# BGP
# Next-Hop Capabilities

draft-decraene-idr-next-hop-capability-00

Bruno Decraene Orange

# Introduction: Recaps on RFC 5492 "Capabilities Advertisement with BGP-4"

- RFC 5492 (BGP capabilities) advertises capabilities of the BGP peer.
  - BGP session related
  - (BGP) control plane capabilities
- The BGP peer may not be the BGP Next Hop:
  - Route Reflection (iBGP)
  - Route Server (eBGP)
  - Next Hop unchanged (not setting Next Hop Self)
- Hence not a way to advertise capability of the BGP Next-Hop.
  - Forwarding planes capabilities

# Next-Hop Capabilities encoding

1. New non-transitive BGP Attribute
2. Carries set of Next-Hop Capabilities
3. A Next-Hop Capability is encoded as a TLV

- In short:
  - same encoding as BGP capabilities but carried in a non transitive attribute

# Next-Hop Capabilities operation

- We want the capability to be removed when the Next-Hop is changed.
- → For compliant peers:
  - if Next-Hop unchanged: attribute SHOULD be passed unchanged
  - if Next-Hop changed: attribute MUST be removed
    - new one may be attached to reflect capabilities of the new Next-Hop
- → For non compliant peers:
  - As the attribute is non-transitive attribute, it will be removed (as per RFC 4271).

# Error handling

- Error condition: lengths mismatch
  - attribute length mismatch the sum of (capabilities lengths+2)
- Error handling: “attribute discard”
  - Assuming implementations do not allow changing route preference based on Next-Hop Capabilities…
  - Is this a safe assumption? Otherwise “treat as withdraw”?
    - or “attribute discard” on eBGP, and “treat as withdraw” in iBGP?

# 1 generic BGP Next-Hop Capabilities Attribute vs N BGP attributes (1 per application)

- Why defining a generic attribute?
- For IDR / implementations: doing the work once
  - single attribute used / single doc
  - single spec/coding/tests
- For the application: incremental deployment
  - non-transitive attribute required
  - a new non-transitive attribute would be unknown hence removed by existing implementations

# First application proposed: "Entropy Label" Next-Hop Capability

- Capability sent if either:
  - BGP Next-Hop can process Entropy Label
  - BGP Next-Hop will perform a MPLS SWAP and not have to process Entropy Label
- When received, means: may send packets with a MPLS entropy label for this Next Hop/NLRI
- Based on the ELC BGP attribute defined in section 5.2 of [RFC6790] but then deprecated.
- Do we want to also advertise the Readable Label Depth?
  - number of labels readable by transit LSR for ECMP load-balancing hashing
    - as defined in draft-ietf-mpls-spring-entropy-label
  - could be RLD of NH or RLD from NH to egress (NLRI).
  - In the Value field? In a different NH Capability? (as RLD is independent of ELC)

# Next

- Feedback & comments welcomed.

Thank you