Network Working Group
Internet-Draft
Intended status: Informational
Expires: October 29, 2015

N. Kumar
R. Asati
Cisco
M. Chen
X. Xu
Huawei
A. Dolganow
Alcatel-Lucent
T. Przygienda
Ericsson
A. Gulko
Thomson Reuters
D. Robinson
id3as-company Ltd
April 27, 2015

BIER Use Cases
draft-ietf-bier-use-cases-00.txt

Abstract

Bit Index Explicit Replication (BIER) is an architecture that provides optimal multicast forwarding through a "BIER domain" without requiring intermediate routers to maintain any multicast related per-flow state. BIER also does not require any explicit tree-building protocol for its operation. A multicast data packet enters a BIER domain at a "Bit-Forwarding Ingress Router" (BFIR), and leaves the BIER domain at one or more "Bit-Forwarding Egress Routers" (BFERs). The BFIR router adds a BIER header to the packet. The BIER header contains a bit-string in which each bit represents exactly one BFER to forward the packet to. The set of BFERs to which the multicast packet needs to be forwarded is expressed by setting the bits that correspond to those routers in the BIER header.

This document describes some of the use-cases for BIER.

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 http://datatracker.ietf.org/drafts/current/.

Table of Contents

1. Introduction 2
2. Specification of Requirements 3
3. BIER Use Cases 3
3.1. Multicast in L3VPN Networks 3
3.2. BUM in EVPN 4
3.3. IPTV and OTT Services 5
3.4. Multi-service, converged L3VPN network 6
3.5. Control-plane simplification and SDN-controlled networks 7
3.6. Data center Virtualization/Overlay 7
3.7. Financial Services 8
4. Security Considerations 9
5. IANA Considerations 9
6. Acknowledgments 9
7. References 9
7.1. Normative References 9
7.2. Informative References 9
Authors' Addresses 10

1. Introduction

Bit Index Explicit Replication (BIER) [I-D.wijnands-bier-architecture] is an architecture that provides optimal multicast forwarding through a "BIER domain" without requiring intermediate routers to maintain any multicast related per-

flow state. BIER also does not require any explicit tree-building protocol for its operation. A multicast data packet enters a BIER domain at a "Bit-Forwarding Ingress Router" (BFIR), and leaves the BIER domain at one or more "Bit-Forwarding Egress Routers" (BFERs). The BFIR router adds a BIER header to the packet. The BIER header contains a bit-string in which each bit represents exactly one BFER to forward the packet to. The set of BFERs to which the multicast packet needs to be forwarded is expressed by setting the bits that correspond to those routers in the BIER header.

The obvious advantage of BIER is that there is no per flow multicast state in the core of the network and there is no tree building protocol that sets up tree on demand based on users joining a multicast flow. In that sense, BIER is potentially applicable to many services where Multicast is used and not limited to the examples described in this draft. In this document we are describing a few use-cases where BIER could provide benefit over using existing mechanisms.

2. Specification of Requirements

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

3. BIER Use Cases

3.1. Multicast in L3VPN Networks

The Multicast L3VPN architecture [RFC6513] describes many different profiles in order to transport L3 Multicast across a providers network. Each profile has its own different tradeoffs (see section 2.1 [RFC6513]). When using "Multidirectional Inclusive" "Provider Multicast Service Interface" (MI-PMSI) an efficient tree is build per VPN, but causes flooding of egress PE's that are part of the VPN, but have not joined a particular C-multicast flow. This problem can be solved with the "Selective" PMSI to build a special tree for only those PE's that have joined the C-multicast flow for that specific VPN. The more S-PMSI's, the less bandwidth is wasted due to flooding, but causes more state to be created in the providers network. This is a typical problem network operators are faced with by finding the right balance between the amount of state carried in the network and how much flooding (waste of bandwidth) is acceptable. Some of the complexity with L3VPN's comes due to providing different profiles to accommodate these trade-offs.

With BIER there is no trade-off between State and Flooding. Since the receiver information is explicitly carried within the packet,

there is no need to build S-PMSI's to deliver multicast to a sub-set of the VPN egress PE's. Due to that behaviour, there is no need for S-PMSI's.

Mi-PMSI's and S-PMSI's are also used to provide the VPN context to the Egress PE router that receives the multicast packet. Also, in some MVPN profiles it is also required to know which Ingress PE forwarded the packet. Based on the PMSI the packet is received from, the target VPN is determined. This also means there is a requirement to have a least a PMSI per VPN or per VPN/Ingress PE. This means the amount of state created in the network is proportional to the VPN and ingress PE's. Creating PMSI state per VPN can be prevented by applying the procedures as documented in [RFC5331]. This however has not been very much adopted/implemented due to the excessive flooding it would cause to Egress PE's since *all* VPN multicast packets are forwarded to *all* PE's that have one or more VPN's attached to it.

