

Improving performance through BGP Graceful Shutdown [draft-ietf-grow-bgp-gshut](#)

Job Snijders

job@ntt.net



What is BGP Graceful Shutdown?

- A simple procedure to reduce the negative impact of shutting down BGP sessions
- Can be combined with [RFC 8203](#) “Shutdown Communication”
- Can be part of the operational procedure as outlined in [draft-ietf-grow-bgp-session-culling](#)
- Graceful Shutdown is a **“*Make Before Break*”** mechanism
- Does not help against unplanned outages
- **Not** to be confused with BGP Graceful Restart (which is somewhat psychopath 😊)

Context

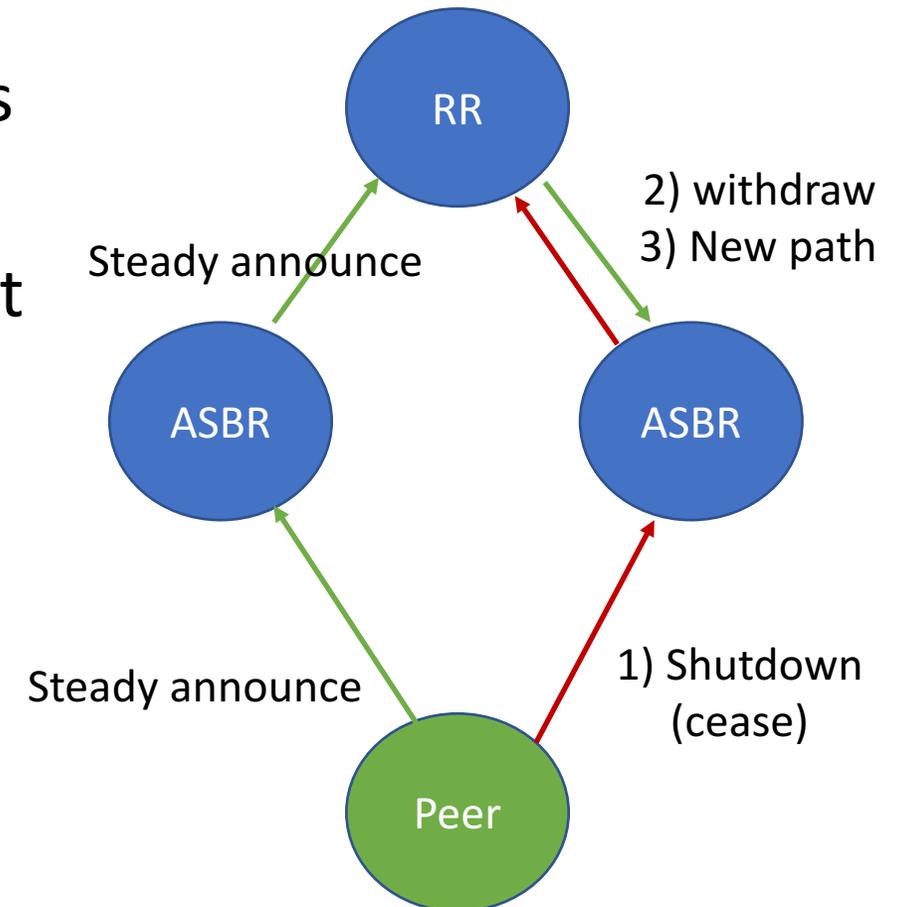
When organizations established direct EBGP sessions, I assume two parties are interested in exchanging traffic with as little packet loss as possible and low latency in an economically sound way.

If you are not looking to minimize packet loss, “Graceful Shutdown” is not for you.

