IETF WG for Inter-Domain Routing (aka BGP)

June-2024

IDR chairs – Sue Hares and Jeffrey Haas

Who are we?

Sue Hares

- 80s – ISP + Factory networks
- 90's – NSFNET + NANOG (Merit) + GateD (Freeware to 200 companies)
- '00-'09 – Gated NextHop (100 Companies)
- '10-'13 – D.E. at Futurewei
- '13-'24 – IDR Consultant (Huawei) + PhD

Jeffrey Haas

- Worked at small tier-3 ISP in the late 90's.
- Worked for Sue at NextHop on BGP and IETF for several years.
- These days, implementations and specifications as D.E. at Juniper.

Isn’t BGP done already?

- The core BGP protocol and main extensions used by Internet Service Providers is fairly stable.
  - Even then, SPs want new features on occasion. For example, Large BGP Communities! (RFC 8092)
  - Security extensions are also popular work.
- BGP is heavily used for VPNs and BGP services (multicast, EVPN, etc.) these days.
- Work is standardized in IDR in IETF.
- We have a wiki! https://wiki.ietf.org/en/group/idr

You can participate in standards

- IETF is an open standards organization.
- Work is split between mailing lists and in-person meetings three times a year.
- … and sorry, work doesn't always happen fast.

- If you want to learn more about participating in IETF, see us after the talk.

BGP in 2024

- Intent-based (Color) Routing - CAR, CT, CPR
- Service Routing / BGP-LS +++
- BGP YANG Model - Awaiting implementations
- Fixing “Stuck” BGP sessions
- BGP QUIC
- Flow Specification version 2 - Breaking it into “chunks”
- Old is New – Graceful Restart, BFD Strict Mode, IPsec tunnels

Intent-based Routing in BGP

Color indicates Intended Service Level

Intent-based Routing in BGP

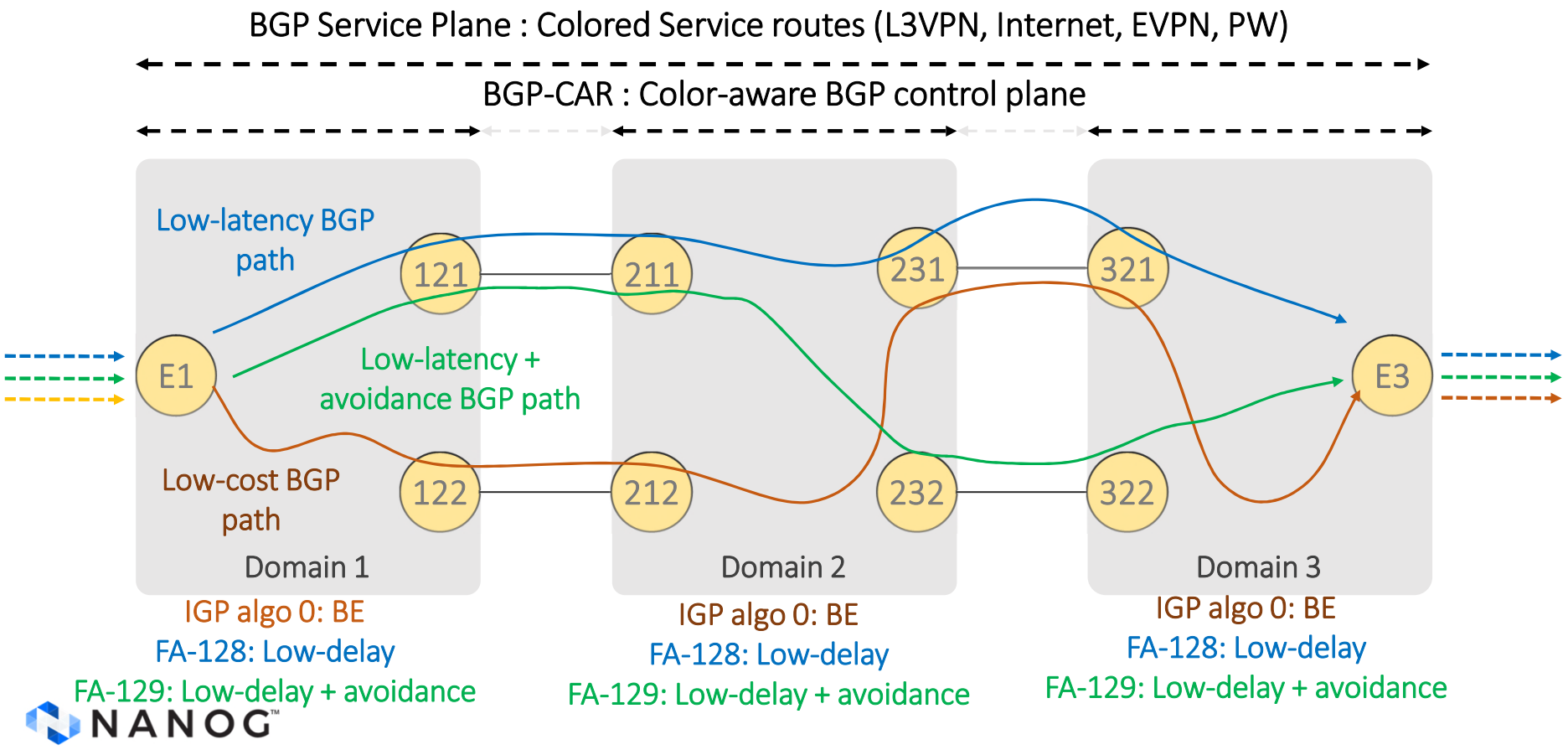
- Intent – User being able to signal desired intent
  - Service routes into color-based (intent) Transport
  - Service routes marked with colors
- CAR (Color-Aware Routing) and CT (Classful Transport)
  - Each have a new address families (SAFIs for AFI/SAFI)
  - Two drafts “functionally equivalent” but operationally different (IDR chairs)
  - Spent 4 years refining – due to IDR WG interest
  - 3 use cases the authors (from IETF-114)
- CPR (Colored Prefix Routing)
  - Intent for SRv6 – done through SID assignment
  - draft-ietf-idr-cpr-02 (informational) [Past WG LC in May]

