Carry Congestion Status in BGP Community

draft-li-idr-congestion-status-community-06

Zhenqiang Li
China Mobile

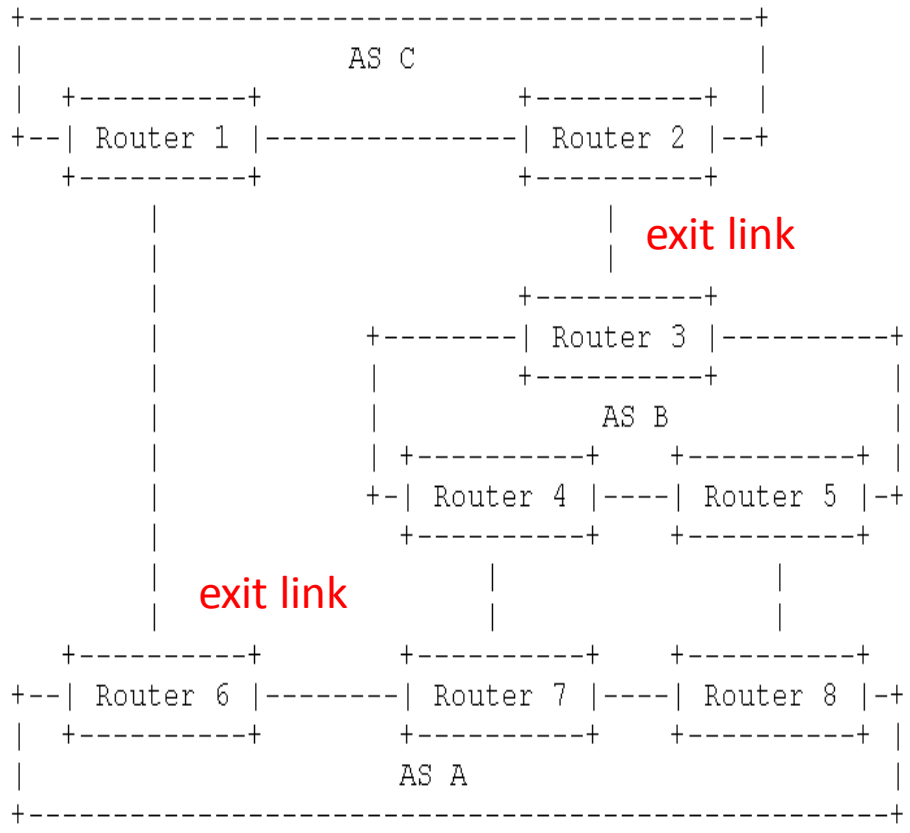
Jie Dong
Huawei Technologies

Scenario to be addressed

- In some circumstances, we need to know the link congestion status, both the link bandwidth and its utilization
  - Egress Peer Engineering using BGP-LU [I-D.gredler-idr-bgplu-epe] needs this kind of information to aid MPLS routers to do traffic-engineering on inter-AS links
  - Constrained Multiple BGP Paths [I-D.boucadair-idr-constrained-multiple-path] wanted to use the link congestion information as a metric to select the appropriate path among multiple BGP paths. The purpose was to do proactive reaction to the congestion events.
  - To aid a router to perform unequal cost load balancing, Cisco introduced Link Bandwidth Extended Community in [link-bandwidth-community] to carry the cost to reach the external BGP neighbor. The cost can be either configured per neighbor or derived from the bandwidth of the exit link

