Projecting Future IPv4 Router Requirements from Trends in Dynamic BGP Behaviour

Geoff Huston

Centre for Advanced Internet Architectures
Swinburne University of Technology
Melbourne, Australia
gih@swin.edu.au

Grenville Armitage

Centre for Advanced Internet Architectures
Swinburne University of Technology
Melbourne, Australia
garmitage@swin.edu.au

Abstract—**For the past 17 years the Internet has used the Border Gateway Protocol (BGP) to manage inter-domain routing. The dynamic behaviour of BGP in operational networks has rarely been studied to date. Yet the processing load of BGP is of intense interest to router vendors and internet service providers, as the stability of the Internet depends on routers being able to keep up with BGP state changes. We analyse 146 million BGP update messages logged over the entire year of 2005 to discover evidence of an aggressively increasing demand for processing power in route engines out to 2010.**

Keywords - routing, border gateway protocol, BGP, prefix updates, prefix withdrawals, dynamic behaviour

I. INTRODUCTION

In recent years there have been concerns expressed about the ability of the Internet's inter-domain routing protocol to continue to cope with growth in demand for Internet service [1]. In 2001 there were predictions that routing requirements would begin to exceed router capabilities sometime between 2003 and 2005 [2]. While this has not eventuated, concerns about the scaling properties of the inter-domain routing environment persist. In this paper we take a closer look at the dynamic behaviour of the Internet's current inter-domain routing protocol to better understand the current and near-future demands that are being placed on routing hardware

Today's inter-domain routing is handled by the border gateway protocol version 4 (BGP-4), first deployed in 1994 [3] [4]. BGP retains much of the same architecture it had when version 1 was first documented 1989 [5]. BGP remains a classic distance vector protocol, using an explicitly enumerated path vector as a combined path metric and loop detector. The impending introduction of 32-bit autonomous system (AS) numbers to BGP (up from the current 16-bit numbers) could be argued as one of the larger forthcoming changes to the BGP protocol since the introduction of explicit address prefix masks (Classless Inter-Domain Routing, or "CIDR") in 1994 [6], and even this change is a relatively minor change to the protocol. (Extensions for MPLS signalling with BGP [7] are somewhat tangential to the inter-domain routing use of BGP.)

While the Internet historically to take some comfort in its ability to perform feats of rapid deployment of innovative technologies up and down the protocol stack to address various forms of growing pains, these days the lower layers of the protocol stack are accreting significant levels of inertia, and it's the upper levels of the stack that are left to carry the innovation burden. Routing is, perhaps unfortunately, an inhabitant of one of these lower levels of the protocol stack, while much of the innovative agenda is taking place at the application level. Consequently, we are unlikely to see much change occurring in the basic design (and implementation) of BGP in coming years.

Given that BGP itself will be with us for at least the next 3 to 5 years, it is well worth identifying the demands BGP is likely to place on routers (and hence router vendors and internet service providers) over this period. The work load of running BGP is not simply confined to storing large routing tables of prefixes that summarize the span of reachable addresses in the Internet. BGP-speaking routers must handle the second-by-second incremental fluctuations in the routing tables as small and large ISPs around the planet announce and withdraw routing information to the rest of the Internet. In this paper we analyse 146 million BGP update messages logged over the entire year of 2005 to discover evidence of an aggressively increasing demand for processing power in route engines out to 2010.

The rest of the paper is structured as follows. Section II provides background on BGP itself, the dynamic routing update process, and a summary of aggregate statistics from 2005. Section III explores the dynamic message statistics in more detail, while section IV discusses our predictions out to 2010. Section V concludes the paper.

II. BACKGROUND

A. The Border Gateway Protocol (BGP)

BGP is a distance vector protocol, using an explicitly enumerated path vector as a combined path metric and loop detector, as distinct from a link-state routing protocol or a map-based routing protocol. BGP is a distributed computation that uses address prefixes as its basic unit of routing. Each BGP speaker maintains a set of tables (Routing Information Bases, or RIBs) - one for each BGP-speaking neighbour and one for its own internal use for forwarding. BGP keeps a copy of all prefixes and associated routes that have been advertised by its peers (Adjacency-RIB-IN). It selects the "best" of these routes to use for its local forwarding decisions (Local-RIB), and sends

a copy of this "best" route to all its peers (Adjacency-RIB-OUT). Like any distance-vector routing protocol, BGP operates as a loosely synchronized distributed computation based on partial information forwarding.

