be changed. In this paper we will discuss and answer these and related questions.

We investigate trends in the global routing system at all levels of its conceptual hierarchy. We analyze both extensive (bulk) measures, such as the number of prefixes or ASes, as well as intensive quantities, such as the fraction of root prefixes or transit ASes. We describe data format and availability in Sections 2.1 and 3, and data anomalies in Section 4. Our analysis section begins with a discussion of evolution of bulk measures and continues with a description of AS connectivity, prefix evolution, interaction between prefixes and ASes, IP address space allocation dynamics, and routing flux. A companion paper [4] studies complexity of routing policies using BGP atoms introduced in [5]. We also plan to compare BGP table changes with BGP updates (preliminary results are in [6].)

1.2 TERMINOLOGY

Bulk measures. These measures reflect the Internet routing system's numeric growth. They include counts of prefixes, distinct ASes, AS paths, and AS tokens,

Backbone tables. BGP tables with enough routes to reach the majority of IP addresses in the routing system. As of late 2001, such a table has at least 90K routes.

BGP AS graph. The ASes and AS paths present in a set of backbone BGP tables form a graph that we call the *BGP AS graph*. ASes are nodes of the graph, connected to each other if they are adjacent in some AS path in one of the BGP tables.

Traceroute AS graph. The traceroute AS graph is that derived from converting traceroute IP path to AS path using origin ASes for best-match prefixes for IP addresses. CAIDA's macroscopic IP topology data is collected by skitter [7], a tool based on active probing of forward IP paths using traceroute-like methodology. We will compare traceroute AS graph to the BGP AS graph based entirely on routing tables.

Transit vs. non-transit. A *transit AS* is one that carries someone else's traffic; a non-transit AS is just a source or sink for traffic. BGP routing data gives us no indication of the volume of traffic between any two ASes or even if there is any traffic between them. It shows only an advertised possibility of using an unspecified connection between known ASes to send traffic from a source to a particular destination. A transit AS on the BGP AS graph is an AS with positive outdegree; a non-transit is an AS with outdegree 0. A non-transit AS is always the last AS in any AS path in which it appears.

Origin count. Number of lines in BGP table(s) where a given AS is in the origin (end-of-AS-path) position.

Transit count. Number of lines where a given AS is in a non-origin position (excluding prepended AS copies).

Multiorigin prefix. The vast majority of prefixes in the BGP table are originated by only one AS, but there are a few (1%) that are *multiorigin*, i.e. originated by more than one AS. Multiorigin prefixes pose difficulty for the mapping of IP addresses to ASes used in construction of traceroute AS graph.

Vacuum. We use this term to refer to when a prefix or AS appears or disappears from a table snapshot rather than changing its category, e.g., from transit to non-transit.

Refinement. The number of globally routed prefixes currently grows noticeably more slowly than the number of ASes, resulting in a gradual decrease in the average number of prefixes per AS. We call this phenomenon *AS refinement.* Similarly, the average number of IP addresses per prefix decreases, resulting in prefix size refinement.

Churn. Churn is a process of change in which appearance and disappearance of objects, or transitions between different object types have comparable rates. In that case the total variation is measured by sum of absolute magnitudes, rather than by taking the difference of contributions.

2 BGP DATA

2.1 SOURCES

We use Internet interdomain BGP routing tables from the University of Oregon's RouteViews project [8]. By late 2001, 47 participating peer ASes were contributing 55 tables to RouteViews.

Daily RouteViews tables from November 1997 to March 2001 are available from the National Laboratory for Applied Network Research [9][2]. RouteViews itself has archived BGP tables every two hours since April 2001 [8].

2.2 INTERNALS OF A BGP ROUTING TABLE

A BGP table entry has the following format:

Table 1: BGP Routing Table Entries

Network	Next Hop	AS Path
12.0.0.0	204.29.239.1	6066 3549 7018
12.0.48.0/20	204.29.239.1	6066 3549 209 1742
	213.200.87.254	3257 13646 1742

The first field is the target network prefix. The second field is the IP address of the next hop router – the peer who

[2]Tables between 30 Nov 2000 and 22 Feb 2001 are incomplete.

advertised the route. Finally, the AS path appears; the last AS listed in the path is the origin AS for the prefix. We omit other parameters that have missing, null, or almost constant values.

A BGP table is ordered numerically by network prefix. If a prefix has multiple routes, as in the second line of Table 1, the prefix itself is not repeated.[3] This address block 12.0.48.0/20 (line 2) is originated by AS 1742. Two peers advertise reachability to this address block (prefix) via distinct AS paths. The first AS path (line 2 of Table 1) contains 4 ASes (3 hops) and the other paths contain 3 ASes (2 hops). Note that this block is a subset (or *more specific* prefix) of address block 12.0.0.0/8, but their origin ASes differ and their AS paths diverge after a few hops.

2.3 PREFIX TAXONOMY

Prefixes in the routing table are contiguous intervals of IP addresses, usually represented by an IP network address and CIDR length/netmask. One prefix is *more specific* than another one if it is a subset of the other. For example, the prefix 172.16.243.0/24 is a more specific of the prefix 172.16.0.0/16 but is unrelated to the prefix 172.30.1.0/24. Using these relationships we categorize prefixes in the BGP routing tables as:

standalone – has no subsets or supersets in the table

root – a least specific prefix with subsets in the table

more specific – a subset of some other prefix

A *top* prefix is one that is either a standalone or a root prefix. Note that all these definitions are relative and depend on the particular prefix set.

3 DATA ANALYSIS PREREQUISITES

In this section we discuss the selection of the BGP routing tables and prefixes from which we derive our analyses.

As of November 2001, RouteViews has tables from 55 routers, a five-fold growth since November 1997 when there were only 11 peers. If an analysis spans a long time interval, some participants have likely joined and/or dropped out during the interval; in this case one must either use fewer tables or use tables from different [sets of] peers throughout the interval.

3.1 BGP TABLE SELECTION

BGP routing tables available from RouteViews[4] vary in size from two prefixes to more than 106K prefixes [10]. There are 33 tables with over 100K prefixes, specifically between 103K and 106K prefixes; we call these *full backbone* tables. Several backbone providers filter their routes [11], resulting in tables with 89K or 97K prefixes. Six RouteViews peers contribute filtered tables. We call tables in this and previous set *backbone tables*, inclusively [5]. The distinction between filtered and full backbone tables has existed in RouteViews data since 1998 when prefix counts for the two table types were 50K and 52-53K, respectively.

In this paper, analysis of yearly changes is based on five RouteViews snapshots sampled on 08 November 1997 (10 tables), 01 November 1998 (13), 2000 (18), 2001 (26), and 31 October 1999 (14 tables). The data contains all available backbone tables, except for 2001 when we selected 26 tables common to RouteViews of September to December 2001, to make results compatible with our route flux analysis (Section 10). Table 3 presents an overview of this data.

Table sizes vary noticeably by day, with changes up to a few percent (Figure 1.) There is a daily influx and loss of prefixes, part of what we refer to as *churn*.

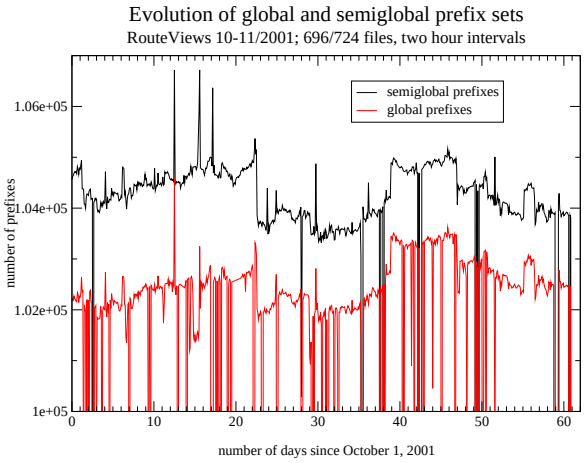

Figure 1: Daily fluctuation in prefixes in RouteViews table

3.2 SEMIGLOBAL PREFIXES

We call a prefix *globally routed* if it is common to a comprehensive set of BGP tables. Our preliminary analysis of globally routed prefixes showed that results depend heavily on the particular BGP tables chosen for analysis. This dependence introduced instability in any metric we analyzed.

We were able to achieve good stability by abandoning the notion of globally routed prefixes and defining and using the concept of *semiglobals* instead. We call a prefix *semiglobal* if it is found in more than half of the RouteViews backbone tables. This choice of prefix set effectively rules out local prefixes seen in only one or two of the tables as well as retaining a prefix even if it is dropped by a few

[3] Networks that align on classful boundaries (/8 for Class A, /16 for Class B, /24 for Class C space) are shown without their prefix mask length.

[4] 01 November 2001 data.

[5] The remaining contributors to RouteViews are legacy peers. One of these currently carries half of the full table and others carry fewer than 10K routes. We omit these smaller tables from analyses.

peers. Semiglobals smooth out variations that appear when using globally routed prefixes on different sets of tables and yield more robust results, especially when trends are small or inconclusive.

Figure 1 shows the size of global and semiglobal prefix sets for 26 full-size backbone peers in 724 snapshots taken once every 2 hours between 1 October and 30 November 2001. The data collection worked effectively most of the time. 711 of the files include more than 100K semiglobal prefixes. We use these 711 tables for detailed analysis of flux in the routing system.

The curve for global prefixes mirrors the semiglobal curve in its upper range. The only significant difference is a number of downspikes in the global prefix curve, caused by a peer being temporarily unavailable or not contributing a full table. This leaves only 632 tables of acceptable quality, 1/8 less than the count of semiglobal-containing tables.