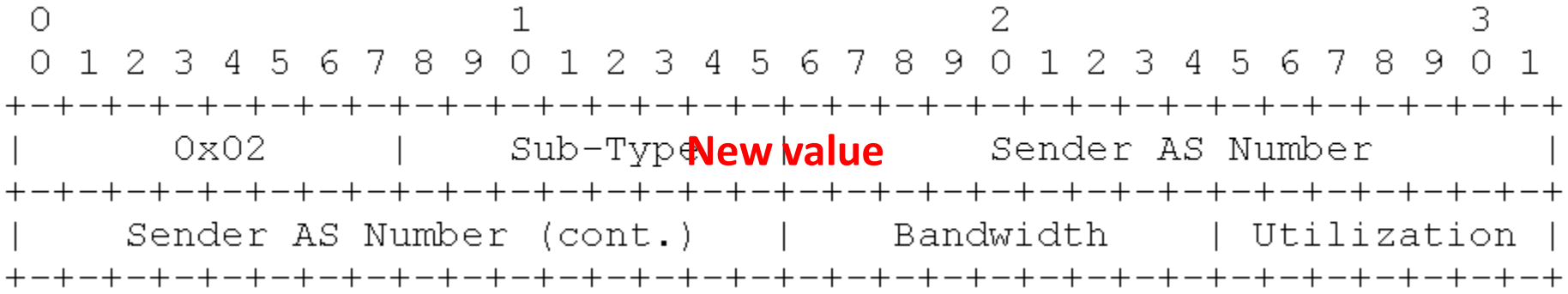
Scenario to be addressed

- General Benefits of knowing the link congestion status
- For the BGP speakers within one AS.
  - If the exit routers, Router 7 and 8, tell their iBGP peers in AS A the congestion status of the exit links, the peers in turn can steer some outgoing traffic toward the less loaded exit link.
- For the BGP speakers across multiple ASes.
  - Due to cost or network performance, AS A prefers AS B to access AS C.
  - If AS A knows the link between router 2 and 3 is congested, it can steer some traffic towards AS C from AS B to the directly connected link between router 1 and 6.
- The province network of China Mobile wants to do AS outgoing load balance by using the congestion status information of the exit links.

