BIER
Internet-Draft
Intended status: Standards Track
Expires: April 2, 2020

Z. Zhang
Juniper Networks
N. Warnke
Deutsche Telekom
I. Wijnands
Cisco Systems
D. Awduche
Verizon

September 30, 2019

Tethering A BIER Router To A BIER incapable Router
draft-zzhang-bier-tether-04

Abstract

This document specifies optional procedures to optimize the handling of Bit Index Explicit Replication (BIER) incapable routers, by attaching (tethering) a BIER router to a BIER incapable router.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on April 2, 2020.

Table of Contents

1. Introduction

Consider the scenario in Figure 1 where router X does not support BIER.

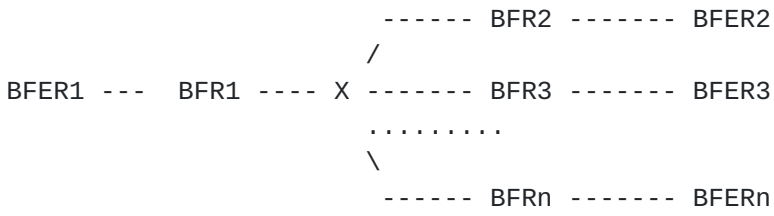

Figure 1: Deployment with BIER incapable routers

For BFR1 to forward BIER traffic towards BFR2...BFRn, it needs to tunnel individual copies through X. This degrades to "ingress" replication to those BFRs. If X's connections to BFRs are long distance or bandwidth limited, and n is large, it becomes very inefficient.

A solution to the inefficient tunneling from BFRs is to attach (tether) a BFRx to X as depicted in Figure 2:

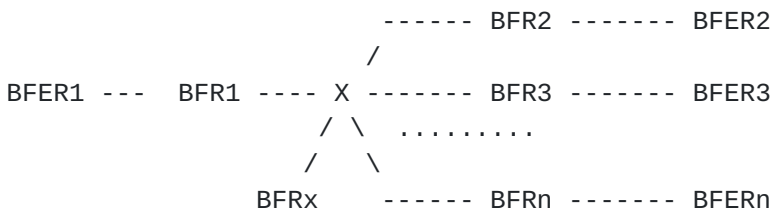

Figure 2: Tethered BFRx

Instead of BFR1 tunneling to BFR2, ..., BFRn directly, BFR1 will get BIER packets to BFRx, who will then tunnel to BFR2, ..., BFRn. There could be fat and local pipes between the tethered BFRx and X, so ingress replication from BFRx is acceptable.

For BFR1 to tunnel BIER packets to BFRx, the BFR1-BFRx tunnel need to be announced in Interial Gateway Protocol (IGP) as a forwarding adjacency so that BFRx will appear on the Shortest Path First (SPF) tree. This needs to happen in a BIER specific topology so that unicast traffic would not be tunneled to BFRx. Obviously this is operationally cumbersome.

Section 6.9 of BIER architecture specification [RFC8279] describes a method that tunnels BIER packets through incapable routers without the need to announce tunnels. However that does not work here, because BFRx will not appear on the SPF tree of BFR1.

There is a simple solution to the problem though. BFRx could advertise that it is X's helper and other BFRs will use BFRx (instead of X's children on the SPF tree) to replace X during its post-SPF processing as described in section 6.9 of BIER architecture specification [RFC8279].

2. Additional Considerations

While the example shows a local connection between BFRx and X, it does not have to be like that. As long as packets can arrive at BFRx without requiring X to do BIER forwarding, it should work.

Additionally, the helper BRFx can be a transit helper, i.e., it has other connections (instead of being a stub helper that is only connected to X), as long as BFRx won't send BIER packets tunneled to it back towards the tunnel ingress. Figure 3 below is a simple case:

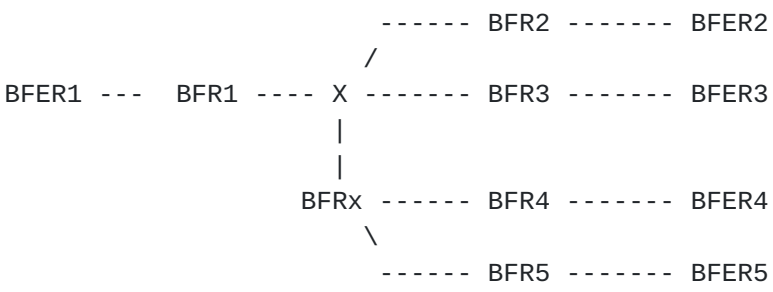

Figure 3: A Safe Transit Helper

In the example of Figure 4, there is a connection between BFR1 and BFRx. If the link metrics are all 1 on the three sides of BFR1-X-BFRx triangle, loop won't happen but if the BFRx-X metric is 3 while other two sides of the triangle has metric 1 then BFRx will send BIER packets tunneled to it from BFR1 back to BFR1, causing a loop.

