BIER: The Best Multicast Solution for Segment Routing

Jeffrey Zhang, Juniper DE

NANOG 90, 12-FEB-2024

PREVIOUSLY IN NANOG81: MULTICAST IN SR NETWORKS

Agenda

- SR principles and Multicast Options
- Controller Signaled P2MP
  - PCEP/BGP-signaled SR-P2MP
  - BGP-signaled mLDP
- BGP Signaled Multicast
  - Both IP multicast and P2MP
  - With or without controllers
- E2E Inter-region Multicast
- Multicast with Classful Transport

Multicast Options in SR Networks

- BIER
  - If you care about effective replication with no-state inside the network, and,
  - Most routers support BIER
- Traditional Multicast (PIM/P2MP/IR)
  - If it works well for you
    - You don't need controller, and,
    - You don't mind running PIM/mLDP/RSVP in your SR network for multicast
      - Perfectly ok to run PIM/mLDP/RSVP for multicast while running SR unicast
- Controller Signaled Multicast
  - If you need controller-calculated trees, and/or,
  - You want to remove PIM/mLDP/RSVP
  - Note that you will still have per-tree/tunnel state inside the network

MULTICAST: STATE VS. REPLICATION EFFICIENCY

- Traditionally, multicast requires per-tree states for efficient replication:
  - PIM, mLDP/RSVP-TE P2MP: per-tree state, different signaling protocols
- Ingress Replication: no per-tree state in the core but replication done at ingress (not efficient)
- BIER is the new technology wonder that achieves efficient replication w/o incurring per-tree state
  - Ideal multicast solution for an SR network (but independent of SR)
  - W/o per-tree state, scaling and convergence are only related to the network topology, not to the number of flows/tunnels
  - There are no special use cases for BIER – any multicast scenarios and operators can benefit from BIER

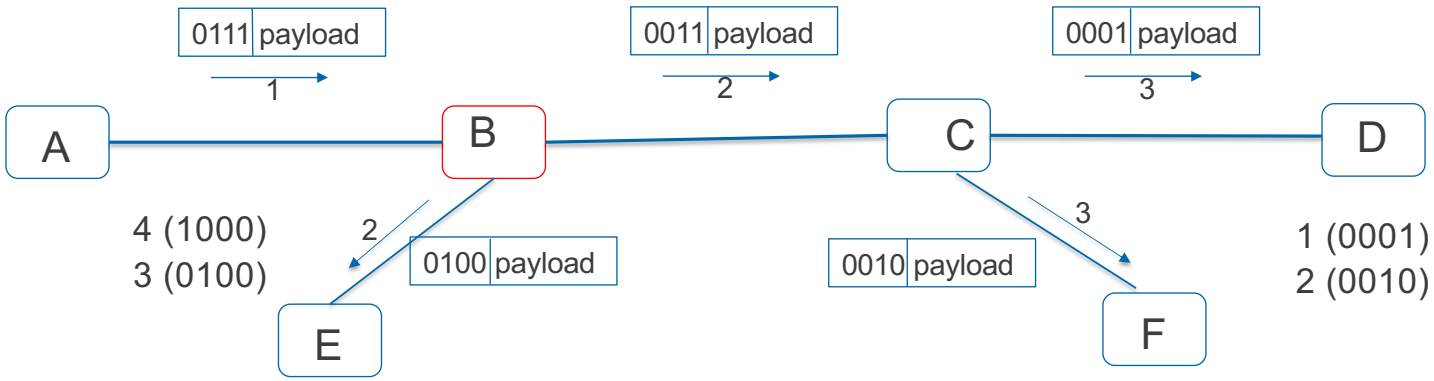
BIT INDEXED EXPLICIT REPLICATION

- Each packet carries a BitString, indicating edge routers (within a BIER domain) that need to receive the multicast packet
  - Each edge router has a unique ID, mapping to a bit in the BitString
  - Large domains with potentially long BitString can be handled in various ways of using small BitString
- Each hop looks at the BitString and replicates the packet to its neighbors that are on the paths to the set of edge routers
  - This can be done very efficiently
- Removes per-tree/tunnel multicast state in the core

