Revised Default Values for the BGP 'Minimum Route Advertisement Interval'
draft-jakma-mrai-02

Status of this Memo

Abstract

This document briefly examines what is known about the effects of the BGP MRAI timer, particularly on convergence. It highlights published work which suggests the MRAI interval as deployed has an adverse effect on the convergence time of BGP.

It then recommends revised, lower default values for the MRAI timer, thought to be more suited to today's Internet environment.

Table of Contents

1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

2. Background

The proper functioning of the [BGP] routing protocol is of great importance to the Internet. Issues regarding matters of its stability and convergence have been documented widely, such as in [BGP-STAB], [bgp-converge] and [Potaroo0607].

One such issue is the effect of 'Minimum Route Advertisement Interval' (MRAI).

2.1. The MRAI Timer

The Minimum Route Advertisement Interval (MRAI) timer is specified in RFC4271 [BGP]. This timer acts to rate-limit updates, on a per-destination basis. [BGP] suggests values of 30s and 5s for this interval for eBGP and iBGP respectively. The MRAI must also be applied to withdrawals according to RFC4271 [BGP], a change from the earlier RFC1771.

Some implementations apply this rate-limiting on a per-peer basis, presumably an adequate approximation. Some implementations apply it to withdrawal methods (often called "WRATE" in the literature). Some implementations do not apply MRAI at all.

2.2. Known effects of the MRAI timer on convergence

The MRAI timer serves to suppress messages which BGP would otherwise send out to describe transitory states, and so allow BGP to converge with significantly fewer messages sent. This beneficial effect of the MRAI timer, in terms of # of messages, increases as the timer is increased until an optimum value is reached, after which the beneficial effect stabilises. [bgp-converge] [mrai-final]

In terms of convergence time, a similar beneficial effect is seen as the MRAI increases to near the same optimum value. However as the timer value is increased past this point, the convergence time increases again linearly. The scale of this increase is significantly worse with WRATE, i.e. applying the MRAI to withdrawals has an adverse effect on BGP convergence time. [bgp-converge] [mrai-final]

The optimum MRAI timer value is dependent on several factors, most particularly the topology in its layout and propagation times. The optimum value will differ between different subsets of the Internet. [mrai-final]

It is believed to be infeasible to try directly calculate this value. However a useful approximation can be made from the diameter of the topology if it is known, along with some assumptions about the the topology, such as the latency between nodes. [mrai-internet]

The interaction between extensions to BGP designed to improve convergence, such as those that allow propagation of additional and/or backup paths, and the MRAI timer is as yet unknown. However, it seems reasonable to speculate these extensions might have the effect of leading to a lower optimum MRAI than would be indicated by an approximation based on the diameter of the BGP topology. Further work on these questions would be useful.

2.3. Interaction with Flap-Damping

As the MRAI helps eliminate some updates, it interacts with flap-damping [BGP-DAMP]. The lower the MRAI timer, the greater the risk of crossing below the threshold of the optimum value. If that threshold is crossed, there will be an increased number of updates somewhere within the BGP system, and hence an increased risk of paths being dampened which otherwise would not.

So, in presence of significant flap-damping deployment and given the uncertainty of what the optimum is, it is reasonable to err towards selecting a value of the MRAI timer significantly higher than the optimum.

However, given that flap-damping increasingly is discouraged [RIPE-378] in Internet routing, this particular need to be conservative in the choice of MRAI timer value may be less important.

2.4. Current Status of the MRAI

The current recommended value of 30s may be far higher than is optimal for the Internet, based on observations of certain parameters related to its topology. In [mrai-final] it is suggested that the optimal value may be between 5s ('semi-safe') to 15s ('safe'). The estimation of the 'safe' value here is of no relevance if WRATE is universally deployed, as in such a case the 'semi-safe' value will be the same as the 'safe' value. Further empirical work by the same authors [mrai-internet] suggests that the optimal, Internet MRAI may be below 5s.

Further, [BGP-STAB] and [Potaroo0607] argue that operational conditions (e.g. different routers using different MRAI values) mean the MRAI is having an adverse effect even on the number of messages sent, and so further exacerbating convergence problems in the global BGP system, such as path hunting. The [BGP-STAB] document goes further still and argues that MRAI be deprecated in favour of some better way of damping BGP UPDATES, however there are no clear proposals before the IDR as of this writing for such changes to BGP.

