RFC 7752bis
(an update to BGP-LS specification)

Ketan Talaulikar, Jan Medved (Cisco Systems)

Hannes Gredler (RtBrick)

Stefano Previdi, Adrian Farrel, Saikat Ray

Recap

- BGP-LS (RFC7752) has been implemented and deployed widely
- draft-ketant-idr-rfc7752bis-00 was submitted before IETF 104 Prague to clarify/update
  - error handling and fault management aspects
  - some TLVs, their mandatory/optional nature, verification, etc.
  - use of Instance ID in BGP-LS
  - certain aspects with interpretation/handling and propagation of BGP-LS information
  - Handling the growth of BGP-LS Attribute
- This draft will obsolete RFC7752

Thanks for the feedback and inputs based on development and deployment experience

Updates in version 01

- Advertisement of the OSPF LAN subnet was not described by RFC7752; this has been clarified
  - Done by Prefix NLRI origination by the DR (psuedonode)
- Reduced the private use code point space based on feedback
- Other minor editorial changes to clarify/correct

IPR declared on RFC7752 also applies to this document and has been filed as a 3rd party disclosure by Adrian

Ordering of TLVs in BGP-LS Attribute

- RFC7752 said "In order to compare NLRIs with unknown TLVs, all TLVs MUST be ordered in ascending order by TLV Type"

- Some implementation(s) have (mis)interpreted this to imply TLVs in BGP-LS Attribute also MUST be ordered

- Bis draft proposes "The TLVs within the BGP-LS Attribute need not be ordered in any specific order."

- Received a proposal to change this to : For BGP-LS Attribute
  - Producer should/must include TLVs in sorted order. Receiver must not consider TLVs appearing in unsorted order as malformed.
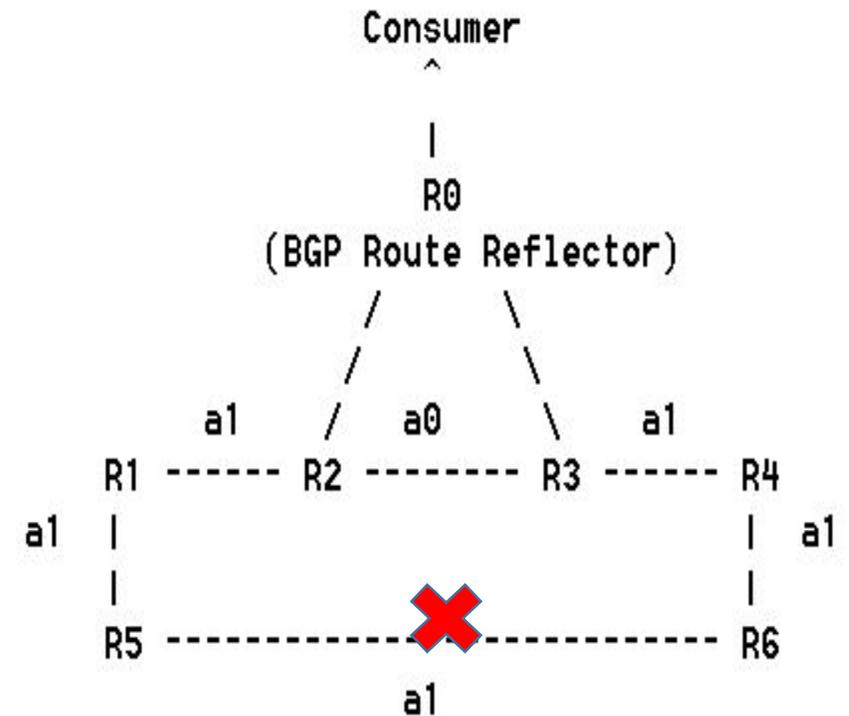
- Feedback?

