IETF 123 – Madrid
July 2025
IDR Working Group

draft-krierhorn-idr-upa-00
Unreachable **P**refix **A**nnouncement

Jakub Horn, Serge Krier (Cisco Systems)

Mihai Ciurea (Swisscom)

Problem Statement

- Underlay route aggregation/summary is very useful for large SRv6 networks
  - Same applies to VxLAN and MPLSoUDP
- ISSUE : Aggregate hides failure of egress PE to ingress PE
  - results in overlay BGP convergence
  - slow to send MP_UNREACH for all service NLRIs
- GOAL
  - trigger BGP PIC mechanism to inform ingress PE about egress PE failure behind the aggregate
- Integrates with IGP mechanism

  draft-ietf-lsr-igp-ureach-prefix-announce

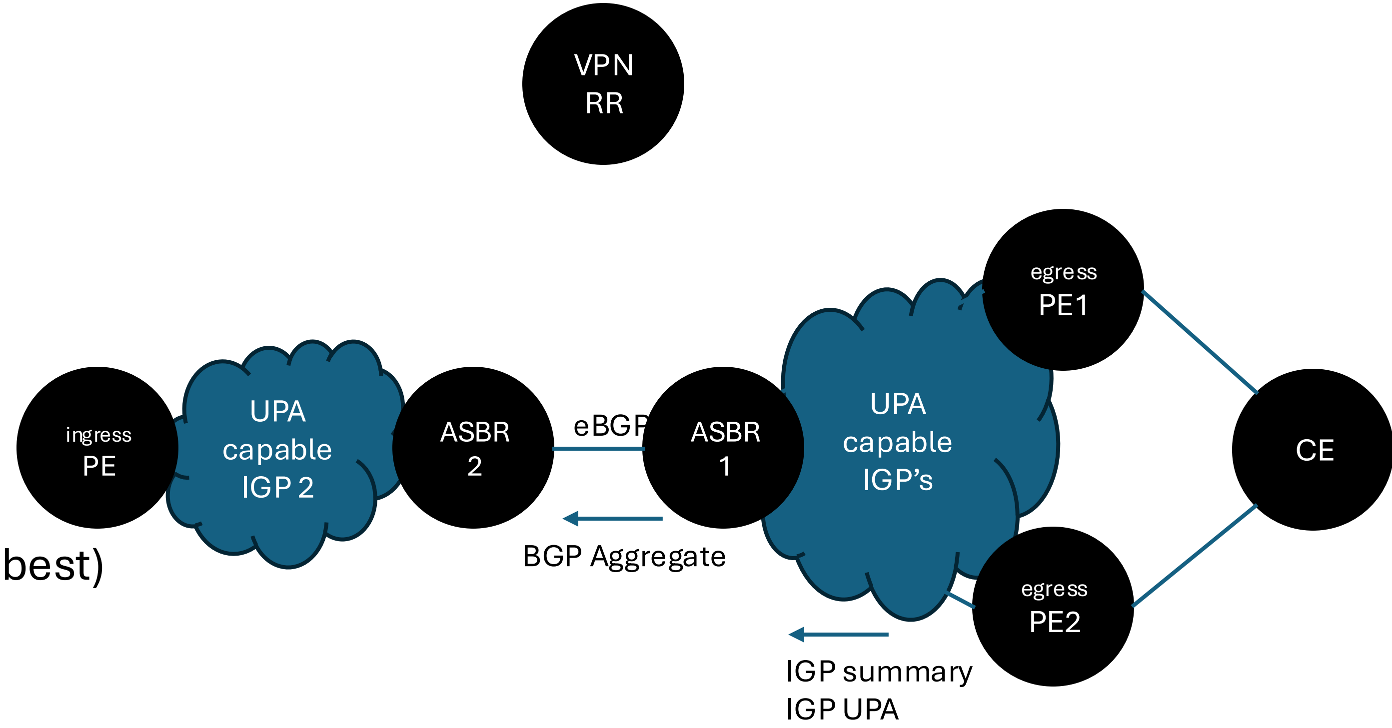
Regular BGP Overlay Convergence

without BGP UPA

Egress PE1 failure

- PE1 to VPN RR session timeout
  - Service Layer
- RR withdraws of PE1 originated prefixes
  - Service Layer
- iPE will reprogram via alternate PE2 paths (next best)
  - PIC not triggered

