LET'S BUILD TOMORROW TODAY
Stateless Multicast with Bit Indexed Explicit Replication

IJsbrand Wijnands – Principal Engineer

BRKIPM-2239
Agenda

• Introduction
• Solution Overview
• Encapsulation
• Sets and Area’s
• Forwarding
• ECMP
• Deployment Scenarios
• Conclusion
Introduction
BIER history

• A team was formed to investigate solutions for multicast in the context of Segment Routing.

• Encoding a Sourced routed Multicast tree path using MPLS labels is difficult.

• The packet header would get very large, and its very hard to parse such header.
The BIER Epiphany

- Only encode the end-receivers in the packet header.
  - Not the intermediate nodes.

- Assign end-receivers a Bit Position from a Bit String.
  - The smallest identifier possible.

- Encode the Bit String in the packet header.
  - Using some sort of encapsulation.

- Create a Bit Forwarding Table on all BIER nodes to allow multicast packet forwarding using the Bit String in the packet.
  - Derived from the RIB, SPF based.

- We call it, Bit Indexed Explicit Replication (BIER).
Tech Fund

- Chambers Tech Fund to prototype the idea Jan 2014.
- Core TF team; Greg Shepherd, Neale Ranns, IJsbrand Wijnands.
- Goal, build a prototype on a real router platform.
- We have a working prototype code on XRVR, CRS and ASR9K
IETF

- The BIER idea was presented in a BOF at the IETF in Hawaii.
  - November 2014.

- A new BIER Working Group has been formed (bier@ietf.org)

- Vendors collaborating (co-authoring) with us;
  - Ericsson
  - Alcatel-Lucent
  - Juniper
  - Huawei
IETF drafts

- draft-ietf-bier-problem-statement-00
- draft-ietf-bier-architecture-00
- draft-ietf-bier-encapsulation-mpls-00
- draft-ietf-bier-use-cases-00
- draft-ietf-l3vpn-mvpn-bier-00
- draft-ietf-ospf-bier-extensions-00
- draft-przygienda-bier-isis-ranges-01
- draft-eckert-bier-te-arch-00
- draft-xu-idr-bier-extensions-00
Solution Overview
1. Assign a unique Bit Position from a BitString to each BFER in the BIER domain.
2. Each BFER floods their Bit Position to BFR-prefix mapping using the IGP (OSPF, ISIS)
Basic Idea BIER

1. Assign a unique Bit Position from a BitString to each BFER in the BIER domain.
2. Each BFER floods their Bit Position to BFR-prefix mapping using the IGP (OSPF, ISIS)
Bit Index Forwarding Table

<table>
<thead>
<tr>
<th>BM</th>
<th>Nbr</th>
</tr>
</thead>
<tbody>
<tr>
<td>0111</td>
<td>B</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BM</th>
<th>Nbr</th>
</tr>
</thead>
<tbody>
<tr>
<td>0011</td>
<td>C</td>
</tr>
<tr>
<td>0100</td>
<td>E</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BM</th>
<th>Nbr</th>
</tr>
</thead>
<tbody>
<tr>
<td>0001</td>
<td>D</td>
</tr>
<tr>
<td>0010</td>
<td>F</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BM</th>
<th>Nbr</th>
</tr>
</thead>
<tbody>
<tr>
<td>0011</td>
<td>C</td>
</tr>
<tr>
<td>0100</td>
<td>B</td>
</tr>
</tbody>
</table>
Bit Index Forwarding Table

- D, F and E advertise their Bit positions in the IGP (flooded).
- Based on shortest path route to RID, the Bit Mask Forwarding Table is created.
Forwarding Packets
Forwarding Packets

```
<table>
<thead>
<tr>
<th>Nbr</th>
<th>0111</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nbr</td>
<td>0011</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>0100</td>
<td>E</td>
</tr>
<tr>
<td>Nbr</td>
<td>0011</td>
<td>&amp;0111</td>
</tr>
<tr>
<td></td>
<td>0010</td>
<td>&amp;0100</td>
</tr>
<tr>
<td>Nbr</td>
<td>0001</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>0010</td>
<td>F</td>
</tr>
</tbody>
</table>
```