Handling Unreachable IGP Nodes

- When BGP-LS Producers continue to advertise link-state objects based on stale LSA/LSPs of unreachable nodes in IGP, then a BGP-LS Consumer may get a wrong or inconsistent topology view

- BGP-LS propagation happens based on BGP best path algorithm which can result in NLRI with stale information being preferred over another with newer and consistent information
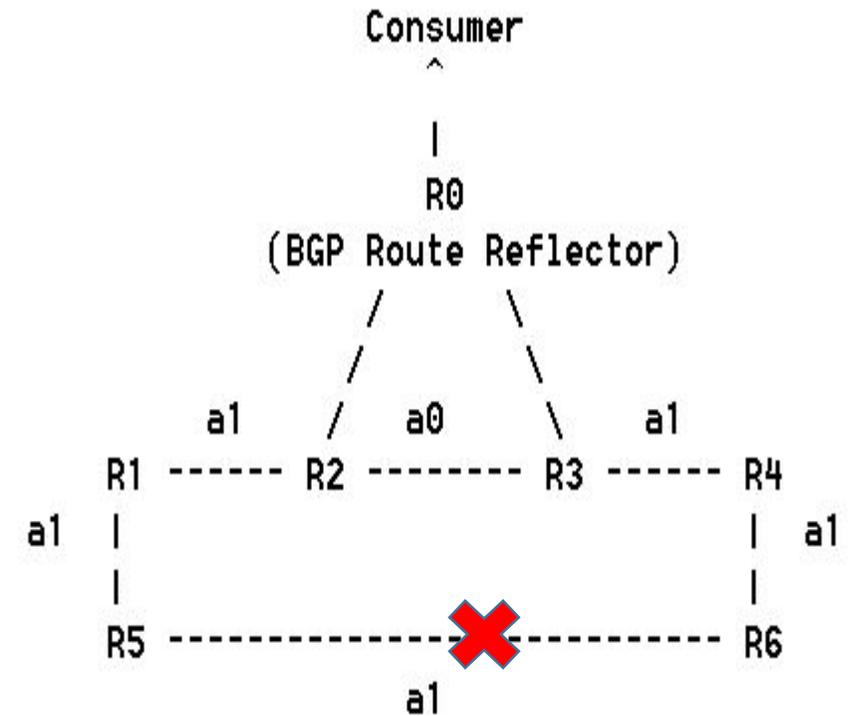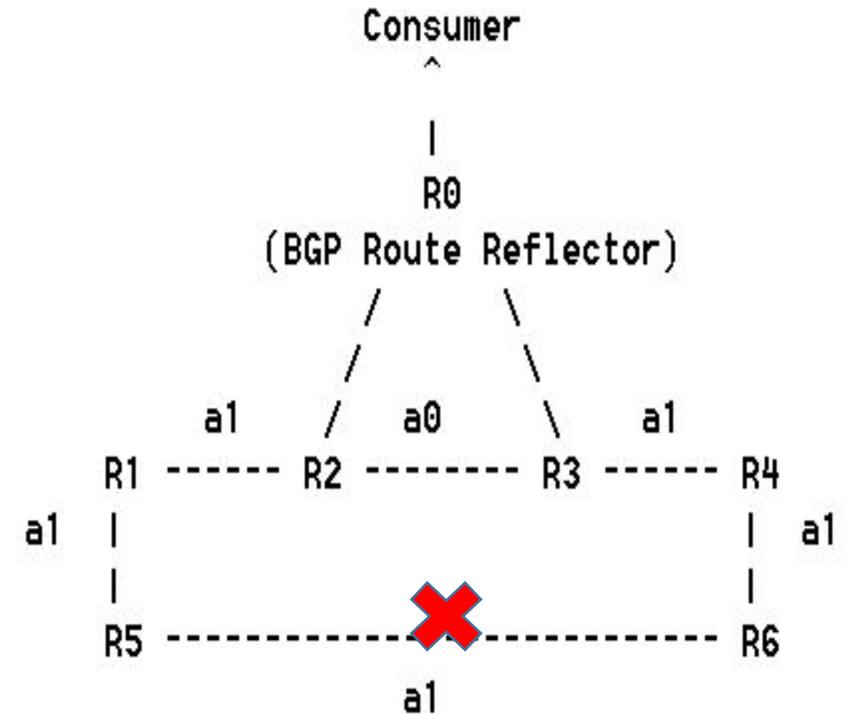
Proposed Solution in draft – (A)