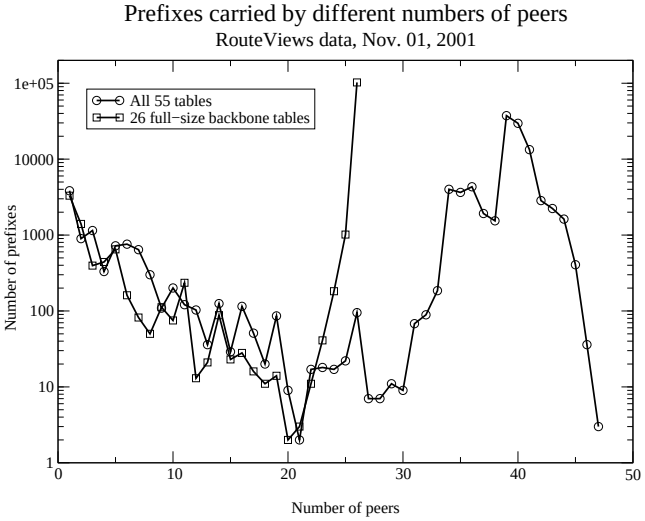

Figure 2: Number of prefixes carried as a function of number of RouteViews peer tables

Figure 2 shows the visibility of prefixes by peers for RouteViews data from 01 November 2001. The figure plots two curves: the 26 full sized tables that we use for analysis; and all 55 tables contributing to RouteViews on 01 November 2001. The large spike at 26 peers (x-axis) and 102K prefixes (y-axis) represents the globally routed prefixes. However this spike only appears when full sized backbone tables are selected. It spreads to a bell-shaped curve when no such restriction is in place, as shown in the longer curve (circle-symbols). The robustness in the use of semiglobal prefixes alleviates the sensitivity to peer selection. Even the bell curve of Figure 2 yields almost the same semiglobal set of prefixes common to 28 or more peers. However, to guarantee the stability of our prefix set and to make inputs from different peers comparable, we will use only prefixes from backbone tables.

Our baseline data set (01 November 2001) consists of 26 full-size backbone tables. It contains 102K global and 104K semiglobal prefixes, from a total of 112K prefixes in 26 tables. There are 113K prefixes in the 55-peer all-prefix set. The number of semiglobals differs from that of globals by one or two percentage points (Figure 1), a small price to pay for the stability realized with the use of semiglobals.

4 ANOMALIES IN BGP DATA

There are many anomalies in BGP data, ranging from counterintuitive observations (e.g., multiorigin prefixes) to rarely used features, to obvious misconfigurations. Although each anomaly is fairly infrequent, we list them all to demonstrate the variety and avoid misinterpretation of results. We prefer to have reasonable estimates for marginal objects rather than ignore their presence, overestimate their prevalence, or treat an anomaly as a matter of consequence within the routing system.

4.1 PREPENDING

BGP's approach to inter-domain routing is qualitative, in that there are no real quantitative metrics that can propagate beyond an AS boundary. However, an AS can indicate its preference for a route by inserting extra ASes (usually its own) into an AS path that it propagates. This prepending will make this AS path appear longer and thus reduce its chance of being selected. In the November 2001 RouteViews data, repeated ASes (i.e., prepending) constitute 5.5% of all AS tokens; they appear in 6.5% of all lines.

4.2 MULTIORIGIN PREFIXES

Multiorigin prefixes, which constitute about 1% of the prefixes in the global table, were defined in Section 1.2. Our analysis suggests that multiorigin prefixes in any single snapshot of BGP table are often actually not multiorigin but prefixes captured in a transition between one origin and another before the moment of convergence [12].

4.3 NON-ORIGIN (TRANSIT-ONLY) ASES

Some ASes do not announce their own networks, presumably intending to reduce visibility of some of their transit infrastructure. (Their owners use other AS numbers for originating customer networks.) The number of non-origin ASes in November 2001 and December 2001 full peer tables is 81. Their transit count, i.e. the number of AS paths with that AS in the middle, in all selected tables, can be anywhere between 1 and 10,000, and many are in the thousands, which means that hundreds of routes use these ASes for transit. Most of them, however (about 60% in 1997-2001; in particular, 45 in 2001) have indegree 1 and outdegree 1, i.e. there is no AS path branching in these ASes. Such a 'transit-only AS' may actually be used to render AS-prepending undetectable by AS comparison.

4.4 ROUTING LOOPS

A routing loop is an AS path that has one or more ASes repeated in non-contiguous positions, with other ASes in between. Apparent loops may arise as a side effect of typographical errors in hand configured prepending, i.e. padding of AS paths with excessive copies of one's own AS, used to deflect traffic away from one's own network. Current administrators typically use autoconfiguration and BGP communities to implement such policy, and as a result routing loops in BGP tables are rare. On any given day, there are several loops observable in the approximately 4.4 million routes collected by RouteViews. For example, the 1 November 2001 collection of all peer tables has 147 lines with routing loops of the form A B C B, where A is a backbone ISP, B is an AS that it owns, and C is a consumer Internet provider.[6]

4.5 TANGLES

Tangles are apparent routing loops when a prefix is reached through two AS paths containing two ASes in opposite order ($A\ B$ in one path and $B\ A$ in another, possibly with other ASes in between.) A small number of tangles (about 10) is present in almost any snapshot of several BGP tables.

4.6 RAMIFICATION

Ramification is observed when two or more AS paths reaching the same destination prefix converge at the same AS and diverge again. The presence of ramification and tangles implies that an AS may consist of sub-parts with distinct routing policies. Most large ASes are ramified for at least some prefix. Backbone providers contributing several tables to the RouteViews project routinely contain thousands of ramified routes.

4.7 AS SETS

When BGP4 was introduced in 1995, router memory was expensive and the pressure to enforce route aggregation high. In response, BGP added the notion of AS sets (unordered AS tuples) in AS paths, to enable loop avoidance for aggregates that merge routes with different AS paths. This radical approach to aggregation never gained much popularity. 01 November 2001 RouteViews data has only 10 different AS sets, present in 328 instances or tokens, out of a total of 15.5 million AS tokens in that day's table.

[6] These 147 loops were present across 104 prefixes seen by 16 RouteViews peers; all involved ASes were in the U.S.

4.8 PRIVATE ASES

Private AS numbers are those between 64512-65534. RouteViews data for 01 November 2001 (all peers) contains 15 private ASes and 01 December had 30 private ASes, two of which were non-origin (transit-only) ASes. Private ASes can leak into the global routing system from confederations [2] and from edges of the Internet, where they are used between providers and customers who want to speak BGP but do not have a registered AS number.[7]

4.9 PRIVATE ADDRESSES

The 01 November 2001 RouteViews data (all peers) includes more than 80 prefixes in [13] private address space blocks. All are in the /28-/32 prefix range and are locally carried (one or two peers), so they are not semiglobals.

4.10 INADVERTENT TRANSIT

Ideally a customer who is multihomed uses each of his upstream providers for transiting his own traffic, but does not become a transit provider for traffic going from one upstream provider to another. Inadvertent transit through customer ASes is due to a common BGP misconfiguration: a customer announcing its upstream routes to another upstream provider. In the BGP AS graph (Section 1.2) this type of misconfiguration appears as a transit network with outdegree 1 and small (near 1) indegree. On 01 November 2001, 26 RouteViews tables have 818 (42% of 1963) transit ASes with outdegree 1, and 258 have both indegree and outdegree 1. Some of these ASes may be providing inadvertent transit; it would require personal communication with those providers' engineers (which we have not done) to determine which are actually doing so [14].

5 DYNAMICS OF BULK MEASURES

The size of BGP routing tables can be quantified by several metrics. We examine prefixes, prefix length, prefix type, ASes and their type, and related values. Table 2 shows values for November and December 2001 using a common set of 26 peers with full-size tables.

The data in the Table 2 shows that small ASes (those with one prefix) contributed few prefixes to the global routing table, despite the large proportion of ASes that they represent (40% of ASes contribute 5% of prefixes.) Large ASes (those with 100 or more prefixes) constituted 1% of all ASes, yet contributed 1/3 of semiglobal prefixes.

Many bulk measures grew sharply between 1997 and

[7] Leakage of private ASes is not limited to routing tables. They are present in APNIC route registry policy database and in the *aggregator* fields of BGP updates coming from that region.

Table 3: Trends in bulk measures of BGP table size (rightmost two columns are annual growth ratios for last two years)

Measure	1997	1998	1999	2000	2001	00:99	01:00
Addresses	904M	1010M	1068M	1083M	1134M	1.4	4.7
Addresses, more spec.	30.2M	55.0M	65.4M	93.9M	121.4M	43.5	29.2
Semiglobal prefixes	45920	52807	64769	88714	103551	37.0	16.7
Semiglobal /24s	27518	30443	37071	51508	59302	38.9	15.1
More specifics	18848	23647	32077	49200	53686	53.4	9.1
ASes	3060	4318	6107	9116	12155	49.3	33.3
AS links	5302	7874	12037	18196	25179	51.2	38.4
AS links per node	1.73	1.82	1.97	2.00	2.07	1.27	3.78
RouteViews b/b tables	10	13	14	18	39	28.6	216.7

Table 2: Bulk Measures of BGP Routing Tables, Nov.-Dec. 2001

Metric	Nov. 1	Dec. 1
Global prefixes, 26 peers	102000	102394
Semiglobal prefixes, >13 peers	103551	103828
Multiorigin semiglobal prefixes	1078	1121
Smallest blocks, /24s	57.3%	56.9%
More specific prefixes	51.8%	51.1%
ASes	12155	12399
Transit ASes	1963	2001
Links in BGP AS graph	25179	25630
Max prefixes originated by AS	2218	2106
ASes originating one prefix	39.4%	40%
Prefixes from such ASes	4.7%	4.9%
ASes with over 100 prefixes	1.1%	1.0%
Prefixes from such ASes	33.2%	32.3%

2000 and then slowed in 2001. Table 3 shows the IP address space, prefix and AS growth over this period[8].

The number of semiglobal prefixes increased 37% between 1999 and 2000 but only 17% between 2000 and 2001, while the number of ASes grew by 50% and 33%, respectively. The number of AS links grew at faster rate[9]. The average degree (links per node in the AS graph) and IP addresses had commensurable slow growth rates; compared to other other variables, they remained almost invariant.[10] But growth of addresses covered by more specifics was much faster, between prefix and AS growth rates. Prefix growth is close to 2/3 of AS growth, which results in algebraic dependence between prefix and AS counts.

5.1 PREFIX VS. AS GROWTH: AS REFINEMENT

The growth of the Internet depends upon an interplay of economic, social and technological factors. It appears at first that different layers within the Internet's logical structure grow independently. However, we found a simple relationship between the numbers of prefixes and ASes that has held for the last few years with surprising accuracy:

$$P = 200\, A^{2/3} \tag{1}$$

where P is the number of semiglobal prefixes and A is the number of ASes.

