Abstract

Multicast VPN, based on the Virtual Router model using PIM with GRE tunnels, has been in operation in production networks for several years. This document describes some of the experiences gained from implementation and deployment of such Multicast VPNs. Further it describes the practices used by such deployments based on the experience.
1. Introduction

The first proposal for multicast support for BGP/MPLS IP VPNs [RFC4364], informally known as "draft-rosen", was presented in San Diego IETF, 2000. At that time multicast support was not in the L3VPN WG charter. Since 2000 draft-rosen went through several revisions, with the last revision informally known as draft-rosen-08.

At San Diego IETF in 2004 multicast support has been added to the L3VPN WG charter. At the present moment L3VPN WG has several working group documents that specify mechanisms for multicast support in BGP/MPLS IP VPNs ([MVPN-ARCH], [BGP-MPLS-MVPN]).

At the present moment multi-vendor interoperable implementations are known to exist only based on a particular version of draft-rosen, informally known as draft-rosen-06, but not based on the later version of draft-rosen - draft-rosen-08.

While the mechanisms described in draft-rosen-06 form a subset of the mechanisms described in [MVPN-ARCH], the additional mechanisms introduced in draft-rosen-08 are not part of the mechanisms described in [MVPN-ARCH]. Specifically the mechanisms to support PIM-SSM for Inclusive P-Tunnels and inter-AS operations option (B) specified in draft-rosen-08 are not part of the mechanisms described in [MVPN-ARCH]. The mechanisms to support PIM-SSM for Inclusive P-Tunnels and inter-AS operations option (B) specified in [MVPN-ARCH] are not interoperable with the corresponding mechanisms in draft-rosen-08.

While unicast BGP/MPLS IP VPNs solution is based on the aggregated routing architecture [RFC4110], the approach taken by draft-rosen is based on a completely different architecture, namely Virtual Routers [RFC4110].

The differences between unicast BGP/MPLS IP VPNs and draft-rosen are not only in the architecture, but also in the mechanisms:

+ While unicast BGP/MPLS IP VPNs use BGP to exchange (unicast) VPN’s routing information among PEs, draft-rosen uses PIM to exchange (multicast) VPN’s routing information among PEs.
While unicast BGP/MPLS IP VPNs use MPLS in the data plane (with GRE as an option for inter-PE tunnels), \textit{draft-rosen} restricts the data plane to just GRE or IP-in-IP tunnels. (While the draft mentions MPLS as a possible encapsulation, the draft specifies no details on how to support it). Moreover, in terms of implementations multi-vendor implementations exist only for GRE, but not for IP-in-IP.

While unicast BGP/MPLS IP VPNs use LDP or RSVP-TE to set up inter-PE MPLS tunnels, \textit{draft-rosen} uses PIM for setting up (GRE-based) inter-PE tunnels.

While the control plane used by unicast BGP/MPLS IP VPNs is decoupled from its data plane, the control plane and the data plane of \textit{draft-rosen} are coupled in a sense that exactly the same inter-PE tunnels are used to exchange both control and data.

As a result, deploying \textit{draft-rosen} requires a set of control plane and data plane mechanisms among the PEs that are completely different from what is required by (unicast) BGP/MPLS IP VPNs.

2. Control Plane Scalability Considerations

Use of the Virtual Router architecture by \textit{draft-rosen} means that a given VRF on a given PE has to maintain PIM adjacencies with every other VRF belonging to that MVPN on every other PE. Moreover, these adjacencies have to be maintained not on a per PE, but on a per MVPN, per PE granularity. A PE also has to maintain PIM adjacencies with all the locally connected CEs (and these adjacencies are not due to the use of the Virtual Router architecture). However, as long as the average number of a given MVPN sites connected to a given PE is less than the average number of PEs that have sites of that MVPN, then on a given PE the overhead of maintaining PIM adjacencies with other PEs will dominate the overhead of PIM adjacencies between that PE and its locally connected CEs. Note that this overhead grows with the number of PEs in an MVPN, as well as with the number of MVPNs.