3. Risk Evaluation in the Choice of MRAI Time

Though there is an optimum value for the MRAI, it's unlikely that it can be determined empirically or otherwise for the general Internet. It may even not be possible, as the optimum MRAI will differ for different subsets of the Internet. Some degree of guesstimation at a reasonable value for the MRAI is required, which is an exercise in risk; whether to err towards fast convergence at the risk of a disproportionate increase in BGP messaging, or to err to the side of an optimal number of messages at the expense of convergence.

Arguably, economising on bandwidth and control-plane processing power is of less importance than the convergence time of BGP, compared to times past. Presuming this, any new recommendations for the MRAI should seek to err slightly to the side of convergence, rather than erring towards minimising BGP traffic.

Further, if we assume most implementations apply the MRAI to withdrawals, then the Internet BGP topology effectively is WRATE-enabled, and [mrai-final] suggests there is even less benefit to erring toward a higher MRAI.

There may be risks in mismatched MRAI values between speakers in an AS as revised MRAI values are deployed. The MRAI values in RFC4271 [BGP] were deliberately specified to be lower for iBGP than for eBGP, so as to allow interior routing to converge while minimising the effect on eBGP state. So with mismatched values there is an increased risk that the external stability of an AS's routes would be affected by transient, internal states.

This last risk suggests that the existing iBGP/VPN values should be the lower-bound for any conservative revision of the eBGP MRAI value.

The most definite risk of lowering the MRAI is the increased risk of flap-damping, if the value is set too much below the optimum. Therefore, taking into account estimations of that optimum is required. That said, at least one BGP implementation by default does not apply any MRAI at all.

4. Recommendations on the MRAI

The suggested default values for the MinRouteAdvertisementIntervalTimer given in RFC4271 [BGP] are revised to be 5s or less for eBGP connections, and 1s or less for iBGP connections, for use on Internet topologies.

These values may not be suitable for topologies which differ from the Internet, be that in scale, arrangement or otherwise. Such non-Internet, BGP topologies likely would have lower optimum values, assuming they are always significantly smaller in scale than the Internet BGP topology. Hence, implementations SHOULD allow the MRAI value to be configured administratively on a per-AFI/SAFI basis, as well as a per-peer basis.

Given the beneficial effects on convergence time, implementations MAY exempt withdrawals from the MRAI timer.

5. IANA Considerations

There are no requests made to IANA in this document.

6. Security Considerations

This document raises no new security considerations.

7. Acknowledgements

The author would like to thank Manav Bhatia for his helpful review and comments; as well as Robert Raszuk, Samita Chakrabarti, Danny McPherson and Jeffrey Haas for their useful comments; dissenting or otherwise.

The authors of the cited documents are thanked for their contributions to the understanding of BGP, of which this document is a simple summary.

8. References

8.1. Normative References

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

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", RFC 2119, BCP 14, February 2001.

8.2. Informative References

[BGP-STAB]
Li, T. and G. Huston, "BGP Stability Improvements", I-D draft-li-bgp-stability, June 2007.

[BGP-DAMP]
Villamizar, C., Chandra, R., and R. Govindan, "BGP Route Flap Damping", RFC 2439, November 1998.

[Potaroo0607]
Huston, G., "Damping BGP", June 2007, <http://www.potaroo.net/ispcol/2007-06/dampbgp.html>.

[RIPE-378]
Smith, P. and P. Panigl, "RIPE RRG: Recommendations on Route-flap Damping", May 2006, <http://www.ripe.net/docs/ripe-378.html>.

[bgp-converge]
Griffin, T. and B. Premore, "An Experimental Analysis of BGP Convergence Time", In Proceedings of ICNP pages 53-61, November 2001, <http://www.ssfnet.org/Papers/icnp-2001.pdf>.

[mrai-final]
Qiu, J., Hao, R., and X. Li, "An Experimental Study of the BGP Rate-limiting Timer", Bell Labs Technical Memo ITD-03-44604H, June 2003, <http://www.net-glyph.org/~qiu/public_html/mrai_final.pdf>.

[mrai-internet]
Qiu, J., Hao, R., and X. Li, "The Optimal Rate-Limiting Timer of BGP for Routing Convergence", IEICE TRANS. Comm. Vol.E88-B, No. 4, April 2005, <http://rio.ecs.umass.edu/~jqiu/mrai_journal.pdf>.

Author's Address

Paul Jakma
Sun Microsystems
Springfield
Linlithgow, West Lothian EH49 7LR
Scotland

Phone: +44 1506 673150
Email: paul.jakma@sun.com