Table 4: Prefix vs. AS growth

Date	#Peers	#AS	#Prefix	Approx	Error,%
Dec 97	10	3149	46741	42968	-8.78
May 98	9	4195	50262	52022	3.38
May 99	15	5039	57812	58784	1.65
May 00	17	7483	75699	76515	1.07
Dec 00	26	9420	91350	89207	-2.40
Mar 01	28	10399	95898	95285	-0.64
May 01	27	10938	99223	98550	-0.68
Aug 01	36	11653	101538	102799	1.23
Oct 01	26	12048	104721	105109	0.37
Nov 01	26	12155	103551	105730	2.07
Dec 01	26	12399	103828	107140	3.20
Jan 02	26	12469	104451	107544	2.96

Table 4 compares predicted and actual count of semiglobal prefixes for several dates[11] between December 1997 and December 2001. Despite oscillations in the error that reflect the dynamic nature of Internet routing, the accuracy of Equation 1 is within a few percent of the data; occasionally the relative error is less than 1%. Figure 3 overlays this approximation on the plot of AS and prefix growth, demonstrating that Equation 1 indeed provides an exceptionally good fit for semiglobal prefixes despite the variability in growth rates.[12]

Equation 1 also implies that the average number of prefix announcements per AS is shrinking as $200/A^{1/3}$. We call this process *AS refinement* since the number of prefixes per AS is decreasing with time. As of December 2001 there are an average of 8.4 prefixes per AS. Extrapolation shows that the number of ASes could possibly reach the number

[8] Recall that all available RouteViews backbone tables are used in 1997-2000; 2001 data is based on 26 tables out of 39.

[9] Only a fraction of this growth can be attributed to the RouteViews peer set increase. For comparison, 39 tables for 01 November 2001 contain 25510 AS links, which is 1.3% more than 26-table data for this date.

[10] Uneven growth of IP space totals may be partly caused by flux in /8s, see Table 24 and [3].

[11] Sampled at the beginning of the month indicated

[12] Equation 1 can be viewed as a relation between volume A and surface P of a body that expands by coordinate stretch (such as an inflating universe).

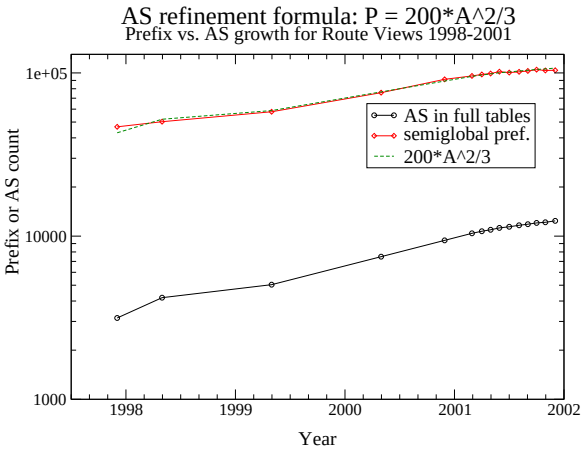

Figure 3: Algebraic dependence between prefix and AS counts

of prefixes when there are 8 million of each. Such evolution would imply many more non-origin ASes than we have today.

Between October and November 2001, the prefix table shrank while the AS count grew (Section 10), resulting in a slight increase in approximation error for equation 1, from 1% in August to 3% in January 2002, in a rare manifestation of a *super-refinement trend*. It will be interesting to see if the prefix count grows faster in the near future, to compensate for the difference between actual and approximated values.

6 Evolution of AS connectivity

6.1 Structure of AS graph

BGP data provides us with a backbone-centric image of the AS graph in which lateral peripheral connectivity is rarely captured. We call the *core* of a graph [15] the set of nodes that can reach cycles of size 3 or more. In the case of Internet graphs most core nodes are bidirectionally connected to each other. The BGP AS graph obtained from one RouteViews snapshot has only 2-3% of its nodes in the core, while the corresponding forward traceroute-derived AS core [15, 7] is 28% of the ASes in the graph. Even if we accumulate BGP data over two weeks and derive an AS graph from that[13], there are still only 4.6% nodes in its core. This striking difference in connectivity captured via BGP versus active probing should give pause to anyone considering using BGP tables to model Internet connectivity with any reasonable veracity [15].

6.2 Implied connections

Whenever a subset (*more specific*) of a prefix has a different origin AS than the origin AS of the covering (or *less specific*) prefix itself, it is assumed that the owner of the covering (less specific, or *superset*) prefix knows how to reach addresses in the subset (more specific) prefix. In particular, if a subset is withdrawn, the superset (less specific) becomes the best match for these addresses and traffic to them will follow the superset route. This *implied connection* to the more specific can thus be added as an arc to the global BGP AS graph (an arc is a path in the BGP AS graph with unspecified intermediate nodes. Most of these implied connection arcs are likely to contain only one AS link.)

Such an *implied connection* is treated as existing in the context of *prefix length filtering*, which is a targeted deletion of subset prefixes from a global routing table. Such an operation is often deemed prudent by experienced Internet engineers to minimize the size and expected size-related churn of the global table[14]. The rationale is that an AS that does not maintain a (provider-customer or other) relation with the owner of a subblock can always deaggregate the superset (less specific) prefix and shed resposibility for forwarding traffic to the subset.

Adding all arcs that connect origin ASes of prefixes to origins of their subsets (more specifics) results in significant extension of the BGP AS graph. For example, for the 26-table 01 November 2001 RouteViews data, the link set augmented by these arcs increases from 25179 links to 30098 links and arcs (a 5000-arc increase). The core of the graph expands by a factor of 3.5, from 285 (2.34%) nodes to 986 (8.11%) nodes. Although still thinner than the graph derived from traceroute AS connectivity, this graph has the advantage of being obtained from essentially the same data, namely BGP tables. [15].

6.3 AS path length trends

AS path length is the primary metric used by BGP in route selection. BGP deems short paths superior, consistent with the assumption that intra-AS backbone networks tend to be overprovisioned (underloaded), while public traffic exchanges are more likely to suffer from congestion. Under these assumptions, choosing shorter AS paths minimizes the likelihood of passing through a traffic exchange with high packet loss. Analysis shows that shorter AS paths have some correlation with lower path RTT [16]. On the other hand, the trend toward shorter paths can also cause BGP to have fewer choices and therefore less selection, limiting potential for better-than-random choice.

We examined AS path length changes for a 3 year period (1999-2001) using 5 full sized backbone peers.[16] The

[13] 162 snapshots taken in 2-hour intervals, 18 November to 01 December 2001, 12570 nodes and 28061 links.

[14] The reduction in table size is less than 15%. In Section 10 we show that the reduction in churn is also small.

[15] This data set is more homogeneous compared to a mixture of sources with uneven spaciotemporal coverage, reliability and operational significance, e.g. combining BGP tables with routing policy databases. However, we will use only BGP AS links (no arcs) to classify ASes as transit or multihomed.

[16] The intersection of RouteViews peers' tables over this 3 year period had only 5 backbone tables in common and therefore limited our choice.

selection of peers includes American and European backbone providers as well as one tier 2 ISP. Figure 4 shows that

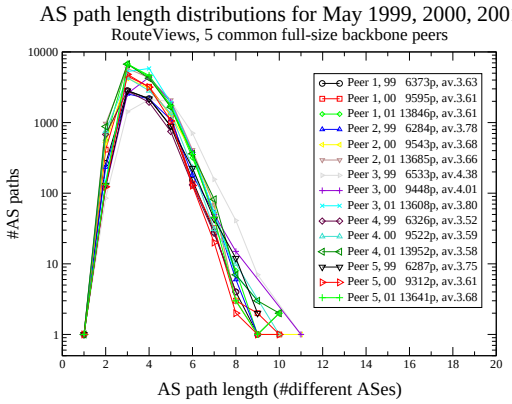

Figure 4: AS path length distributions (RouteViews 1999-2001)

the AS path length, both the average and the overall distribution, remained stable for all 5 peers during this period despite the growth in the number of AS paths.

The distribution in Figure 4 is of unique AS paths and does not take into account how many prefixes were propagated along each path. The data shown is for May of each year; we subsequently analyzed data for November 2001 and observed that the average path length had increased for 2 peers, decreased for another 2 peers, and remained the same for one peer. Again, contrary to current conjectures that AS path lengths are shrinking, we found no significant overall shift in AS path length in any available backbone BGP tables between 1999 and 2001.

6.4 Changes in peering richness

Peering richness is a measure that reflects how many routing choices are potentially operationally available from a given (AS) node in the system. We include in this term relations between ISPs and customers as well as relations among ISPs.

The simplest measure of peering richness is the outdegree of a node in the AS graph. Outdegree indicates how many ASes a given autonomous system accepted routes from via BGP. However this measure does not take into account the diversity of choice that may be highly biased toward a specific AS or limited set of ASes. To account for this bias we introduce a weighted measure of peering richness called *entropy* [17].

$$H = -\sum_k p_k \log_2(p_k)$$

where p_k is the ratio of AS path tokens (i.e. lines in the table) using link k as an exit from this AS, to the AS's non-end-path (transit) count. Entropy measures the uniformity of the spread of AS paths across available links. It has maximum value when every link is used by an equal number of paths and equals zero when only one link is used. Entropy will be close to zero when one link dominates. Numerical values of entropy are computed in bits and can be directly compared with the logarithm of AS node's outdegree. For example, an entropy of 7 corresponds to paths uniformly spread over 128 links.

