Public Review for
Observed Relationships between Size Measures of Internet

Brian E. Carpenter

This paper studies three measures of the size of the Internet (BGP4 routing table, active Autonomous Systems, and a lower bound on the size of the Internet), which appear to have simple relationships over a long period of time. Once we identify the metrics, the execution could be considered straightforward, but the reviewers and I found value in the overall observation. As with most measurement studies, one could express doubts whether this will hold ubiquitously. But then, we should stop doing research in this area, until the complete uber-archive emerges. Overall, the work could inspire researchers : (a) to look beyond number of nodes and edges for growth, and potentially expand it with other metrics, (e.g. what about BGP udpates?), (b) to challenge or verify the relationship between ASes and routing table, (c) understand the inherent reason why these relationships hold and how they were or were not affected by routing practices and policies.

Public review written by
Michalis Faloutsos
University of California,
Riverside

acm sigcomm

Observed Relationships between Size Measures of the Internet

Brian E. Carpenter
Department of Computer Science
The University of Auckland
Auckland, New Zealand
brian@cs.auckland.ac.nz

ABSTRACT

This paper reports some observations on the relationships between three measures of the size of the Internet over more than ten years. The size of the BGP4 routing table, the number of active BGP4 Autonomous Systems, and a lower bound on the total size of the Internet, appear to have fairly simple relationships despite the Internet's growth by two orders of magnitude. In particular, it is observed that the size of the BGP4 system appears to have grown approximately in proportion to the square root of the lower-bound size of the globally addressable Internet. A simple model that partially explains this square law is described. It is not suggested that this observation and model have predictive value, since they cannot predict qualitative changes in the Internet topology. However, they do offer a new way to understand and monitor the scaling of the BGP4 system.

Categories and Subject Descriptors

C.2.2 [**Network Protocols**]: Routing protocols; C.2.3 [**Network Operations**]: Public networks

General Terms

Measurement

Keywords

Internet topology, BGP, inter-domain routing

1. INTRODUCTION

The Internet has grown dramatically since it converted to TCP/IP on January 1, 1983, and much of this growth has occurred since the Border Gateway Protocol version 4 (BGP4) [17] became the principal wide-area (inter-provider) routing protocol in early 1994. BGP4 is usually described as a path vector protocol (i.e., a modified distance vector protocol), whose main difference from its predecessors is that it allows routing prefixes of variable length to be aggregated together or filtered out, according to policy rules defined independently by each Internet service provider [2]. These mechanisms, combined with administrative policies for the hierarchical assignment of address space, are intended to limit the growth of the BGP system. Prior to BGP4, only certain fixed prefix lengths were allowed, and there were no particular mechanisms to limit growth.

In view of current concerns about future scaling issues in the routing system [14], it is of interest to review available historical data, in order to understand how the BGP system has scaled thus far. Given the changes introduced by BGP4, the middle of 1994 is an appropriate time to start such a review. This paper therefore examines the relationship between three measures over a period of more than ten years: the size of the BGP4 routing table, the number of active *Autonomous Systems (ASes)*, and a lower bound on the size of the globally routeable public part of the Internet.

BGP4 distributes routing information for IP address prefixes; in other words, one route applies to all traffic towards a contiguous set of IP addresses that all start with the same N bits. Where two prefixes of lengths N and M ($N < M$) match in the first N bits, the longer prefix has priority in routing lookups. Prefix lengths are typically in the range from 8 to 24 bits in the case of IP version 4. (Adequate historical data do not yet exist for IP version 6, so this paper considers only IPv4.) Each prefix that is separately advertised by a BGP4 speaker potentially causes one entry in the forwarding table of every other BGP4 router. In practice this does not occur. Route advertisements are filtered administratively according to policy rules set individually by network operators, and routes for contiguous prefixes may be aggregated. The consequence is that the number of BGP4 routes actually held in the forwarding table of a given BGP4 router depends on where it lies in the network. Nevertheless, this table size is widely quoted as a measure of the *default free zone* or *DFZ*, i.e., the size of a hypothetical routing table for the notional backbone of the Internet. In BGP4, the terms "prefix count", "routing table size", "number of routes", and "DFZ size" are all equivalent. Under any name and measured at any point, this value is a measure of the scaling challenge faced by operators. As discussed in [14], excessive growth of this table is a primary concern for the future viability of Internet routing. Its relationship to the growth of the Internet is therefore an important issue.