A BGP peer session uses TCP as its reliable transport protocol. The reliable data transfer between BGP speakers implies that periodic re-flooding of the route tables, as used by the interior routing protocol RIP2, for example, is not required by BGP. BGP is a far more parsimonious protocol where once a BGP session has been set up and the initial route set is exchanged, then the subsequent protocol traffic is limited to notification of a prefix that is no longer reachable, or when the characteristics of the local "best" route have changed and the local BGP instance wants to inform its neighbouring peers, or the appearance of a new prefix. This information is passed in a BGP update message. This protocol message contains a collection of route attributes, and a list of prefixes that share this attribute set (announcements) and a set of prefixes that are no longer reachable (withdrawals).

If the entire network is perfectly stable, with no changes of any form, then BGP would be a very quiet protocol, with only the intermittent (30 second by default) exchange of keepalive messages to indicate any activity at all. On the other hand, a large dynamic network where prefixes are appearing and disappearing, and where paths are created and lost, such as in the Internet, is capable of generating a relatively impressive set of updates in very small time intervals.

B. Dynamic behaviour – the impact of BGP state changes

Each received update represents work to be undertaken. The incoming update message causes a change in the Adjacency-RIB-IN. If the information is a prefix withdrawal, then a comparison needs to be made with the local-RIB. If there is a match, the this implies that the current "best" route has been removed. In this case all other Adjacency-RIB-INs need to be scanned and a new "best" route installed into the local-RIB, as well as loading new pending announcement messages in the Adjacency-RIB-OUTs to reflect this local change of best path. These pending messages are then sent to the BGP-speaking peers. If there are no other candidate routes in the other Adcacency-RIB-IN's then the route is withdrawn from the local-RIB and a withdrawal message is passed to the BGP-speaking peers. If the incoming update message is an announcement, then the BGP engine has to update the Adjacency-RIB-IN and then compare this route to the current best path in the Local-RIB. If this new route represents a "better" path, then the Local-RIB is updated and announcement messages are queued in all the Adjacency-RIB-OUTs, and new update messages are passed to the peers.

In terms of protocol workload and routing stability it is not fundamentally the size of the BGP routing table that is the critical scaling and stability issue - it's the dynamic characteristics of BGP update messages. The longer the delay in processing update messages the longer the time for the entire system to converge upon a stable routing state that reflects optimised paths across the inter-domain space, and the larger the number of intermediate messages that are generated during this process of convergence, which in turn compounds the problem of increasing processing loads. At the extreme case the local BGP engine will exhaust its incoming BGP message buffer and fail to process updates. At this stage there is the potential for inconsistent information to be embedded in the routing system, leading to loops and black holes in the routing system. This is the point at while the routing could be said to have "collapsed".

Looking at the BGP update rate, and in particular the relative rates of growth of the BGP routing table as compared to the rates of growth of update messages, and updated prefixes can give us a helpful indicator of the pressures for growth in the routing system, and also an indicator of what size router we'll need to use to cover the Internet's routing system in the coming years.

C. Routing statistics from 2005

Table 1 summarises the IPv4 Internet's vital statistics during 2005. These are derived from a stream of one-hourly 'snapshots' of the routing table taken from the boundary of AS1221 (Telstra Pty Ltd).

Table 1 Summary of IPv4 BGP Data over 2005

Prefixes	148,000 - 175,400	+18%	+26,900
Prefix Roots	72,600 - 85,500	+18%	+12,900
More Specifics	77,200 - 88,900	+18%	+14,000
Addresses	80.6 - 88.9 (/8s)	+10%	+8.3 /8s
ASNs	18,600 - 21,300	+14%	2,600

Table 1 indicates that the use of aggregates in the routing system has not improved. The average size of advertisements is getting smaller in terms of address span per routing table entry, the span of originating addresses per AS is getting smaller, the average AS path length is constant at around 3.5 AS hops, the number of AS's is increasing, and the interconnection degree of AS's is getting higher. The implication is that the granularity of the inter-domain routing system continues to get finer and the density of interconnection is getting greater. For a distance vector protocol such as BGP is not heartening news.