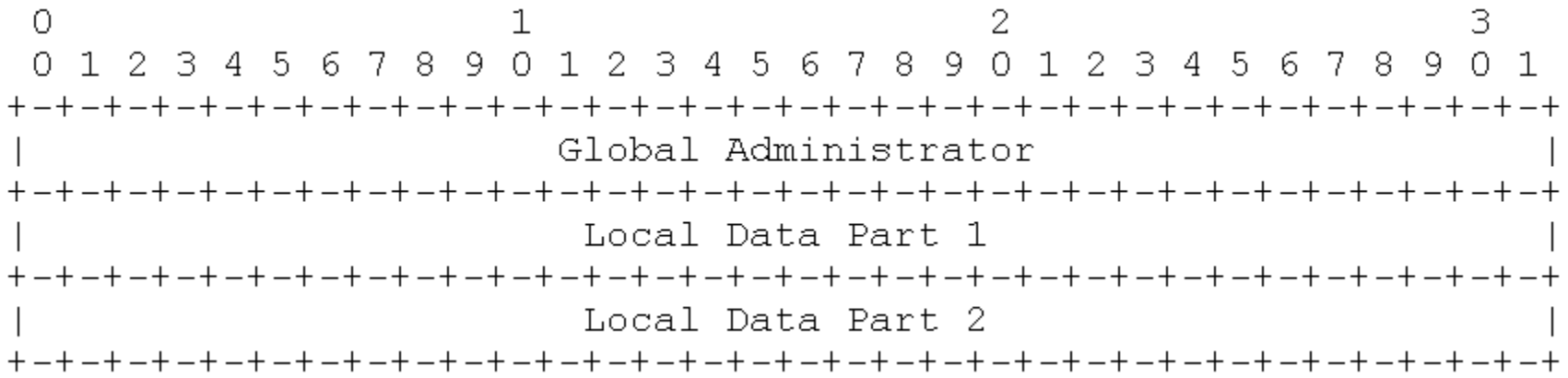


Solution Alternatives

- Option 1: Extended Community

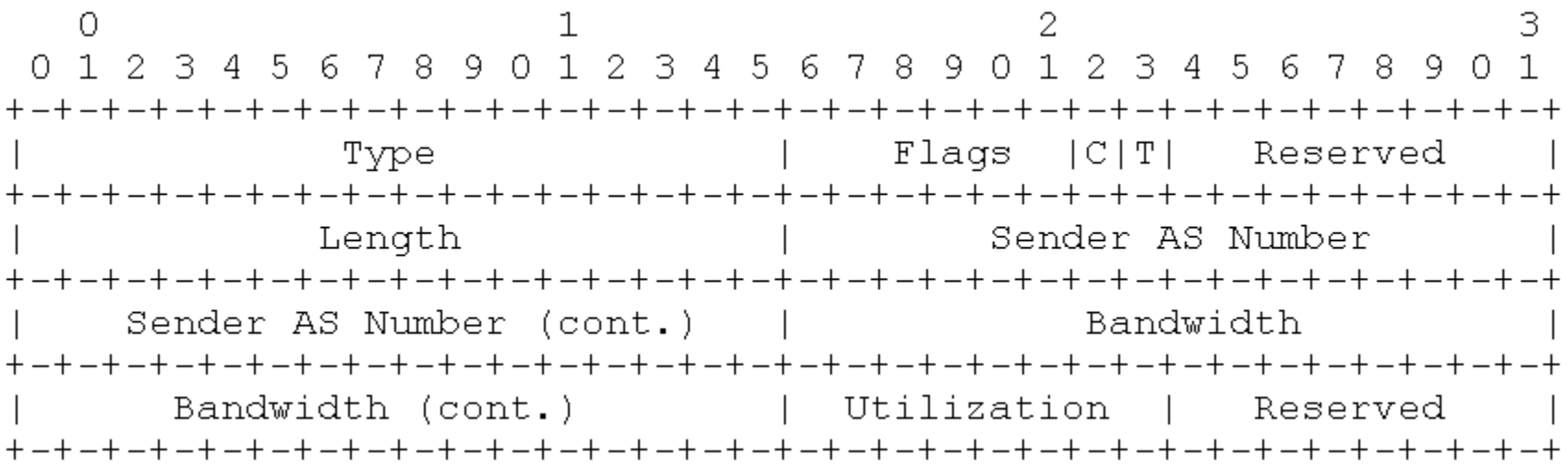


- Option 2: Large Community



Solution Alternatives

- Option 3: Community Container



- **Type:** A new type is to be defined for the congestion status community
- **Sender AS Number**: in 4 octets, the AS number of the BGP speaker who generates this community
- **Bandwidth**: in 4 octets, with unit of mbps

Deployment Considerations

- To avoid route oscillation
  - the exit router SHOULD set a threshold. The exit router generates BGP update messages with congestion status community only when the link utilization change reaches the threshold.
  - The method similar to BGP Route Flap Damping is RECOMMENDED for the implementations to further reduce the BGP update messages trigered by link utilization change.
  - To reduce the update churn introduced, when one BGP router needs to re-advertise a BGP path due to attribute changes, it SHOULD update its Congestion Status Community at the same time.

Deployment Considerations

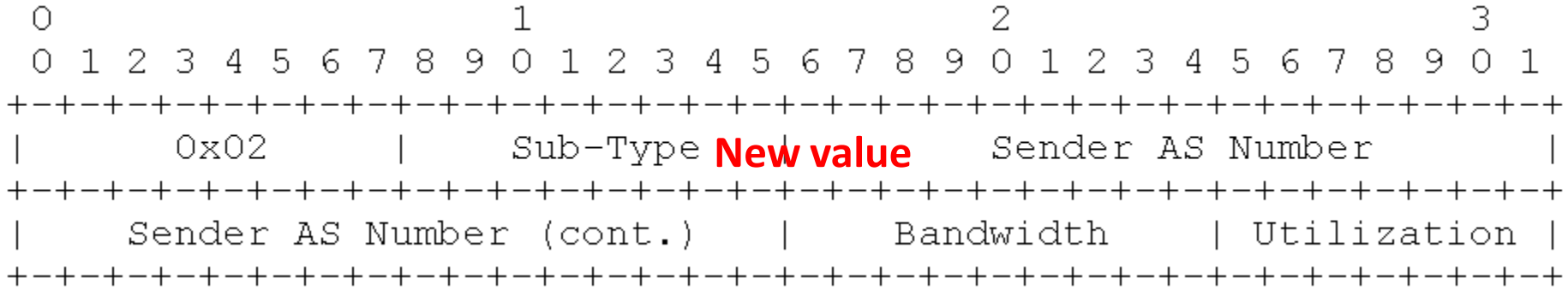
- To avoid traffic oscillation
  - Route policy can be set at the exit router. Congestion status community is only conveyed for some specific routes or only for some specific BGP peers.
  - If the congestion status community is used by a SDN controller, the controller can steer the Internet access traffic among all the exit links from the perspective of the whole network.
- To avoid forwarding loops
  - the reception of such community over IBGP session SHOULD NOT influence routing decision unless tunneling is used to reach the BGP Next-Hop.

