IETF 102 – Montreal
July 2018
IDR Working Group

draft-xu-idr-neighbour-autodiscovery-09

Xiaohu Xu, Chao Huang, Guixin Bao (Alibaba)
Ketan Talaulikar, Satya Mohanty (Cisco Systems)
Kunyang Bi, Shunwan Zhuang (Huawei)
Jeff Tantsura (Nuage Networks)
Nikos Triantafillis
Jinghui Liu (Ruijie Networks)
Zhichun Jiang (Tencent)
Shaowen Ma (Juniper Networks)

Problem Statement

- BGP is used as the only routing protocol in DCs using RFC7938 design
- Operational complexity involved in provisioning of hop-by-hop per link eBGP peering between BGP nodes
- When doing peering using loopbacks (e.g. due to ECMP links or when using IPv6 link-local addresses or unnumbered links) need to also provision static route for reachability

Requirements

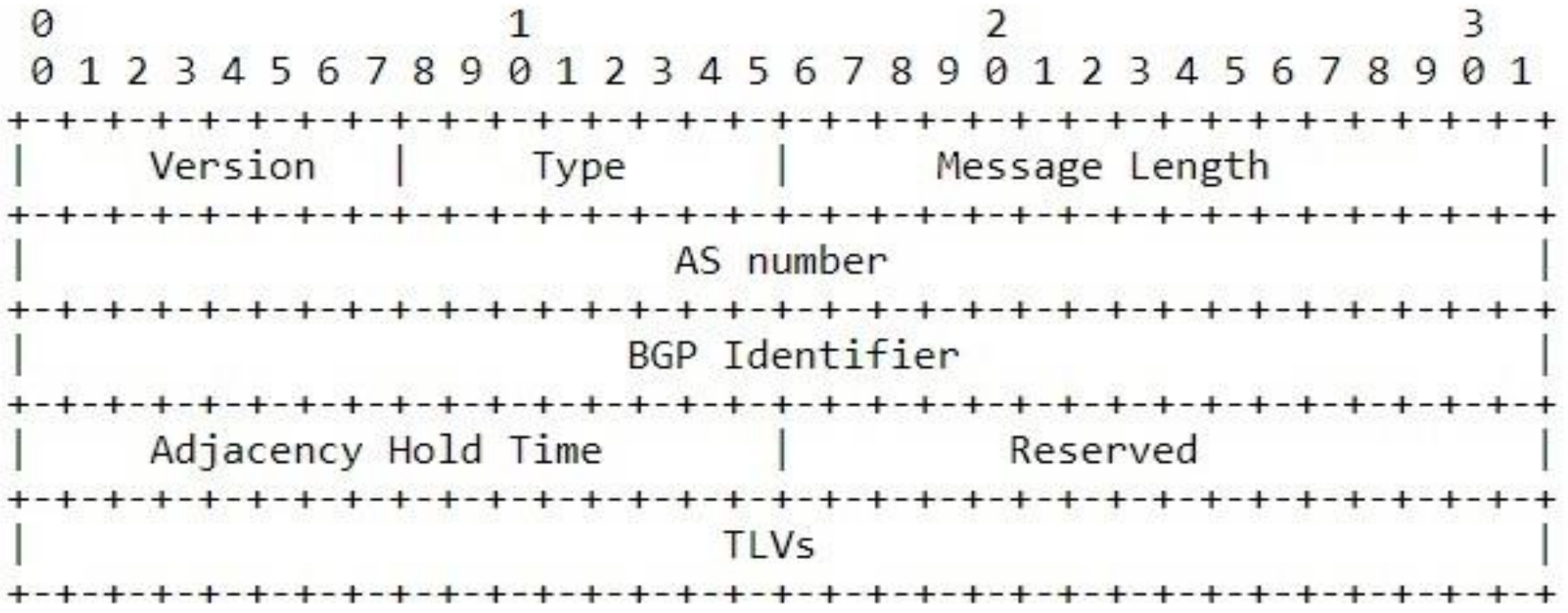
- Need a neighbour discovery mechanism that runs on top of IPv4/IPv6
  - Is media independent; works on IPv4, IPv6 and dual stack
  - Needs to support authentication mechanism for security purposes
- Keep it simple and focus on current BGP requirements
  - We have LLDP and BFD widely deployed; leverage them
  - Make mechanism extensible for signalling of information for BGP
- Auto-discovery and bootstrap for BGP TCP Sessions between directly connected nodes
- Separate discovery and liveness for BGP neighbours
  - Discovery and maintenance of adjacency is the core part
  - Use of liveness mechanism is optional; continue to leverage BGP KA, BFD and Fast External Failover features
- Minimal changes for integration with BGP Peer FSM and no changes in BGP protocol operations

