

IETF 102

BIER WG Montreal

BIER multicast overlay for http response - Debashish

Based on multicast http using bier where we described the requirements.

In this draft we referenced it. The multicast overlay is formed by the BFIR and BFER of the bier layer and the additional SR (service handler) and PCE elements. Service handler is added to BFIR/BFER.

Greg: you show bier te in the pce. Is that a requirement?

Debashish: its not a requirement just a possibility.

Greg: what would the PCE do?

Debashish: Will get to that shortly.

Realization of use case over IPMC

HTTP response multicast maybe realized over IP multicast

Require support for group formation, maintain group state and igmp signaling to join a group.

For a few receivers:

Many of the bitrates may not be required and dropped by the CNAP

Extremely high and undesirable amount of ip mcast signaling protocol activity (pim/igmp)

Greg: whats the assumption that the control plane will be high with fewer receivers?

Debashish: at the snap level it will subscribe for everyone. Multicast formed for each encoded bit rate. On the client side subscribing for a particular bit rate. The other groups may not be joined. As the users are requesting another group they will keep changing and sending requests. It will happen at the edge and not propagate with large number of receivers.

Greg: it will happen at the edge but it won't propagate.

SR Service router terminates application level protocols, extracts the URI to determine the PATH ID via PCE request

PCE keeps track of all service execution points and how to reach them (can be part of bier-te).

Interface functions to the BFIR where the path id is mapped to the bier header

Advantages of realization over BIER:

Eliminates any dynamic multicast signaling between cnap and snap

Avoid sending of any unnecessary data block, which in the IP multicast

solution is pretty much unavoidable

Next steps:

We suggest to include an additional applicability statement documenting how bier can be applied to aggregate http responses over a bier infrastructure.

Tony: Even if you decide to coalesce a bier domain gives you penalty for 1:1.

There's no bit cost. Also how the magic of http is accomplished is not described which isn't in scope of bier anyway.

Rachel: Multicast ABR, they are doing it in the application layer. HTTP also being worked on in BBF for streaming live video and also is similar to one being described here. On the edge they have devices to convert unicast to multicast. Might be useful show how bier can be useful in this video scenario. Strongly suggest you be consistent with the frameworks in different orgs.

Jeffrey: multicast over bier discussion or a use case discussion?

Debashish: Multicast over bier flow overlay. More of an applicability document.

Greg: if service handler and bier is the same physical router then there's a new challenge. Getting bier in hardware is the first challenge. Have you considered a deployment model where bier is the virtual overlay and the service handler can be separated?

Debashish: we haven't thought from a deployment perspective if we want to keep them separate.

Cisco guy: using acronym SR is not a great use of terms in this context.

Greg: perhaps SH for service handler instead of SR.

Akbar: a coauthor. The feedback is we should continue this effort?

Chairs: yes, virtual bier overlays are a good discussion. Having this document and engaging people is of high value.

Bier pim signaling draft - Hooman

Update:

Discovering of EBBR on IBBR examples have been moved to appendix A

Exampels includes SPF, indirect next-hop etc

Next steps:

Solution considered complete

Asking to go to LC

9 read. 7 ready for LC.

Chairs will take to the list.

Tony: this draft is much more readable. Who thinks this draft needs to be even more readable in a new rev?

No response

Andrew: anything specific driving that question?

Tony: the jump in quality was just stunning. Perhaps it could need even more.

Draft-ietf-bier-mld-01 Stig

Document is close to being finished

How to use igmp/mld to use signaling over bier

Added text about MTU

No other changes

Uses configurable multicast groups for all BMLD queriers/nodes

The overlay is quite simple depends on how you implement it except for the mcast groups used for sending reports and queries. Can use the existing link local addresses for igmp queries/reports. But with bier it's not link local, more of a tunnel. Rather think of it as a network for multicast routers. Draft says that the mcast address for queries and reports should be configurable. But tricky to deploy this on every router. Question is it useful to go to iana to ask for a range of addresses or a single well known address. Most people have a single bier domain and need two addresses: queries and reports. Perhaps we should have well known addresses for these.

Greg: challenge is if something leaks for these link local addresses. Leak would be a configuration problem. Even if having multiple sets/areas the mask would prevent it from going outside. The mcast address is hidden from the bier domain. Why not just use the existing addresses?

