BGP Neighbor Discovery
draft-xu-idr-neighbor-autodiscovery

Xiaohu Xu@China Mobile
Ketan Talaulikar@Cisco
Kunyang Bi@Huawei
Jeff Tantsura@Nvidia
Nikos Triantafillis@AWS

IETF123, Madrid

Backgrounds

- The Optical Circuit Switch (OCS) is becoming an emerging choice for hyper-scale data center networks for various purposes, including but not limited to: 1) reducing power consumption; 2) decreasing CapEx.
- In a generic 3-stage CLOS network, spine switches are replaced by OCS devices, which enable the dynamic reconfiguration of a full mesh of leaf switch interconnects by reconfiguring the MEMS mirror arrays.
  - For more details, see Google's SIGCOMM'22 paper about its OCS-based data center networks (https://dl.acm.org/doi/pdf/10.1145/3544216.3544265).
- BGP neighbor discovery among leaf switches in the above scenario would be much helpful.

Solution Overview

- Learning from the LDP [RFC5036] mechanism, UDP-based Hello messages are used to discover directly connected BGP neighbors.
  - The peering address (i.e., loopback address) and other information (e.g., local ASN) are exchanged through **multicast** BGP Hello messages initially.
- The route to a neighbor's peering address is automatically installed, pointing to the links over which the neighbor is discovered.
  - There is no need for any IGP or static route.
  - The links are IP unnumbered.
- BGP sessions are then dynamically established over the discovered peering addresses without any further changes to the existing BGP mechanisms as per [RFC4271].

Comparison with Related Solutions

- Neighbor discovery using LLDP (draft-acee-idr-lldp-peer-discovery).
  - The advertisement of a loopback address as a peering address is allowed, but reachability to it is beyond the scope of this document—**Like a carrot on a stick!**
  - N links between a pair of peers require N BGP sessions—**There's still room for improvement!**
- Link-Local Next Hop Capability for BGP (draft-ietf-idr-linklocal-capability)
  - Permits the use of IPv6 link-local address as the only next hop address for a route—**Not intended for the BGP neighbor discovery.**

Implementation Status

- The BGP neighbor discovery protocol has been implemented by at least one well-known network vendor, to the best of my knowledge.
- The protocol is now being implemented by China Mobile Cloud on our SONiC-based NOS and by at least three gray-box switch vendors.

Next Step

- The document is ready for the WG adoption call.
- The OCS-related use case would be added in the next revision.