With BIER, the destination PE's are identified in the multicast packet, so there is no flooding concern when implementing [RFC5331]. For that reason there is no need to create multiple BIER domain's per VPN, the VPN context can be carry in the multicast packet using the procedures as defined in [RFC5331]. Also see [I-D.rosen-l3vpn-mvpn-bier] for more information.

With BIER only a few MVPN profiles will remain relevant, simplifying the operational cost and making it easier to be interoperable among different vendors.

3.2. BUM in EVPN

The current widespread adoption of L2VPN services [RFC4664], especially the upcoming EVPN solution [I-D.ietf-l2vpn-evpn] which transgresses many limitations of VPLS, introduces the need for an efficient mechanism to replicate broadcast, unknown and multicast (BUM) traffic towards the PEs that participate in the same EVPN instances (EVIs). As simplest deployable mechanism, ingress replication is used but poses accordingly a high burden on the ingress node as well as saturating the underlying links with many copies of the same frame headed to different PEs. Fortunately enough, EVPN signals internally P-Multicast Service Interface (PMSI) [RFC6513] attribute to establish transport for BUM frames and with that allows to deploy a plethora of multicast replication services that the underlying network layer can provide. It is therefore relatively simple to deploy BIER P-Tunnels for EVPN and with that distribute BUM traffic without building of P-router state in the core required by PIM, mLDP or comparable solutions.

Specifically, the same I-PMSI attribute suggested for mVPN can be used easily in EVPN and given EVPN can multiplex and disassociate BUM frames on p2mp and mp2mp trees using upstream assigned labels, BIER P-Tunnel will support BUM flooding for any number of EVIs over a single sub-domain for maximum scalability but allow at the other extreme of the spectrum to use a single BIER sub-domain per EVI if such a deployment is necessary.

Multiplexing EVIs onto the same PMSI forces the PMSI to span more than the necessary number of PEs normally, i.e. the union of all PEs participating in the EVIs multiplexed on the PMSI. Given the properties of BIER it is however possible to encode in the receiver bitmask only the PEs that participate in the EVI the BUM frame targets. In a sense BIER is an inclusive as well as a selective tree and can allow to deliver the frame to only the set of receivers interested in a frame even though many others participate in the same PMSI.

As another significant advantage, it is imaginable that the same BIER tunnel needed for BUM frames can optimize the delivery of the multicast frames though the signaling of group memberships for the PEs involved has not been specified as of date.

3.3. IPTV and OTT Services

IPTV is a service, well known for its characteristics of allowing both live and on-demand delivery of media traffic over end-to-end Managed IP network.

Over The Top (OTT) is a similar service, well known for its characteristics of allowing live and on-demand delivery of media traffic between IP domains, where the source is often on an external network relative to the receivers.

Content Delivery Networks (CDN) operators provide layer 4 applications, and often some degree of managed layer 3 IP network, that enable media to be securely and reliably delivered to many receivers. In some models they may place applications within third party networks, or they may place those applications at the edges of their own managed network peerings and similar inter-domain connections. CDNs provide capabilities to help publishers scale to meet large audience demand. Their applications are not limited to audio and video delivery, but may include static and dynamic web content, or optimized delivery for Massive Multiplayer Gaming and similar. Most publishers will use a CDN for public Internet delivery, and some publishers will use a CDN internally within their IPTV networks to resolve layer 4 complexity.

In a typical IPTV environment the egress routers connecting to the receivers will build the tree towards the ingress router connecting to the IPTV servers. The egress routers would rely on IGMP/MLD (static or dynamic) to learn about the receiver's interest in one or more multicast group/channels. Interestingly, BIER could allows provisioning any new multicast group/channel by only modifying the channel mapping on ingress routers. This is deemed beneficial for the linear IPTV video broadcasting in which every receivers behind every egress PE routers would receive the IPTV video traffic.

With BIER in IPTV environment, there is no need of tree building from egress to ingress. Further, any addition of new channel or new egress routers can be directly controlled from ingress router. When a new channel is included, the multicast group is mapped to Bit string that includes all egress routers. Ingress router would start sending the new channel and deliver it to all egress routers. As it can be observed, there is no need for static IGMP provisioning in each egress routers whenever a new channel/stream is added. Instead, it can be controlled from ingress router itself by configuring the new group to Bit Mask mapping on ingress router.

