BGP Link Bandwidth Extended Communities


IETF 120 IDR Meeting
July 22, 2024

Presenters:
Reshma Das (dreshma@juniper.net)
Satya Mohanty (smohanty@zscaler.com)

On behalf of authors
Agenda

- Current Draft Status
- Current Landscape
- Goal
- Proposed Solution
- Current Trend
- Next Steps
Current Draft Status

- IDR adopted since 2009, **expired** since 2018:
- In the draft, link bandwidth extended community is defined as **non-transitive** extended community
  - The value of the high-order octet of the extended Type Field is 0x40
  - The value of the low-order octet is set as 0x04
- The current version talks only about single domain, Intra-AS Case
  - Handling of EBGP is missing (NH self/unchanged)
- **Problem:**
  - Limited Interop as some implementations use transitive LBW and some others use non-transitive LBW extended community.
  - Both transitive/non-transitive versions are deployed in the field
- IDR chairs requested the WG to look into this
Current Landscape

- There are multiple drafts attempting to address this problem using extended communities:
    - New EVPN specific extended community which is **transitive** and uses a **new** format[1] to carry LBW
    - This draft happened as the adopted IDR draft defines LBW community as non-transitive
      (Refer: Appendix-A)
  - [https://www.ietf.org/archive/id/draft-ietf-bess-ebgp-dmz-03.html](https://www.ietf.org/archive/id/draft-ietf-bess-ebgp-dmz-03.html)
    - Works over the adopted LBW draft
    - Introduces neighbor level knob to advertise and accept non-transitive LBW extended community across an EBGP session
    - This uses an ambiguity in the base RFC-4360 (Ext-Comm) that also needs to be addressed in a new 4360-bis, keeping into consideration the DMZ draft use cases
    - Introduces another new extended community with **new** format[2]
    - Recommends both transitive/non-transitive extended community usage
- MNH uses a new per next hop scoping to carry LBW
  - [https://www.ietf.org/archive/id/draft-ietf-idr-multinexthop-attribute-00.html#section-5.4.4.1](https://www.ietf.org/archive/id/draft-ietf-idr-multinexthop-attribute-00.html#section-5.4.4.1)
  - Out of scope for this discussion
Goal

• Achieve interop of existing LBW extended community with minimum changes to procedures.
• Keep the changes to procedure simple.
  • Assumption: Receiver runs upgraded code to be able to interop.
  • It is desirable that RR is transparent, doesn't modify LBW when reflecting
• Cover all applicable use cases. (IntraAS, InterAS)
• Define Error handling
• Revive and retain existing draft (draft-ietf-idr-link-bandwidth)
Proposed Solution (1/2)

A. Sender (originating link bandwidth community) :
   An originator of the link bandwidth community SHOULD be able to originate either a transitive or a non-transitive link bandwidth extended community. Implementation SHOULD provide configuration to set the transitivity type of the link bandwidth community. No more than one link bandwidth extended community SHALL be attached to a route.

B. Receiver (receiving link bandwidth community) :
   A BGP receiver MUST be able to process link bandwidth community of both transitive or non-transitive type. The receiver SHOULD NOT flap or treat the route as malformed based on the transitivity of the link bandwidth community.
Proposed Solution (2/2)

C. Conflict Management:
   If a receiver receives a route with more than one link bandwidth community then it SHOULD:
   1. Prefer the lowest value of the attached link bandwidth community (Irrespective of the transitivity).
   2. Prefer the transitive link bandwidth extended community when choosing between transitive and non-transitive types that have the same value.
   3. Implementations MAY provide knobs to change the preference in (1) and (2)

D. Re-advertisement with Next hop Self :
   Follow the same procedures as A.

E. Re-advertisement with Next hop unchanged :
   A BGP speaker that receives a route with link bandwidth community, re-advertises or reflects the same without changing its next hop SHOULD NOT change the link bandwidth extended community in any way.
Current Trend

- Deployments need both transitive and non-transitive version to be supported by all vendors.
- Juniper implementation now supports sending/receiving both transitive and non-transitive form of LBW extended community.
- Incremental support from a few other vendor implementations for both transitive and non-transitive versions are seen.
- Ongoing discussions with other vendors who are interested in LBW community.
Next Steps

• Revive the expired document
  • Modify and extend the existing draft to ensure backward compatibility and interoperability of solutions
  • Expand the draft to include various use cases
  • Invite participation
• Request Interim to discuss further
• Requesting other vendor participation in Hackathon to demonstrate interop
• Standardize the solution
• Follow on work: 4360-bis
Thank you.
Example Topology for Transitive LBW Community

- RR22 and RR11 are service RRs connecting PE22 and PE11/PE12
- RR22 and RR11 is connected via multi-hop EBGP session.
- The load balancing point for service routes in this topology is PE22 towards PE11 and PE12.
- The link bandwidth extended community cannot be advertised over EBGP peers as it is defined to be optional non-transitive.
- So, the link bandwidth community needs to preserve transitive capability also.