The RPF Vector TLV

draft-ietf-pim-rpf-vector-03

Status of this Memo

By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with Section 6 of BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt.

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

This Internet-Draft will expire on April 4, 2007.

Copyright Notice

Copyright (C) The Internet Society (2006).

Abstract

This document describes a use of the PIM Join Attribute as defined in draft-ietf-pim-join-attributes[I-D.ietf-pim-join-attributes] which enables PIM to build multicast trees through an MPLS-enabled network, even if that network’s IGP does not have a route to the source of the tree.
# Table of Contents

1. Introduction .............................................. 3  
2. Use of the RPF Vector TLV ................................. 4  
   2.1. Attribute and shared tree joins ....................... 4  
   2.2. Attribute and Bootstrap messages .................... 5  
   2.3. The Vector Attribute ................................. 5  
      2.3.1. Inserting a Vector Attribute in a Join ............ 5  
      2.3.2. Processing a Received Vector Attribute .......... 5  
      2.3.3. Vector Attribute and Asserts .................... 5  
3. Vector Attribute TLV Format ............................... 7  
4. IANA Considerations ...................................... 8  
5. Security Considerations .................................. 8  
6. Acknowledgments ......................................... 8  
7. References ............................................... 8  
   7.1. Normative References ................................ 8  
   7.2. Informative References .............................. 9  
Authors' Addresses ......................................... 9  
Intellectual Property and Copyright Statements ............... 10
1. Introduction

It is sometimes convenient to distinguish the routers of a particular network into two categories: "edge routers" and "core routers". The edge routers attach directly to users or to other networks, but the core routers attach only to other routers of the same network. If the network is MPLS-enabled, then any unicast packet which needs to travel outside the network can be "tunneled" via MPLS from one edge router to another. To handle a unicast packet which must travel outside the network, an edge router needs to know which of the other edge routers is the best exit point from the network for that packet's destination IP address. The core routers, however, do not need to have any knowledge of routes which lead outside the network; as they handle only tunneled packets, they only need to know how to reach the edge routers and the other core routers.

Consider, for example, the case where the network is an Autonomous System (AS), the edge routers are EBGP speakers, the core routers may be said to constitute a "BGP-free core". The edge routers must distribute BGP routes to each other, but not to the core routers.

However, when multicast packets are considered, the strategy of keeping the core routers free of "external" routes is more problematic. When using PIM-SM[RFC4601], PIM-SSM[I-D.ietf-ssm-arch] or PIM-BIDIR[I-D.ietf-pim-bidir] to create a multicast distribution tree for a particular multicast group, one wants the core routers to be full participants in the PIM protocol, so that multicasting can be done efficiently in the core. This means that the core routers must be able to correctly process PIM Join messages for the group, which in turn means that the core routers must be able to send the Join messages towards the root of the distribution tree. If the root of the tree lies outside the network’s borders (e.g., is in a different AS), and the core routers do not maintain routes to external destinations, then the PIM Join messages cannot be processed, and the multicast distribution tree cannot be created.

In order to allow PIM to work properly in an environment where the core routers do not maintain external routes, a PIM extension is needed. When an edge router sends a PIM Join message into the core, it must include in that message a "Vector" which specifies the IP address of the next edge router along the path to the root of the multicast distribution tree. The core routers can then process the Join message by sending it towards the specified edge router (i.e., toward the Vector). In effect, the Vector serves as an attribute, within a particular network, for the root of the tree.

This document defines a new TLV in the PIM Join Attribute message[I-D.ietf-pim-joint-attributes]. It consists of a single
Vector which identifies the exit point of the network.

2. Use of the RPF Vector TLV

Before we can start forwarding multicast packets we need to build a forwarding tree by sending PIM Joins hop by hop. Each router in the path creates a forwarding state and propagates the Join towards the root of the forwarding tree. The building of this tree is receiver driven. See Figure 1.

------------------ BGP ------------------
|[S]---( Edge 1)--(Core 1)---( Core )--(Core 2)---( Edge 2 )---[R]
<--- (S,G) Join

Figure 1

In this example, the 2 edge routers are BGP speakers. The core routers are not BGP speakers and do not have any BGP distributed routes. The route to S is a BGP distributed route, hence is known to the edge but not to the core. The Edge 2 router determines the interface leading to S, and sends a PIM Join to the upstream router. In this example, though, the upstream router is a core router, with no route to S. Without the PIM extensions specified in this document, the core router cannot determine where the send the Join, so the tree cannot be constructed.

To allow the core router to participate in the construction of the tree, the Edge 2 router will include an attribute field in the PIM Join. In this example, the Attribute field will contain the IP address of Edge 1. Edge 2 then forwards the PIM Join towards Edge 1. The intermediate core router do their RPF check on the Attribute (IP address of Edge 1) rather than the Source, this allows the tree to be constructed.