BGP4 also recognizes that the network is divided administratively by using the concept of Autonomous Systems. An AS is essentially a small set of networks that present themselves as a single entity to BGP4. Each AS has a unique label, commonly referred to as an *AS number*. Paths in BGP4 are in fact *AS paths*, i.e., a route including a sequence of AS numbers leading to the AS that is directly connected to the required destination prefix. A given AS may advertise multiple prefixes, because it provides connectivity for all of them; in other words, more than one BGP4 route may lead to the same AS. The total number of active ASes, although not a primary concern for BGP4 scaling, is also a useful indicator of the size and growth characteristics of the BGP4 system.

In [14] and elsewhere, it is common to consider the growth of the BGP4 system as a function of time, and to express concern when the growth appears to be faster than linear, or faster than the perceived growth in router capability (memory size, processing speed, etc.). However, there is no reason that BGP4 growth should have any particular relationship to the calendar. The present study takes a different approach: how does the BGP4 system grow in relation to the total size of the public Internet? This relationship should have a physical basis independent of the passage of time, so it is susceptible to observation and analysis.

BGP4 can by definition only route globally the public address space of the Internet, i.e., all public addresses used by directly accessible hosts, and all active addresses on the public side of network address translators. Therefore, it is relevant to the present study to measure the number of publicly addressable nodes. Ideally, we would need an exact count of the number of IP addresses to and from which packets may be directly routed over the global Internet. This number would characterise the total size of the network as far as the global BGP4 routing system is concerned, but it is not easy to measure.

There are numerous important studies of the BGP4 topology, several of which are cited in Section 5, but they do not yet provide long-term historical data on the size of the network. Data on the number of active hosts in the Internet have been collected in various ways, starting with a manual count of the four ARPANET hosts in December 1969. However, since the Internet grew beyond a few thousand hosts, and progressively lost its transparency due to firewalls and network address translators, such direct counts have lost their meaning. This was already the case by 1994, the starting date for the present study.

The total number of IP addresses handed out is not a useful measure for several reasons:

- In the early years of the Internet, address blocks were handed out directly to user sites. Today, addresses are handed out in a hierarchy: very large blocks to Regional Internet Registries, large blocks to national registries, medium blocks to Internet service providers and large users, small blocks to small users, single addresses to domestic subscribers. As a result of this administrative complexity, it is impossible in practice to collect accurate administrative data about which address blocks are actually in use.

- Even if it is known that a certain address block has been allocated to a given user, it is highly unlikely that all of it is being used. Although the various Regional Internet Registries have policy guidelines for the usage rate of address space, these guidelines have come into existence only in recent years, and they are not rigidly enforceable. Therefore they have no statistical value in a study extending over fourteen years.

- In many cases, very large numbers of allocated addresses are used exclusively inside a company network, cannot be accessed from outside, and have no impact on the public part of the Internet. (Note that the widespread use of private address space hidden behind network address translators similarly has no impact on the size of the public Internet visible to BGP4, and can be ignored for the purposes of this paper.)

Thus it is not possible to construct a history of the real size of the public Internet by counting addresses handed out by the registries.

It is possible to observe from BGP4 data which address blocks are actively advertised as being reachable via the routing system. However, the second and third arguments above still apply: it is impossible to know what fraction of the advertised address space is actually occupied by reachable nodes. Also, such data do not exist in an adequately long historical series.

It is also impossible to measure the size of the Internet by probing or scanning to discover which addresses are active. Such probes or scans are widely blocked, and are in any case likely to be treated as denial of service attacks or as invasions of privacy. No such data are available as a historical series.

Thus, neither the address registry system, nor the actual BGP4 data, nor direct probes, can be used to measure the actual size of the public network. Therefore this paper uses a different measure of the size of the public Internet, namely a measure of the number of entries in the Domain Name System (DNS).

