Performance-based BGP Routing Mechanism

draft-xu-idr-performance-routing-00

Xiaohu Xu (Huawei)

Hui Ni (Huawei)

Mohamed Boucadair (France Telecom)

Christian Jacquenet (France Telecom)

Ning So (Tata Communications)

Yongbing Fan (China Telecom)

IETF89, London

Motivation

- **Network latency is widely recognized as one of the major obstacles for migrating business applications to the cloud.**
  - **Cloud-based applications usually have very clearly defined and stringent network latency requirements.**
- **Service providers with global reach aim at delivering low-latency network connectivity services to their cloud service customers as a competitive advantage.**
- **Performance routing paradigm is meant to use network latency information as an input to the route selection process.**
- **It's expected that the performance routing paradigm could coexist with the vanilla routing paradigm.**
  - **Service providers could thus provide low-latency routing services while still offering the vanilla routing services depending on customers' requirements.**

Proposed Solution: Rationale

- **Enhance BGP with the ability to disseminate network latency information via a dedicated attribute and take that information as an input to the route selection process.**
- **The solution is designed to be backward compatible with existing BGP implementations and have no impact on the stability of the overall routing system.**
- **This document focuses exclusively on BGP matters.**
  - **All those BGP-irrelevant matters such as the mechanisms for measuring network latency are outside the scope of this document.**

Performance Routing Capability

- **"Performance" (low latency) routes SHOULD be exchanged by means of a specific SAFI (TBD) and also be carried as labeled routes.**
  - **Performance routes can then be looked as specific labeled routes associated with the network latency attribute.**
- **A MP-BGP speaker that advertises "performance" routes SHOULD use the Capabilities Optional Parameter [RFC5492] to inform its peers about the performance route computation capability.**
- **A MP-BGP speaker that implements the Performance Routing Capability MUST support the BGP Labeled Route Capability [RFC3107].**
  - **A BGP speaker that advertises the Performance Routing Capability to a peer using BGP Capabilities advertisement [RFC5492] does not have to advertise the BGP Labeled Route Capability to that peer.**

Performance Route Advertisement

- **Network latency metric is attached to the performance routes as NETWORK_LATENCY path attribute.**
- **Originating performance routes**
  - **A BGP speaker SHOULD be configurable to enable or disable the origination of performance routes.**
- **Distributing a performance route learnt from a BGP peer**
  - **If this BGP speaker has set itself as the NEXT_HOP of such route, the value of the NETWORK_LATENCY path attribute is increased by adding the network latency from itself to the previous NEXT_HOP of such route. Otherwise, the NETWORK_LATENCY path attribute of such route MUST NOT be modified.**
- **To keep performance routes stable enough, a BGP speaker SHOULD use a configurable threshold for network latency fluctuation to avoid sending any UPDATE which would otherwise be triggered by a minor network latency fluctuation below that threshold.**

Performance Route Selection

- **Performance route selection only requires the following modification to the tie-breaking procedures of the BGP route selection decision (phase 2, [RFC4271]):**
  - **Network latency metric comparison SHOULD be executed just ahead of the AS-Path Length comparison step.**
- **The Loc-RIB of performance routing paradigm is independent from that of the vanilla routing paradigm.**
  - **Accordingly, the performance routing table is independent from that of the vanilla routing table.**
  - **Whether performance routing or vanilla routing paradigms would be used for a given packet is a local policy issue which is outside the scope of the document.**

Deployment Considerations

- **It is strongly RECOMMENDED to deploy the performance-based BGP routing mechanism across multiple ASes which belong to a single administrative domain.**
- **Within each AS, it is RECOMMENTED to deliver a packet from a BGP speaker to the BGP NEXT_HOP via tunnels, typically TE LSP tunnels.**
  - **If a TE LSP is used between iBGP peers, it is RECOMMENDED to use the latency metric carried in Unidirectional Link Delay Sub-TLV [draft-ietf-ospf-te-metric-extensions] [draft-previdi-isis-te-metric-extensions] to calculate the cumulative link latency associated with the TE LSP and use that cumulative link latency to approximately represent the network latency. Thus, there is no need for frequent measurement of network latency between IBGP peers.**

Next Steps

- **Comments?**