2.1. Attribute and shared tree joins

In the example above we build a source tree to illustrate the attribute behavior. The attribute is however not restricted to source tree only. The tree may also be constructed towards a Rendezvous Point (RP) IP address. The RP IP address is used in a similar way as the Source in the example above. PIM Attribute procedures defined for sources are equally applicable to (*,G) and (*,*,RP) joins unless otherwise noted.
2.2. Attribute and Bootstrap messages

The RPF vector does not apply to BSR bootstrap messages. To allow BSR messages to be forwarded across a core where the RP IP address is not routable in the core a solution has the developed in BSR.

2.3. The Vector Attribute

2.3.1. Inserting a Vector Attribute in a Join

In the example of Figure 1, when the Edge 2 router looks up the route to the source of the multicast distribution tree, it will find a BGP-distributed route whose "BGP next-hop" is Edge 1. Edge 2 then looks up the route to Edge 1 to find interface and PIM adjacency which is the next hop to the source, namely Core 2.

When Edge 2 sends a PIM Join to Core 2, it includes a Vector Attribute specifying the address of Edge 1. Core 2, and subsequent core routers, will forwarding the Join along the Vector (i.e, towards Edge 1) instead of trying to forward it towards S.

Whether an attribute is actually needed depends on whether the Core routers have a route to the source of the multicast tree. How the Edge router knows whether or not this is the case (and thus how the Edge router determines whether or not to insert an attribute field) is outside the scope of this document.

2.3.2. Processing a Received Vector Attribute

When processing a received PIM Join which contains a Vector Attribute, a router must first check to see if the Vector IP address is one of its own IP addresses. If so, the Vector Attribute is discarded, and not passed further upstream. Otherwise, the Vector Attribute is used to find the route to the source, and is passed along when a PIM Join is sent upstream. Note that a router which receives a Vector Attribute must use it, even if that router happens to have a route to the source. A router which discards a Vector Attribute may of course insert a new Vector Attribute. This would typically happen if a PIM Join needed to pass through a sequence of Edge routers, each pair of which is separated by a core which does not have external routes. In the absence of periodic refreshment, Vectors expire along with the corresponding (S,G) state.

2.3.3. Vector Attribute and Asserts

In a PIM Assert message we include the routing protocol’s "metric" to the source of the tree. This information is used in the selection of the assert winner. If a PIM Join is being sent towards a Vector,
rather than towards the source, the Assert message must have the
metric to the Vector instead of the metric to the source. The Assert
message however does not have an attribute field and does not mention
the Vector.

A router may change its upstream neighbor on a particular multicast
tree as the result of receiving Assert messages. However a Vector
Attribute should not be sent in a PIM Join to an upstream neighbor
which is chosen as the result of processing the Assert messages.
Reachability of the Vector is only guaranteed by the router that
advertises reachability to the Vector in it’s IGP. If the assert
winner upstream is not our real preferred next-hop, we can’t be sure
this router knows the path to the Vector. In the worst case the
assert winner has a route to the Vector that is on the same interface
where the assert was won. That will point the RPF interface to that
interface and will result in a O-list being NULL. The Vector
attribute is not inserted if the RPF neighbor was chosen via an
assert process and the RPF neighbor is different from the RPF
neighbor that would have been selected via the local routing table.
In all other cases the Vector has to be included in the Join message.
3. Vector Attribute TLV Format

```
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|F|S| Type      | Length        |   Encoded-Unicast address
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-.......
```

**F bit**

- Forward Unknown TLV. If this bit is set the TLV is forwarded regardless if the router understands the Type.

**S bit**

- Bottom of Stack. If this bit is set then this is the last TLV in the stack.

**Type**

- The Vector Attribute type is 0.

**Length**

- Length depending on Address Family of Encoded-Unicast address.

**Value**

- Encoded-Unicast address, see PIM-SM [RFC4601]

```
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|F|S| Type      | Length        |   Encoded-Unicast address
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-.......
```

**F bit**

- Forward Unknown TLV. If this bit is set the TLV is forwarded regardless if the router understands the Type.

**S bit**

- Bottom of Stack. If this bit is set then this is the last TLV in the stack.
The Vector Attribute type is 0.

Length
------
Length depending on Address Family of Encoded-Unicast address.

Value
-----
Encoded-Unicast address, see PIM-SM

4. IANA Considerations

An attribute type needs to be assigned. For now we propose the value 0.

5. Security Considerations

Security of the RPF Vector Attribute is only guaranteed by the security of the PIM packet, so the security considerations for PIM join packets as described in PIM-SM [RFC4601] apply here.

6. