Satya R Mohanty, Aaron Millisor
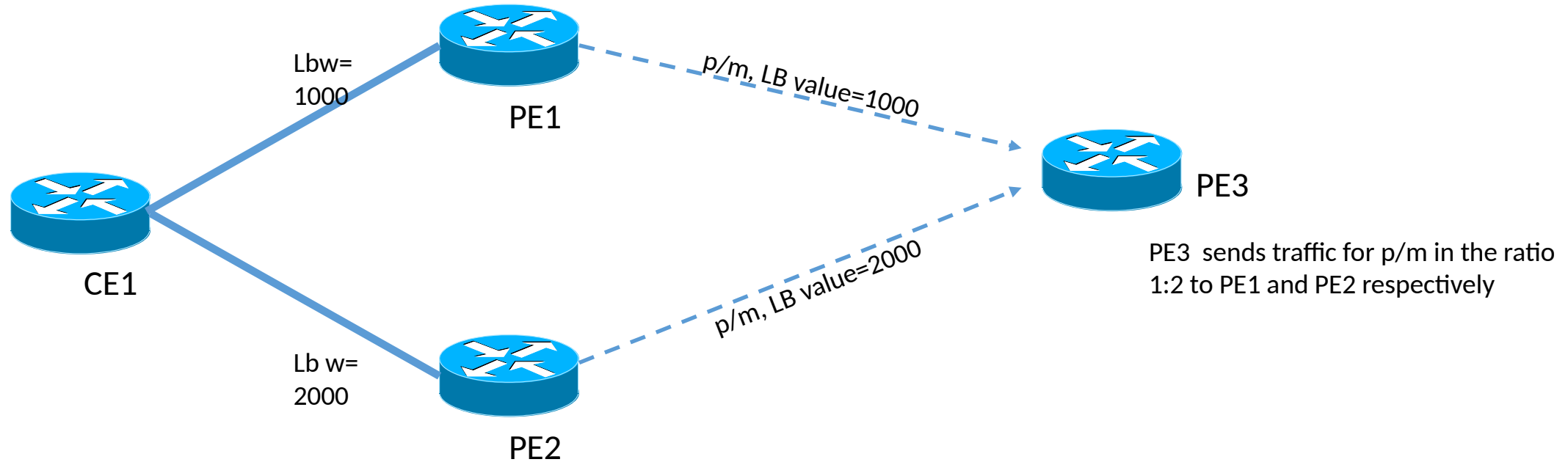Cisco Systems

Arie Vayner
Google

Cumulative DMZ Link BW in Data Center

- DMZ link Bandwidth Extended Community

- Requirement for Accumulated DMZ in EBGP Network

- Requirement for New knobs

# DMZ Link Bandwidth Extended Community

- Defined in draft-ietf-idr-link-bandwidth-07.txt

- The LB extended community encodes the value of the EBGP egress's link bandwidth

- Useful in load balancing across Inter-AS Links where the same prefix/net is advertised into the receiving AS via different egress-points or next-hops

DMZ Link Bandwidth and load balancing

Lbw= 1000

PE1

p/m, LB value=1000

CE1

PE3

PE3 sends traffic for p/m in the ratio 1:2 to PE1 and PE2 respectively

Lb w= 2000

p/m, LB value=2000

PE2

- LB extended community is Optional Non-Transitive.
- Is Not set when the next-hop is set to Self
- All the N multi-paths need to be associated with a DMZ Link Bandwidth attribute. Otherwise the attribute wont be downloaded to RIB.
- FIB Hashing takes into account the DMZ Link Bw value to load-balance the traffic