Publicly addressable nodes comprise both servers and subscriber devices acting as clients or peer-to-peer nodes. Servers are necessarily listed in the DNS. Subscriber devices are also normally allocated DNS names automatically by their Internet service providers. Even if the latter devices are in fact network address translators serving a group of hosts, they appear in the public Internet as a single publicly visible node with one IP address. There is unfortunately no way to estimate the fraction of active public IP addresses that are not counted because they do not have corresponding DNS names. The number of distinct hosts listed in the DNS, known as the *domain count*, is thus a lower bound on the size of the publicly addressable Internet that the routing system must be able to handle. There is no better measure available.

2. THE DATASETS

Reliable data series are available for the BGP4 routing table size back to 1994, and for the number of active ASes back to 1997. The active ASes are considered to be those actually advertised as unique AS numbers within the BGP4 table. Many other AS numbers are assigned, but are not advertised in any BGP4 path.

Both these data series are observed from a BGP4 router in an autonomous system known as AS65000 [8]. The raw data are presented in the first two columns of Table 1, extracted from the original series by choosing the nearest sample to 00:00 hours on the 15th day of the month indicated. The ratios between the BGP4 size and AS count are also shown in the third column.

The raw datasets contain very many more data points, but for the current study, only one data point per year was considered, with the month chosen to match the third dataset described below.

It should not be forgotten that observed from a different vantage point in the BGP4 system, these data would have been different, although not dramatically so. This is because the actual BGP4 advertisements arriving at a given BGP4 router depend on the BGP4 filtering policies in place [2]. There is no single view of the BGP4 data that has more validity than any others. The AS65000 vantage point was chosen largely because it provides a conveniently formatted

Table 1: BGP4, AS count, and domain count history

Year, Month	BGP4 size AS65000	Unique AS count AS65000	Ratio	BGP4 size RIS	Unique AS count RIS	Ratio	Domain Count
1994-07	18468						3864000
1995-07	27717						8200000
1996-07	36851						16729000
1997-07	46948	2473	19.0				26053000
1998-07	52199	3695	14.1				36739000
1999-07	62318	5287	11.8				56218000
2000-07	83921	7942	10.6				93047785
2001-07	103095	11283	9.1				125888197
2002-07	111940	13283	8.4				162128493
2003-01	118231	14355	8.2	126063	14729	8.6	171638297
2004-07	138726	17498	7.9	152223	17908	8.5	285139107
2005-07	163442	20001	8.2	173151	20448	8.5	353284187
2006-07	189700	22569	8.4	200003	23087	8.7	439286364
2007-07	228856	25762	8.9	251515	26160	9.6	489774269
2008-07	273992	28811	9.5	307181	29261	10.5	570937778

data series extending over more than ten years. The University of Oregon Route Views on-line archive [21] is also available, but the available data are identically equal to the AS65000 data for certain years, and are therefore not considered to be independent. The relevant data available from the RIPE Routing Information Service (RIS) [18] begin in 2003, so are insufficient for the purposes of this paper. As a cross-check, the RIS data are also shown in Table 1. For the years 2003-2008, the BGP4 size observed by RIS exceeds the AS65000 values by 8.3% on average, and the AS count observed by RIS exceeds the AS65000 values by 2.1%. The discrepancies between the two data sources are to be expected, but would not substantially change the results described below.

There is only one data series that logs the long-term history of the domain count, i.e., the number of distinct hosts in the Domain Name System, known as the ISC Internet Domain Survey [11]. The intent of this survey is to count the number of distinct IP addresses that have been assigned a DNS name, by processing all DNS data on-line at the date of each survey. It counts only domains that are currently supported by a DNS server. The data are collected every six months over a period of days, by scanning the entire DNS. The DNS, unlike the BGP4 system, constitutes a single distributed database, so the observation point does not matter. The data are processed to eliminate duplicate names for the same address. The result is a measure of the total number of distinct hosts with a corresponding name and IP address. For this study one value per year has been used, and the month of collection was used to determine which BGP4 and AS data were used. Data from July were used, except for one year for which only January data were available. Since the date plays no part in the subsequent analysis, this does not affect the validity of the following results in any way.

The measurement method was changed in January 1998, so the earlier "old survey" data have been scaled to allow for this change. The ISC web site did not show a scaled value for 1994, so for the present study that value has been scaled using the same factor applied in 1995. The resulting data are shown in the final column of Table 1, and are used in the remainder of this paper as the relevant values of the domain count. The results presented must be interpreted in

the knowledge that the measured domain count is a lower bound on the actual size of the public network.