With BIER in OTT environment, these edge routers in CDN domain terminating the OTT user session connect to the Ingress BIER routers connecting content provider domains or a local cache server and leverage the scalability benefit that BIER could provide. This may rely on MBGP interoperation (or similar) between the egress of one domain and the ingress of the next domain, or some other SDN control plane may prove a more effective and simpler way to deploy BIER. For a single CDN operator this could be well managed in the Layer 4 applications that they provide and it may be that the initial receiver in a remote domain is actually an application operated by the CDN which in turn acts as a source for the Ingress BIER router in that remote domain, and by doing so keeps the BIER more descrete on a domain by domain basis.

3.4. Multi-service, converged L3VPN network

Increasingly operators deploy single networks for multiple-services. For example a single Metro Core network could be deployed to provide Residential IPTV retail service, residential IPTV wholesale service, and business L3VPN service with multicast. It may often be desired by an operator to use a single architecture to deliver multicast for all of those services. In some cases, governing regulations may additionally require same service capabilities for both wholesale and retail multicast services. To meet those requirements, some operators use multicast architecture as defined in [RFC5331]. However, the need to support many L3VPNs, with some of those L3VPNs scaling to hundreds of egress PE's and thousands of C-multicast

flows, make scaling/efficiency issues defined in earlier sections of this document even more prevalent. Additionally support for ten's of millions of BGP multicast A-D and join routes alone could be required in such networks with all consequences such a scale brings.

With BIER, again there is no need of tree building from egress to ingress for each L3VPN or individual or group of c-multicast flows. As described earlier on, any addition of a new IPTV channel or new egress router can be directly controlled from ingress router and there is no flooding concern when implementing [RFC5331].

3.5. Control-plane simplification and SDN-controlled networks

With the advent of Software Defined Networking, some operators are looking at various ways to reduce the overall cost of providing networking services including multicast delivery. Some of the alternatives being consider include minimizing capex cost through deployment of network-elements with simplified control plane function, minimizing operational cost by reducing control protocols required to achieve a particular service, etc. Segment routing as described in [I-D.ietf-spring-segment-routing] provides a solution that could be used to provide simplified control-plane architecture for unicast traffic. With Segment routing deployed for unicast, a solution that simplifies control-plane for multicast would thus also be required, or operational and capex cost reductions will not be achieved to their full potential.

With BIER, there is no longer a need to run control protocols required to build a distribution tree. If L3VPN with multicast, for example, is deployed using [RFC5331] with MPLS in P-instance, the MPLS control plane would no longer be required. BIER also allows migration of C-multicast flows from non-BIER to BIER-based architecture, which makes transition to control-plane simplified network simpler to operationalize. Finally, for operators, who would desire centralized, offloaded control plane, multicast overlay as well as BIER forwarding could migrate to controller-based programming.

3.6. Data center Virtualization/Overlay

Virtual eXtensible Local Area Network (VXLAN) [RFC7348] is a kind of network virtualization overlay technology which is intended for multi-tenancy data center networks. To emulate a layer2 flooding domain across the layer3 underlay, it requires to have a mapping between the VXLAN Virtual Network Instance (VNI) and the IP multicast group in a ratio of 1:1 or n:1. In other words, it requires to enable the multicast capability in the underlay. For instance, it requires to enable PIM-SM [RFC4601] or PIM-BIDIR [RFC5015] multicast

routing protocol in the underlay. VXLAN is designed to support 16M VNIs at maximum. In the mapping ratio of 1:1, it would require 16M multicast groups in the underlay which would become a significant challenge to both the control plane and the data plane of the data center switches. In the mapping ratio of n:1, it would result in inefficiency bandwidth utilization which is not optimal in data center networks. More importantly, it is recognized by many data center operators as a unaffordable burden to run multicast in data center networks from network operation and maintenance perspectives. As a result, many VXLAN implementations are claimed to support the ingress replication capability since ingress replication eliminates the burden of running multicast in the underlay. Ingress replication is an acceptable choice in small-sized networks where the average number of receivers per multicast flow is not too large. However, in multi-tenant data center networks, especially those in which the NVE functionality is enabled on a high amount of physical servers, the average number of NVEs per VN instance would be very large. As a result, the ingress replication scheme would result in a serious bandwidth waste in the underlay and a significant replication burden on ingress NVEs.

With BIER, there is no need for maintaining that huge amount of multicast states in the underlay anymore while the delivery efficiency of overlay BUM traffic is the same as if any kind of stateful multicast protocols such as PIM-SM or PIM-BIDIR is enabled in the underlay.