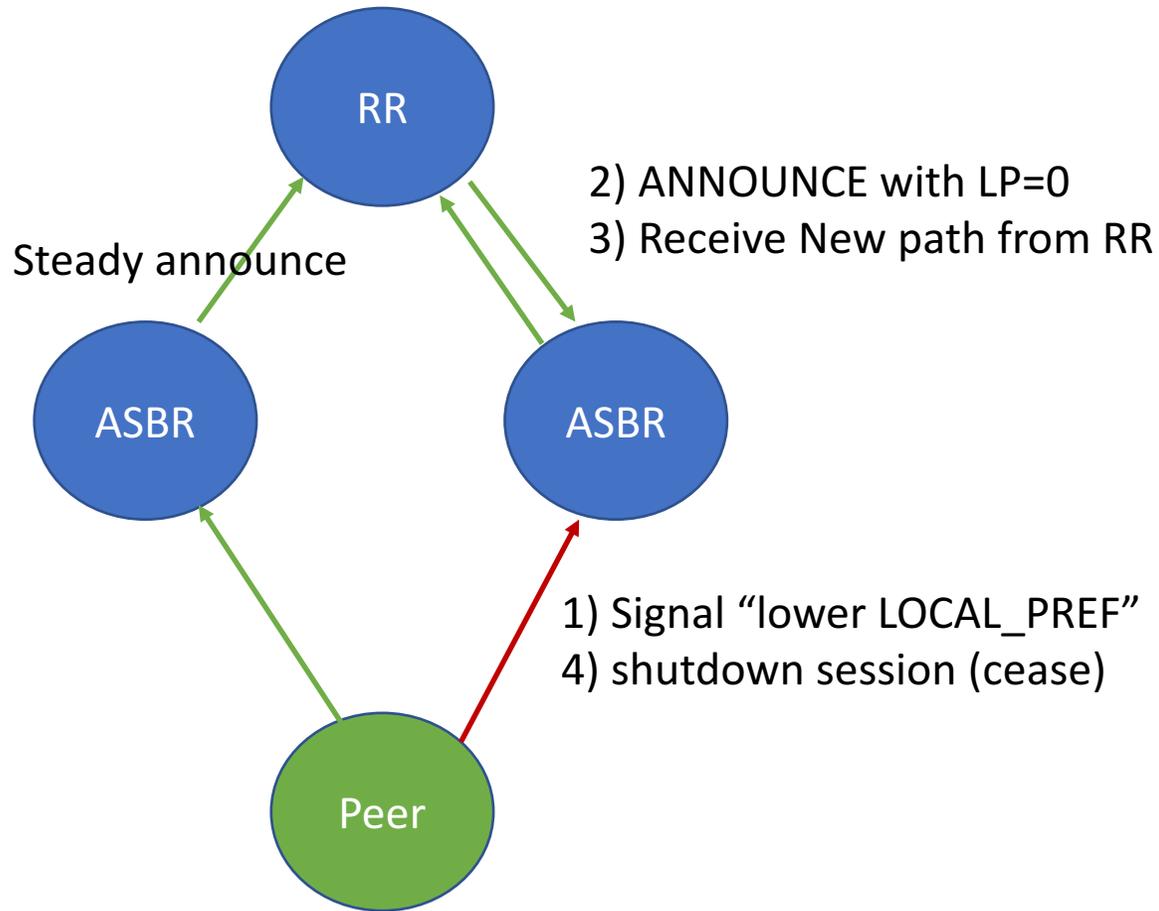
When does blackholing happen with vanilla shutdown?

- Lack of an alternative route on some routers
- Transient routing inconsistency
- A route reflector may only propagate its best path
- The backup ASBR may not advertise the backup path because the nominal path is preferred

Admittedly, the above scenarios usually are short periods of blackholing, but why accept that if they can easily be prevented?