3. OBSERVATIONS

Unsurprisingly, all three measures have grown greatly since 1994. The domain count is the most spectacular, exceeding 570 million in 2008 (in contrast to the Internet's host count of 562 in 1983). However, clearly the BGP4 size and AS count have grown much more slowly. During the fourteen years considered, the domain count grew by a factor of approximately 148, and the BGP4 table coincidentally grew by a factor of 14.8 (one order of magnitude less). This can be viewed as an indicator of notable success in the scaling of the routing system.

The relationship between the BGP4 table size and the number of active ASes observed at AS65000 is straightforward - to a reasonable approximation it is linear (see the solid line in Fig. 1). It would also be possible to interpret this plot as two separate linear segments, with the point of inflection in 2004, but in the absence of any underlying physical model to explain it, this interpretation is not pursued. In this figure, and the following two, the limited data available from RIS are shown as a dashed line.

Linear and logarithmic plots of the BGP4 table size against the domain count show nothing interesting. However, it appears that there is an approximately linear relationship between the BGP4 table size and the *square root* of the domain count (see Fig. 2).

Prompted by this observation, the final plot is of the number of active ASes against the square root of the domain count, which is remarkably linear (see Fig. 3).

We can make three gross observations from these plots.

- As has already been observed [19], there are approximately 9 BGP4 entries for each active AS today. This ratio has changed little in the past eight years. However, it was significantly higher, approximately 19 BGP4 entries per active AS in 1997, for the first year or two of observation.

- From 1994 until quite recently, the number of BGP4 entries was approximately 9 times the square root of
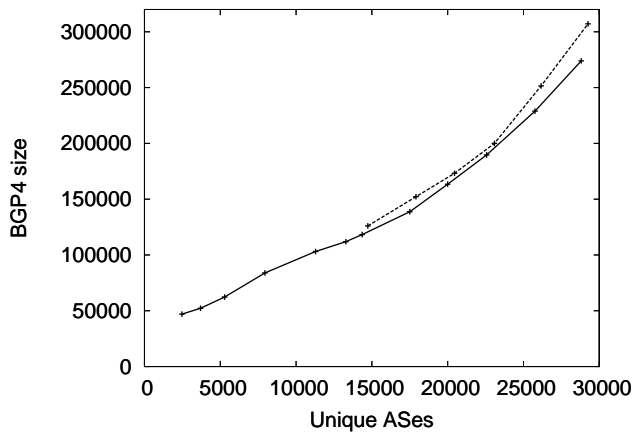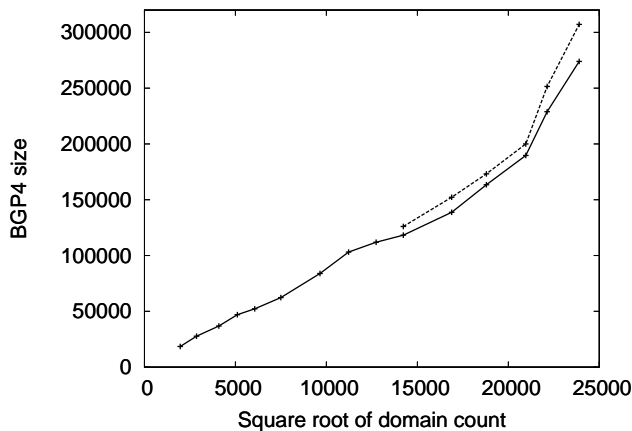
Figure 1: Routes vs AS count (slope = 8.06)

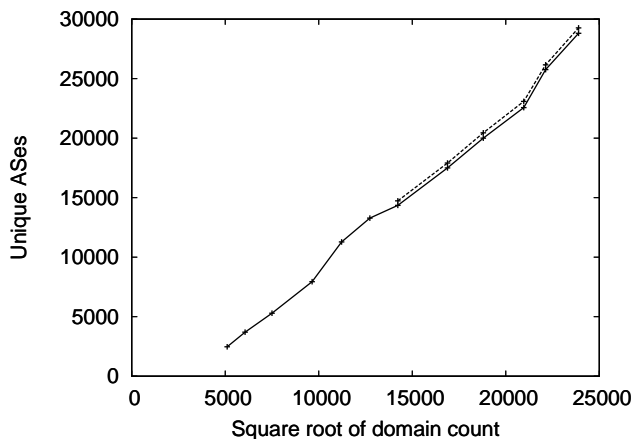**Figure 2: Routes vs square root of domain count (slope = 10.19)**

