RIPE Routing-WG Recommendation for Coordinated Route-flap Damping Parameters

Tony Barber
Sean Doran
Daniel Karrenberg
Christian Panigl
Joachim Schmitz

Document ID: ripe-210
Version 1.1
Date: 12 May 2000

Obsoletes: ripe-178
Obsoleted by: ripe-229

Abstract

This paper recommends a set of route-flap damping parameters which should be applied by all ISPs in the Internet and deployed as default values by BGP router vendors.

Table of Contents

1. Introduction

Route-flap damping is a mechanism for (BGP) routers which is aimed at improving the overall stability of the Internet routing table and offloading core-routers CPUs.

In the Routing WG session of RIPE26 Christian Panigl asked whether people are interested to participate in a BOF on route flap damping. The BOF session was held after the plenary session of RIPE26.

The discussion was continued in the Routing WG session of RIPE27 and led to a task-force directed to write a proposal document for coordinated route-flap damping parameters.

1.1 Motivation for route-flap damping

In the early 1990s the massive growth of the Internet with regard to the number of announced prefixes (often due to inadequate prefix-aggregation), multiple paths and instabilities started to do significant harm to the efficiency of the core routers of the Internet. Every single line-flap at the periphery which makes a routing prefix unreachable has to be advertised to the whole core Internet and has to be dealt by every single router by means of updates of the routing-table.

To overcome this situation a route-flap damping mechanism was invented in 1993 and has been integrated into several router code since 1995 (Cisco, ISI/RSd, GateD Consortium). It significantly helps now with keeping severe instabilities more local.

And there’s a second benefit: it’s raising the awareness of the existence of instabilities because severe route/line-flapping problems lead to permanent suppression of the unstable area by means of holding down the flapping prefixes.

Route-flap damping is at its best value and most consistent and helpful if applied as near to the source of the problem as possible. Therefore flap-damping should not only be applied at peering and upstream boundaries but even more at customer boundaries (see 1.4 and 1.5 for details).

1.2 What is route-flap damping ?

When BGP route-flap damping is enabled in a router, the router starts to collect statistics about the announcement and withdrawal of prefixes. Route-flap damping is governed by a set of parameters with vendor-supplied default values which may be modified by the router manager. The names, semantic and syntax of these parameters differ between the various implementations. However, the behavior of the damping mechanism is basically the same:

If a threshold of the number of pairs of withdrawals/announcements (=flap) is exceeded in a given time frame (cutoff threshold) the prefix is held down for a calculated period (penalty) which is further incremented with every subsequent flap. The penalty is then decremented by using a half-life parameter until the penalty is below a reuse threshold. Therefore, after being stable up for a certain period the hold-down is released from the prefix and it is reused and re-advertised.

Pointers to some more detailed and vendor specific documents:

- Cisco BGP Case Studies: Route Flap Damping at:
  http://www.cisco.com/warp/public/459/16.html
- ISI/RSd Configuration: Route Flap Damping at:
  http://www.isi.edu/div7/ra/RSd/doc/dampen.html
- GateD Configuration: Weighted Route Damping Statement at:
  http://www.gated.org/gated-web/code/doc/manuals/config_guide/bgp/weighted_route_dampening

See also "4. References"

1.3 "Progressive" versus "flat&gentle" approach

One easy approach would be to just apply the current default-parameters which are treating all prefixes equally ("flat&gentle") everywhere, however, there is a major concern to penalise longer prefixes (=smaller aggregates) more than well aggregated short prefixes ("progressive"), because the number of short prefixes in the routing table is significantly lower and it seems in general that those are tending to be more stable and also are tending to effect more users.

Another aspect is that progressive damping might increase the awareness of aggregation needs, however, it has to be accompanied by a careful design which does not force a rush to request and assign more address space than needed.

Because a significant number of important services is sitting in long prefixes (e.g. root name servers) the progressive approach has to exclude the strong penalisation for those long but "golden" prefixes.

With this recommendation we are trying to make a compromise and call it therefore "graded damping".

1.4 Motivation for coordinated parameters

There is a strong need for the coordinated use of damping parameters because of several reasons:

Coordination of "progressiveness":

Penalties are not coordinated throughout the Internet, route-flap damping could even lead to additional flapping or inconsistent routing because longer prefixes might already be re-announced through some parts of the Internet where shorter prefixes are still held down through other paths.