- BGP-LS producer should withdraw link-state objects when the associated node becomes unreachable in IGP processing on the producer node
  - IGP on R2 marks the LSA/LSPs of R4 and R6 as unreachable after running it's SPF following the link failure
  - R2 withdraws the BGP-LS NLRIs corresponding to LSA/LSPs originated by R4 and R6
  - Similarly R3 withdraws BGP-LS NLRIs corresponding to LSA/LSPs originated by R1 and R5
  - R0 has NLRIs from only R2 for R1 & R5 and only from R3 for R4 & R6.
  - No NLRIs for stale LSA/LSPs exist in BGP
  - BGP RR and other BGP routers do normal BGP path computation and propagation


```
              Consumer
                 ^
                 |
                 |
                R0
        (BGP Route Reflector)
              /       \
             /         \
    a1      /  a0       \    a1
R1 ------ R2 -------- R3 ------ R4
a1 |                          | a1
   |                          |
R5 ------------- ✖ --------- R6
        a1
```

Alternate Solution – (B)

- BGP-LS Producer advertises information from IGP LSDB without regards to reachability; NLRIs advertised may contain stale info
  - BGP RR in this case MUST do ADD_PATHS; if not then it is possible that stale info gets selected by best path and propagated to consumers
  - All BGP propagators along the way (multiple RRs, eBGP, etc. if used) MUST also do ADD_PATHS
  - We need an "originator" attribute/information to be carried through all BGP propagation methods
  - Consumer gets multiple copies of the same NLRI associated with each originator when there are multiple originators (for redundancy) or multiple paths
  - Consumer needs to run IGP SPF on the topology from each originator perspective to determine "valid" NLRIs
  - Consumer then consolidates "valid" NLRIs to form a topology depending on use-case (e.g. performing TE computation)

```
                    Consumer
                        ^
                        |
                        R0
                (BGP Route Reflector)
                     /        \
                    /          \
            a1     /    a0       \     a1
    R1 ------ R2 -------- R3 ------ R4
  a1 |                      |   | a1
     |                      |   |
     |                    ✗ |   |
    R5 ------------------------- R6
            a1
```

Comparison

Solution (A)

- Pros
  - No change to BGP propagation or path computation rules; no additional features are mandated

- Cons
  - Does not provide a "complete" IGP LSDB view to consumer (i.e. NLRIs for unreachable nodes is not presented)
  - Creates dependencies between NLRI origination & SPF processing, which the current RFC does not specify

Need feedback/inputs from WG on the two solutions

Solution (B)

- Pros
  - Provides a "complete" IGP LSDB view to consumer

- Cons
  - Mandates use of ADD_PATHS and propagation of "originator" whenever consumer has multiple feeds via BGP (due to multiple originators or multi-path)
  - Mandates consumer applications to do IGP computation from each originator perspective to determine "valid" NLRIs
  - Increases BGP scale for carrying multiple paths for topology
  - RR reflects all churn towards other nodes and consumers along the propagation paths (since it is not doing best path to select amongst redundant info)

IANA Registry Allocation Rules

- Discussions ongoing between AD, WG chairs and Designated Experts regarding IANA allocation rules for BGP-LS

- Goal is to make the process faster/easier

- This draft will capture those decisions in upcoming versions

Next Steps

- Make some editorial fixes for format and to comply with RFC editor guideline

- Please continue to provide feedback from existing implementations and deployments

- Review, discuss on IDR mailing list

- Requesting WG adoption