Figure 3: AS count vs square root of domain count (slope = 1.35)

the domain count. This relationship possibly broke down as recently as 2007, with the ratio now being over 10, but it is impossible to determine whether this change is significant without further data.

- Consistently with these two observations, the number of active ASes was quite similar to the square root of the domain count, being close to equal in 2001-2006. It appears that this relationship too may have broken down in 2007, with a perceptible increase in the relative number of active ASes.

To somewhat refine these observations, a statistical fit was performed for the AS65000 data, using the least-squares fit algorithm of the gnuplot package. If we let B be the BGP4 size, A be the AS count, and D be the domain count, then the data fit the relationships:

$B = 8.06A + 15013$, where the standard error of the coefficient is 5.7%.

$B = 10.19\sqrt{D} - 10563$, where the standard error of the coefficient is 5.8%.

$A = 1.35\sqrt{D} - 4615$, where the standard error of the coefficient is 2.2%.

It would also be possible to seek more complex polynomial relationships between the data sets. However, this would serve little purpose, in the absence of an underlying theoretical model to be tested. Correlation coefficients were not calculated, since it is trivially true that growth measurements of the Internet are correlated.

4. DISCUSSION

Firstly, we observe that the plots show great similarity between the observations made at AS65000 and the more limited data from RIS. We may infer that the results are not due to some peculiarity of the observation point.

It is well known that all ASes in BGP4 are not the same; at the very least they fall into two categories (stubs, and transit providers) but the reality is considerably more complex [16]. It is also well known that the topology of the BGP4 system shows the power law or scale-free behaviour that is characteristic of many types of network [15], [5], essentially because some service providers are more popular than others, which leads to the power law effect. However, this is beside the point for the present paper, which considers the number of prefixes originated and not the topology of the resulting routes. The power law effect in the topology reflects commercial and organisational relationships among the ASes, but it does not determine how many active ASes exist or how many address prefixes must be routed.

The fact that all three plots presented above are roughly linear may be viewed as another aspect of scale independence in the Internet, averaged across all ASes.

First we may ask why there is a roughly linear relationship between BGP4 size and AS count. Why does each active AS contribute on average about 9 routing entries (or more accurately, about 8 entries above a baseline of about 18000)? There is nothing in the BGP4 protocol to determine this value. A related question is why the ratio changed rather abruptly from 19 in 1997 to 11.8 in 1999 (and to 9 by 2001).

The BGP4/AS ratio is a measure of how effectively address prefixes in the Internet are aggregated at the level of individual ASes. In an idealized world, all ASes would originate advertisements for exactly one address prefix: all hosts reachable directly from that AS would have addresses taken

from a single address block. However, this ideal is unattainable for various reasons. In the original Internet, IP address blocks (at that time divided into Class A, B and C network numbers) were allocated chronologically on demand from a central pool. As a result, when BGP was introduced, there was no reason why the address blocks conglomerated into a single AS should have any particular relationship. In fact, the smaller blocks (originally allocated as Class C networks) became known as the "swamp" or even sometimes as "toxic waste," because they could not reasonably be aggregated in BGP4. As classless ("CIDR") addressing [6] was deployed starting in 1994, new address blocks were progressively allocated in a way designed to favour better aggregation, by assigning suitably large blocks to Internet service providers, and by tightening the criteria for assigning smaller blocks directly to user sites. Although the old swamp assignments could not be withdrawn, as the Internet grew rapidly, their relative impact could be expected to decline.

We can speculate that the observed BGP4/AS ratio of 19 in 1997 represents the residual effect of the swamp, and that by 2001, the swamp had become "background radiation," with the BGP4/AS ratio of approximately 9 representing some sort of equilibrium state resulting from the application of CIDR by the various Internet address registries. In a sense we can use this observation to measure the success of CIDR and BGP4: together, they at least halved the number of routes per AS compared with the historical situation.