TERMINOLOGIES

- BFR: BIER Forwarding Router
- BFIR: Ingress BFR
- BFER: Egress BFR
- BFR-ID: each BFIR/BFER has a 16-bit ID number – maps to a bit in the BitString
- BFR-prefix: each BFR has a (loopback) address for BIER signaling and forwarding
- BIRT: BIER Routing Table
- BIFT: BIER Forwarding Table
- BSL: BitStringLength
- F-BM: A bit-mask associated a BFR NBR, representing all BFERs reachable through that NBR

BIER FORWARDING



B's BIFT

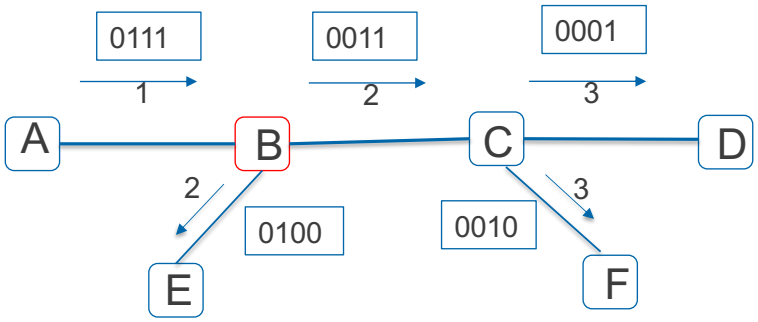
BFER ID	BFR-NBR	F-BM
1 (0001) D	C	0011
2 (0010) F	C	0011
3 (0100) E	E	0100
4 (1000) A	A	1000

F-BM is the property of a BFR-NBR
It represents all BFERs that are reachable via that BFR-NBR

e.g. C's F-BM 0011 = 0001 | 0010
D F

ROUTE LOOKUP & PACKET FORWARDING

- Send one copy to one BFR-NBR that is on the path to a subset of the BFERs that needs to receive the packet
  - Send another copy to another BFR-NBR for another subset, and so on
- How do you find out each subset?
  - Start with the lowest set bit in the packet's BitString – use that bit's index to look up the BIFT
    - The row identifies the BFR-NBR (to send a copy to), and a F-BM that indicates all BFERs that are reachable by the BFR-NBR
      - Set the copy's BitString to (packet's BitString & F-BM)
      - Change the original packet's BitString to (packet's BitString & ~F-BM)
    - Repeat the procedure to send another copy to another NBR
- The process repeats N times, where N is the number of neighbors that need to forward the packet to all BFERs that need to receive the packet



BFER ID	BFR-NBR	F-BM
1 (0001) D	C	0011
2 (0010) F	C	0011
3 (0100) E	E	0100
4 (1000) A	A	1000

- Incoming packet with BitString 0111
  - Need to reach E,F,D
- Lowest set bit is the 1st (right most) so use index one to look up BIFT
- Row 1 (index 1) has F-BM 0011 and BFR-NBR C
  - Send a copy to C, with BitString 0011 (to reach F,D)
    - 0011 == 0111 & 0011 (BitString & C's F-BM)
    - Notice that this takes care two bits (all those BFERs to be reached via C)
  - Change incoming packet's BitString to 0100
    - 0100 == 0111 & 1100 (BitString & ~ C's F-BM)
- Now the lowest set bit is the 3rd so use index 3 to repeat the above
  - Send a copy to E, with BitString 0100

BIER ARCHITECTURE: THREE LAYERS

- Multicast Flow Overlay: signals BFERs to BFIRs for multicast flows
  - E.g. BGP-MVPN, EVPN
  - To the overlay, BIER is an underlay “tunnel”
- BIER Signaling & Forwarding
  - BIER signaling exchanges information needed to calculate the BIFTs
  - Signaling and forwarding states are tied only to the network, not to flows
    - No per-tree/tunnel state
- Routing Underlay
  - Packets can be routed according to different routing underlay “topologies”
  - Utilizes underlay ECMP/FRR for free