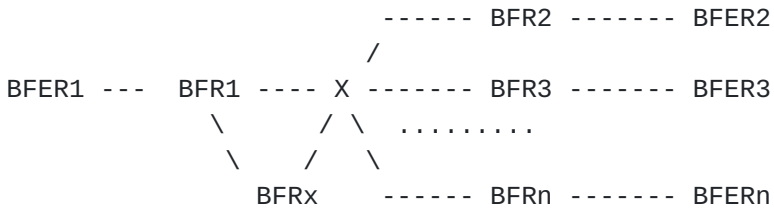

Figure 4: Potential looping situation

This can easily be prevented if BFR1 does an SPF calculation with the helper BFRx as the root. For any BFERn reached via X from BFR1, if BFRx's SPF path to BFERn includes BFR1 then BFR1 must not use the helper. Instead, BFR1 must directly tunnel packets for BFERn to X's BFR (grand-)child on BFR1's SPF path to BFERn, per section 6.9 of [RFC8279].

Notice that this SPF calculation on BFR1 with BFRx as the root is not different from the SPF done for a neighbor as part of Loop-Free Alternate (LFA) calculation. In fact, BFR1 tunneling packets to X's helper is not different from sending packets to a LFA backup.

Also notice that, instead of a dedicated helper BFRx, any one or multiple ones of BFR2..N can also be the helper (as long as the connection between that BFR and X has enough bandwidth for replication to multiple helpers through X). To allow multiple helpers to help the same non-BFR, the "I am X's helper" advertisement carries a priority. BFR1 will choose the helper advertising the highest priority among those satisfying the loop-free condition

described above. When there are multiple helpers advertising the same priority and satisfying the loop-free condition, any one or multiple ones could be used solely at the discretion of BFR1. However, if multiple ones are used, it means that multiple copies may be tunneled through X.

The situation in Figure 5 where a helper BFRxy helps two different non-BFRs X and Y also works. It's just a special situation of a transit helper.

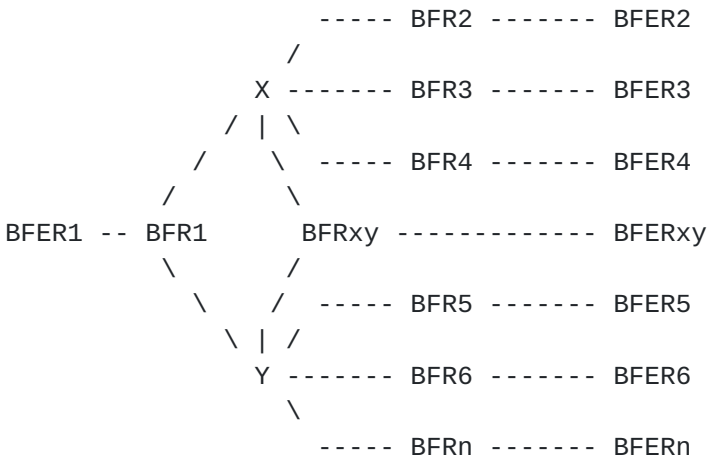

Figure 5: One Helper for multiple helped

3. Specification

The procedures in this document apply when a BFRx is tethered to a BIER incapable router X as X's helper for BIER forwarding.

3.1. IGP Signaling

Suppose that the BIER domain uses BIER signaling extensions to ISIS [RFC8401] or OSPF [RFC8444]. The helper node (BFRx) MUST advertise one or more BIER Helped Node sub-sub-TLVs (one for each helped node). The value is BIER prefix of the helped node (X) followed by a one-octet priority field, and one-octet reserved field. The length is 6 for IPv4 and 18 for IPv6 respectively.

The post-SPF processing procedures in Section 6.9 of the BIER architecture specification [RFC8279] are modified as following for BIER tethering purpose.

At step 2, the removed node is added to an ordered list maintained with each child that replaces the node. If the removed node already

has a non-empty list maintained with itself, add the removed node to the tail of the list and copy the list to each child.

At the end, the calculating node BFR-B would use a unicast tunnel to reach next hop BFRs for some BFERs. The next hop BFR has an ordered list created at step 2 above, recording each BIER incapable node replaced by their children along the way. For a particular BFER to be reached via a tunnel to the next hop BFR, additional procedures are performed as following.

- Starting with the first node in the ordered list of incapable nodes, say N1, check if there is one or more helper nodes for N1. If not, go the next node in the list.

- Order all the helper nodes of N1 based descending (priority, BIER prefix). Starting with the first one, say H1, check if BFR-B could use H1 as LFA next hop to reach the BFER. If yes, H1 is used as the next hop BFR for the BFER and the procedure stops. If not, go to the next helper in order.