An alternative interpretation is that the baseline of about 15000 BGP4 routes represents the residual swamp, and that the equilibrium state adds about 8 routes per AS.

Nevertheless, we do not have an explanation of why the ratio settled at a value of approximately 8 or 9 in recent years. This value must be the result of some combination of business and organisational factors with the CIDR policy. An additional complication is that the data [8] show that as many as 42% of ASes originate only one prefix, so the average value is very far from being a typical value. In fact the data also show that the maximum number of prefixes originated by a single AS varies widely over time (in July 2008 it was 171; a year earlier it was 557). This is clearly a heavy-tailed distribution, but one with a reasonably stable mean. This stability over periods of years is a more interesting observation than the actual value of the ratio.

The second question is why there is a roughly square law relationship between BGP4 size and the domain count. Why is the BGP4 size roughly 10 times the square root of the domain count? It is perhaps more meaningful to ask: why does the number of active ASes have a linear relationship with the square root of the domain count? Indeed, the years 1997-2000 during which the BGP4/AS ratio was rapidly decreasing are exactly the years during which the AS count was increasing towards equality with the square root of the domain count. To make this clear, the values on which Fig. 3 is based are shown in Table 2. (The square root values are shown to two decimal places, since that is how they were used in plotting the figures.)

We may speculate that the low ratio of active ASes in 1997-2000 was, like the high BGP4/AS ratio at that time, a result of the then incomplete deployment of classless address assignment. However, it is reasonable to regard the linear relationship between the AS count and the square root of the domain count as a characteristic of the BGP4 routing system throughout a decade of continuous rapid growth. Note

Table 2: AS and domain count history

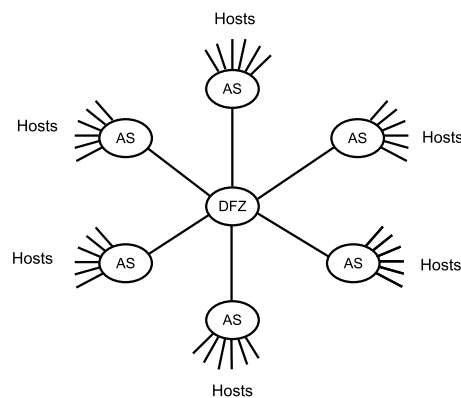| Year, Month	Unique AS count	Square root of Domain Count	Ratio
1997-07	2473	5104.20	0.48
1998-07	3695	6061.27	0.61
1999-07	5287	7497.87	0.71
2000-07	7942	9646.10	0.82
2001-07	11283	11219.99	1.01
2002-07	13283	12732.97	1.04
2003-01	14355	14230.25	1.01
2004-07	17498	16886.06	1.04
2005-07	20001	18795.86	1.06
2006-07	22569	20959.16	1.08
2007-07	25762	22130.84	1.16
2008-07	28811	23894.30	1.21

Figure 4: Idealized square law topology

that the linear fit ($A = 1.35\sqrt{D} - 4615$) shows that the rough equality in the 2001-2006 data is incidental, but nevertheless, a square law relationship is apparent.

To understand why a square law relationship might exist, consider a grossly simplified model of the topology of the Internet as a star of stars. In this model, the majority of ASes act as concentrators for hosts, and the central star is the DFZ (see Fig. 4). In fact, the data [8] show that currently, 86% of the active ASes are purely originators of routes (known as stubs [15], [16]), and 14% also provide transit, of which only about 0.35% are pure transit systems. The star of stars is therefore a reasonable, but imperfect, model of reality. It is imperfect because it ignores the meshed nature of the central part of the actual BGP4 topology and any effects of the power-law aspect of its connectivity. However, it is related to the topologies observed and discussed in [12].

With N concentration points each supporting N hosts, the total host count will be N^2. If each concentration point supports kN hosts, the total host count will be kN^2. Setting $D = kN^2$ and $A = N$, the above linear fit becomes $N = 1.35\sqrt{kN^2} - 4615$. For large N, this gives $k = 0.55$. (Since D as measured is a lower bound, this value is also a lower

bound.) Thus the apparent relationship between host count and AS count suggests that during the years in question, the growth of the public Internet has been absorbed both by growth in the number of ASes, and by proportionally lower growth in the number of hosts reachable via each AS.