Stig: more like a purist architectural thing.

Greg: for simplicity. I would map what we have and plug into bier domain and follow mask as set.

Stig: draft says you may use the link local addresses but also says it must be configurable.

Greg: once this progresses we will get larger feedback from larger community.

Stig: now wondering about the pim link local document.

Jeffrey: mentioned multiple instances, multiple sub domains. What's the relationship of instance and subdomain. What's an instance.

Stig: draft isn't clear enough on that.

Jeffrey: could instance be a routing instance.

Stig: yes that could be one possibility. When you send an igmp query on bier if you have those instances you decide which instance/group address to use for query. In the appendix we talk about the data center a bit and group address can signify which instance this is.

Jeffrey: I'll read it more closely and discuss with him offline.

draft-venaas-bier-mtud-01. MTU discovery. Stig.

Gave an MTU overview.

For bier it would be bad to have in-flight fragmentation in the middle of a bier domain. Would have to reencapsulate the fragments.

We could do mtu discovery to avoid fragmentation in bier. A bier router could send a response packet to a BFIR saying packet is too big. Or we could use something Stig is proposing in this draft using probes. But what do you do when there is a change and you wait for new probe, just drop packets? He's proposing a sub-domain mtu. Normally when a link flaps there is no mtu change and no signaling needed and no dropping packets.

Greg: we have a bier domain which is single administered. If we have an mtu mismatch it's a configuration matter. As an operator I wouldn't want some topology change to mess up link config. I wouldn't want my content to fail due to a config problem. I would want to know that with feedback if exceeding mtu. Let the operator know there's a problem on a link but keep forward packets elsewhere.

Stig: could be links in domain can't handle a big mtu.

Greg: DC's should be configuration. Should be only Enterprise case or when crossing admin domains. Response for mtu mismatch should not impact those who are not impacted. If we can't fragment we should only be dropping at link with the problem. Bier shouldn't be crossing administrative domains.

Stig: could be a concern with this draft because it's the weakest link mtu.

Alia: agree this is a config thing most of the time. You could have an acceptable mtu size for bier to mitigate the failures.

Stig: Right, perhaps try to adapt whatever the mtu is and if it's less I'll use whatever the minimum is. Must be this tall to forward. Should announce minimum size before encap is what the draft currently says. But you need to know what the encaps must be. Maybe you announce mtu available after encap.

Jeffrey: have you considered making this generic independent of bier?

Stig: useful to make it bier specific because bier may just be deployed in parts of the network and helpful to know what you can send.

Jeffrey: if you announce generic mtu then bier can figure out everything it needs to know.

Stig: only problem is if you have some small links not part of bier domain it will impact what you think you can transfer via bier.

Jeffery: once you finish spf you know.

Stig: then it becomes a path mtu then domain mtu at that point.

Tony: may want to treat some of this in the oam document. Indicate on which receivers set you blew the mtu. Could punt this, if you are really worried about this put it in the bier payload. Get routers to look at a df bit. Today most people don't even probe in today's mcast deployments.

Stig: useful to have mtu discovery. Already have a wg document about mtu discovery. Is it worth looking at alternatives to just using probes for mtu?

Les co-author: for me the fundamental question is this useful at all? If its useful we are adding a value of advertising something easier to discover the mtu for bier usage. Should we be doing this at all in context of probe draft and this one?

Stig: if you get a new receiver joining on a smaller link then that receiver wouldn't get any packets until a probe discovers the mtu is now smaller. Path mtu discovery has a long history of not working to well.

Chairs will take to the list

BIER p2mp mldp signaling - Hooman

Extending signaling to mldp.

Operators want a lean core.

Would like to choose bier for multicast in core for simplicity.

Stitching the legacy mcast protocols to a bier core is challenging.

Need a solution to stitch legacy protocols to a bier core. One step going forward in the mpls domain. Not trying to propose mldp neighboring thru the bier core. Trying to use protocols already there to make this happen.

Proposal is similar to pim signaling. Need a label in bier domain to represent a p2mp lsp. BTL (bier domain tree label) assigned by bier router closest to the source.

Greg: Why bier domain "tree" label? trying to slice the term tree from bier.

Avoid the tree. Focus on flows.