These IPv4 trends for 2005 are a source of some concern. How big can the Internet grow in the coming years? Will we continue to be able to deploy routers in the default-free routing zone of the Internet that can comfortably route the Internet? Can we add additional functionality into the routing system and still stay within comfortable limits of the capability of the routing system and the routers? What router capacities are required to support the Internet for the next 3 to 5 years? Answering such questions requires a more detailed examination of BGP behaviour over the year.

III. DYNAMIC BGP ACTIVITY DURING 2005

A BGP measurement point was set up inside AS1221, and all BGP protocol messages ("updates") passed within that network were time-stamped and logged. Internal routing changes were eliminated from the logs leaving roughly 146 million exterior IPv4 BGP updates for analysis. Our aim is to

identify trend data from the assembled 2005 update logs and make some predictions about overall BGP capacity requirements in the coming years.

A. Update messages per Day

Figure 1. shows that the number of update messages appears to have almost doubled for 2005, growing from ~260K per day at the start of 2005 to ~550K per day by the end of the year. Considering that even by the end of the year there were 170K prefixes in the global routing table, to have this routing population generate 550K updates messages per day is an impressive achievement. This growth rate vastly exceeds growth rate in the routing table size. Either the Internet is far less stable than we'd like to believe, or some other factor is driving up the BGP update rate. The increasing density of interconnection in the inter-domain space may be relevant to this very high growth rate.
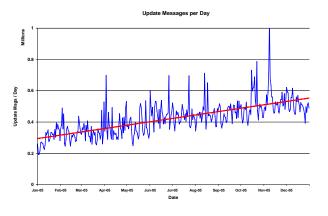

Figure 1.   BGP Update messages per Day

It also appears that BGP has 'good' days and 'bad' days. For example, a single day in November recorded 1 million update messages. This high level of variation indicates a degree of instability in the Internet that is not normally evident at the user level where most users tend to see a relatively stable and reliable Internet service, at least from a routing perspective.

B. Prefixes per Update Message

During 2005 the daily update rate doubled while the size of the routing table itself grew by only 18%. Each BGP update message contains a number of prefixes, so it is reasonable to ask whether the number of prefixes in each update message is increasing or decreasing on average. Figure 2. shows the daily average number of prefixes per update message over 2005.

On average there were between 8.1 and 8.3 prefixes per originating AS across 2005. If prefixes are managed such that each AS has a single coherent routing policy we would expect to see a relatively consistent number of prefixes in each BGP update message. Figure 2. suggests this is not the case, revealing quite high levels of daily variation. In addition, a least squares best fit indicates an downward trend from 2.4 prefixes per update message at the start of the year to 2.3 prefixes per update message at the end of the year. (The high 'spikes' of this measure on some individual days indicates some

form of BGP session resets, where a number of peering sessions may have been reset on a day and the resultant reconstruction of the BGP peering session would normally use dense packing of a large number of prefixes in each update message.)

Averaging a little over 2 prefixes per update message appears to indicate a use of fine-grained routing policies at a level finer than an AS. It would appear that the 'unit' of a BGP routing policy is more fine-grained than an AS, and is now heading towards the level of each advertised prefix having individual routing policies and individual attributes. This implies that the efforts of BGP to compress the update load by grouping prefixes into bundles is no longer as effective as it may have been in the past as a measure of assisting in making BGP an efficient routing protocol.
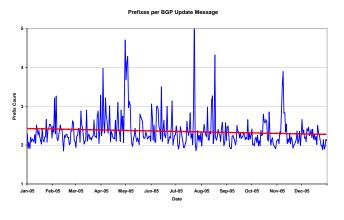

Figure 2.   Average number of Prefixes per Update Message

The evidence suggests we should look more closely at the update and withdrawal rates of individual prefixes, rather than looking at the level of BGP protocol update messages.