A: 0111
AND 0001
  0101
  0101
B: 0101
AND 0001
  0001
C: 0001
AND 0100
  0100
D: 0010
F: 0010
AND 0100

E: 0100
AND 0001
  0001
AND 0100
  0100
AND 0111
  0101
  0101
Forwarding Packets

- As you can see from the previous slides, the result from the bitwise AND (&) between the Bit Mask in the packet and the Forwarding table is copied in the packet for each neighbor.

- This is the key mechanism to prevent duplication.

- Look at the next slide to see what happens if the bits are not reset.

- If the previous bits would not have been reset, E would forward the packet to C and vice versa.
Encapsulation
How many Bits and Where?

• The number of multicast egress routers that can be addressed is depending on the number of Bits that can be included in the BitString.

• The BitString length is depending on the encapsulation type and router platform.

• We identified 5 different encoding options, most attractive below;
  1. MPLS, below the bottom label and before IP header.
  2. IPv6, extensions header.
MPLS encapsulation

• We’ve analyzed the MPLS option, CRS and ASR9K platform.
• Both these platforms can do 256 bits.
• We consider 256 a good starting point.
• Other vendors confirmed they can do 256.
http://www.ietf.org/id/draft-ietf-bier-mpls-encapsulation-01.txt
Sets and Area’s
BIER Sets

- To increase the scale we group the egress routers in Sets.

<table>
<thead>
<tr>
<th>Set</th>
<th>BM</th>
<th>Nbr</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0111</td>
<td>I</td>
</tr>
<tr>
<td>2</td>
<td>0111</td>
<td>I</td>
</tr>
</tbody>
</table>

Note, Bit Positions 1, 2, 3 appear in both Sets, and do not overlap due to Sets.

Note, we create different forwarding entries for each Set.

• To increase the scale we group the egress routers in Sets.
BIER Sets

<table>
<thead>
<tr>
<th>Set</th>
<th>BM</th>
<th>Nbr</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0111</td>
<td>I</td>
</tr>
<tr>
<td>2</td>
<td>0111</td>
<td>I</td>
</tr>
</tbody>
</table>

Note, we create different forwarding entries for each Set

- There is no topological restriction which set an egress belongs to
BIER Sets

- If a multicast flow has multiple receivers in different Sets, the packet needs to be replicated multiple times by the ingress router, for each set once.
- Is that a problem? We don’t think so…
- The Set identifier is part of the packet.
- Can be implemented as MPLS label.
BIER Area

- A bit Mask only needs to be unique in its own area.
- ABR’s translate Bit Masks between area’s.
- Requires a IP lookup and state on the ABRs.
- This is very similar for ‘Segmented Inter-AS MVPN’.

![Diagram showing BIER Area with ABRs and Bit Masks]

BM | Nbr
---|---
0:10 ABR

BM | Nbr | BM | Nbr
---|---|---|---
0:01 A 0:01 B

BM | Nbr
---|---
0:10 ABR
The BFR-id
The BFR-id

• A BFER is uniquely identified by a two tuple {Set ID, Bit Position}
• The number of Bit Positions is depending on the length of the support BitString in the network.
• To make the BFER identifier independent of the BitString length we defined the BFR-id as a number between [1,65535]
• We auto-generate the BFR-id into a {Set ID, Bit Position}, based on the BitString Length.
The BFR-id

- **Formula:**
  
  \[
  SI = \frac{(BFR\text{-}id - 1)}{\text{BitStringLength}} \\
  BP = ((BFR\text{-}id - 1) \mod \text{BitStringLength}) + 1
  \]

- **Example BFR-id = 129.**
  
  BitString Length 128 -> \{SI = 1, BP = 1\}
  
  BitString Length 256 -> \{SI = 0, BP = 129\}
The BFR-id

• By decoupling the BFR-id from the SetID and Bit Position tuple, the BFR identifier is agnostic to the supported BitString length in the network.