Security Considerations

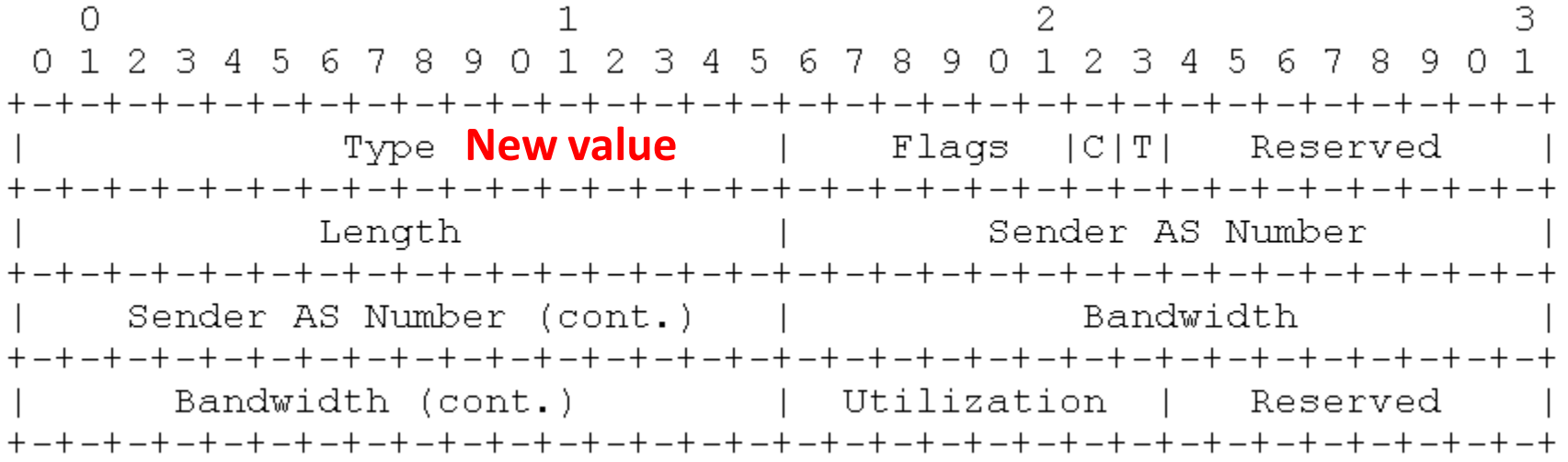
- The BGP receiver has the right to trust the congestion status communities or not.
  - The BGP receiver may choose to only trust the congestion status communities generated by some specific ASes or containing bandwidth greater than a specific value.
  - You can filter the congestion status communities at the border of your trust/administrative domain. Hence all the ones you receive are trusted.
  - You can record the communities received over time, monitor the congestion e.g. via probing, detect inconsistency and choose to not trust anymore the ASes which advertise fake news.

IANA Requirements

- Solution 1: Extended Community



- Solution 3: Community Container



Next steps

- Accepted by the working group?

Thanks