C. Prefix Update and Withdrawal Rates

Figure 3. shows the number of prefixes updated and withdrawn per day on separate lines. Again a high level of daily variation is visible, but this time with clear differentiation between full BGP session resets without backup paths (high withdrawal and update counts) and BGP re-routing (high update count without a corresponding high withdrawal count)
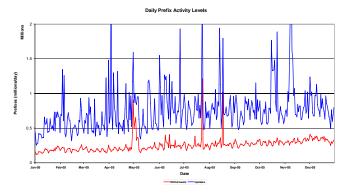

Figure 3.   Daily average prefix count of updates and withdrawals

Figure 4. shows prefix updates per day with an exponential curve best fit trend line. The overall growth trend ranges from 570K updates per day to some 850K updates per day over the year. That is a very high growth rate in the context of the Internet's routing table having only 170K unique prefixes by the end of 2005. Some prefixes are evidently generating a disproportionate number of daily updates.
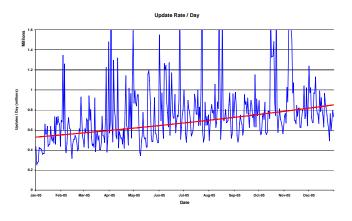

Figure 4.   Prefix Update Counts

Figure 5. shows prefix withdrawals per day with an exponential curve best fit trend line. The withdrawal count grows from 160K per day to some 340K per day by the end of the year
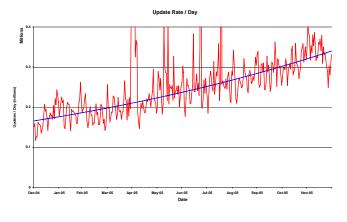

Figure 5.   Prefix Withdrawal Counts

D.   Extrapolating from 2002 – 2006 BGP Table growth

The next question is to relate these prefix update and withdrawal rates against the BGP table size, and look at the likely trends of the load of the BGP protocol in terms of prefix update and withdrawal rates against the trend of the projections of growth of the BGP table itself. The BGP table size over the period from 2002 until the start of 2006 is shown in Figure 6.

In Figure 6. the raw data of hourly snapshots (the blue line) has been smoothed as part of the first step in generating a trend projection. The next step is to take the first order differential of the smoothed data series (Figure 7.). The linear approximation of the first order differential can be fitted to a trend of an O(2) polynomial trend in the BGP table size. This allows a trend

projection in the BGP table over the next 3 - 5 years using this O(2) polynomial (Figure 8.).

If current trends in BGP continue for the next 3 - 5 years then this model predicts the BGP routing table will grown from ~ 176K entries at the end of 2005 to 275K entries at the end of 2008 and some 370K prefixes by the end of 2010.
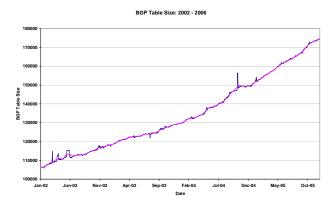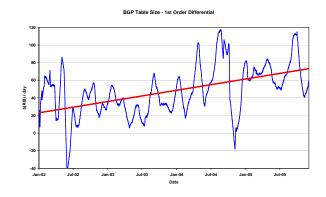

Figure 6.   BGP Prefix Table Size

Figure 7.   First order differential of BGP Table Size
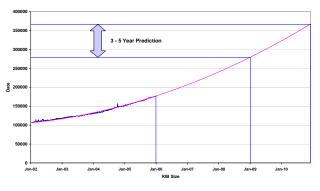

Figure 8.   BGP Table Size Projection

IV.   PREDICTED UPDATE AND WITHDRAWAL RATES TO 2010

It is possible to use this predictive model to also forecast the amount of BGP update activity. As our starting point we use the trend of the number of prefix updates and withdrawals

per BGP routing table entry across 2005 (Figure 9.). These trend lines can then be applied to the BGP projection model (Figure 10.).