• This is very useful during migration.

• If a network supports multiple BitString lengths, an egress router only needs one BFR-id, and is reachable via each BitString length.
BIER Forwarding
BIER Forwarding

• We define two different forwarding methods for BIER.
  1. Neighbor based forwarding.
  2. Bit Indexed forwarding
BIER Forwarding, neighbor based.

- A packet BitString is matched against each Neighbor in the BFT.
- This model works well for systems that have multi-cores.
- It works less well for serialized processing as it requires a neighbor walk for each packet.
BIER Forwarding, Bit Indexed

- We translate the BFT neighbor table into a table sorting on Bit Position (so not by neighbor)
- We walk the Bit Mask in the packet and Index into the FIB table.
BIER Forwarding, Bit Indexed

- We walk the Bits in the packet, as soon as we hit a ‘1’, we copy the packet, index into the FIB table with the position of the Bit.
- The Bit Mask entry is reverse ‘&’ with the Bit Mask in the packet.
- This resets the Bits that were processed.
BIER Forwarding, Bit Indexed

- We walk the Bits in the packet, as soon as we hit a ‘1’, we copy the packet, index into the FIB table with the position of the Bit.
- The Bit Mask entry is reverse ‘&’ with the Bit Mask in the packet.
- This resets the Bits that where processed.
BIER Forwarding, Bit Indexed

- Walking the bits in a Bit String takes less clock cycles compared to walking a list of neighbors.

- For that reason its faster to walk the Bit String and index into the neighbor table.

- The table is a NxN bit matrix, where N is the Bit String length.

- Bits that where already processed are reset so we don’t processes them if they appears later in a Bit String. This way we avoid multiple copies being forwarded.
BIER Forwarding, Bit Indexed

- Walking the Bit String in the packet is basically a repeating of ‘Find First Set’ Bit operation.
- Could be optimized to record the last position.
  - Checkout Bruijn sequence (Count the consecutive zero bits)
- Is supported in compilers ffs and can be done in HW fairly easily.
ECMP
ECMP

• It is possible the same Bit Position is reachable over different interfaces if there are ECMP paths.

• We distinguish two different ECMP behaviors
  • ECMP via parallel interfaces to a single neighbor.
  • ECMP via different neighbors.

• The handling is different for both cases.

• If the ECMP path is going to a single neighbor, the Bit Mask is the same for each candidate path, no special processing needed.

• If the ECMP path is to different neighbors, the Bit Mask will be different for each neighbor because the Bit String in the packet is the end result of the lookup in BIER, which is different for each nbr.
Duplicate bit positions need to be resolved, ECMP logic needs to select based on Hash. In the example we selected C.
ECMP

- We distribute the Bit Positions over multiple tables.
- Each Bit Position only appears once in each table.
- Table selection is based on entropy, done before Bit Index lookup
ECMP

• The number of tables is depending on the number of ECMP paths to different neighbors.

• If the max number of ECMP paths is 4, but there is a Bit Position reachable via 3 paths, this will cause unequal distribution.
Deployment Scenario's
Native BIER
Native BIER

- With Native BIER there is NO PIM involved, just IGMP and BIER.
- The Source and Receiver(s) are connected to BIER router.
  - There are no RP’s.
  - There is no equivalent of PIM modes, like sparse, ssm, bidir etc..
- We speak of ‘single’ sender and ‘multi’ sender, which is basically the same solution.
- The overlay can be BGP or SDN based.
Native BIER

- E and F announce their Group membership via overlay to all other routers.
- A BIER router connected to the Source can immediately start sending.
Native BIER