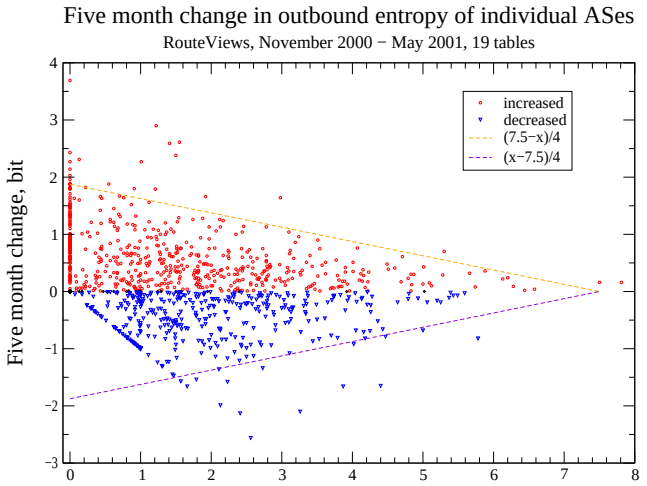

Figure 5:
Changes in AS peering richness, 28 November 2000 – 03 May 2001

Figure 5 shows a scatter plot of the entropy of outbound link distribution and its change in five months from 28 November 2000 to May 2001. The average richness of outbound connections did not change over this five-month period despite the fact that many individual ASes significantly changed their richness. The area in which most changes in peering richness occur is bounded by straight lines forming a triangle. ASes rich in downstream connections (high entropy) are likely to be more stable than those with fewer connections and less diversity.

We have also analyzed navigation complexity measured as an average sum of AS entropies along the path ('the number of directions to an average destination') and found that it did not significantly change in the same period (13.33 bits to 13.48 bits.)

6.5 AS churn

Tables 5-7 present statistics of AS churn for 1999-2001.[17] They show that growth in transit ASes mostly originates from existing non-transit ASes (380 in 1999-2000 and 508 in 2000-2001, Table 5), rather then from totally new ASes. The number of totally new transit ASes (252, 283 ASes for 2000 and 2001 growth, respectively) is close to the number of transit ASes becoming non-transit (188 and 243).

[17] Section 7 describes these tables; recall that appearance and disappearance of ASes within each class are counted as a transition from or to the category *vacuum*.

Table 5: AS churn in 1999/2000

	Tran.	Non-tr.	Vacuum	Sum
Transit	908	188	57	1153
Non-tr.	380	4136	438	4954
Vac	252	3252	0	3504
In	54.81	69.44	14.13	6107r
Sum	1540	7576	495	9116c

Table 6: AS churn in 2000/2001: transit, non-transit, and vacuum

	Tran.	Non-tr.	Vacuum	Sum
Transit	1172	243	125	1540
Non-tr.	508	6056	1012	7576
Vacuum	283	3893	0	4176
In	51.36	54.59	27.23	9116r
Sum	1963	10192	1137	12155c

Both transit and non-transit ASes had smaller net change rate in 2000/2001 than in 1999/2000. The growth of non-transit ASes slowed more than the growth of transit ones. Loss rates in both categories have increased by 3-4%. (Table 7.) The total AS growth rate dropped from 49% to 33% between those two years. Transitions between AS categories imply that comparison between prefixes originated by transit and non-transit ASes involves contributions from groups whose membership changes through both influx and loss, so that contributing ASes are different in each time sample.

7 LONG-TERM PREFIX EVOLUTION

In this section and the following we study prefix counts in association with various attributes of prefixes and originating ASes, to identify portions of the Internet that seem to contribute a disproportionately high number of prefixes or instability to the global routing system.

Objects that are small or peripheral are often considered the major cause of the rapid growth and instability in global routing tables. These objects include: long prefixes (/21-/24); subset (more specific) prefixes; ASes originating a few prefixes; and non-transit networks, particularly multihomed ones.

In contrast to prevailing wisdom, we will show that most growth and churn in the BGP table comes from large ASes that originate dozens or hundreds of prefixes, and that small ASes do not contribute more than their fair share to table growth and instability. In fact by some metrics, small ASes contribute even less than their fair share of growth and churn.

7.1 /24 PREFIXES

A /24 network has at most 256 IP addresses; canonically it is the smallest globally routed address block size. Longer prefixes (smaller networks) do not typically cross AS boundaries. Table 8 shows[18] that the fraction of /24s in the backbone routing tables has been stable since 1997 despite varying periods of Internet growth and stagnation. At present, most full backbone tables contain 57% /24 prefixes; in the filtered tables it is approximately 55%.

Table 7: Summary of AS churn in 1999/2000 and 2000/2001, transit vs non-transit

99/00	Tran.	Non-tr.
Out	21.25	16.51
In	54.81	69.44
Net	33.56	52.93

00/01	Tran.	Non-tr.
Out	23.90	20.06
In	51.36	54.59
Net	27.47	34.53

Table 8: Percentage of semiglobal /24s and more specifics

Year	1997	1998	1999	2000	2001
/24s	59.9	57.6	57.2	58.0	57.3
more specifics	41.0	44.8	49.5	55.5	51.8

7.2 MORE SPECIFICS AND LONG-TERM PREFIX CHURN

Approximately half of all semiglobal prefixes are more specifics (subsets). We examined tables from November 1997-2001 and found (Table 8) that the share of more specifics grew through the end of 1990s and then dropped in 2001. These results are consistent with those in [3]. We suggest that the reason for this drop is the Internet economy slump. As the Internet 'bubble' burst in late 2000 and continued to deflate throughout 2001, many companies failed and dropped off the Internet. Most of them did not have their own address blocks, but rather used subsets from ISPs aggregates. Growth in the share of more specifics is one of many trends that reversed at this point.

Table 9: Semiglobal prefix breakdown, November 2001

Prefix type	Number	Perc	IP addr.	Perc
standalone	44264	42.75%	590.6M	52.1%
root	5601	5.41%	543.5M	47.9%
more specifics	53686	51.84%	121.4M	10.7%
total prefixes	103551	100.00%	1134M	100%

Table 9 shows the prefix breakdown by type in November 2001 for all semiglobal prefixes in the 26 full backbone tables. The fraction of root prefixes grew from 4.64% in 1999 to 5.41% in 2001, suggesting that the number of more specifics per root prefix has been steadily decreasing. The trend for root prefix tree refinement (decrease of average number of subset prefixes per root) is comparable to that of AS refinement. Despite an 8:1 ratio of standalone to root prefix counts, roots cover approximately as many IP addresses as standalones.

We have defined *churn* as the total number of prefixes that either appear or disappear in a given interval. Churn

[18] Data is described in Section 3 and Table 3.

is an absolute measure of variation in the prefix set, and a metric that can capture changes even when the number of prefixes remains nearly constant.

Table 10: Prefix change, May-August 2001

	Stand alone	Root	More specific	Vacuum	Total
Standalone	38676	544	576	2450	42246
Root	377	4374	40	307	5098
More sp.	917	41	42188	8746	51892
Vacuum	4232	420	9206	0	13858
Total	44202	5379	52010	11503	101591

Tables 10 and 11 show transitions between the three types of prefixes: standalones, roots and more specifics in all backbone tables including filtered tables. Table 10 illustrates changes in prefix status between May 2001 and August 2001. Values in the table capture movement of prefixes from one category to another during the three-month interval. Downward diagonal values represent prefixes that did not change categories. Row labels identify prefix categories in May 2001 and column labels mark their categories in August 2001. For example, the entry in row 1, column 2 means that 544 prefixes moved from standalone in May to root in August.

Note that the *Vacuum* category, which measures the appearance or disappearance of prefixes, accounts for most of the change. However a significant fraction of the change arises from transitions between different prefix groups. This crosstalk between groups is essentially noise that prevents us from correctly measuring each prefix type's contribution to the overall churn of BGP tables.

Table 11: Prefix change, August – November 2001

	Stand alone	Root	More specific	Vacuum	Total
Standalone	39530	607	1460	2605	44202
Root	396	4532	146	305	5379
More sp.	604	33	41835	9538	52010
Vacuum	3738	433	10178	0	14349
Total	44268	5605	53619	12448	103492

Table 11 presents the same data for the period August 2001 to November 2001. This data illustrates a reversal of the trend seen from May to August, when prefixes mostly shifted from more specifics to standalones. During this latter interval from August to November, the direction of the flow reversed and more prefixes moved from standalones to more specifics. The condition can be characterized as dynamic equilibrium between more specifics and standalones. We did not analyze this data to determine if these trends were influenced by events involving hundreds of prefixes at once, e.g. deaggregation, or the disappearance of large blocks in the /8 to /12 range.

Table 12 shows the percentage of prefixes moving in and out of the prefix categories and the resulting net change for May to November 2001. All percentages are given with respect to the group size at the beginning of the comparison interval. The percentage of prefixes shown as moving into a group is the ratio of the number of new prefixes to the original group size.

Table 12: Prefix movement in/out of categories (values = %change)

	Stand alone	Root	More specific
May - August 2001			
Out	8.45	14.20	18.70
In	13.08	19.71	18.93
Net	4.63	5.51	0.23
August - November 2001			
Out	10.57	15.75	19.56
In	10.72	19.95	22.66
Net	0.15	4.20	3.09
November, 2000 - 2001			
Out	23.37	32.50	46.72
In	49.12	62.51	55.70
Net	25.75	30.02	8.98

From May to August 2001 the net change in the total number of semiglobal prefixes was 2.37%; from August to November 2001 it was 1.87%.[19]. We see no growth of more specifics in the May-August 2001 data and no growth of standalones in the August-November 2001 data.

Table 12 shows that during the interval of observation the churn in each category was much higher than its numeric growth, particularly so for more specifics. The turnover is almost 20% of prefixes in a three-month interval, while the number of prefixes in the set changed only 3%. In the entire year from November 2000 to Novembers 2001, the prefix loss for top prefixes (32%) was about half of their growth (62.5%). Further, prefix loss for *more specifics* was comparable to their growth, resulting in a much smaller net increase than recentlhy conjectured.

8 PREFIX AND AS EVOLUTION COMBINED

8.1 SMALL ASES

Networks (ASes) that originate only one prefix into the global routing system constitute 40% of all ASes, but contribute only 5% of the prefixes in the routing table. These networks cannot be the cause of the major growth in the routing table, and we show in Section 10.2 that they are also not responsible for significant churn.

[19] Nov 2000 (18 tables), May 2001 (33), Aug 2001 (42), Nov 2001 (39); all available backbone tables, including filtered tables. Percentages for 2000-2001 churn with 26 tables in 2001 differ from the 39-table data by less than 0.14%.

8.2 TRANSIT AND NON-TRANSIT ASES

As of late 2001, transit ASes make up 1/6 of all ASes in the BGP routing tables. This fraction has slowly decreased over the last three years as shown in Table 13 in spite of the fact that the absolute number of transit ASes has increased slowly over time.