Another observation from the data in Table 1 is that the lower bound on the average number of hosts covered by each BGP4 route increased from 209 in July 1994, to 1221 in July 2001, and to 2084 in July 2008. This clearly represents good scaling. However, it contrasts with the trend in the average address span covered by each BGP4 route, i.e., the number of addresses theoretically reachable through the prefix for which the route is advertised. According to the data available at [8], this average span was 10849 in July 2001 (earlier data are not available) and 6843 in July 2008. In other words, during a period in which the real number of hosts reachable via each BGP4 route appears to have increased by a factor 1.8, the theoretical address span of each route has decreased by a factor 1.6. On the one hand, prefixes now tend to be more "filled up" with active hosts; on the other hand, shorter prefixes are being advertised in BGP4. Stated otherwise, the effective utilisation of routed address space has increased from 11% in 2001 to 33% in 2008 (both of these values being lower bounds).

For the routing system to continue to scale at a reasonable rate compared to technological progress in router design, it is highly desirable that the gradients of all three curves plotted here do not increase. There is no scientific basis for extrapolating these entirely observational data sets, but the most recent data do suggest a slight increase in gradient. We can speculate that this is largely due to increased pressure from customers for multihoming support and load balancing, which cause deaggregation of BGP4 announcements [1]. However, more time and more data will be necessary to confirm or deny this change in gradient and its cause.

Ideally we would derive some practical prescriptions for containing the scaling problem of BGP4 from this discussion. As mentioned, there is considerable concern about this problem in the Internet technical community [14]. However, there is little practical value in advising Internet service providers to maximise address space utilisation, to improve BGP4 aggregation, or to minimise the number of prefixes originated by each AS. These are already well known goals, which are often in conflict with commercial constraints on the service providers. We can now additionally suggest that, to the extent that the star of stars model is accurate, it has been responsible for limiting the scale of BGP4 roughly in proportion to the square root of the total size of the public Internet. This is a markedly sub-linear and desirable scaling property. Thus we can conclude that it is highly desirable to maintain the current situation, in which the large majority (86%) of ASes are origin-only , so that the transit ASes may continue to function in the gross topology similarly to a central star.

5.   RELATED WORK

Perhaps the first important review of BGP4 growth was [9], at a time when rapid, apparently exponential, growth in the BGP4 and AS counts was a growing concern, with the focus of attention being on the factors causing the lengthening of the average address prefix advertised in BGP4, i.e., deaggregation. Papers by Oliveira *et al* [15], [16] dig into the AS topology in detail.

Xu *et al* [22] and Meng *et al* [13] studied details of the relationship between IP address allocation policy and BGP4 table size. Their conclusions indicate that BGP4 growth during 1998 to 2004 was largely driven by fragmentation (deaggregation) of recently allocated address blocks.

Bu *et al* [1] studied BGP4 growth with respect to the causes of fragmentation, concluding that multihoming and load balancing contribute the most. They assert that the increase in routeable IP addresses contributes "little" to routing table growth, whereas this paper suggests more precisely that the resultant growth is related to the square root of the number of routeable addresses.

Li *et al* [12] noted that many ISPs adopt a design they describe as *heuristically optimal* with a meshed core and a hierarchical edge, which is what this paper suggests in fact describes the Internet as a whole, approximated as a star of stars.

Other important efforts to study the topology are also reported, such as [4], [3], [7], and [20], but these do not provide long historical data series.

An immediate impetus for this paper was provided by the Internet Architecture Board's Routing and Addressing Workshop [14] and the Internet Research Task Force's Routing Research Group [10], both of which provide links to many works in progress.

6.   CONCLUSION

The study of historical observations over more than a decade has shown that the data on BGP4 table size and the number of active ASes are roughly consistent with a linear relationship between the two. Additionally, the data on the DNS domain count, taken as a lower bound on the number of active public IP addresses, are consistent with a square law relationship with both the BGP4 and AS counts. In particular, the active AS count has a strikingly linear relationship with the square root of the domain count. While these observations do not have predictive value, they span fourteen years and two orders of magnitude of growth of the globally addressable Internet, and can be used as ongoing measures of the scaling of the Internet routing system.