- If none of the helper nodes of N1 can be used, go to the next node in the list of incapable nodes.

If the above procedure finishes without finding any helper, then the original BFR next hop via a tunnel is used to reach the BFER.

3.2. BGP Signaling

Suppose that the BIER domain uses BGP signaling [I-D.ietf-bier-idr-extensions] instead of IGP. BFR1..N advertises BIER prefixes that are reachable through them, with BIER Path Attributes (BPA) attached. There are three situations regarding X's involvement:

(1) X does not participate in BGP peering at all

(2) X re-advertises the BIER prefixes but does not do next-hop-self

(3) X re-advertises the BIER prefixes and does next-hop-self

With (1) and (2), the BFR1..N will tunnel BIER packets directly to each other. It works but not efficiently as explained earlier. With (3), BIER forwarding will not work, because BFR1..N would try to send BIER packets to X though X does not advertise any BIER information. If Tunnel Encapsulation Attribute (TEA) [I-D.ietf-idr-tunnel-encaps] is used as specified in [I-D.zzhang-bier-multicast-as-a-service] with (3), then it becomes similar to (2) - works but still not efficiently.

To make tethering work well with BGP signaling, the following can be done:

- Configure a BGP session between X and its helper BFRx. X re-advertises BIER prefixes (with BPA) to BFRx without changing the tunnel destination address in the TEA.

- BFRx advertises its own BIER prefix with BPA to X, and sets the tunnel destination address in the TEA to itself. X then re-advertises BFRx's BIER prefix to BFR1..N, without changing the tunnel destination address in the TEA.

- For BIER prefixes (with BIER Path Attribute) that X re-advertises to other BFRs, the tunnel destination in the TEA is changed to the helper BFRx.

With the above, BFR1..N will tunnel BIER packets to BFRx (following the tunnel destination address in the TEA), who will then tunnel packets to other BFRs (again following the tunnel destination address in the TEA). Notice that what X does is not specific to BIER at all.

4. Security Considerations

This specification does not introduce additional security concerns beyond those already discussed in BIER architecture and OSPF/ISIS/BGP extensions for BIER signaling.

5. IANA Considerations

This document requests a new sub-sub-TLV type value from the "Sub-sub-TLVs for BIER Info Sub-TLV" registry in the "IS-IS TLV Codepoints" registry:

Type	Name
TBD1	BIER Helped Node

This document also requests a new sub-TLV type value from the OSPFv2 Extended Prefix TLV Sub-TLV registry:

Type	Name
TBD2	BIER Helped Node

6. Acknowledgements

The author wants to thank Eric Rosen and Antonie Przygienda for their review, comments and suggestions.

7. Normative References

[I-D.ietf-bier-idr-extensions]
Xu, X., Chen, M., Patel, K., Wijnands, I., and T. Przygienda, "BGP Extensions for BIER", draft-ietf-bier-idr-extensions-07 (work in progress), September 2019.

[I-D.ietf-idr-tunnel-encaps]
Patel, K., Velde, G., Ramachandra, S., and E. Rosen, "The BGP Tunnel Encapsulation Attribute", draft-ietf-idr-tunnel-encaps-13 (work in progress), July 2019.

[I-D.zzhang-bier-multicast-as-a-service]
Zhang, Z., Rosen, E., and L. Geng, "Multicast/BIER As A Service", draft-zzhang-bier-multicast-as-a-service-00 (work in progress), October 2018.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[RFC8279] Wijnands, IJ., Ed., Rosen, E., Ed., Dolganow, A., Przygienda, T., and S. Aldrin, "Multicast Using Bit Index Explicit Replication (BIER)", RFC 8279, DOI 10.17487/RFC8279, November 2017, <https://www.rfc-editor.org/info/rfc8279>.

[RFC8401] Ginsberg, L., Ed., Przygienda, T., Aldrin, S., and Z. Zhang, "Bit Index Explicit Replication (BIER) Support via IS-IS", RFC 8401, DOI 10.17487/RFC8401, June 2018, <https://www.rfc-editor.org/info/rfc8401>.

[RFC8444] Psenak, P., Ed., Kumar, N., Wijnands, IJ., Dolganow, A., Przygienda, T., Zhang, J., and S. Aldrin, "OSPFv2 Extensions for Bit Index Explicit Replication (BIER)", RFC 8444, DOI 10.17487/RFC8444, November 2018, <https://www.rfc-editor.org/info/rfc8444>.

Authors' Addresses

Zhaohui Zhang
Juniper Networks

EMail: zzhang@juniper.net

Nils Warnke
Deutsche Telekom

EMail: Nils.Warnke@telekom.de

IJsbrand Wijnands
Cisco Systems

EMail: ice@cisco.com

Daniel Awduche
Verizon

EMail: daniel.awduche@verizon.com