To illustrate the overhead consider an example where a PE has 1,000 VRFs, and each of these VRFs corresponds to an MVPN that on average is present on 100 PEs. The average number of PIM adjacencies that the PE would need to maintain with other PEs is 100,000. The average rate of PIM Hellos that the PE would need to process is 3,300 per second.
3. Join and Prune Latency

One consequence of using unreliable transport for PE-PE MVPN multicast routing information exchange with PIM is that if the first PIM Join sent by a PE to another PE is lost, then this results in Join latency of at least up to the duration of the PIM Join retransmission. This duration is usually at least 30 seconds. Losing subsequent PIM Joins may further increase this Join latency.

Similarly if the last PIM Prune sent by a PE to another PE is lost, then this results in the receiver PEs receiving unwanted data until the corresponding multicast state on the sender/upstream PE times out, which, as estimated in [PORT], is roughly 3 minutes. That has negative implications on the efficient use of bandwidth within the Service Provider(s).

Moreover, if the last PIM Prune sent by a PE to another (upstream) PE is lost, then this results in the upstream PE maintaining the corresponding multicast state until that state times out, which as estimated in [PORT] is roughly 3 minutes.

Additional issue of increased Join latency when switching to S-PMSIs is described in the next section.

4. Possible packet losses when switching to S-PMSI

Signaling switching to S-PMSI is done by periodically (every 60 secs) sending a UDP message that contains the identity of the provider multicast tree used by an S-PMSI as well as the customer multicast stream bound to that S-PMSI.

One consequence of using unreliable transport (UDP) for signaling switching to S-PMSI is that if the first S-PMSI advertisement is lost, then this results in losses of multicast data for up to 57 secs (MDT-INTERVAL - MDT-DATA-DELAY).

Moreover, if all the PEs in a given MVPN do not always store all the S-PMSI advertisements for that MVPN, irrespective of whether these PEs have downstream receivers for the multicast traffic carried by these S-PMSIs, then it may result in multicast data losses (Join latency) on average of 30 secs and up to 60 secs (MDT-INTERVAL timer) on the PEs. This is in the absence of any S-PMSI advertisement losses. Losing an S-PMSI advertisement may further increase the duration of the losses (increases this Join latency).
5. Limitations when Handling Anycast Customer RP (C-RP)

Large enterprises may have multiple data centers where Anycast RP’s and sources of multicast traffic are located. Such MVPN customers might require multicast applications to be simultaneously sourced from each data center and delivered via corresponding Anycast RP’s to different sets of branch locations. Each data center and each branch location may be in its own MVPN site.

When an MVPN uses anycast RP, where several customer’s RPs (C-RPs) share the same anycast address and are reachable via more than one PE, then with the mechanisms provided by draft-rosen for a given customer’s multicast group (C-G) at any given point in time only one of these C-RPs can be used to deliver (multicast) traffic to receivers in other sites of that MVPN. That eliminates one of the main benefits of anycast RP – the ability to share load among multiple RPs.

6. Limitations when Handling Multi-homed Multicast Sources

When an MVPN site contains a given multicast source, and the site is connected to two or more PEs, then at any given point in time only one of these PEs could be used to forward multicast traffic from the source to the receivers in other sites of that MVPN. This is in contrast to unicast, where unicast traffic could be forwarded to the source via all of these PEs.

7. Mandatory I-PMSI

Mechanisms used by draft-rosen mandate that each MVPN must have its own I-PMSI. This is even if most/all of the multicast data is sent using S-PMSIs. The overhead of maintaining I-PMSI, both in the control and in the data plane, is especially significant in the inter-AS scenario, as in this scenario the number of point-to-multipoint tunnels required for a single I-PMSI is equal to the number of PEs that span that I-PMSI.
8. QoS

If a service provider uses DiffServ based QoS, the service provider needs two different mechanisms for providing such QoS for its VPN customers – one for unicast BGP/MPLS VPNs using MPLS and another for multicast VPNs using draft-rosen. This is because the former uses MPLS based DiffServ mechanisms such as mapping the MPLS EXP bits to forwarding classes while the latter uses IP based DiffServ mechanisms such as mapping the DSCP bits to forwarding classes.

Moreover, certain QoS mechanisms that are available for (unicast) BGP/MPLS IP VPNs are not available for draft-rosen. Specifically, MPLS Traffic Engineering and MPLS DiffServ Traffic Engineering that are available for unicast VPN traffic are not available for draft-rosen.