BITSTRING LENGTH

- Typical BitStringLength is 256
- What if there are more than that many edge routers?
  - Partition all BFERs into multiple sets, transparent to overlay
    - E.g. 512 BFERs with BitStringLength 256 will have two sets
    - BFIR sends one copy for BFERs in each set
      - Each set has a Set ID
      - Bit 1 in set 1 means BFER #1 but in set 2 means BFER #257
  - Partition a large BIER domain to multiple small sub-domains
    - Small BitStringLength in each sub-domain
    - Overlay signaling stitches the sub-domain “tunnels” together
      - Seamless MPLS Multicast, aka Inter-area Tunnel Segmentation
- Multiple BitString Lengths in a BIER domain
  - E.g. transition from length1 to length2

BIER SUB-DOMAIN

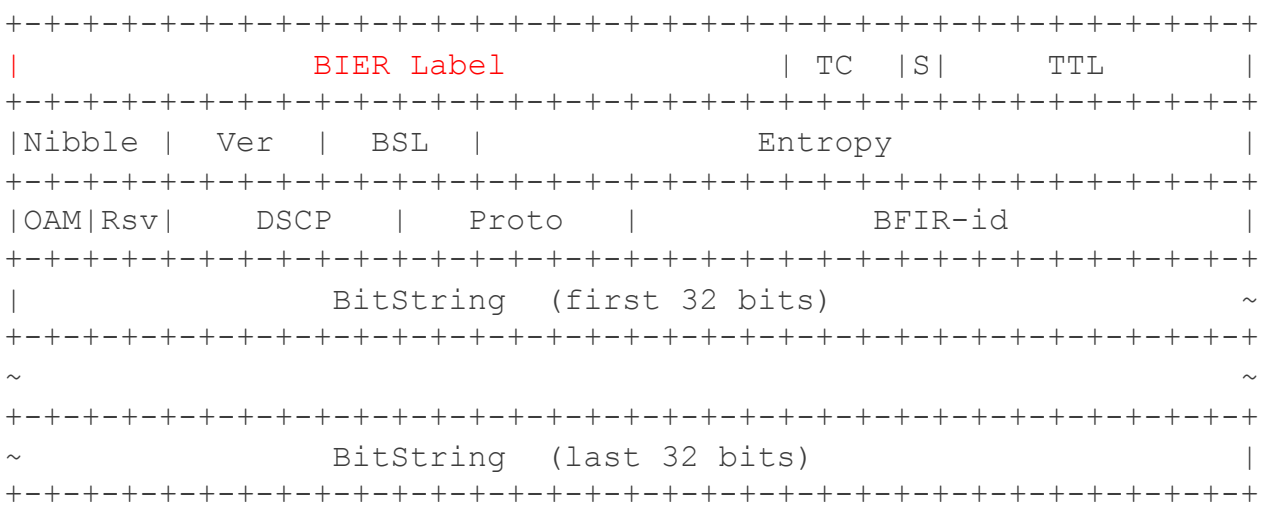
- A set of BIER routers for a “purpose”
  - Same set of BIER routers can belong to multiple sub-domains
- A sub-domain could be
  - for a topology as in Multi-Topology routing
  - a sub-set of a BIER domain or a topology
    - e.g., an IGP area
  - a combination of the above
- Each <sub-domain, bitstring-length, set> has its own BIFT

BIER ENCAPSULATION

- Each packet has a BIER header
  - The BIER header contains a BitString (among other things)
    - Indicating the BFERs that need to receive the packet
  - Other fields in the BIER header
    - Payload type: IPv4/IPv6/Ether/MPLSupLabel/MPLSdownLabel
    - Entropy, BFIR-id, etc.
- BIER header follows an outer encapsulation header
  - MPLS
  - Ethernet
  - Other encapsulations can be used as well