BGP-CAR Use Cases

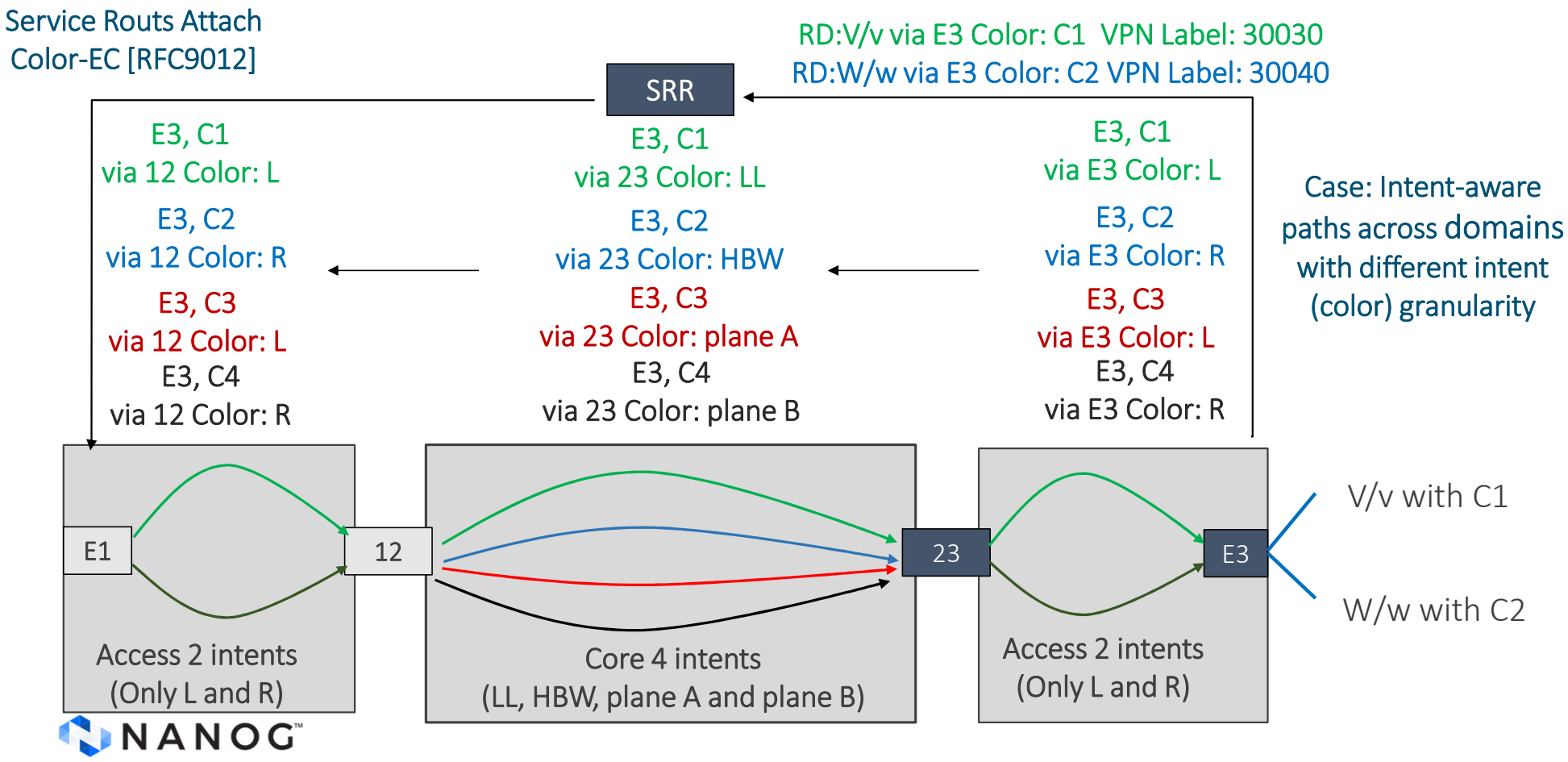
draft-ietf-idr-bgp-car

BGP-CAR: Inter-Domain Multiple color-aware paths (intent)

Base case: Intent-aware paths to a specific transport endpoint (e.g., PE loopback)