3.7. Financial Services

Financial services extensively rely on IP Multicast to deliver stock market data and its derivatives, and critically require optimal latency path (from publisher to subscribers), deterministic convergence (so as to deliver market data derivatives fairly to each client) and secured delivery.

Current multicast solutions e.g. PIM, mLDP etc., however, don't sufficiently address the above requirements. The reason is that the current solutions are primarily subscriber driven i.e. multicast tree is setup using reverse path forwarding techniques, and as a result, the chosen path for market data may not be latency optimal from publisher to the (market data) subscribers.

As the number of multicast flows grows, the convergence time might increase and make it somewhat nondeterministic from the first to the last flow depending on platforms/implementations. Also, by having more protocols in the network, the variability to ensure secured delivery of multicast data increases, thereby undermining the overall security aspect.

BIER enables setting up the most optimal path from publisher to subscribers by leveraging unicast routing relevant for the subscribers. With BIER, the multicast convergence is as fast as unicast, uniform and deterministic regardless of number of multicast flows. This makes BIER a perfect multicast technology to achieve fairness for market derivatives per each subscriber.

4. Security Considerations

There are no security issues introduced by this draft.

5. IANA Considerations

There are no IANA consideration introduced by this draft.

6. Acknowledgments

The authors would like to thank IJsbrand Wijnands, Greg Shepherd and Christian Martin for their contribution.

7. References

7.1. Normative References

[I-D.rosen-l3vpn-mvpn-bier]
Rosen, E., Sivakumar, M., Wijnands, I., Aldrin, S., Dolganow, A., and T. Przygienda, "Multicast VPN Using BIER", draft-rosen-l3vpn-mvpn-bier-02 (work in progress), December 2014.

[I-D.wijnands-bier-architecture]
Wijnands, I., Rosen, E., Dolganow, A., Przygienda, T., and S. Aldrin, "Multicast using Bit Index Explicit Replication", draft-wijnands-bier-architecture-04 (work in progress), February 2015.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

7.2. Informative References

[I-D.ietf-l2vpn-evpn]
Sajassi, A., Aggarwal, R., Bitar, N., Isaac, A., and J. Uttaro, "BGP MPLS Based Ethernet VPN", draft-ietf-l2vpn-evpn-11 (work in progress), October 2014.

[I-D.ietf-spring-segment-routing]
Filsfils, C., Previdi, S., Bashandy, A., Decraene, B., Litkowski, S., Horneffer, M., Shakir, R., Tantsura, J., and E. Crabbe, "Segment Routing Architecture", draft-ietf-spring-segment-routing-01 (work in progress), February 2015.

[RFC4601] Fenner, B., Handley, M., Holbrook, H., and I. Kouvelas, "Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)", RFC 4601, August 2006.

[RFC4664] Andersson, L. and E. Rosen, "Framework for Layer 2 Virtual Private Networks (L2VPNs)", RFC 4664, September 2006.

[RFC5015] Handley, M., Kouvelas, I., Speakman, T., and L. Vicisano, "Bidirectional Protocol Independent Multicast (BIDIR-PIM)", RFC 5015, October 2007.

[RFC5331] Aggarwal, R., Rekhter, Y., and E. Rosen, "MPLS Upstream Label Assignment and Context-Specific Label Space", RFC 5331, August 2008.

[RFC6513] Rosen, E. and R. Aggarwal, "Multicast in MPLS/BGP IP VPNs", RFC 6513, February 2012.

[RFC7348] Mahalingam, M., Dutt, D., Duda, K., Agarwal, P., Kreeger, L., Sridhar, T., Bursell, M., and C. Wright, "Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks", RFC 7348, August 2014.

Authors' Addresses

Nagendra Kumar
Cisco
7200 Kit Creek Road
Research Triangle Park, NC 27709
US

Email: naikumar@cisco.com

Rajiv Asati
Cisco
7200 Kit Creek Road
Research Triangle Park, NC 27709
US

Email: rajiva@cisco.com

Mach(Guoyi) Chen
Huawei

Email: mach.chen@huawei.com

Xiaohu Xu
Huawei

Email: xuxiaohu@huawei.com

Andrew Dolganow
Alcatel-Lucent
600 March Road
Ottawa, ON K2K2E6
Canada

Email: andrew.dolganow@alcatel-lucent.com

Tony Przygienda
Ericsson
300 Holger Way
San Jose, CA 95134
USA

Email: antoni.przygienda@ericsson.com

Arkadiy Gulko
Thomson Reuters
195 Broadway
New York NY 10007
USA

Email: arkadiy.gulko@thomsonreuters.com

Dom Robinson
id3as-company Ltd
UK

Email: Dom@id3as.co.uk