Jeffrey: there has been some discussion on list and offline. mldp over a core that doesn't run mldp itself has already been specified. Same thing can be done here. Instead of rsvp-te p2mp using a bier tunnel. The concern here is the provider may not want to run mldp session over that core. That mldp session between IBBR/EBBR is not through the core its over the core. It's a tcp session between bbr and ipbr. I don't think it's a concern but to some operators it may be concern. Not much difference between this and existing solution.

Hooman: with regards to label assignment the upstream router will assign label. We feel this is a stepping stone for mobile backhaul going to 5G are trying to get rid of mldp.

Jeffrey: its not through the core its over the core. We can talk to the operators and see.

Andrew: It doesn't have to be a problem but it may be a problem. Both are viable solutions and depends on how operators want to optimize their environments.

Tony: not leaning either way. Boils down to operator tooling. We don't have any control precedence for what you are doing here. What tooling do the operators have? Is all my tooling already across mldp?

Andrew: we sell bullets/guns but don't say how you use them.

Hooman: thoughts of operators are leaning towards bgp for signaling because of thoughts of sdn. We still need tracking of ibbr. 7060 doesn't explain how to track IBBR. The bier header is going to be built on fec and opec that is advertised from ibbr to ebbr. Majority of draft is tracking this. Multiple ways to signal.

Jeffrey: tracking of ibbr/ebbr is similar to existing mldp in band signaling or pim signaling. You already have everything already.

Andrew: are you on a mode of introducing bier to a network to simply part of a network? Are you on a way of mpls as a control plane. Is mpls dead?

Remove it and not introduce it. Different operators will want to run different solutions. Do we want to have both options?

Jeffrey: I'm for simplification and moving forward. Here you are still using mpls over core.

Hooman: example is for security reasons ACL's block unknown ports.

Greg: if we will move forward with both options we would appreciate having them in the same document.

Andrew: I agree.

bier penultimate hop popping - Jeffrey

An mvpn/evpn deployment with enough p/pe routers capable of bier to benefit from using bier as a provider tunnel. We know how to deal with bier incapable P routers. What if some pe routers are not bier capable.

A bier incapable ingress pe would have to send traffic via bier to bier pes and via traditional tunnels to incapable PEs. Complicated and inefficient.

What if an incapable egress pe pretends that it supports bier but request the upstream bfr to pop the bier header? Transparent to other PEs.

Bier php:

A bier incapable router signals bier info but requests other bfrs to pop the bier header and send traffic natively. Those bfrs do not have to be directly connected. There could be incapable p routers in between, traffic could be tunneled.

PHP requested via:

Php sub-sub-tlv in bier sub-tlv. Mpls and non mpls encap

Implicit null label as label range base. In MPLS encap sub-sub-TLV

7 have read draft. 4 thinks it should be adopted.

Chairs will take to list.

Bier brownfield migration frameworks - Tony.

Bier migration options, helping customer to roll out bier.

Non-judgmental, customers pick and choose options for themselves.

How to mix core with bier and non bier routers.

How do I upgrade my PE's. How do I brownfield bier?

Not including bier overlays like mvpn.

Different solutions to get around or through non-BFR's.

Greg: Work came out of discussions in London to provide a migration path.

Mike: This is exactly what we need. Overlays are out of scope. Do you see a use for something similar for overlays?

Tony: yes. We have a ton of drafts which aren't helpful in transforming networks. What are my options to transform my service.

Greg: should we integrate overlays to this document?

Mike: that or create a standalone document. Combined would be too long.

Tony: two different crowds, should be two different documents.

Greg: we can include tradeoffs in the document. Reach out to various folks with drafts and create an overlay migration draft.

mvpn using segment routing and bier for high reachability mcast deployment -
Jingrong

Problem:

Pim has high reachability: only requires the mcast source address

Mldp has a high reachability: only requires the mldp root ip address be reachable

Bier has a poor reachability: Segmented mvpn introduces per flow states on boundary routers. It has limited scope for igp to advertise the bier info.

Sending to many sets (SI's) is something like ingress replication + bier.

Considering a highly reachable SR(IR)+BIER for un-segmented mvpn.

SR for high reachability, stateless

BIER for bandwidth saving, stateless too.

A hybrid p2mp tunnel, SR+BIER, or BIER over SR+hop by hop.

MVPN is deployed end to end

Use SRGB label for BIFT building.