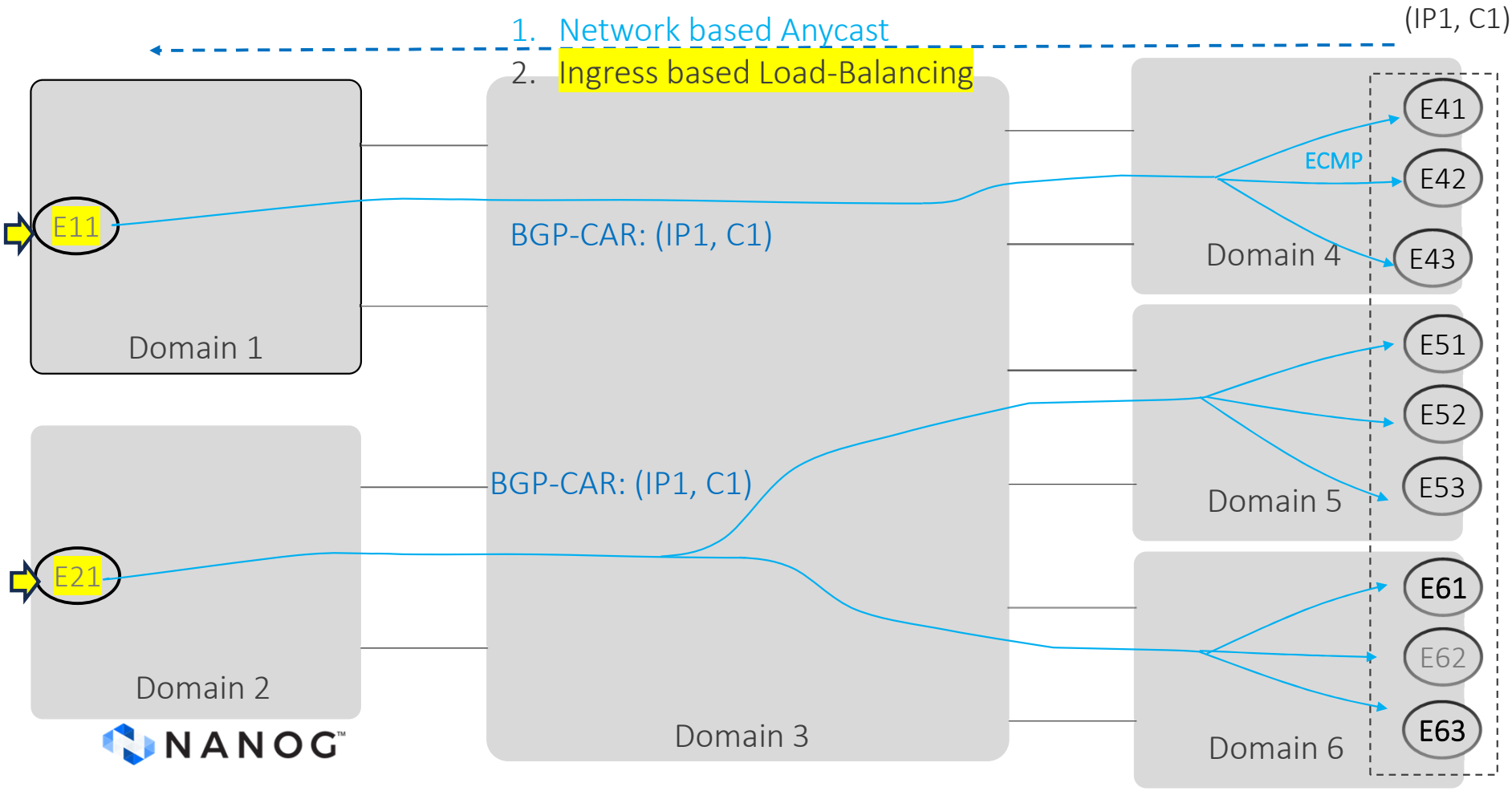


BGP-CAR: 4 intents to 2 colors (Transport)



NANOG™

BGP-CAR: Transport Anycast – Color + IP Address (Anycast)