Coordination of hold-down and reuse-threshold parameters:

If an upstream or peering provider would be damping more aggressively (e.g. triggered by less flaps or applying longer hold-down timers) than an access-provider towards his customers, it will lead to a very inconsistent situation where a flapping network might still be able to reach "near-line" parts of the Internet. Debugging of such instabilities is then much harder because the effect for the customer leads to the assumption that there is a problem "somewhere" in the "upstream" Internet instead of making him just call his ISP’s hot-line and complain that he can’t get out any longer.

Further, after successful repair of the problem, the access-provider can easily clear the flap-damping for his customer on his local router instead of needing to contact upstream NOCs all over the Internet to get the damping cleared.

1.5 Aggregation versus damping

Of course, if a customer is just using Provider Aggregatable addresses, the aggregating upstream provider doesn’t need to apply damping on these prefixes towards his customer, because instabilities of such prefixes wouldn’t propagate into the Internet. However, if a customer insists on announcing prefixes which can’t be aggregated by its provider, damping should be applied for the reasons given in Section 1.4. Reasons might be dual-homing (to different providers) of a customer or customers reluctant to renumber into the provider’s aggregated address range.

1.6 "Golden Networks"

Even though damping is strongly recommended, in some cases it may make sense to exclude certain networks or even individual hosts from damping. This is especially true if damping would cut off the access to vital infrastructure elements of the Internet. Most prominent examples are root name servers.

At least in principle, there should be enough redundancy for root name servers. Though, in fact we are still facing a situation where, at least outside the USA, large parts of the Internet are seeing all of them through the same one or two backbone/upstream links (sea cable) and any instability of those links which is triggering damping would unnecessarily prolong the inaccessibility of the root name servers for an hour (at least those sitting in a /24 or longer prefix). Therefore we decided to define those "golden networks". Probably we could remove the exemptions for the A, D and H servers, which are sitting in a /16. We might consider this for a new version of the recommendation. Our recommendation is just dealing with a minimum set of "golden networks" which of course might be extended by local decision.

Still, these must be exceptions resulting from strong needs - the rule should be to apply coordinated route flap damping throughout.

2. Recommended damping parameters

2.1 Motivation for recommendation

At RIPE26 and 27 Christian Panigl presented the following network backbone maintenance example from his own experience, which was triggering flap damping in some upstream and peering ISPs routers for all his and his customers /24 prefixes for more than three hours because of too "aggressive" parameters:

scheduled SW upgrade of backbone router failed:

```
- reload after SW upgrade       1 flap
- new SW crashed                1 flap
- reload with old SW            1 flap
                                ------
                                3 flaps within 10 minutes
```

which resulted in the following damping scenario at some boundaries with progressive route-flap damping enabled:

```
Prefix length:      /24     /19     /16
suppress time:      ~3h     45-60'  <30'
```

Therefore, in the Routing-WG session at RIPE27, it was agreed that suppression should not start until the 4th flap in a row and that the maximum suppression should in no case last longer than 1 hour from the last flap.

It was agreed that a recommendation from RIPE would be desirable. Given that the current allocation policies are expected to hold for the foreseeable future, it was suggested that all /19s or shorter prefixes are not penalised harder (longer) than current Cisco default damping does (see: 2.3).

Those suggestions in mind, Tony Barber designed the following set of route-flap damping parameters which have proved to work smoothly in his environment for a couple of months.

2.2 Description of recommended damping parameters

Basically the recommended values do the following with harsher treatment for /24 and longer prefixes:

- don't start damping before the 4th flap in a row (suppress-value = 3000)
- /24 and longer prefixes: max=min outage 60 minutes
- /22 and /23 prefixes: max outage 45 minutes but potential for less because of half life value - minimum of 30 minutes outage
- all else prefixes: max outage 30 minutes, min outage 10 minutes

If a specific damping implementation does not allow configuration of prefix-dependent parameters the softest set should be used:

- don't start damping before the 4th flap in a row - max outage 30 minutes, min outage 10 minutes