Requirement for Cumulative link Bandwidth (Unfolded Clos)
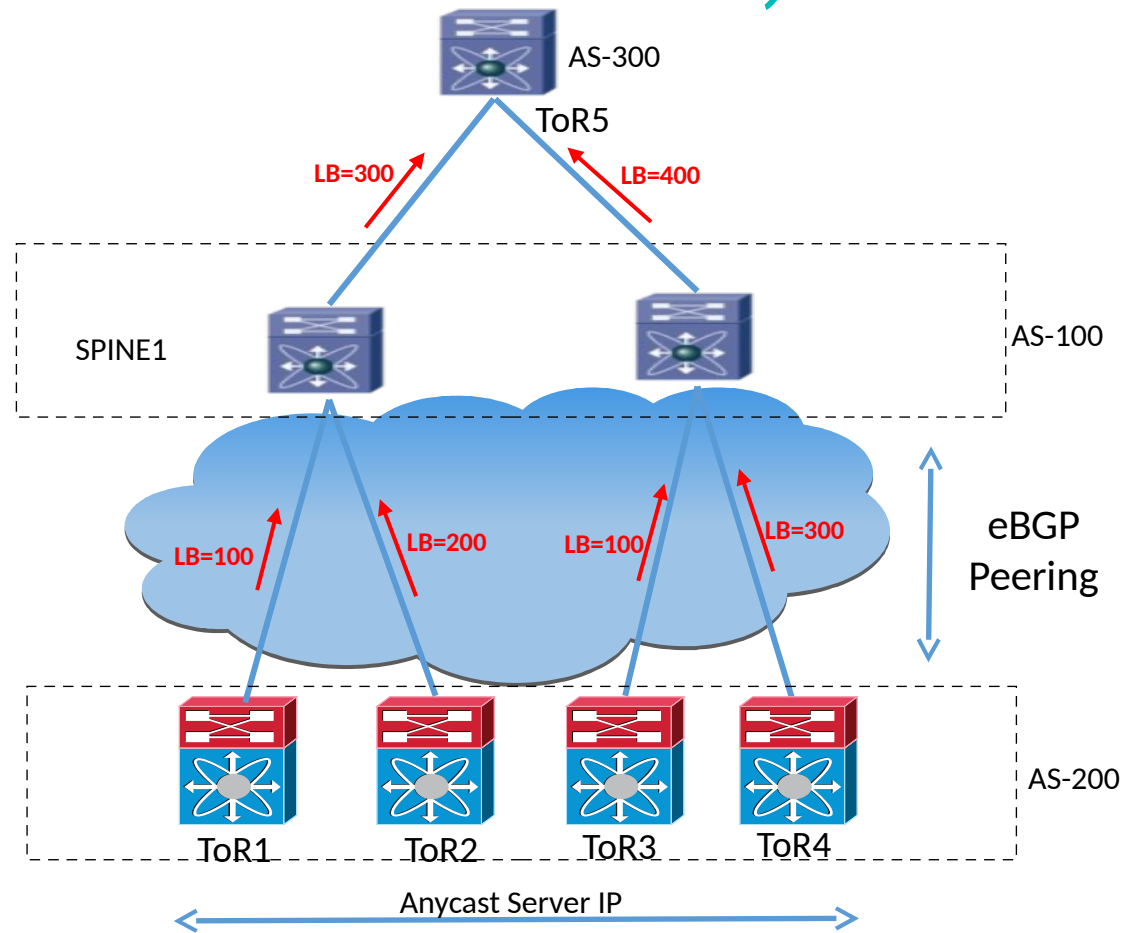
- TOR1 and TOR2 generate link bw extcom with LB value 100 and 200
- Spine1 adds up ("accumulates") both and sends 300 to TOR5
- Spine2 similarly adds up and sends 400 to TOR5

Enhancements to existing DMZ

DMZ Link BW in the data center

- Accumulate the DMZ across all eligible multi-paths

- Although the LBw extended community is still Optional Non-Transitive, it needs to be be advertised to EBGP peers. Similarly the extended community will need to be received from an EBGP peer.

- Use of new Configuration knobs to enable this behavior for this "draft violation"

- Existing customers can continue to deploy link aggregation or cumulative link aggregation as before. They do not need these knobs

Enhancements to existing DMZ

DMZ Link BW in the data center

Existing Knobs and Requirement for New Ones [Cisco IoS XR BGP Example]

- bgp multipath as-path ignore onwards

    This knob ignores all criteria starting from as-path for multi-paths to be eligible. [Existing knob]

- bgp bestpath as-path multipath as-path relax

    This knob allows load balancing across ASes. However, while the AS path need not be same the total length should be same. [Existing knob]

- ebgp-send-extcommunity-dmz cumulative

    This neighbor specific knob enables advertisement to the EBGP peer. [This is newly implemented for this ask]

- ebgp-recv-extcommunity-dmz

    This neighbor specific knob enables receiving the community from an EBGP peers. [This is newly implemented for this ask]

What is not attempted

DMZ Link BW in the data center

- The directionality aspect is not taken care of here.

- A receiving peer's contribution towards the accumulated bandwidth is not factored out in the advertisement towards that peer.

- So, deploying the feature in a symmetric way needs to be carefully thought out, otherwise, it may lead to churn because of continuous updates.