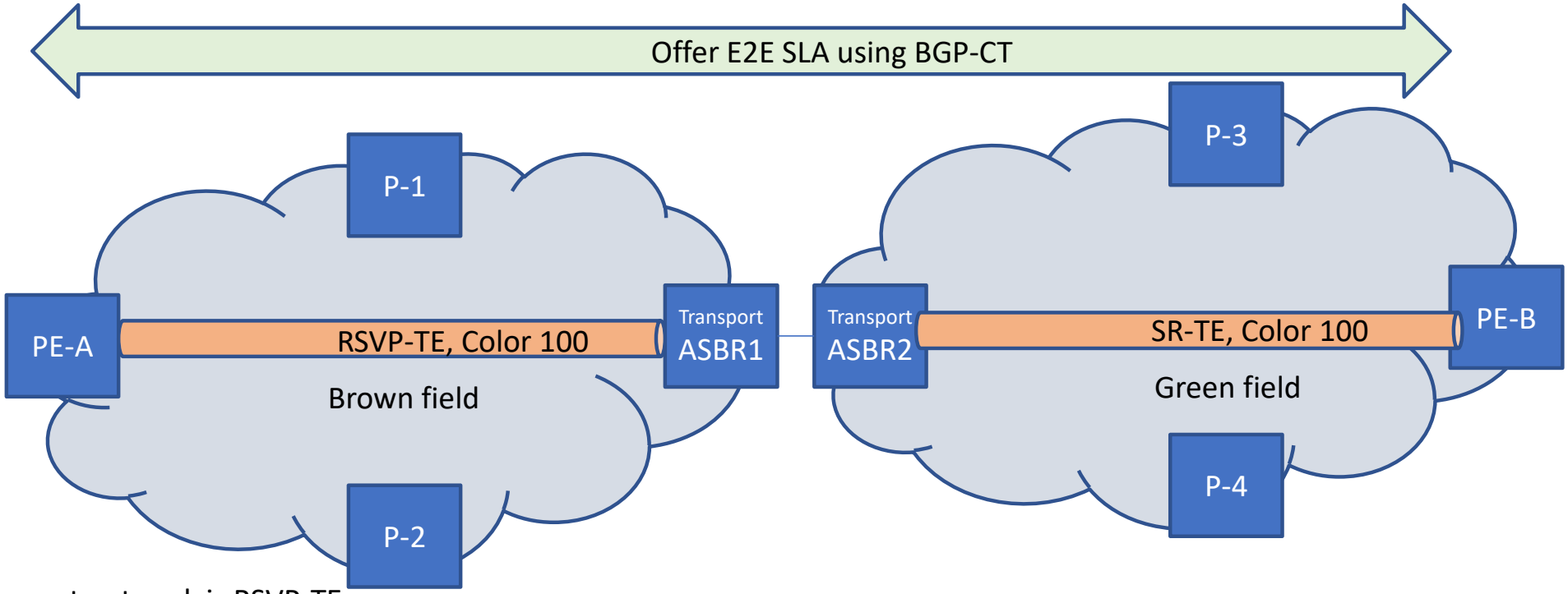


BGP-CT Use Cases

draft-ietf-idr-bgp-ct

BGP CT: Customer Use Case 1 – AT&T

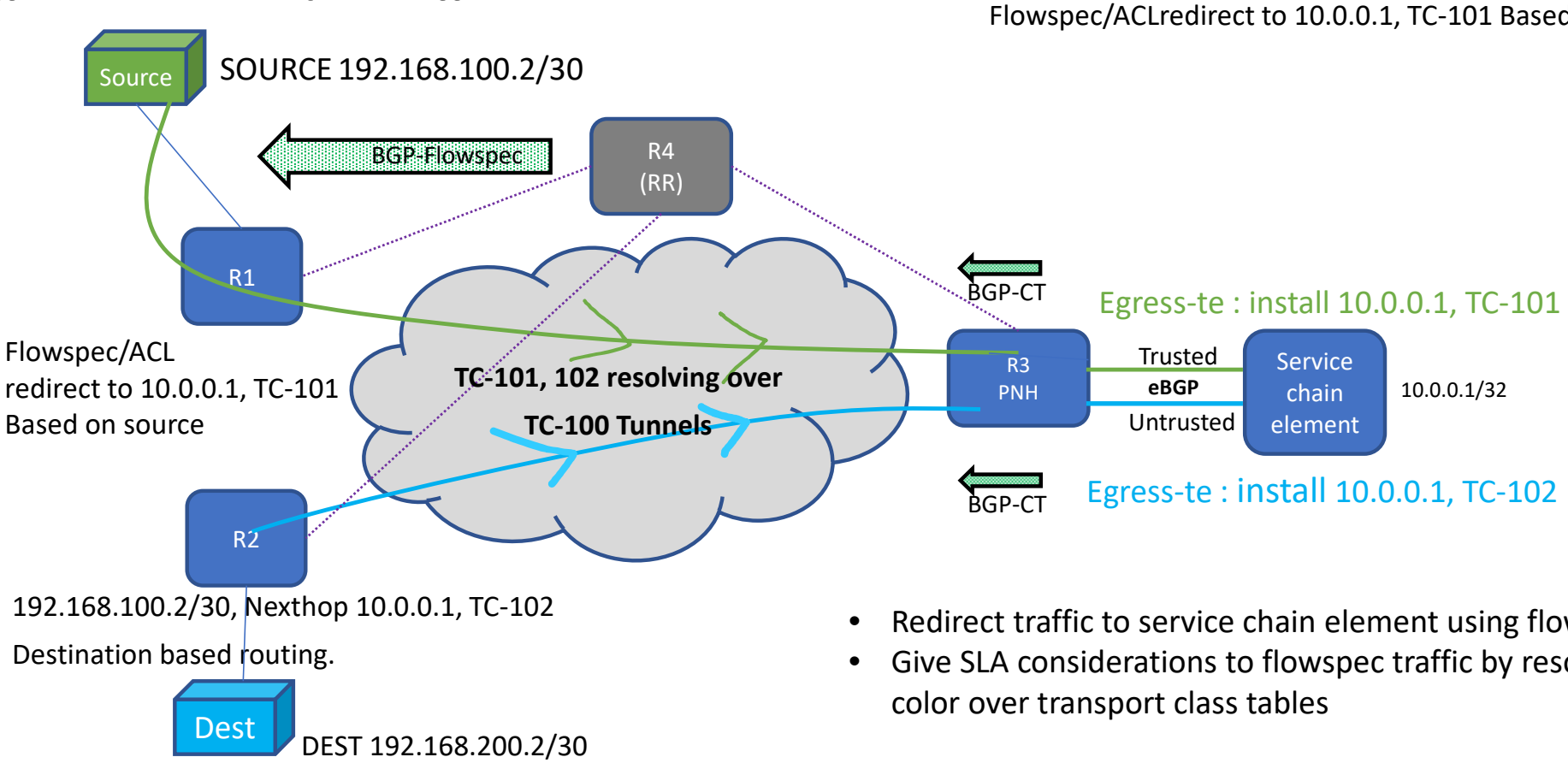
RSVP-TE/SR-TE coexistence during migration