2.3 Example configuration for Cisco IOS

```
! Parameters are :
! set damp <half-life-time> <reuse-at> <suppress-at>
!          <max-suppress-time>
! There is a 1000 penalty for each flap
! Penalty decays at granularity of 5 seconds
! Unsuppressed at granularity of 10 seconds
! damping info kept until penalty becomes < half of reuse limit.
!
! Cisco/IOS value-ranges:
!
!   <half-life-time> (range is 1-45 minutes).
!   <reuse-value> (range is 1-20000).
!   <suppress-value> (range is 1-20000).
!   <max-suppress-time> (range is 1-255 minutes ).
!
!-----------------------------------------------------------------------
! ENABLE BGP DAMPenING using "graded" route-map
!-----------------------------------------------------------------------
router bgp 65500
NO bgp damp
bgp damp route-map graded-flap-damping
!
!-----------------------------------------------------------------------
! DEFINE "graded" route-map
!-----------------------------------------------------------------------
NO route-map graded-flap-damping
!
! don't damp Candidate Default Routes
! OPTIONAL (not part of recommendation)
! prefix-list default-networks lists the Candidate Default Routes
!
!route-map graded-flap-damping deny 5
! match ip address prefix-list default-networks
!
! don't damp root name server nets
!
route-map graded-flap-damping deny 10
 match ip address prefix-list rootns
!
!    - /24 and longer prefixes: max=min outage 60 minutes
!
route-map graded-flap-damping permit 20
 match ip address prefix-list min24
 set damp 30 750 3000 60
!
!    - /22 and /23 prefixes: max outage 45 minutes but potential for less
!      because of shorter half life value - minimum of 30 minutes outage
!
route-map graded-flap-damping permit 30
 match ip address prefix-list max22-23
 set damp 15 750 3000 45
!
!    - all else prefixes: max outage 30 minutes min outage 10 minutes
!
```

```
route-map graded-flap-damping permit 40
 set damp 10 1500 3000 30
!
!-----------------------------------------------------------------------
! DEFINE PREFIX-LISTS
!-----------------------------------------------------------------------
!
! OPTIONAL default-networks
!no ip prefix-list default-networks
!ip prefix-list default-networks description Candidate Default Routes
!ip prefix-list rootns permit ...
!
no ip prefix-list rootns
ip prefix-list rootns description Root-nameserver networks
ip prefix-list rootns permit 198.41.0.0/24
ip prefix-list rootns permit 192.112.36.0/24
ip prefix-list rootns permit 198.17.208.0/24
ip prefix-list rootns permit 192.5.4.0/23
ip prefix-list rootns permit 192.36.148.0/24
ip prefix-list rootns permit 192.203.230.0/24
ip prefix-list rootns permit 198.41.0.0/24
ip prefix-list rootns permit 195.8.96.0/19
ip prefix-list rootns permit 198.41.3.0/24
ip prefix-list rootns permit 210.176.0.0/16
ip prefix-list rootns permit 216.33.64.0/19
ip prefix-list rootns permit 205.188.128.0/17
!
no ip prefix-list min24
ip prefix-list min24 description Apply to /24 and longer prefixes
ip prefix-list min24 permit 0.0.0.0/0 ge 24
!
no ip prefix-list max22-23
ip prefix-list max22-23 description Apply to /22 and /23 prefixes
ip prefix-list max22-23 permit 0.0.0.0/0 ge 22 le 23
!
```

2.4 No BGP fast-external-fallover (Cisco IOS)

In Cisco IOS there is a BGP configuration parameter "fast-external-fallover" which when on (default) leads to an immediate clearing of a BGP neighbor whenever the line-protocol to this external neighbor goes down. If it is turned off the BGP sessions will survive short line-flaps as they will use the longer BGP keepalive/hold timers (default 60/180 seconds). The drawback of turning it off - and currently it has to be done for a whole router and can not be selected peer-by-peer - is that the switch-over to an alternative path will take longer. We are recommending to turn off fast-external-fallover whenever possible:

```
! router bgp 65501
  no bgp fast-external-fallover
!
```

Alternatively it might be considered to stay with "BGP fast-external-fallover" and to turn off "interface keepalives" on flappy lines, to overcome the immediate BGP resets on any significant CRC error period.