- PREFIX dependent convergence



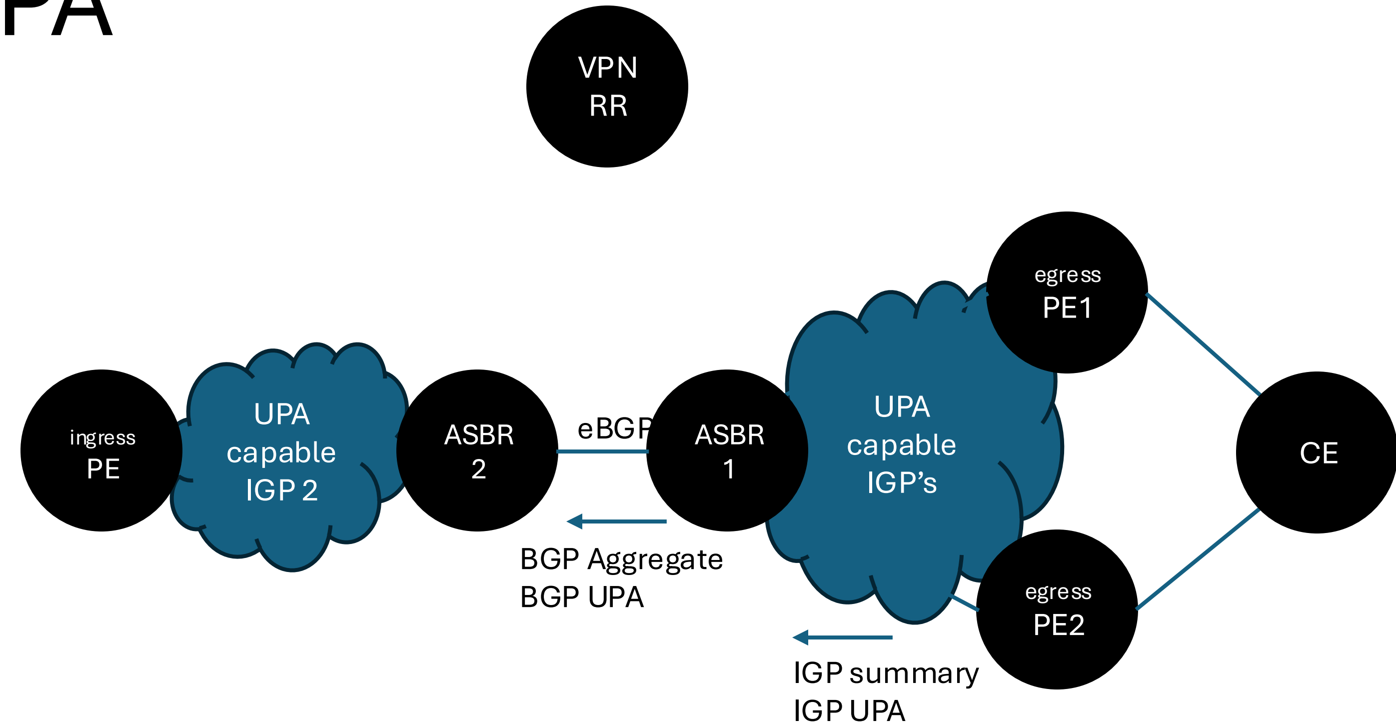
Convergence With BGP UPA

Egress PE1 failure

- ASBR1 will see PE1 failure
  - Service Layer
  - Faster because in same IGP

- ASBR1 generates UPA prefix for PE1 to ASBR2
  - ASBR2 advertises it to iPE via IGP UPA

- iPE triggers BGP PIC upon reception of UPA route



Several options considered to create UPA

1. MP_*_UPA (new attribute codepoint)
   - new msg type needed (??) because bgp update (type 2) needs either MP_REACH OR MP_UNREACH
   - No Backward compatibility and difficult to deploy
2. MP_ REACH + UPA community
   - For speaker without UPA community support looks looks like regular MP_REACH !! Danger !!
3. MP_UNREACH + UPA community
   - speaker without UPA support will ignore, not propagate
   - speaker with UPA support will react and propagate to UPA enabled peers
4. MP_UNREACH without community
5. New safi
   - Similar to (1)

Proposing Option 3 as best for backward compatibility and simplicity

Timer

- UPA announcement
  - Mid point BGP [UPA-aware] speaker will propagate
  - Processed by ingress PE to trigger BGP PIC
- Timer usage
  - Mid point BGP can receive multiple UPA for same prefix
    - Within time period subsequent UPA for same prefix are not propagated
  - Ingress PE holds UPA prefix until full BGP convergence in overlay
- What if egress PE becomes reachable again ?
  - UPA state cleared by "classic" MP_UNREACH (w/o UPA ExtCom)
  - Handled via UPA timer

Applicability/Scope

- should be off by default
- scope is not Internet
- very specific to transport router reachability

Next Steps

- Requesting WG review and feedback
- Feedback on signalling options
- Feedback on egress PE restored reachability