Table 13: Transit and non-transit ASes and originated prefixes

Data type	1999	2000	2001
transit AS	18%	17%	16%
non-transit AS	81%	83%	84%
prefixes of non-tr. ASes	42%	43%	46%

Transit and non-transit ASes contribute approximately equal numbers of prefixes even though there are five times as many non-transit ASes as transit ASes. This relative contribution has equalized since two years ago when transit ASes dominated the table in proportion 4:3. It appears that the Internet has grown primarly at the periphery, although as mentioned earlier, we cannot assert this based only upon BGP evidence since we lack the ability to capture comprehensive peripheral connectivity.

8.3 MULTIHOMED NETWORKS

A multihomed network is a network that accepts traffic from more than one upstream provider. In the BGP AS graph, multihomed networks are nodes with indegree ≥ 2. BGP gives only a lower bound on multihoming, although a consistent one across peer selection. The fraction of multihomed networks in the AS graph generated by 26 selected backbone tables differs by only 0.3% from the graph of all 39 backbone tables.

Table 14: Percentage of nontransit and multihomed AS in 1997-2001

Year	1997	1998	1999	2000	2001
nontransit AS	79.1	78.8	81.1	83.1	83.9
multihomed AS	44.0	50.9	57.1	58.4	60.9
nontransit m/h	30.5	37.1	43.7	45.8	48.8

Table 14 shows evolution of multihomed and non-transit networks in 1997-2001. The fraction of non-transit multihomed ASes in every year is about 3% less than what one would predict by multiplying the fractions of non-transit ASes and multihomed ASes. The data suggests that non-transit ASes are less likely to be multihomed. Multihomed transit ASes make up 75% of all transit ASes, whereas multihomed non-transit make up only 58% of all non-transit ASes. For the November 2001 RouteViews data, multihomed (transit and non-transit) networks originated 76% of all prefixes and 74% of the more specific prefixes; therefore, multihoming and more specifics are independent notions.

Table 15 shows that the fractions of non-transit multihomed ASes grew in 1999-2001, but their share of prefixes stabilized in 2000-2001 at 30%. These ASes contribute fewer prefixes than the proportion of total ASes they represent; they are not a primary cause of the large size of backbone BGP tables.

Table 15: Non-transit multihomed AS statistics

Contributions	1999	2000	2001
ASes	43.65%	45.84%	48.75%
Prefixes	27.46%	29.37%	29.74%

8.4 SUBSET (MORE SPECIFIC) PREFIXES IN MULTIHOMED & NON-TRANSIT ASES

Table 16: Semiglobal prefixes originated by AS groups

origin AS type	%ASes	%prefixes	Pf/AS
Transit multihomed	12.11%	46.64%	32.8
Transit single-homed	4.04%	6.67%	14.1
Nontransit multihomed	48.75%	29.74%	5.2
Nontransit single-homed	35.10%	15.90%	3.9

We examine the connection between prefix types and transit versus non-transit networks, using RouteViews data from 01 November 2001 (26 full-size tables.) Table 16 contains the breakdown of semiglobal prefixes by origin AS type[20]. We see that on the average, a transit multihomed AS originates 6 times as many prefixes as a non-transit multihomed AS, and a transit single-homed AS originates 3.5 times as many as a non-transit single-homed AS. The prominence of the same ASes in providing transit for others and in originating prefixes is a manifestation of *cuspidality*, i.e. increase of dependence between two conceptually independent variables towards the extreme end of their ranges. This phenomenon is clearly visible in Figure 6, both in the constellation of Tier 1 ASes (upper right corner) and among mainstream transit ASes (wedge-shaped area stretching from the center to the upper right.)[21]

Table 17 compares the prefix type breakdown for each of four AS groups obtained by combining the attributes: single/multihomed and transit/non-transit. The split of prefixes into top and subset (more specific) is almost the same for transit ASes as non-transit single-homed ASes. The only difference is that transit single-homed ASes originate a larger share of root prefixes than non-transit single-homed ASes, an indication that the majority of transit single-homed ASes are *bona fide* small transit providers (whose indegree may be undercounted by BGP) rather than inadvertent transit entities.

Further, transit multihomed ASes originate fewer subsets than standalones (43% vs. 49% of all their prefixes.) The three remaining classes of ASes originate relatively larger shares of subset blocks (57-60% of their total). Nevertheless, more subset prefixes are originated by transit

[20] Percentages add to 98.95%; remaining 1% are multiorigin prefixes.

[21] RouteViews peer ASes are aligned at transit counts of about 100K. ASes with k prefixes are aligned at origin counts $y = 26k$ for small k.

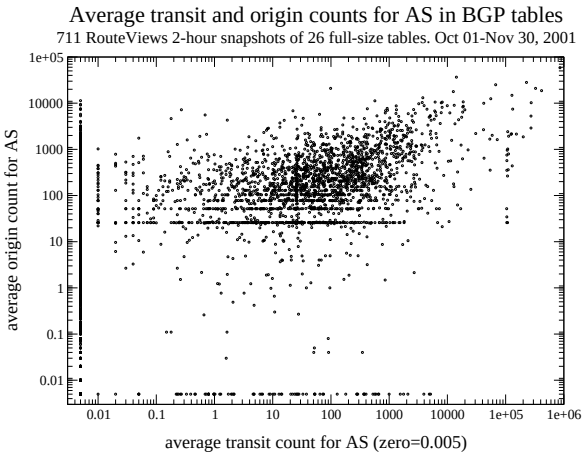

Figure 6: Average number of transit and origin (end-path) positions for an AS in a BGP table. Non-transit ASes are at $x = 0.005$; transit-only at $y = 0.005$.

multihomed ASes than by any other group. Therefore we cannot conclude that non-transit ASes are flooding the prefix table with more specifics.

Table 17: Non/transit, single/multihomed ASes, Nov 01 prefixes

Transit multihomed ASes		
standalone	23801	49.28%
root	3723	7.71%
more specific	20777	43.02%
total	48301	100.00%
Transit single-homed ASes		
standalone	2608	37.74%
root	344	4.98%
more specific	3958	57.28%
total	6910	100.00%
Non-transit multihomed ASes		
standalone	10715	34.80%
root	999	3.24%
more specifics	19080	61.96%
total	30794	100.00%
Non-transit single-homed ASes		
standalone	6507	39.51%
root	452	2.74%
more specifics	9509	57.74%
total	16468	100.00%

8.5 MULTIHOMING AND TRAFFIC ENGINEERING

Internet lore suggests that some ASes maximize the utilization of several links by announcing parts of their networks on selected connections rather than across all possible links. To protect against outages, an aggregate prefix that covers the whole network is announced on all links. Splinter blocks associated with individual connections become subsets of the aggregate. Traffic engineering may be implemented by announcing these more specifics on outbound connections based on traffic loads.

The number of links on which a prefix is originated is derived from the last hop of AS paths reaching that prefix. 01 November 2001 RouteViews[22] data has multihomed ASes originating on a single link 22.2% of all prefixes, 15.5% of all top prefixes, and 28.6% of all more specifics.

More specifics are announced on one link more frequently than top prefixes despite the fact that multihomed ASes do not have any preference for more specifics as we showed above. In the absence of traffic engineering, the fraction of top prefixes announced on one link should be expected to equal the fraction of more specifics. The fact that this condition does *not* hold suggests that a sizable minority (up to 13%) of more specifics announced on one link may be traffic engineered splinter blocks.[23]

Table 18: Link utilization re. traffic engineering

Prefixes announced on	Percent
1 connection by singly homed AS	22.6%
2 connections by doubly homed AS	15.8%
3 connections by triply homed AS	31.1%
2 connections by triply homed AS	7.3%

The discussion above should be contrasted with Table 18[24], which shows that the majority of semiglobal prefixes are announced on all available connections, confirming that traffic engineering of this flavor is only marginally present in today's Internet. We conclude that most of the more specific prefixes are not involved in traffic engineering, and BGP-based traffic engineering techniques contribute a small number of prefixes (at most 1/16th) to the BGP table.

9 DYNAMICS OF IP ADDRESS SPACE

Address space can be obtained from one of three registries, ARIN (Americas, S.Africa), RIPE (Europe, North Africa) and APNIC (Asia-Pacific). Responsibility for a block is transferred either to an ISP (allocation, Figure 7) or to an end customer (assignment, Figure 8). RIPE only allocates address space and maintains a policy of recycling address blocks. About 80% of all transactions have been through ARIN (as of March 2001).

Figures 7 and 8 show allocation and assignment of address space over the last 20 years. Figure 7 shows that total allocated space has grown at approximately at the same pace since the mid-1990s. Assignment leveled off in mid-1990s after reaching 1.3 billion addresses.[25] Allocated addresses are approaching 900 million as of 2001.

[22] 26 full-size backbone tables, 103551 semiglobal prefixes. The data for 39 tables has 103492 semiglobals and produces almost the same statistics (within 0.2% of these numbers).

[23] Subtraction of percentages makes sense since the number of top prefixes and more specifics in the semiglobal set are close to each other.

[24] 3 and "triply" means here "3 or more".

[25] This number includes IANA assigned addresses such as 10/8, 14/8, 17/8 and 224/4 (multicast), a total of 285 million addresses.