The projections of BGP activity from this model indicate a growth rate of some 1.7 million prefix updates per day by the end of 2008 and 2.8 million prefix updates per day by the end of 20010. A similar growth trend is forecast for prefix withdrawal rates, to 0.9 million withdrawals per day by the end of 2008 and 1.6 million withdrawals by the end of 2010. This implies a CPU processing load that will increase by a factor of 4 over this 3 to 5 year period. Table 2 summarises the projections.
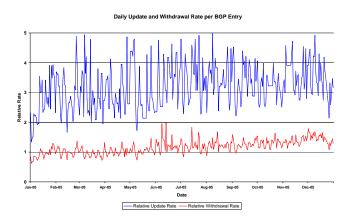

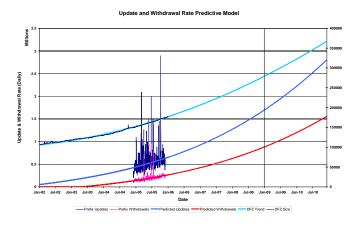Figure 9. Prefix Update and Withdrawal Rates per BGP Table Entry

Figure 10. Prefix Update Rate Projection

Table 2 BGP activity projections to 2010

Date	BGP Table Size	Daily Prefix Updates	Daily Prefix Withdrawals
End 2005	176,000	700,000	400,000
End 2008	275,000	1,700,000	900,000
End 2010	370,000	2,800,000	1,600,000

V.   CONCLUSION

There are many factors that impact on the growth in demand for processing and storage capacity of BGP speaking routers. Nevertheless it is evident there are some accelerating factors within BGP suggesting that the 'load' of BGP, in terms of processing update messages and in terms of processor cycles (update-related processing) is growing faster than the memory requirements and the forwarding decision structure (table size-related aspects). Router engine processing capacity will need to grow substantially to cope with the projected BGP load over the coming 3 to 5 years.

Finer levels of granularity of routing information in the routing system, denser levels of interconnectivity in the network, and greater levels of policy discrimination in the routing system are all evident. These factors are combining to create a system increasingly sensitive to perturbation and increasingly challenged to discover and stabilise on new converged state following each dynamic change. These BGP 'load' factors appear to be growing far faster than the number of advertised prefixes in the BGP Routing Table. In addition, the level of routing overhead (updates and withdrawals) appears to grow faster than the routing system itself.

The ratio of peak capacity to average capacity in the routing system is also a significant issue. BGP is a very chaotic system in terms of burstiness of traffic, and the peak per-second rate of BGP updates can be some 1,000 times greater than the daily average. Consequently, BGP routers must handle very short term peak loads well (rather than extended average loads) to preserve acceptable convergence in the routing system.

Future work is required to evaluate how the routing system may cope with adding additional functionality, such as the additional processing required to improve the overall security in BGP through the attachment of authenticable attributes of BGP updates, or the addition of further policy-based functions to direct route propagation that increase the workload required per received BGP update.

REFERENCES

[1] G. Huston, "Commentary on Inter-Domain Routing in the Internet," RFC 3221, Internet Architecture Board, December 2001

[2] G. Huston, "The State of BGP Routing," IETF Plenary presentation, March 2001 (online at http://www3.ietf.org/proceedings/01mar/slides/plenary-2/)

[3] Y. Rekhter, T. Li, "A Border Gateway Protocol 4 (BGP-4)," RFC 1771, Internet Engineering Task Force, March 1995

[4] Y. Rekhter, T.Li, S. Hares, "A Border Gateway Protocol 4 (BGP-4)," RFC 4271, Internet Engineering Task Force, January 2006

[5] K. Lougheed, Y. Rekhter, "A Border Gateway Protocol (BGP)," RFC 1105, Internet Engineering Task Force, June 1989

[6] V. Fuller, T. Li, J. Yu, K. Varadhan, "Classless Inter-Domain Routing (CIDR): an Address Assignment and Aggregation Strategy," RFC 1519, Internet Engineering Task Force, September 1993

[7] E. Rosen, Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)," RFC 4364, Internet Engineering Task Force, February 2006