Furthermore, we may conclude that in addition to the well-known goals of maximising address utilisation and BGP4 aggregation, and minimising the number of prefixes originated by each AS, it is also highly desirable that the large majority of ASes should continue to be origin-only systems providing no transit. This will tend to maintain the observed $\sqrt{D}$ scaling of the BGP4 system.

7.   ACKNOWLEDGMENTS

8. REFERENCES

[1] T. Bu, L. Gao, and D. Towsley. On characterizing BGP routing table growth. *Computer Networks*, 45(1):45–54, 2004.

[2] M. Caesar and J. Rexford. BGP routing policies in ISP networks. *IEEE Network*, 19(6):5–11, Nov.-Dec. 2005.

[3] S. Carmi, S. Havlin, S. Kirkpatrick, Y. Shavitt, and E. Shir. A model of Internet topology using k-shell decomposition. *Proc Natl Acad Sci U S A*, June 2007.

[4] DIMES project. http://www.netdimes.org/.

[5] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On Power-Law Relationships of the Internet Topology. In *Proceedings of the ACM SIGCOMM 1999 Conference*, pages 251–262, August 1999.

[6] V. Fuller, T. Li, J. Yu, and K. Varadhan. Classless Inter-Domain Routing (CIDR): an Address Assignment and Aggregation Strategy. *Internet RFC 1519, ISSN 2070-1721*, 1993.

[7] Y. He, G. Siganos, M. Faloutsos, and S. V. Krishnamurthy. A Systematic Framework for Unearthing the Missing Links: Measurements and Impact. In *NSDI*. USENIX, 2007.

[8] G. Huston. Web site http://www.potaroo.net/.

[9] G. Huston. Analyzing the Internet BGP Routing Table. *The Internet Protocol Journal*, 4(1):2–15, 2002.

[10] Internet Research Task Force. Routing Research Group, http://www.irtf.org, 2008.

[11] Internet Systems Consortium. New Domain Survey, https://www.isc.org/solutions/survey.

[12] L. Li, D. Alderson, W. Willinger, and J. Doyle. A FirstPrinciples Approach to Understanding the Internet's Router level Topology. In *Proceedings of the ACM SIGCOMM 2004 Conference*, pages 3–14, August 2004.

[13] X. Meng, Z. Xu, B. Zhang, G. Huston, S. Lu, and L. Zhang. IPv4 address allocation and the BGP routing table evolution. *ACM SIGCOMM Computer Communication Review.*, 35(1):71–80, 2004.

[14] D. Meyer, L. Zhang, and K. Fall. Report from the IAB Workshop on Routing and Addressing. *Internet RFC 4984, ISSN 2070-1721*, 2007.

[15] R. Oliveira, D. Pei, W. Willinger, B. Zhang, and L. Zhang. Observing the Evolution of Internet AS Topology. In *Proceedings of the ACM SIGCOMM 2007 Conference*, pages 313–324, August 2007.

[16] R. Oliveira, B. Zhang, and L. Zhang. In Search of the elusive Ground Truth: The Internet's AS-level Connectivity Structure. In *SIGMETRICS'08*, pages 217–228, June 2008.

[17] Y. Rekhter and T. Li. A border gateway protocol 4 (BGP-4). *Internet RFC 1654, ISSN 2070-1721*, 1995.

[18] RIPE Network Coordination Centre. http://www.ripe.net/.

[19] P. Sherbin and G. Huston. RRG emails http://www.ops.ietf.org/lists/rrg/2007/msg00246.html, August 2007.

[20] G. Siganos and M. Faloutsos. Analyzing BGP Policies: Methodology and Tool. In *INFOCOM*, 2004.

[21] University of Oregon Route Views project. http://www.routeviews.org/.

[22] Z. Xu, X. Meng, L. Zhang, S. Lu, and C. J. Wittbrodt. Impact of IPv4 Address Allocation Practice on BGP Routing Table Growth. In *Computer Communications, CCW 2003*, pages 172–178, 2003.