When B learns about a new source, it can immediately start sending.
MVPN over BIER
MVPN over BIER

- BIER replaces PIM, mLDP, RSVP-TE or IR in the core.
- BIER represents a full mesh (P2MP) connectivity between all the PE’s in the network.
- There is no need to explicitly signal any MDT’s (or PMSI’s).
- With MVPN there are many profiles,
  - This is partly due to the tradeoff between ‘State’ and ‘Flooding’.
  - Different C-multicast signaling options.
- MVPN over BIER, there is one profile.
  - BGP for C-multicast signaling.
- No need for Data-MDTs.
MVPN over BIER

- The BGP control plane defined for MVPN can be re-used.
- Big difference, there is no Tree per VPN…!!!
- The BIER packets needs to carry Source ID and upstream VPN context label
Conclusion
Advantages

• Packets forwarded via BIER follow the unicast path towards the receiver, inheriting unicast features like FRR and LFA.

• There is no per multicast flow state in the network.

• Multicast convergence is as fast as unicast, there is no multicast state to re-converge, signal, etc.

• Nice plugin for SDN, it’s only the ingress and egress that need to exchange Sender and Receiver information.

• The core network provides a many-2-many connectively between all BIER routers by default following the IGP.

• No Multicast control protocol in the network.
Disadvantages

• The Bit String length has an upper bound and may not cover all deployment scenarios.

• Using sets to increase the number of egress routers may cause the ingress to replicate the packet multiple times.

• Using area’s requires the ABR to have state.
```plaintext
router bier
  address-family ipv4
  encapsulation mpls 256 source Loopback0
  !
  bfr-id 1
  !
router bier
  address-family ipv4
  encapsulation mpls 256 source Loopback0
  !
  bfr-id 2
  !
router bier
  address-family ipv4
  encapsulation mpls 256 source Loopback0
  !
  bfr-id 3
  !
router bier
  address-family ipv4
  encapsulation mpls 256 source Loopback0
  !
  bfr-id 4
  !
```
multicast-routing vrf red/blue
  address-family ipv4
  bgp auto-discovery bier
  leaf-info-required
  !
  mdt bier-signaling ipv4
  !
  router pim
  vrf red/blue
  address-family ipv4
  rpf topology route-policy bier
  mdt c-multicast-routing bgp

multicast-routing vrf red/blue
  address-family ipv4
  bgp auto-discovery bier
  leaf-info-required
  !
  mdt bier-signaling ipv4
  !
  router pim
  vrf red/blue
  address-family ipv4
  rpf topology route-policy bier
  mdt c-multicast-routing bgp

multicast-routing vrf green/blue
  address-family ipv4
  bgp auto-discovery bier
  leaf-info-required
  !
  mdt bier-signaling ipv4
  !
  router pim
  vrf green/blue
  address-family ipv4
  rpf topology route-policy bier
  mdt c-multicast-routing bgp

multicast-routing vrf green/blue
  address-family ipv4
  bgp auto-discovery bier
  leaf-info-required
  !
  mdt bier-signaling ipv4
  !
  router pim
  vrf green/blue
  address-family ipv4
  rpf topology route-policy bier
  mdt c-multicast-routing bgp
Key Takeaways

• Interested in BIER on your router platform? Talk to your Cisco Account Team

• If you want to participate in defining the technology, -> IETF
Complete Your Online Session Evaluation

• Give us your feedback to be entered into a Daily Survey Drawing. A daily winner will receive a $750 Amazon gift card.

• Complete your session surveys though the Cisco Live mobile app or your computer on Cisco Live Connect.

Don’t forget: Cisco Live sessions will be available for viewing on-demand after the event at CiscoLive.com/Online
Continue Your Education

• Demos in the Cisco campus
• Walk-in Self-Paced Labs
• Table Topics
• Meet the Engineer 1:1 meetings
• Related sessions
Thank you