- Current network is RSVP-TE
- Looking to introduce SR-TE in the newer network
- Need to maintain E2E SLA across both networks and RSVP-TE needs to understand color. Transport class provides route resolution accordingly

BGP CT: Customer Use Case 2

Offer SLA to Flowspec Traffic over RSVP-TE/SR tunnels



- Redirect traffic to service chain element using flowspec
- Give SLA considerations to flowspec traffic by resolving color over transport class tables

Flowspec redirect traffic to “BGP-CT EPE” end-points

BGP CT: Customer Use Case 3 – APAC

Network slicing across heterogenous color domains

“Currently, we have separate networks for domestic and international and they are independent so as the TE policies. We intentionally maintain the autonomy and modularity for administrative purposes. When we need inter-domain intent awareness, we would need the same level of flexibility in the proposed solution.

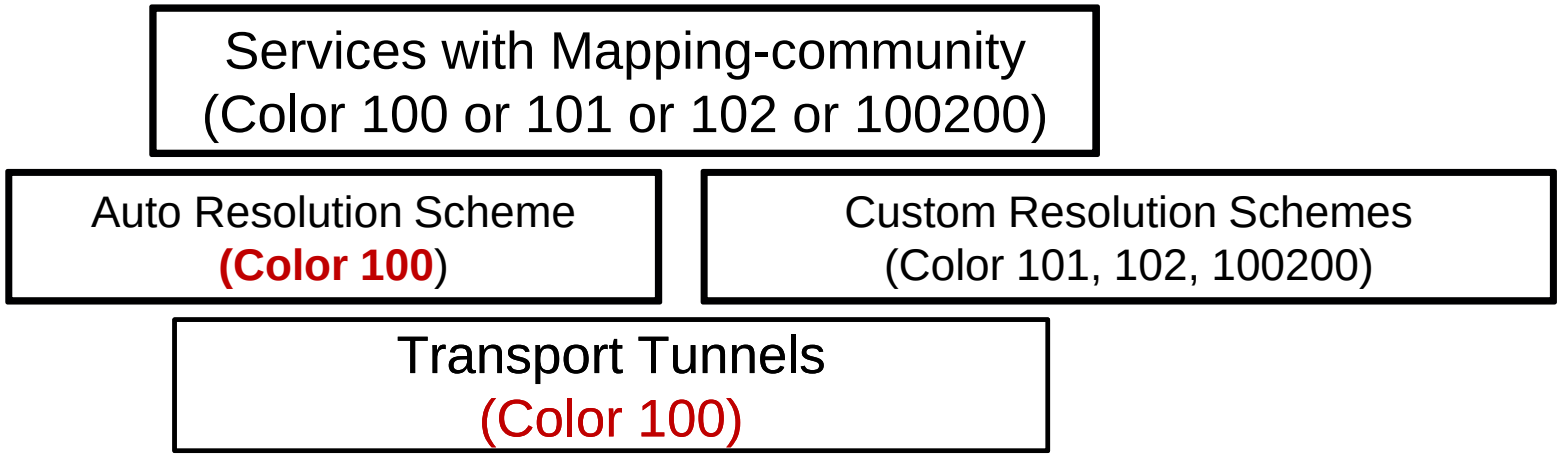
I would also like to highlight, service provider networks usually have more meshed paths in the core and aggregation domains where more granular intents can be realised. However, the access network domain will have less number of paths (either left or right in a ring / partial mesh / hub and spoke – in regional remote areas) where we would need only a few discrete transport classes / colours.

Hence, requirement for remapping of transport classes / colours within a single AS shouldn’t be considered as a corner case in my opinion.”

- Moses Nagarajah (Telstra Networks)

BGP CT: Customer Use Case 3 – Solution

Heterogenous Color Granularity, Customizing resolution schemes



Service Layer

- Carve out a service route mapping-community space across the AS domains
- Each mapping-community in this space is an “abstract value” identifying an SLA (e.g. color: 0:100200)

Transport Layer

- Customized resolution-schemes for BGP-CT family routes in relevant AS domains to use available colors
- Mapping-community **transport-target:0:101** (**Medium Red**) and **transport-target:0:102** (**Light Red**) can be
  - Custom Mapped to **tc-100 [Red]** in **AS Metro Domain A and C**
  - Strictly Mapped to **tc-101[Medium Red]** and **tc-102[Light Red]** respectively in **AS Core Domain B**