9. Protection/Restoration

Certain protection/restoration mechanisms that are available for (unicast) BGP/MPLS IP VPNs are not available for draft-rosen. Specifically, none of the MPLS Fast Re-route mechanisms that could be used in conjunction with unicast VPN traffic are available for draft-rosen.

10. Security Considerations

The Security Considerations section of RFC4797 discusses security issues in the context of unicast BGP/MPLS IP VPNs when PE-PE tunnels are realized via GRE. These considerations are applicable in the context of draft-rosen as well. However, draft-rosen presents additional security considerations, above and beyond of what is covered in RFC4797. The following describes some of them.

Inability to restrict joining an I-PMSI (Default MDT) of a given MVPN to only the PEs that have VRFs of that MVPN may result in (a) leaking multicast traffic originated within that MVPN to the receivers that are not supposed to be part of that MVPN, and (b) various forms of packet spoofing, as described below. Once an attacker joins an I-PMSI of a given MVPN, the attacker, in addition to the ability to receive multicast traffic originated within that MVPN, can also create attacks based on packet spoofing.

The implications of packet spoofing in the context of draft-rosen are more significant than in the context of RFC4797, as in the context of RFC4797 spoofing can impact only the data traffic, while in the context of draft-rosen spoofing can impact not only the data, but...
also the control traffic associated with the exchange of MVPN routing information among PEs. This is because in the context of \textit{draft-rosen} the same GRE tunnels that are used to exchange MVPN data traffic among PEs are also used to exchange MVPN multicast routing information among PEs.

Securing control traffic associated with the exchange of MVPN routing information among PEs by applying security mechanisms specified in [RFC4601] to PIM that is used to exchange MVPN routing information among PEs is problematic for the following reasons.

One option specified in [RFC4601] states that "A PIM router SHOULD provide an option to limit the set of neighbors from which it will accept Join/Prune, Assert, and Hello messages. Either static configuration of IP addresses or an IPsec security association may be used." Use of the first option, static configuration, in the context of \textit{draft-rosen} to authenticate PIM messages used to exchange MVPN routing information among PEs may be problematic, as it would require a given PE for each MVPN present on the PE to have an apriori knowledge of all the other PEs with whom the PE has at least one MVPN in common. That gets especially problematic in the inter-provider scenario where PEs in different providers may need to become PIM neighbors, as it would require a provider to share this information with other providers.

Another option specified in [RFC4601] is to use IPsec. As stated in [RFC4601] "To use IPsec, the administrator of a PIM network configures each PIM router with one or more security associations (SAs) and associated Security Parameter Indexes (SPIs) that are used by senders to authenticate PIM protocol messages and are used by receivers to authenticate received PIM protocol messages." However, neither [RFC4601], nor \textit{draft-rosen} provide an automated mechanism for establishing such security associations. In fact [RFC4601] "assumes that manual configuration of SAs is performed". Use of this option in the context of \textit{draft-rosen} to authenticate PIM messages used to exchange MVPN routing information among PEs is problematic, as it involves manual configuration of SAs on PEs that have at least one MVPN in common. It is especially problematic in the inter-provider scenario where it may involve PEs in different providers.

At the CE-PE interfaces each PE must install packet filters to filter out any UDP traffic on port 3232 that the PE may receive from any of its attached CEs, as UDP port 3232 is used for the S-PMSI messages exchanged among PEs. Failure to filter out such traffic can cause the creation of extra multicast states on the Service Providers routers.

Additional security considerations that are applicable in the context
of draft-rosen are described in [RFC4023].

11. IANA Considerations

This document does not require any action on the part of IANA.

12. Intellectual Property Statement

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

13. Copyright Notice

Copyright (C) The IETF Trust (2008).

This document is subject to the rights, licenses and restrictions contained in BCP 78, and except as set forth therein, the authors retain all their rights.

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.
14. Disclaimer of validity:

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79. Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

15. Acknowledgements

We would like to thank Maria Napierala for pointing to the problems with draft-rosen of supporting MVPN customers who use Anycast RP, and MVPN customers who have (multicast) sources in multihomed sites.

Many thanks to Leonard Giuliano for his review and comments.

16. Normative References

17. Non-normative References


18. Author Information

Rahul Aggarwal
Juniper Networks, Inc.
e-mail: rahul@juniper.net

Yakov Rekhter
Juniper Networks, Inc.
e-mail: yakov@juniper.net