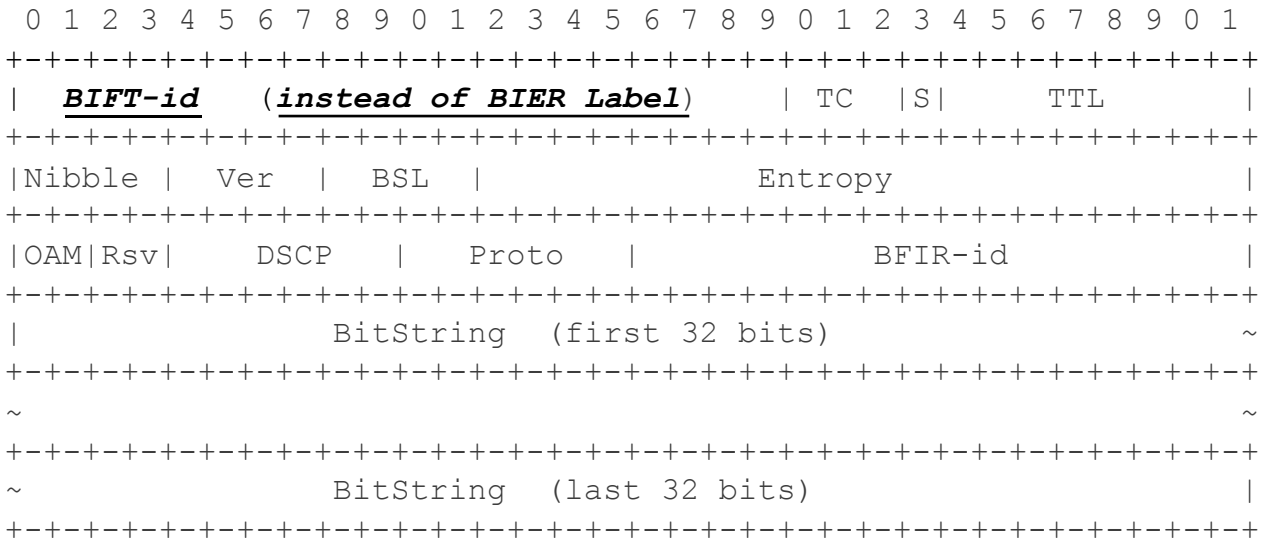
MPLS ENCAPSULATION

- BIER header follows a BIER label
  - Advertised by a traffic-receiving BFR specifically for BIER purpose
    - Indicates that a BIER header follows
    - Also maps to a (sub-domain, bitstring-length, set) tuple, or a BIFT



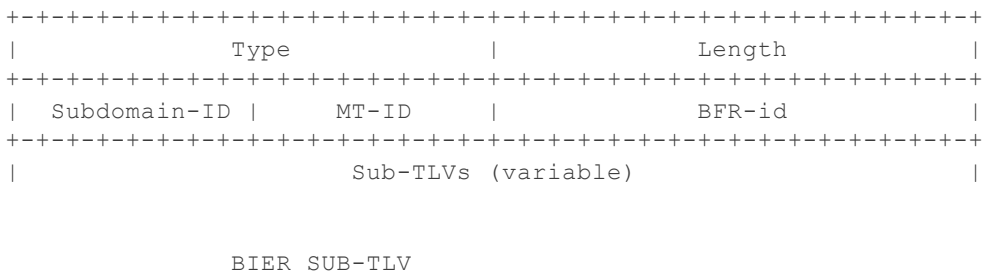
NON-MPLS ENCAPSULATION

- The 32-bit preceding the BIER header interpreted as a generic BIFT-id followed by TC/S (unused) and TTL fields
- EtherType AB37 for Ethernet encapsulation
- Even with non-MPLS encap, BIER payload could be MPLS

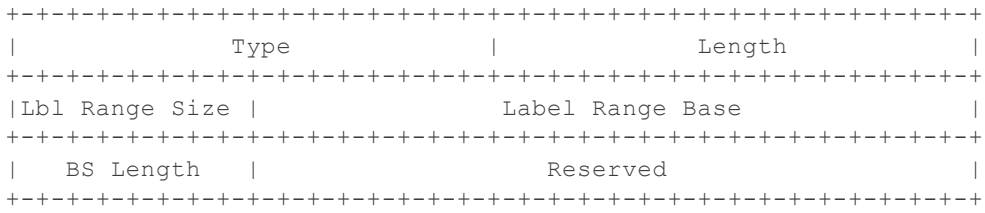


IGP SIGNALING

- Each BFR signals its BFR-ID
  - In each sub-domain that it is in
  - Via BIER Sub-TLV in the Extended Prefix TLV for the BFR's loopback prefix
  - BFR-ID could be 0 for core BFRs
- Each BFR signals a BIER base label or BIFT-ID
  - For each <sub-domain, bsl, set> tuple
  - Via MPLS Encap or non-MPLS Encap Sub-sub-TLV in the BIER Sub-TLV
    - One base label or BIFT-ID in the range for each set