Graceful Shutdown triggers “path hunting”



- Initiated by the operator on the router before maintenance by sending the GRACEFUL_SHUTDOWN well-known community (65535:0 as per IANA)
- Receiving EBGP peer sets LOCAL_PREFERENCE to 0 and selects paths to route traffic away from the initiator, (similar to setting overload in an ISIS)
- When BGP session goes down, minimizes impact to traffic because alternate paths have already been installed

Normal vs Graceful

- Operator types “shutdown” on Router A
- Router A sends “CEASE NOTIFICATION”
- Router B generates a BGP WITHDRAW message for all prefixes received from Router A
- Until Router B’s ASN reconverges there may be micro blackholes

- Operator types “shutdown” on Router A
- Router A sends UPDATES for all prefixes with community 65535:0
- Receiving Router lower’s LOCAL_PREFERENCE for all routes and UPDATES the rest of the ASN
- Router A drinks cup of coffee while router B’s AS reconverges
- Router A shuts down BGP session (CEASE NOTIFICATION)

Usage Guidelines

- To support receiving graceful shutdown, update your routing policy to
 - Match the GRACEFUL_SHUTDOWN well-known community (65535:0)
 - Set the LOCAL_PREF attribute to a low value, like 0
 - **Do not strip** the community when propagating the prefix
- To send graceful shutdown, update your routing policy to
 - Send the GRACEFUL_SHUTDOWN well-known community (65535:0) before you start maintenance
 - When to/from traffic from the peer has stopped, start maintenance and use BGP shutdown communication (usually just a few minutes)
 - Remove the GRACEFUL_SHUTDOWN well-known community when you are done

GRACEFUL_SHUTDOWN signals:

“Hello everyone, if you consider this path your ‘best path’, please start considering this path the ‘worst path’ and if you find anything better install that into your FIB. This path will disappear within a few minutes.”

Configuration Example – Simple to Implement

```
!  
route-policy AS64497-ebgp-inbound  
    if community matches-any (65535:0) then  
        set local-preference 0  
    endif  
end-policy  
!  
router bgp 64496  
    neighbor 2001:db8:1:2::1  
    remote-as 64497  
    address-family ipv6 unicast  
        send-community-ebgp  
        route-policy AS64497-ebgp-inbound in  
    !  
!
```

IOS XR

Arista/Brocade/IOS/Quagga/FRR

```
!  
ip community-list standard  
gshut 65535:0  
!  
route-map ebgp-in permit 10  
    match community gshut  
    set local-preference 0  
    continue  
!
```

Making use of GRACEFUL SHUTDOWN on IOS XE

```
neighbor 10.0.0.1  
  remote-as 65000  
  graceful-maintenance  
  activate
```

A single neighbor

```
neighbor-group test  
  graceful-maintenance  
  activate
```

A neighbor group

What does this look like in AS 2914 / NTT Communications Global IP Network?

- Since we assume networks connect to us for mutual benefit, it makes sense to facilitate those networks to perform maintenance without introducing packet loss. Why are we otherwise connected, if not to forward packets?
- NTT honors the well-known 65535:0 BGP community on all EBGP sessions. (<https://www.us.ntt.net/support/policy/routing.cfm#experimental>) This includes:
 - Customer EBGP sessions
 - Peering partner EBGP sessions
- **For any prefix sent to AS 2914 with community 65535:0, NTT will lower the LOCAL_PREFERENCE to 0 to facilitate maintenance. The low LP triggers the search for alternative paths within an ASN.**

What about you?

- Would you allow NTT (and others) to perform hitless maintenance by honoring the 65535:0 GRACEFUL_SHUTDOWN community?
- This deviates from the standard practice of "*not honoring any communities from peering partners*" – but does that 'standard' actually make sense in context of cooperation and preventing packet loss?
- The receiving side implementation is trivial, this can be done on anything from an IOS/Brocade box up to and including the most modern operating systems.

GTT (AS 3257), Github (AS 36459), Nordunet (AS 2603), Coloclue (AS 8283), Amsio (AS 8315), BIT (AS 12859) deployed support... who else? 😊

The science behind gracefully shutting down BGP sessions

- Avoiding disruptions during maintenance operations on BGP sessions:
<https://inl.info.ucl.ac.be/system/files/ucl-ft-bgp-shutdown-inl.pdf>
(August 2008)
- Requirements for the Graceful Shutdown of BGP Sessions
<https://tools.ietf.org/html/rfc6198> (April 2011)