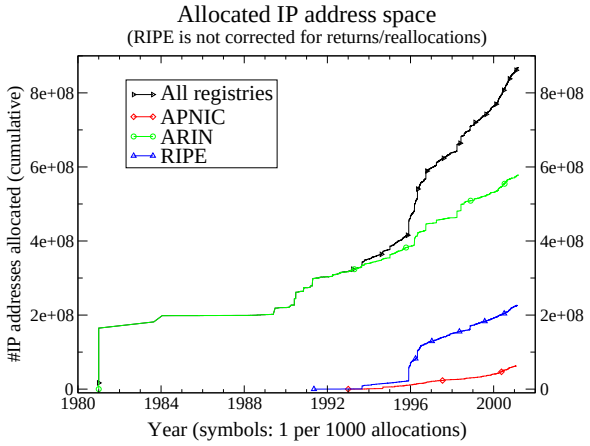

Figure 7: Allocated address space by registry

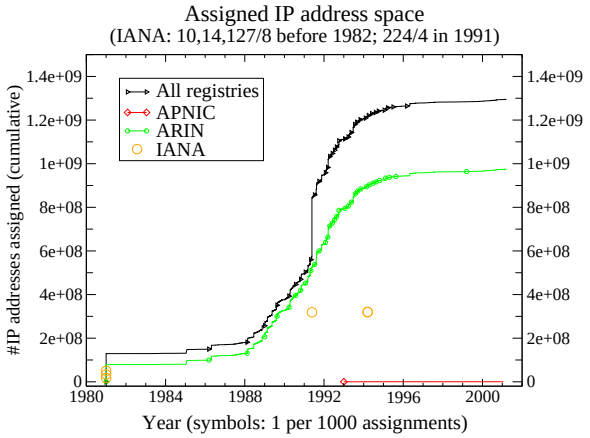

Figure 8: Assigned IP address space by registry

The number of routed IP addresses grew at a rate of a few percent (1-7%) a year since 1998 (Table 3). Combined with much faster growth in the number of announced prefixes, this resulted in prefix *refinement* i.e. steady decrease in the average number of IP addresses per prefix.

9.1 IP SPACE COVERAGE BY PREFIX GROUPS

We have so far studied subset (more specific) and top (root and standalone) prefixes. We will now examine the hierarchy of prefixes, where a given prefix A may have a subset B, which has in turn a subset C, etc. At the top of this hierarchy are prefixes that have no supersets; at the bottom are those that have no subsets.

Table 19: Prefix depth and address space coverage

	2000			2001	
depth	prefixes	addresses	depth	prefixes	addreses
0	39514	1083.33M	0	49865	1134.14M
1	39839	93.91M	1	45058	121.36M
2	8500	8.04M	2	7799	9.16M
3	832	0.43M	3	791	0.77M
4	28	0.02M	4	36	0.04M
5	1	0.00M	5	2	0.00M

Table 20: Prefix depth vs. height in Nov 2001

Depth	ht.0	ht.1	ht.2	ht.3	ht.4	ht.5
0	44264	4743	727	120	10	1
1	43072	1816	156	12	2	
2	7516	265	16	2		
3	768	21	2			
4	34	2				
5	2					

Table 21: Prefix depth vs. height and address space coverage

Depth	ht.0	ht.1	ht.2	ht.3	ht.4	ht.5
0	590.63	328.79	154.30	58.62	1.77	0.03
1	79.84	35.06	5.56	0.88	0.02	
2	7.57	1.20	0.38	0.01		
3	0.70	0.07	0.00			
4	0.04	0.00				
5	0.00					

DEFINITION. The number of supersets (less specifics) of a prefix is called its *depth*. The length of the longest chain of subsets (next subset in the chain being the subset of the previous) is called prefix *height*.

We can think of depth as a specificity level. In particular, top (root and standalone) prefixes have depth 0. Standalones are both top and bottom prefixes. Subsets of a prefix make up a (binary) tree; prefix height is the tree's height, i.e. the length of a longest path from the root to the leaves.

Tables 19-21 compare counts of prefixes in each group with their IP address space coverage. Note that addresses from each depth level are also accounted for in all smaller depths, since respective prefixes are subsets of those with smaller depth. In particular, the entire IPv4 space consumption is given by addresses covered by top (depth 0) prefixes; subsets do not contribute.

Analysis of root prefix length distribution shows that most IPv4 addresses covered by roots are in /8s (235M out of 543M in all roots) and /16s (100.6M). Roots in /9-/15 prefix range contribute roughly equal amounts – 20-30M addresses per each prefix length, whereas /17-/19s cover about 10M each, /20s 2.8M, and roots in /21-/23 prefix range cover almost no addresses at all.

Table 19 compares distribution of subsets (more specifics) by depth in 2000 and 2001. We see that subsets can have any depth up to five[26]. Each depth level roughly corresponds to a factor of 10 drop in covered IP addresses. This factor may change in the future, since growth of subset prefixes is much faster than general growth of routed IP space (Table 3.)

9.2 EVOLUTION OF TOP PREFIXES

As we have shown in Section 7, most of the table growth in the year 2000/2001 is caused by an increase in

[26] Such a set of IP addresses will be covered by six prefixes, thus making sure it is reachable against all odds!

top prefixes. This increase derives from four sources: allocation, deaggregation, expansion and aggregation. We will analyze the relative importance of these four causes, comparing top global prefixes from 19 full-size peers common for 28 November 2000 and 03 May 2001.

We have 8237 new top global prefixes that were not in the 28 November 2000 table but were in the 03 May 2001 table. 4525 prefixes (55%, covering 38.75M addresses in the May 2001 table) are completely new. They have no related prefixes, i.e prefixes with any common addresses (which can only happen if they are more or less specifics) in the 28 November 2000 table. We will label these new prefixes as *allocation*. (Actual allocation of the address block by registries may have occurred long before.)

3712 prefixes (45%) have related prefixes in the 28 November 2000 RouteViews table. Of those, 3306 (40%) are fully covered by global prefixes from the 28 November 2001 table and 406 (5%) are partly covered. We will label partly covered prefixes as obtained by *expansion*, although some types of automatic aggregation (e.g. *autosummarization*) can also result in aggregates that are only partially covered by summarized blocks.

Of the 3306 fully covered prefixes, 2941 (35.7%) have exactly one less specific in the November 2000 table. 150 prefixes (1.8%) have two less specifics. None have 3 or more less specifics. These prefixes are obtained by *deaggregation.*

215 prefixes (2.6%) out of 3305 fully covered prefixes do not have less specifics. These are obtained by *aggregation*. Among the 5% of prefixes that were partly covered in the 28 November 2000 table (which we count as expansion): half of those prefixes were expansion of just one prefix; the other half were expansion of more than one prefix to a larger single block, which can also result from aggregation.

The breakdown of sources of the new top prefixes is:

Allocation	55%
Deaggregation	37.5%
Expansion	5%
Aggregation	2.6%.

9.3 CHURN IN IP ADDRESSES

With respect to the set of top semiglobals of RouteViews 01 November 2001 versus 01 November 2000 data, 9496 blocks are not covered or only partly covered by the RouteViews 01 November 2000 (26 tables), resulting in 119.5M new IP addresses over the year. Among the November 2000 top semiglobals, 3341 are not fully covered by the final 2001 set, resulting in a loss of 68.7M addresses for the year. The yearly net change is approximately a 50M address increase, with loss exceeding net gain. So in terms of address space, the incontrovertible conclusion is that Internet inter-domain routing evolution characteristics have much more to do with churn than with growth.

10 ROUTING FLUX

Our study of routing fluctuation uses the notion of semiglobal prefixes and is based on 724 snapshots sampled at 2 hour intervals in the 61 days between 01 October and 30 November 2001. We performed our analysis with two sets of peers: one with all backbone tables including filtered tables (32 peers); the other with 26 full-sized backbone tables. Both sets were common across the measurement interval. We present analysis results for the 26 peers; the data for 32 peers is essentially the same. Figure 1 plots bi-hourly semiglobal and global prefix counts.

In a small number of snapshots the amount of data collected was too small to yield enough semiglobal prefixes to study. We chose a cutoff of 100,000 prefixes to avoid occasional outages of the collection machine. This cutoff leaves 711 files in our sample.

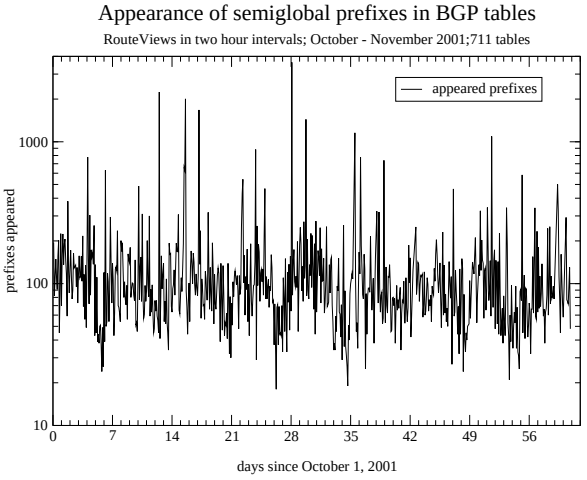

Figure 9: Ssemiglobal prefixes: new appearance and reappearance in RouteViews BGP tables (bi-hourly counts)

If we were to use global prefixes instead, we would have only 696 files that contained data from all 26 peers. We could not use the remaining files because they contain data from 25 or fewer peers. The number of cases in which the count of prefixes dropped below 100,000 was also higher, resulting in only 632 usable data sets. Semiglobals are a much more robust notion than global prefixes, even though their number only slightly exceeds that of globals under normal conditions.

Figure 9 shows the number of prefixes that appear in the table after (at least a 2-hour) absence. The variation has a clear mid-week maximum and weekend drop (01 October 2001 was a Monday). Figure 10 shows that the number of times a prefix can appear in a table after absence, which is a bulk measure of prefix instability, has a fairly long-tailed distribution. The semiglobal prefix counts for our data sample are in the table below.

Oct 01, 2001, 00:00 semiglobals	104555
Nov 30, 2001, 20:00 semiglobals	103815
Union over all 711 tables	137374

We measure stability of the routing tables by examining the number of prefixes that appear and disappear in the BGP snapshots between 01 October and 30 November 2001. We divide prefixes into several groups based on their presence or absence in the tables:

long-lived – present in both the first and last table
persistent – present in all tables
apparently emerged – missing in the first table
apparently disappeared – missing in the last table
transients – missing in both the first and last table

The word *apparent* refers to the fact that prefixes may have disappeared just during the first or last tables in our selected sample. Note that these two categories, *emerged* and *disappeared* are not mutually exclusive since both contain transients, in fact: long-lived + emerged + disappeared - transients = total observed = 137K. Table 22 presents these categories with their corresponding prefix counts.

Table 22: Prefix classification (01 Oct - 30 Nov 2001, Union = 137K; 01 Oct = 104K prefixes)

Category	Number	% Union	% Oct 1
Long-lived	93127	67.79%	89.07%
Persistent	72176	52.54%	69.03%
App. emerged	32819	23.89%	–
App. disappeared	33559	24.43%	–
Transients	22131	16.11%	–

The number of emerged and disappeared prefixes is almost the same, since the table changed (decreased) by only 740 prefixes in 2 months; the churn in the prefix set can be quite high without causing significant change in the table size.