Another, even better alternative would be to use a shorter per-neighbor BGP keepalive timer which has to be applied on both routers (e.g. 10 seconds which gives a hold-timer of 30 seconds):

```
! router bgp 65501
  neighbor w.x.y.z timers 10
!
```

2.5 Clear IP BGP soft inbound (Cisco IOS)

There is a "soft" mechanism for the clearing of BGP sessions available with Cisco IOS. For being able to make use of the "clear ip bgp x.x.x.x soft inbound" command the router which should support it needs to be configured for additional data structures:

```
! router bgp 65501
  neighbor 10.0.0.2 remote-as 65502
  neighbor 10.0.0.2 soft-reconfiguration inbound
!
```

Without the keyword "soft" a "clear ip bgp x.x.x.x" will completely reset the BGP session and therefore always withdraw all announced prefixes from/to neighbor x.x.x.x and re-advertise them (= route-flap for all prefixes which are available before and after the clear). With "clear ip bgp x.x.x.x soft out" the router doesn’t reset the BGP session itself but sends an update for all its advertised prefixes. With "clear ip bgp x.x.x.x soft in" the router just compares the already received routes (stored in the "received" data structures) from the neighbor against locally configured inbound route-maps and filter-lists.

2.6 BGP Route refresh, soft reset enhancement (Cisco IOS 12.0(7)S/T and higher)

There is a new "refresh" mechanism for the clearing of BGP sessions available with newer versions of Cisco IOS. For being able to use this feature inbound both peers need to support it.

You may find out if your neighbor is supporting it with:

```
  Router# sho ip bgp neigh w.x.y.c | include refresh
  Received route refresh capability from peer
  Route refresh request: received 0, sent 0
```

If you and your peer is supporting it you can use

```
  Router# clear ip bgp w.x.y.c in
```

for requesting a route refresh without clearing the BGP session.

For an outbound route refresh without clearing the BGP session use

```
Router# clear ip bgp w.x.y.c out
```

3. Open problems

3.1 Multiplication of flaps through multiply interconnected ASs

Christian Panigl made the following experience with a line upgrade of an Ebone customer:

- It is absolutely positive that through the upgrade process just ONE flap was generated (disconnect router-port from modem A reconnect to modem B), nevertheless the customers prefix was damped in all ICM routers.

- The flap statistics in the ICM routers stated *4* flaps !!!

- The only explanation would be that the multiple interconnections between Ebone/AS1755 and ICM/AS1800 did multiply the flaps (advertisements/withdrawals arrived time-shifted at ICM routers through the multiple lines).

- This would then potentially hold true for any meshed topology because of the propagation delays of advertisements/withdrawals.

- It appears to be (confirmed) buggy behavior of (at least) the Cisco implementation.

- Workaround for scheduled actions like with the given example:

Schedule a downtime for at least 3-5 minutes which should be enough for the prefix withdrawals to have propagated through all paths before reconnection and re-advertisement of the prefix. Avoid clearing BGP sessions as this is usually generating a 30" outage which might easily give the same result.

- A final solution has to be provided by the vendors !

4. References

RIPE/Routing-WG Minutes dealing with Route Flap Damping:

http://www.ripe.net/ripe/meetings/archive/ripe-24/ripe-m-24.txt

http://www.ripe.net/ripe/meetings/archive/ripe-25/ripe-m-25.txt
http://www.ripe.net/wg/routing/r25-routing.html
http://www.ripe.net/wg/routing/r26-routing.html
http://www.ripe.net/wg/routing/r27-routing.html

Curtis Villamizar, Ravi Chandra, Ramesh Govindan
RFC2439: BGP Route Flap Damping (Proposed Standard)
ftp://ftp.ietf.org/rfc/rfc2439.txt

Merit/IPMA: Internet Routing Recommendations
http://www.merit.edu/ipma/docs/help.html

Cisco BGP Case Studies: Route Flap Damping
http://www.cisco.com/warp/public/459/16.html

Cisco Documentation: Configuring BGP / Route Dampening
http://www.cisco.com/univercd/cc/td/doc/product/software/ios121/121cgcr/ip_c/ipcprt2/1cdbgp.htm#xtc

Cisco Documentation: BGP Soft Reset Enhancement
http://www.cisco.com/univercd/cc/td/doc/product/software/ios120/120newft/120t/120t7/sftrst.html

ISI/RSd Configuration: Route Flap Damping
http://www.isi.edu/div7/ra/RSd/doc/dampen.html

GateD Configuration: Weighted Route Damping Statement
http://www.gated.org/gated-web/code/doc/manuals/config_guide/bgp/weighted_route_dampening.html

5. Acknowledgements

The following people have contributed their input to the updated Version 1.1:

Steffen Baur, DFN Network Operation Centre, Germany

Fredrik Rosenbecker, IP-Only Telecommunication, Sweden

Christian Panigl, University of Vienna / ACOnet, Austria (Editor)