IETF 125 – Shenzhen
March 2026
IDR Working Group

BGP Link-State Extensions for BGP-only Networks draft-ietf-idr-bgp-ls-bgp-only-fabric-04

Ketan Talaulikar (ketant.ietf@gmail.com)

Aravind Babu MahendraBabu (aramahen@cisco.com)

Clarence Filsfils (cfilsfil@cisco.com)

Krishna Swamy (kriswamy@cisco.com)

Shawn Zandi (szandi@linkedin.com)

Gaurav Dawra (gdawra.ietf@gmail.com)

Muhammad Durrani (mdurrani@equinix.com)

What does this draft propose?

- **BGP-LS** was originally defined in **RFC 9552** for northbound distribution of **link-state** and **traffic engineering (TE)** information.
- Historically, BGP-LS has mainly focused on distributing information learned from **IGPs**.
- **RFC 9086** expanded by introducing BGP protocol for EPE links.
- This draft defines the mechanism to distribute **network topology** and **TE information** in networks where **BGP is the only routing protocol**.

Why Is This Required?

- Many **DC/MSDC fabrics** (RFC7938) and emerging **AI DC networks** run **BGP as the only routing protocol**.
- In BGP-only environments, controllers often lack a full **underlay link topology** and **TE visibility** (which IGP-based networks can export via BGP-LS).
- Optimized fast-convergence and resilience solutions in DC (especially AI/ML infra) need topology visibility. Refer ongoing discussion in RTGWG.
  - **AI/ML-driven IPFRR and resiliency optimization** (draft-clad-rtgwg-ipfrr-aiml)
  - **Efficient Remote Protection** (draft-clad-rtgwg-efficient-remote-protection)
  - **IP Fast Reroute (IPFRR) in BGP-only networks** (draft-abdelsalam-rtgwg-ipfrr-bgp-only-network)
- Exposing this topology enables **Traffic Engineering (TE)** use cases in BGP-only networks like building planes/rails in AI DCs and end-to-end TE between DCs over WAN.

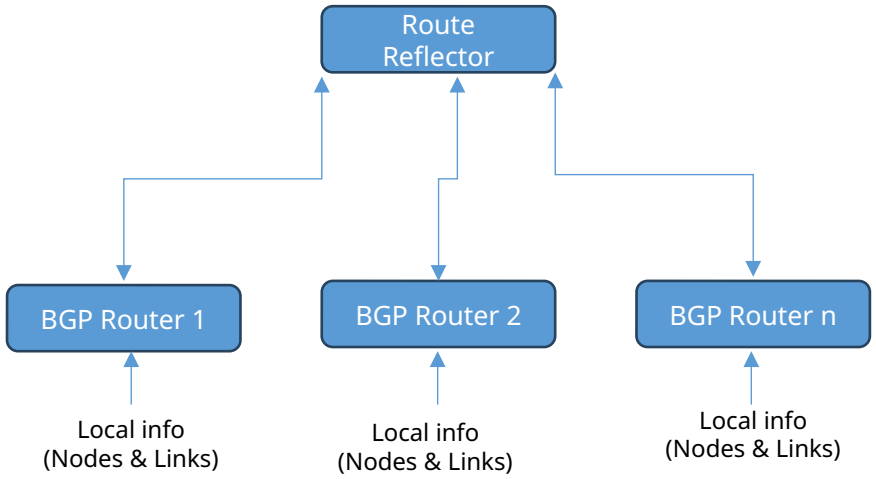
How Is This Done?

- Enable **BGP-LS on each BGP speaker** to advertise that node's local topology using existing NLRI types: **Node, Link, Prefix**.
- Mostly reuse existing **BGP-LS descriptors** for these NLRI types.
  - Specifies procedures for advertisement of NLRI TLVs for BGP fabric
  - New BGP Route Type TLV in the Prefix Descriptor introduced in this draft.
- Reuse existing **BGP-LS Attribute TLVs** for these NLRI types (**no new TLVs introduced at this stage**).

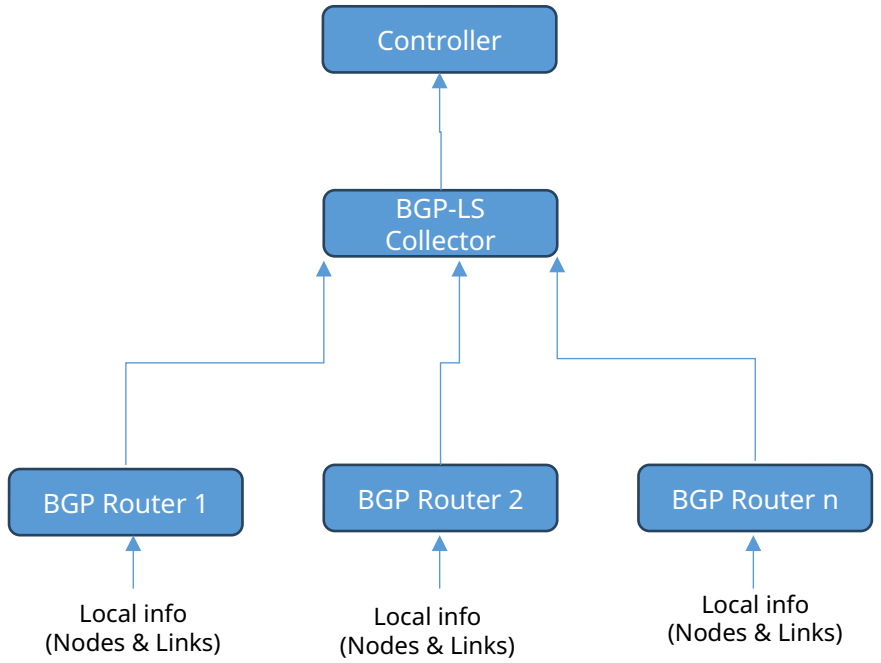
What remains the same?

- The base BGP underlay (IPv4/IPv6) remains unchanged and continues to follow the standard **BGP decision process**.
- Existing BGP peering models and design are retained.
  - Current draft text focuses on **eBGP single-hop** (as in RFC 7938).
  - Other peering models are **not precluded**.
- BGP-LS sessions to controller(s) can be established using reachability provided by the same underlay routing.

Modes of deployment



Distributed topology visibility using RR



Centralized topology visibility using Collector

Topology Information Advertised

- **Node-Level Information**
  - Node Name
  - BGP Router ID
  - ASN
- **Link-Level Information**
  - Link Name
  - Link IDs
  - Link Addresses
  - Bandwidth
  - SR Adjacency SIDs
  - Affinities (Admin Groups/Colours)
  - TE Metric
- **Prefix Information**
  - SR Prefix SID

Other BGP-LS TLVs can be leveraged based on the use case …

Next steps

- Implementations in progress
  - Draft will be updated based on the implementation experience
- Will request code point allocations soon
- Requesting Working Group (WG) review and feedback