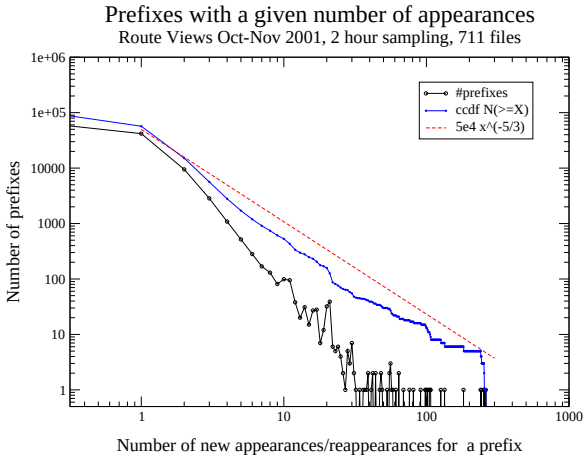

Figure 10: Number of prefixes as a function of their churn

10.1 LIFETIME AND UPTIME

We define the *lifetime* of a prefix as the time from its first appearance in our data sets to its last disappearance. The number of time samples that the prefix was present in the semiglobal set is its *uptime*. The distribution of prefix uptimes shown on Figure 11 reveals a clear dichotomy, with many prefixes being up and running for the whole observation interval, others only for a short while, and small counts for intermediate uptimes. It is interesting to note intermediate uptimes also follow a weekly pattern induced by a similar structure in prefix dis/appearance (Figure 9.)

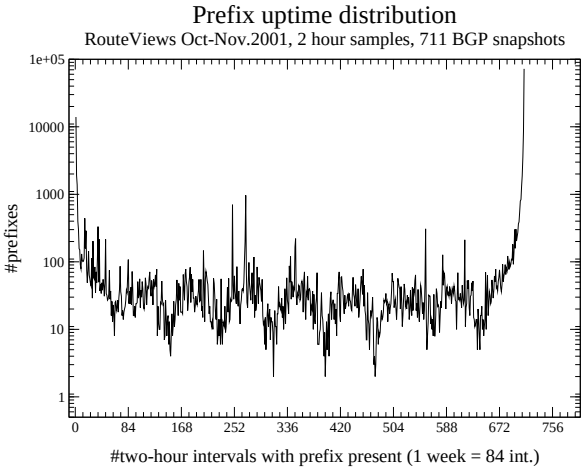

Figure 11: Prefix uptime distribution

10.2 FLIPS

We define a *flip* to be an event in which a prefix is dropped from the routing table and then reannounced after a contiguous interval of non-existence. The number of flips observed depends on the sampling rate and the exact moment that the sample is taken. If a prefix flips in only a few tables but remains stable in most of the tables, it will remain in the semiglobal set and will not be counted as a flip. Semiglobal prefixes thus flip less frequently than global prefixes, one of the reasons for using this definition.

Table 23: Flip statistics for semiglobal prefixes

Percentage of prefixes		
Total prefixes that flip	31007	22.57%
#Long-lived that flip	20951	15.25%
#Transient that flip	3743	2.72%
Percentage of flips		
Total flips of long-lived	35591	60.06%
Total flips of transients	9285	15.67%

Table 23 shows flip statistics for long-lived and transient semiglobal prefixes. Percentages refer to all observed prefixes (137K) and all flips (59K). 68% of prefixes are long-lived and account for 60% of the flips. Transients are 16% of the prefixes and contribute 15.7% of the flips. The frequency of flips is thus mostly independent of prefix category, although long-lived prefixes are slightly more stable.

Table 24 shows flip behavior by prefix length and shows that most transients are in the /21 to /24 range.

Table 24: Flips by prefix length (fp.pf = flipping prefixes)

Len	prefixes	fp.pf	flips	tran	%all	%fp.pf	%flips	%tran
8	28	8	264	11	0.02	0.03	0.45	0.05
9	5	0	0	0	0.00	0.00	0.00	0.00
10	9	0	0	1	0.01	0.00	0.00	0.00
11	12	1	2	0	0.01	0.00	0.00	0.00
12	39	8	12	1	0.03	0.03	0.02	0.00
13	95	13	21	0	0.07	0.04	0.04	0.00
14	236	27	43	5	0.17	0.09	0.07	0.02
15	410	55	81	5	0.30	0.18	0.14	0.02
16	7387	1256	1966	115	5.38	4.05	3.32	0.52
17	1452	203	302	62	1.06	0.65	0.51	0.28
18	2605	401	581	115	1.90	1.29	0.98	0.52
19	7450	1213	1792	203	5.42	3.91	3.02	0.92
20	6852	1737	3981	356	4.99	5.60	6.72	1.61
21	7176	1193	1866	2009	5.22	3.85	3.15	9.08
22	8803	1792	3367	1169	6.41	5.78	5.68	5.28
23	11636	2600	4503	2028	8.47	8.39	7.60	9.16
24	83177	20500	40480	16049	60.55	66.11	68.31	72.52
Tot	137K	31007	59261	22131	100.0	100.0	100.0	100.0

Prefixes of length /24 (or /24's) cause a slightly larger fraction of flips than their proportion of semiglobals (68% vs. 60%). This difference is not huge so we cannot say that /24s are the only source of flips in the table. Recall that /24s are the smallest address blocks that can cross AS boundaries and therefore any smaller block that needs to be globally routed (even individual host routes) must be carried within a /24 (or larger prefix). The /24's thus tend to inherit instability from many smaller blocks and therefore tend to be more unstable as a group.

Table 25: /24 stability by address space class

Class	/24s	fp.pf	flips	%/24s	%fp./24s	%flips
A	11266	1890	3338	13.54	9.22	8.25
B	6666	2034	4587	8.01	9.92	11.33
C	65245	16576	32555	78.44	80.86	80.42
Total	83177	20500	40480	100.00	100.00	100.00

Table 25 shows the stability of /24s in each of the original address classes: A, B, and C. The most unstable portion of /24s lies in traditional class B space (l28.0.0.0-191.255.255.255), consistent with the results of [18]. This instability can be a justification for filtering out long prefixes in this space. However, the share of /24s in class B space is only 8%, which translates to 4.5% of all prefixes in the table. The reduction is therefore almost negligable both in terms of table size and in terms of flux.

The number of flips over the lifetime of a prefix is a good measure of its stability. We will call it *flip rate*. An AS flip rate is defined as the sum of all its prefixes' flips divided by the sum of their lifetimes. Another possible measure of stability is the number of flips per uptime. This measure will make prefixes that are present in the the table over long periods of time appear more stable. We provide this measure for comparison in the tables below

Table 26 shows statistics of flips and AS events associated with prefix (re- and dis-) appearances. We compute flips over time using both lifetime and uptime values. Each row of the table represents flips accumulated by an AS over the lifetime of prefixes that this AS originates. If a prefix changes its origin AS after a period of downtime, this interval of downtime is not counted as part of the prefix lifetime in any AS. Columns that represent the rate of flipping are normalized to 1000 time units (85 days).

Table 26: Flip counts per AS and time slot counts for AS's prefixes (re-, dis-) appearances; wtd=withdrawal, rean=reannouncement. Kupt (Klft) – uptime (lifetime) of 1000 2-hour time slots; ASes not identified (see authors if identification leads to improvement/repair)

AS	Av.#pf	flips	flips/ Kupt	flips/ Klft	pc.fp	Ac.%	new ann	tmp wtd	rean	fin wtd
1	421.72	1466	4.89	4.59	2.47	2.47	34	52	51	16
2	3.60	1444	564.72	335.58	2.44	4.91	1	659	658	2
3	2196.94	1379	0.88	0.82	2.33	7.24	101	351	350	91
4	956.48	924	1.36	1.32	1.56	8.80	68	373	346	45
5	652.29	921	1.99	1.79	1.55	10.35	63	105	109	43
6	397.04	661	2.34	2.18	1.12	11.47	62	162	165	46
7	180.75	613	4.77	4.36	1.03	12.50	5	147	144	7
8	28.91	540	26.27	21.79	0.91	13.41	23	237	238	22
9	308.19	510	2.33	2.20	0.86	14.27	31	195	180	31
10	161.29	497	4.33	4.14	0.84	15.11	21	52	59	19
11	48.03	378	11.07	8.71	0.64	20.58	7	7	7	4
12	244.60	278	1.60	1.56	0.47	25.14	12	76	81	11
13	692.89	190	0.39	0.38	0.32	30.21	55	94	90	33
14	6.90	66	13.46	13.26	0.11	50.01	0	12	12	0
15	239.63	15	0.09	0.09	0.03	75.02	6	10	10	3

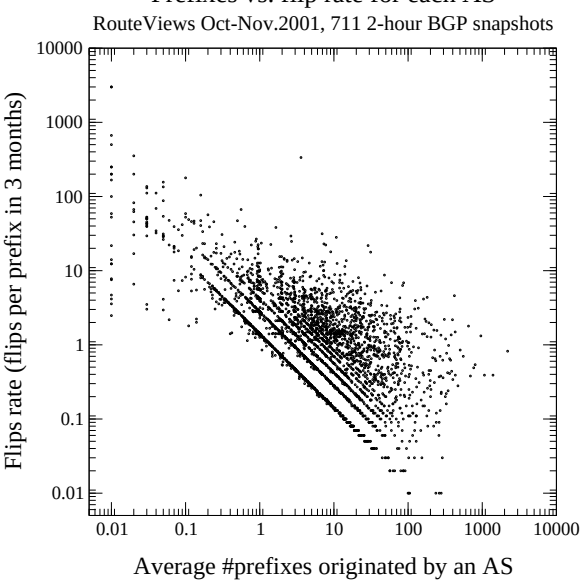

Figure 12: Lifetime flip rate vs. prefixes originated by AS

There are approximately 12,400 ASes in the tables; 5259 (42%) of them originated prefixes that flipped. Figure 12 shows a scatterplot of the number of flips per prefix for an AS versus the number of prefixes that AS originates. The general pattern of the graph suggests that ASes that originate more prefixes tend to exhibit more stability per prefix. Some ASes originate many (100s) of prefixes but flip very infrequently. Those further up to the right from the diagonal contribute most to instability, via a combination of above average flip rate and large number of prefixes.