BGP-LS

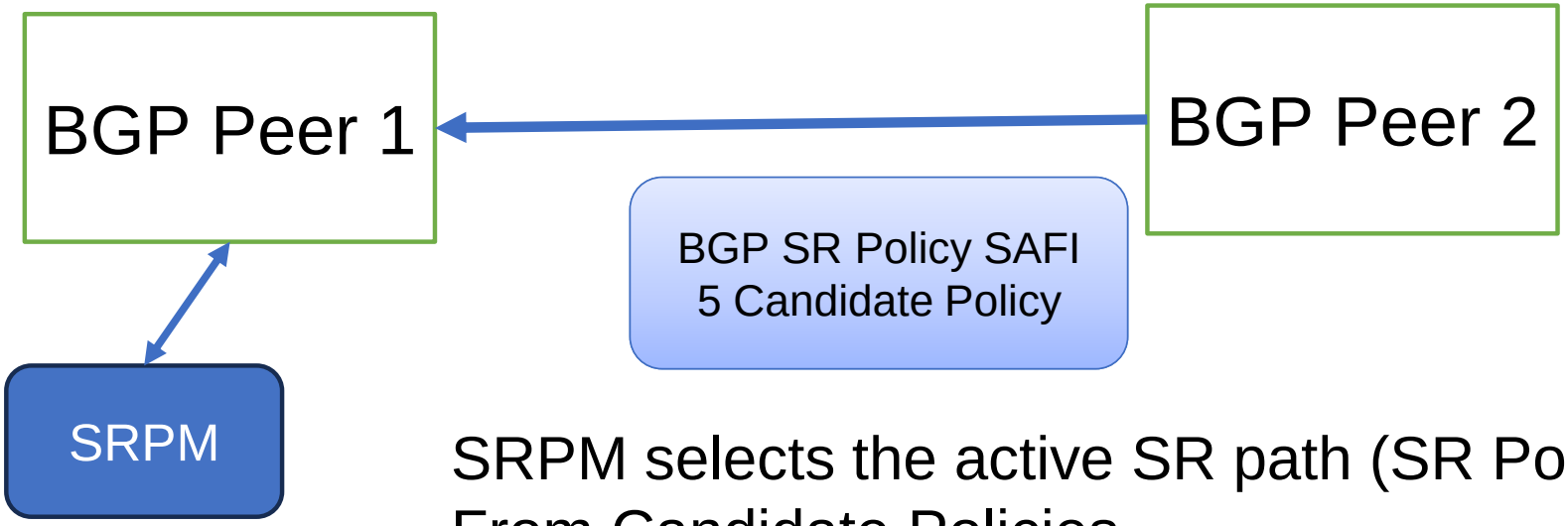
BGP-LS + Service Routing (SR)

- Base BGP-LS Revised (RFC7752 to RFC9552)

- BGP passes SR (Segment Routing) Candidate SR Policy (tunnels)
  - BGP AFI/SAFI – passes info (draft-ietf-idr-sr-policy-safi)
  - SR Processing Model (SRPM) calculates active tunnels
  - SR segment types A and B in use (SR-MPLS, SRv6)
  - Additional SR segment types (C-L, M-0) – not implemented

- SR + BGP-LS - Proposals – 5+ per IETF meeting
  - IDR Wiki (https://wiki.ietf.org/group/idr) contains status
  - Would it help to have a short abstract on each proposal?
  - Operator feedback – Private or Public matters to IDR chairs

Graphic slide for SR Policy Candidate Routes



SRPM selects the active SR path (SR Policy From Candidate Policies.

Like BGP-LS – BGP is just the transport.