What does this draft propose?

- Automated neighbour discovery using UDP Hello Messages on a per link basis for directly connected neighbours only
- Signalling of peering address and ASN so that BGP Peering session can be automatically initiated with discovered neighbour
- BGP session can be setup using loopbacks and reachability established via peering route setup that points over the links over which neighbour is discovered
- Minimal changes to the BGP Peer FSM and no change to BGP route processing

Hello Message Format

- Uses UDP port 179 and sent to link-local multicast address
- Can be used over either IPv4 or IPv6 addresses



Important TLVs

- Peering Address TLV
  - Indicates one or more IPv4 and/or IPv6 peering address(es) to be used
  - Optionally can indicate which AFI/SAFI to be used for which Peering
- Link Attributes TLV
  - Indicates link addresses and link identifiers for describing the link endpoint (so information is learnt for exporting via BGP-LS)
  - Can perform subnet and other policy checking before session setup
- Neighbour TLV
  - Signals discovered neighbours and their adjacency status (1-way, 2-way, reject and established)
  - Used to indicate to neighbour whether the BGP TCP session can be initiated (i.e. when both sides have accepted each other)

Optional TLVs

- Local Prefix TLV
  - Indicates the prefix route to be programmed after neighbour discovery goes to 2-way state to ensure reachability for the neighbour's peering address
  - Required when peering is to be done using loopback interface; not required when doing peering with interface addresses
- Accepted ASN TLV
  - Indicates the list of ASNs to which peering session would be established – local policy
- Cryptographic Authentication TLV
  - Carries the SA ID and authentication information

Adjacency State Machine

- Initial State
  - Initial state when a neighbour is detected
- 1-way State
  - When router accepts the peer and includes it in its own hello message
- Reject State
  - When router rejects the peer due to detection of some config mismatch or violation of local policy
- 2-way State
  - When router detects itself in the neighbour's hello; now ready for TCP session establishment step
  - Adds peering route for the neighbour over the link (i.e. when using loopbacks for peering)
  - Creates the BGP Peer State context for discovered peer and triggers the BGP Peer FSM
- Established
  - When the BGP TCP session is established

Session Management

- Once established, session management is performed as per BGP FSM
- Liveness detection via Keepalives & Hold timer
  - BFD and Fast External Failover also works when enabled
- Established BGP session is NOT brought down due to adjacency hold timer expiry by default
  - This may be optionally enabled in cases where required
- Adjacency hold timer expiry used to clean-up BGP Peer state after the session goes down for auto-discovered peer

Peering Route

- Required only when peering is done using loopback interfaces
- Route programmed with higher Admin Distance than normal BGP routes to prevent oscillation (in case the peering route is also learnt via BGP itself)
- When there are multiple links between neighbours then peering route will have ECMP paths over each of them
- BGP NH for the neighbour resolved over this peering route for reachability
- No need for programming static route or running another protocol when doing Peering over loopback addresses

Next Steps …

- WG adoption call ongoing in IDR
- Solicit WG review and comments/inputs/feedback