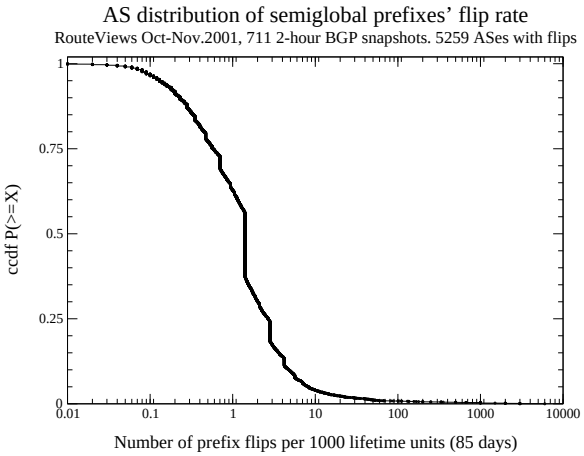

Figure 13: Prefix flip rate distribution (ccdf $P\{x \geq X\}$.)

Half the flips are contributed by 150 ASes (1.2% of all ASes) and 3/4 of the flips come from 1/8 of the flipping ASes. The AS with the maximum number of flips is a large consumer ISP. The second largest contributor is a research network that works on routing stability and injects a few routes targeted for specific backbones. In third place is the largest AS in the BGP table in terms of advertised prefixes. ASes that on average announce 5 or fewer prefixes accounted for 1/5 of all flips in two month period. In the 01 Nov 2001 12:00 snapshot these small ASes constituted 3/4 of the total AS count. Therefore we cannot conclude that small ASes are a major source of routing instability, even if some of them have higher than average flip rate.

The column labeled flips/lifetime is a reasonable measure of the stability of a network. A value of 1 in this column would indicate that a prefix flips once in a three-month period. As Figure 13 shows, the median flip rate is about 1.4 flips per 1000 time slots, or one flip in two months. It shows as a drop in the ccdf caused by one-prefix ASes (recall that they constitute 40% of all ASes). ASes with flip rates lower than this value can be viewed as reasonably stable. Other drops in the ccdf curve (Figure 13) correspond to two flips for one prefix in two months, to one flip in two prefixes from the same AS, etc. We also note that global prefixes flip on average more frequently than semiglobals.

Table 27: Top contributors to BGP table change. a.d.= appearance and disappearance (AS info is anonymized; interested engineers please contact authors)

Type	Av.#pf	app dis	a.d /Kup	a.d /Klf	% all a.d.	acc %	new ann	tmp wtd	Rean	fin wtd
Bb	2196.94	6790	4.35	4.04	3.67	3.67	101	351	350	91
Bb	652.29	5915	12.75	11.49	3.20	6.87	63	105	109	43
CP	421.72	3236	10.79	10.14	1.75	8.62	34	52	51	16
Bb	692.89	3199	6.49	6.47	1.73	10.35	55	94	90	33
Res	3.60	2893	1131.4	672.32	1.56	11.92	1	659	658	2
Bb	956.48	2883	4.24	4.12	1.56	13.48	68	373	346	45
CE	46.22	2589	78.79	77.60	1.40	14.88	3	10	10	6
Bb	397.04	2409	8.53	7.94	1.30	16.18	62	162	165	46
Bb	21.64	1493	97.02	28.93	0.81	16.99	6	2	2	6
Mil	180.75	1448	11.27	10.30	0.78	17.77	5	147	144	7

Table 27 shows the top contributors to the overall BGP table change during our two-month measurement period. This data includes appearance, disappearance and reappearances (flips) of prefixes, which renders total counts higher. The top 10 contributors to total routing system flux are classified as: Bb – backbone provider, CP – content provider, Res – research network, CE – computer engineering company, and Mil – a US military network.

For many ASes, prefix flips and appearance/disappearance come from a large number of AS events, i.e. moments when some prefix from an AS changes status (the last four columns of Table 27), suggesting that they derive from background noise rather than from BGP storms. Less frequent but still observable are cases in which a large number of flips arises from a small number of storm-like events.

11 CONCLUSIONS

We have analyzed trends in evolution of the global Internet interdomain routing system using RouteViews BGP routing tables snapshots of 1997-2001. We taxonomized address blocks expressed by semiglobally routed prefixes (those present in the majority of backbone tables) with regard to having subsets and supersets as

standalone – no subsets, no supersets;

root – have subsets, but no supersets;

subset, or more specific – are subsets of other blocks.

We found that in 1999-2001 many measures of routing system complexity demonstrated slow growth, dynamic equilibrium, and occasional contraction.

We also found that many net change measures reflect contributions of opposite sign, and that variation, or *churn*, should be measured as sum, not a difference of their values.

In particular, we found that:

- The number of semiglobal prefixes was stable in from October to December 2001, compared to 37% growth between November 2000 and November 2001.
- AS path length, both the mean and the overall distribution, did not significantly change in several backbone BGP tables between 1999 and 2001.
- The link/node ratio (average degree), and peering richness of the BGP AS graph did not significantly change between November 2000 and May 2001, although individual ASes often exhibited a high degree of change.
- Prefix set churn in 2001 was much higher than the prefix growth rate. The churn was highest for subset routes, followed by root and standalone blocks. AS and IP address churn was smaller, but still comparable to their net growth.
- Subset (more specific) routes constitute half of the entries in global BGP tables. Their proportion grew

from 50% in November 1999, to 55% in November 2000, and then decreased to 52% by November 2001.

- 57% of the table is composed of /24 prefixes, the smallest globally routable address blocks. This fraction has been almost constant since 1997.
- As of November 2001, being a multihomed network (transit or non-transit) is *not* significantly related to announcing subset (more specific) routes.
- Transit ASes originate more prefixes than non-transit ASes, despite the fact that there are five times as many non-transit as transit ASes. Transit multihomed ASes originate about as many prefixes as all non-transit ASes, and more subset prefixes than non-transit multihomed ASes.
- The number of non-transit multihomed ASes grew from 46% to 49% from 2000 to 2001, but their share of global routes remained stable at around 30%.
- Between November 2000 and May 2001, new address space announcements and deaggregation of existing prefixes were two major sources of new root and standalone prefixes.
- Half of the routing instability in the form of withdrawal/reannouncement events in late 2001 is contributed by 1.2% of all ASes, with government networks, telecoms in developing countries and major backbone ISPs at the top of contributors' list. Small ASes (those originating a few prefixes) do not contribute more than their fair share to the BGP table size and to instability of the global routing system.

We conclude that in the studied period many Internet metrics were stable, and that the Internet's growth and instability originate mostly in large and medium-sized ISPs.

12 ACKNOWLEDGMENTS

Many thanks to: David Meyer of U. Oregon; Bill Woodcock and Sean McCreary of Packet Clearing House; Brad Huffaker, David Moore, and Daniel Plummer of CAIDA; Geoff Houston of Telstra; Sean Doran of Ebone; and Andrew Partan for their helpful feedback and guidance.

REFERENCES

[1] Y. Rekhter and T. Li., "A Border Gateway Protocol 4 (BGP-4), RFC1771," Mar 1995.

[2] John W. Stewart, *BGP4: Inter-Domain Routing in the Internet*, Addison-Wesley, 1999.

[3] Geoff Huston, "Analyzing the Internet's BGP Routing Table," *The Internet Protocol Journal*, vol. 4, Mar 2001, http://www.telstra.net/gih/papers/ipj/4-1-bgp.pdf.

[4] A. Broido and k claffy, "Complexity of global routing policies," 36 p., in preparation.

[5] A. Broido and k claffy, "Analysis of Route Views BGP data: policy atoms," Proceedings of Network-related data management (NRDM) workshop, Santa Barbara, May 25, 2001, 18 p.

[6] A. Broido and k claffy, "Internet stability amid change, Presentation at IETF and ISMA," Dec 2001, http://www.caida.org/outreach/presentations/BGP2001dec.

[7] B. Huffaker, A. Broido, k. claffy, M. Fomenkov, K. Keys, E. Lagache, and D. Moore, "Skitter AS Internet Graph," Oct 2000, http://www.caida.org/analysis/topology/as_core_network/.

[8] David Meyer, "Universtiy of Oregon Route-Views Project," 2001, http://www.antc.uoregon.edu/route-views/ISMA00.

[9] "Nlanr measurement and operations analysis team (moat)," http://moat.nlanr.net/Routing/rawdata.

[10] Sean McCreary, "BGP Core Routing Table Size," 2000, http://www.routeviews.org/dynamics.

[11] J. Rexford, R. Bush, and S. Bellovin, "Some Initial Measurements of Prefix Length Philtres, Presentation at NANOG," Scottsdale, AZ, May 21, 2001, http://research.att.com/~jrex/nanog/lost.html.

[12] Andre Broido, "Multiorigin prefixes in backbone BGP tables," http://www.caida.org/~broido/bgp/multiorigin.html.

[13] Y. Rekhter, B. Moskowitz, D. Karrenberg, G. J. de Groot, and E. Lear, "Address Allocation for Private Internets, RFC1918," February 1996.

[14] R.Mahajan, D.Wetherall, and T.Anderson, "A study of BGP misconfigurations, Presentation at ISMA," Dec 2001, http://www.caida.org/outreach/isma/0112/agenda.xml.

[15] Andre Broido and k claffy, "Internet Topology: connectivity of IP graphs," in *SPIE conference on Scalability and Traffic Control in IP Networks*, Aug 2001, http://spie.org/Conferences/Programs/01/itcom/confs/4526.html.

[16] Bradley Huffaker, Daniel J. Plummer, David Moore, and k claffy, "Distance Metrics in the Internet," February 2002, SAINT2002 Workshop: Measurement Technology for the Internet Applications, Tokyo, Japan. to appear.

[17] Claude Shannon, *The mathematical theory of communication*, 1949, Urbana, Univ.Illinois Press.

[18] C. Labovitz, R. Malan, and F. Jahanian, "Internet Routing Stability," Proceedings of ACM SIGCOMM 1997. http://www.acm.org/sigcomm/sigcomm97/program.html#ab109.