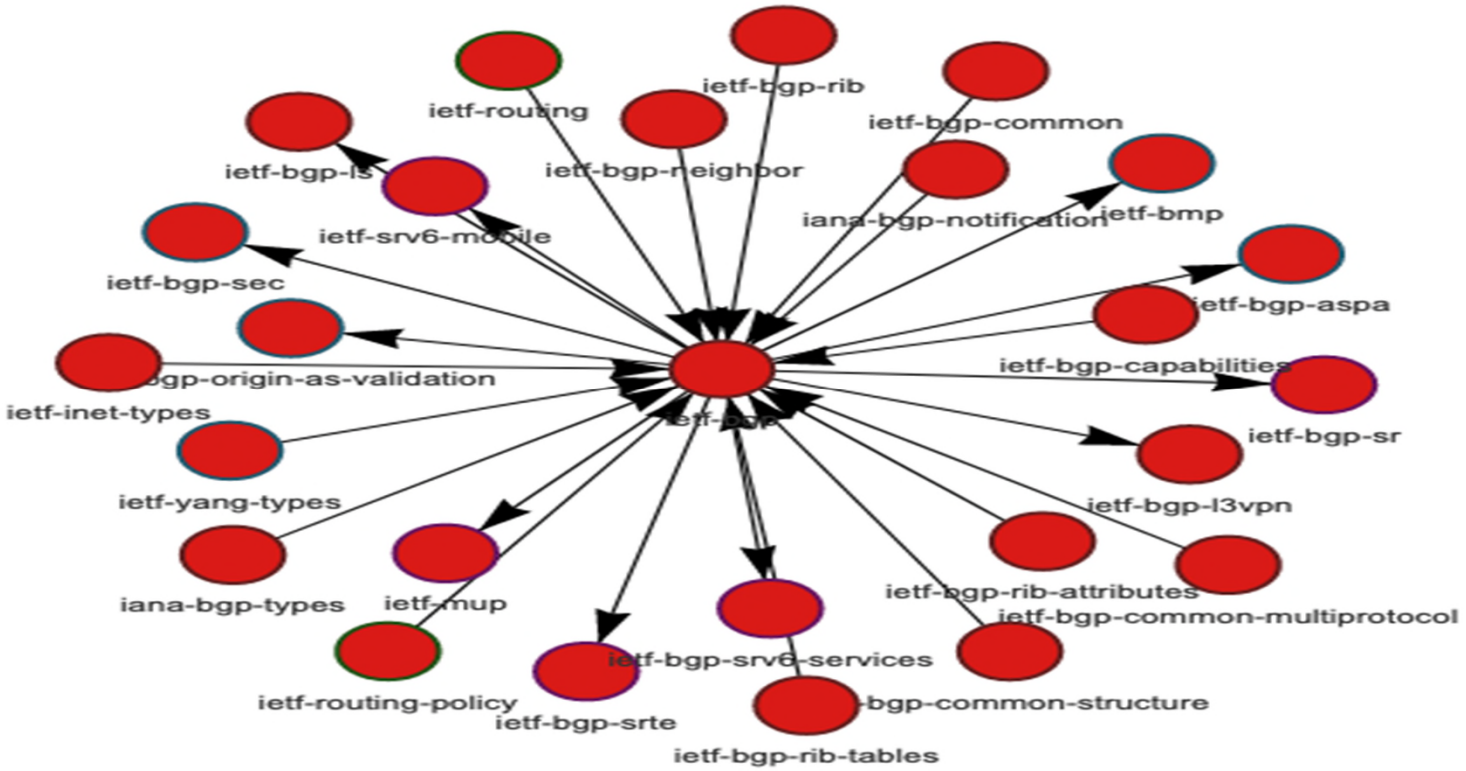
BGP YANG

BGP YANG Model

- Base BGP YANG (draft-ietf-idr-bgp-model)
  - Supports common set of BGP features from operators (RFC4271+++)
  - Basis for future BGP extensions specified in IDR or other WGs
  - Open to vendor augmentation!
  - Multiple implementations in progress

- Additional BGP YANG Models build from this BGP model
  - New BGP features will have new YANG Models
  - As operators, what do you need in YANG Models?

Growing BGP models



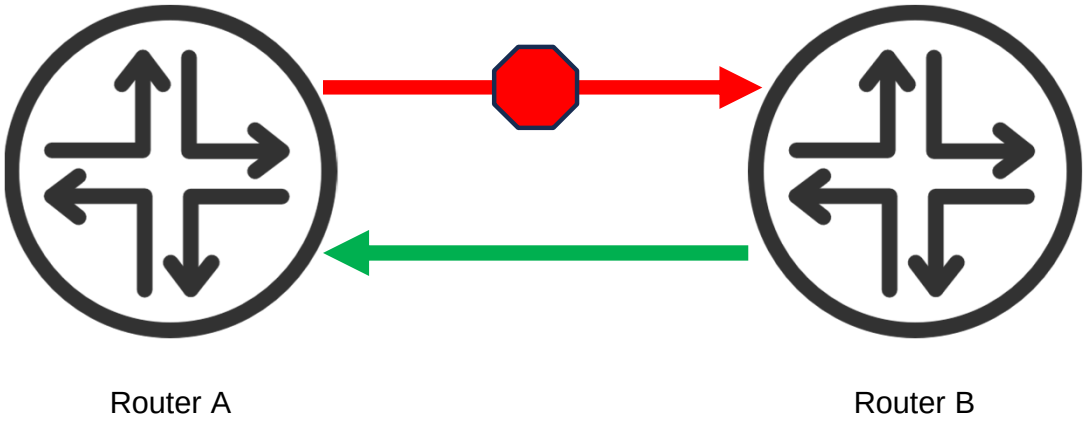
https://www.yangcatalog.org/yang-search/impact_analysis/ietf-bgp@2023-07-05

Stuck BGP Sessions

BGP Sessions Closed on Receive Side Only…

- The BGP HoldTimer is run by a BGP speaker based on the last time it receives a BGP message on the TCP connection.
- The protocol negotiates that time during BGP OPENs to either zero seconds, or at least 3 seconds based on smallest bid by the two speakers.
- Implementations sometimes take liberties with closing sessions.
- … other times, the problems are just bugs!

What happens if the other side gets stuck?



Router A stops being able to send BGP messages to Router B for a “long time”, perhaps far longer than the negotiated HoldTime!

In some circumstances, like a TCP zero-window, Router B is still happily sending BGP messages to Router A!

Consequences to Stuck Sessions

- Regardless of the underlying reason this is happening, BGP is expecting the stuck router to close the BGP session.
- Since it's broken and can linger in this state potentially forever, BGP can get out of sync and blackholes or incorrect routing can result.
- Solution, reset the session.

SendHoldTimer

- Specified by draft-ietf-idr-bgp-sendholdtimer
- Solution is to keep track of last successfully sent BGP message. If the SendHoldTime is exceeded, the session is reset.