Forwarding plane view: BIER over SR+BIER hop-by-hop

Jeffrey: why is SR label stack needed? I don't think bier label has a problem with this. Let's discuss offline.

Bier entropy for DC CLOS – Mike

Presented this also in mboned.

Option in bier for entropy for deterministic load balancing

Due to ECMP hash function inefficiencies its possible to have frequent flow collisions. More flows get placed on one path over the others.

A deterministic path can be found in CLOS networks by part of the 20 bit entropy field. Bit 0 to bit 2 of entropy label can represent a value of 0-7 and can be used to select a deterministic path from 8 equal costs paths.

Jeffrey: you use part of the 20 bit entropy field, what about the rest of it?

Jingrong: leave the other bits unused.

Alia: sympathetic to wanting deterministic hashing. One thing is putting in SALT. All flows don't end of hashing to the same hash with SALT. If you don't include SALT, the first hop all traffic going across nexthop 1. At the next stage its going to nexthop 1. Need to think through this. There is a reason its hard to

put info about how traffic is hashed. It would be lovely to have the determinism but need to think about it carefully.

Tony: this may be a BCP.

Greg: If this affects forwarding of a bier packet it should be done here.

Entropy does point to a forwarding table. The room agrees with Greg.

Alia: look at SALT. Don't be naïve about determinism.

Work on in bier.

11 read 11 adoption

Mvpn using MPLS P2MP and BIER – Mike

Jingrong has presented this a few times and discussed with many of you.

Mvpn based bier to be more flexible to look at not only igp based forwarding but also mdlp/rsvp-te.

Introducing 2 new tunnel attributes and migration for multicast specific paths.

Removed pim built p2mp bier.

Removed live-live protection.

Candidate for the overlay transition framework document.

Tony: would discourage to try to fix hybrid deployments with overlay techniques. Layers will bleed into each other. We have enough underlay techniques.

Jingrong: what do you mean changing the overlay?

Tony: I see you use pim signaling mechanisms.

Mike: we agree and we removed pim signaling.

5 read 4 adoption

Draft-xie-bier-6man-encap-Jingrong

Encaps for bier in non mpls networks

Can can bier ipv6 encap behave like mpls?

Checking existing ipv6 options

First problem is where to put the bier header. We could reinvent new v6 extension header. Not recommended. Or we reuse existing v6 extension header.

Greg: we have a well-defined bier header. When forwarding bier again just use the bier header.

Jingrong: Two separate questions. First is where is the right place to place the bier header. Second question is how encap can run over SRH.

Greg: once we are out of srv6 its bier. At some point its at a bier native router and will forward bier so use bier encoding at that point.

Jeffrey: when using non mpls encap you can't tunnel bier packet? I don't understand that. Why can't you tunnel bier packets?

Jingrong: currently the rfc has two encap. First one is mpls.

Jeffrey: as long as the tunnel can indicate the next payload type you are fine.

Tony: we carry L2 frames over many tunnels. Only problem is tunnel has to indicate tunnel type which is not in the scope of this group.

Jeffrey: don't need to include a l2 header in front of bier header as long as transport tunnel indicates the next header is a bier header. Ipgre tunnel has a payload type. As long as tunnel can indicate next payload type you are set.

Alia: Ethertype is commonly used in a number of overlay protocols to describe the payload type. Which is why we used ethertype. If you look at geneve we have a protocol type. This exists in a number of different tunneling headers to have consistency to make it easy to include an arbitrary payload in a tunnel. Good relationship with iee for ethertypes.

Jingrong: Bier ipv6 hop by hop. Capture the bier header in dest option header.

Greg: if you can come up with reasons why existing solutions don't work then specify them.

Discussion to list.

Segmented MVPN Using IP Lookup for BIER - Jingrong

LIR explicit tracking for segmented BIER is very inefficient.

We can use LIR-pF explicit tracking to reduce join latency.

Jeffrey: This is not specific to bier or explicit tracking. Your trying to solve the problem of a per flow label advertisements in an segmented ABR.

Jingrong: bier needs path flow states.

Jeffrey: Resorting to ip lookup makes the problem worse. You save the mpls state but introduce the ip states in the vrfs.

Jeffrey: I agree if you don't want to do per flow label then you can use ip lookup. But again you are making the problem worse. Scaling problem in forwarding path is even worse.

Will take it offline.