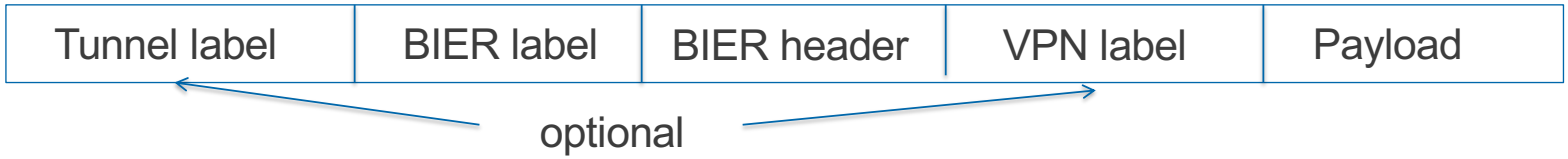
BIER SUB-TLV



MPLS ENCAP Sub-sub-TLV

BIER & MPLS

- BIER label to indicate payload is BIER and identify BIFT to use
  - A sending BFR Imposes outgoing BIER label before sending downstream
  - A receiving BFR Identifies BIFT based on incoming BIER label
- BIER payload may start with an MPLS label stack
- A BIER packet might be transported on a base LSP
  - Tunnel through/around a non-BFR, or tunnel around a failed BFR
- All generic MPLS functionalities



BIER & SR

- BIER is the best multicast solution for SR
  - Same principle of “no per-flow state inside the network”
- But BIER is independent of SR
  - The only relevance is that SR paths (as well as non-SR tunnels) can be used to tunnel BIER traffic between non-adjacent BFRs
- BIER and non-MPLS IPv6
  - BIERin6: generic non-MPLS BIER applicable to both IPv4 and IP6
    - Generic layered architecture: BIER over any transport (L2 or any SR or no-SR tunnel)
    - IETF adopted-solution – in WG Last Call
  - BIERv6: non-MPLS encapsulation tightly coupled with SRv6

BIER RFC/DRAFT

- RFC8279, 8296: architecture and encapsulation
- RFC8401, 8444: ISIS/OSPF signaling
- RFC9272: Underlay Path Calculation Algorithm and Constraints
- RFC8556: MVPN with BIER as provider tunnel
- draft-ietf-bier-evpn: EVPN with BIER as provider tunnel
- draft-ietf-bier-idr-extensions-10: BGP signaling
- draft-ietf-bier-bierin6: generic non-MPLS BIER for IPv4/IPv6
- Brownfield Deployments:
  - draft-ietf-bier-tether, draft-ietf-bier-php,
  - draft-ietf-bier-multicast-as-a-service

CHICKEN AND EGG DILEMMA

- A technology wonder that has not been widely deployed
- New encapsulation and forwarding paradigm requires:
  - Programmable ASIC, or,
  - New generation of pipeline-based ASIC
- Operators have shown great interest but had to back off due to the platform restriction
  - Anyone with scaled multicast requirements would benefit from BIER
- Vendors have been holding back due to unclear customer demand and the ASIC limitation
  - Some did have PoC/production implementation

PIONEERS WILL BREAK THE DILEMMA

- Vendors nonetheless have been investing in BIER
  - Active work in IETF
  - Hardware and software support from several major vendors
- Hardware capability across access/edge/aggregation/core platforms
  - From several major vendors
- Successful interoperability testing in IETF118 (11/2023)
  - A more formal/comprehensive interop is being planned in another venue
- Pioneering vendors and operators will break the chicken and egg dilemma
  - A prime time for BIER and BIER-enabled-multicast is coming

Thanks