- Benefits:
  - Submitted to the IESG for publication.
  - 4 Implementations

TCP User Timeout

- Specified by draft-chen-idr-tcp-user-timeout
- Solution based on TCP feature to reset TCP sessions whenever the data needing to be ACKed has not been ACKed during timeout window.
- Benefit: Can deal with packet loss or TCP bugs leading to dropped ACKs.
- Issue: Not deployed
- Misses: Doesn’t address applications that have working TCP and a non-cooperative BGP speaker.

BGP QUIC

draft-retana-idr-bgp-quic

BGP QUIC

- Primary new feature is using QUIC streams to separate address families for more dynamic behavior and better error handling when there's issues in an individual family.
- Potential benefits from QUIC security model in some circumstances. (Certificates.)
- Establishes a new pattern for stream-based control plane protocols.
  - … where we'll find out what interesting problems we might have…

BGP Flowspec, version 2

draft-ietf-idr-flowspec-v2, et al.

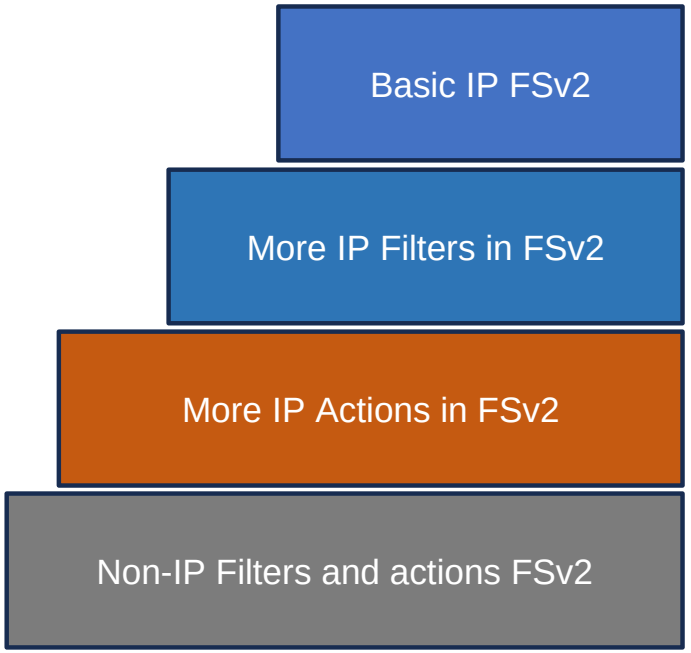
Breaking FSv2 into “chunks”

- NANOG 85 presentation on Flow Specification (FS)
  - Can’t add to FSv1 due to encoding issues
  - Create FSv2 that allows user-ordering of filters and action sequences
- IDR FSv2
  - Is technically correct (draft-ietf-idr-flowspec-v2)
  - Can run in parallel with FSv1 (ships-in-night)
  - Implementers – stated took much for single upgrade
- IDR FSv2 in “chunks”
  - Proposing bite-size chunks for implementers
  - IDR defining of minimal bite-size chunk for DDOS in June (draft-hares-idr-fsv2-ip-basic-02.txt)
  - Other “chunks” optional

FSv2 Chunks

We need your feedback!

- Minimal set = Basic IP FSv2
  - Current IPv4/v6 filters + Current actions
  + User Ordering of Filters
- More IP Filters FSv2
  - Easy addition of new IP Filters with out disturbing basic filters
- More IP actions FSv2
  - Extended Community actions – with defined order and interaction
  - Community attribute with FSv2 TLV – with user ordering of Actions + Dependency
- Non-IP filters FSv2
  - Non-IP: L2VPN, SFC, MPLS, tunnels developed as add on to basic
  - Do not disturb the basic function



A few other things...

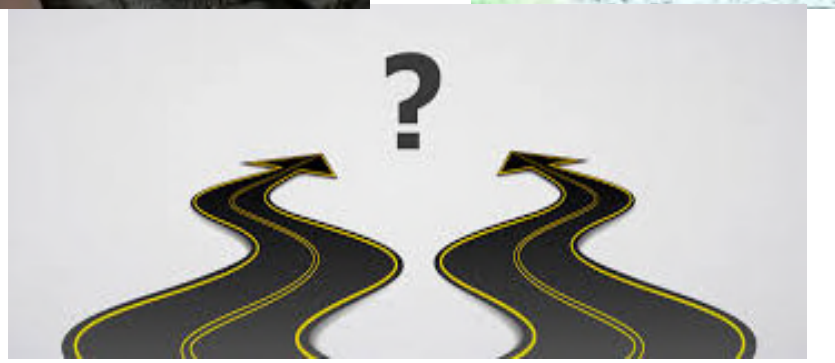
Old is new

- BGP RFC4271 – to Full standard. Any opinions?
- Long lived graceful restart (RFC9494)
  - updates L3VPNs (RFC 6368)

- BFD link to BGP State Machine
  - BFD down subcode (RFC 9384)
  - BFD strict (draft-ietf-idr-bgp-bfd-strict-mode)

- Secure VPNs
  - RFC9012 – obsoleted IP-SEC tunnel type
  - SD-WAN – new tunnel type for IPsec hybrid links (IPsec + MPLS) [draft-ietf-idr-sdwan-edge-discovery]

Are we on the Right Track?



Thank you

NANOG™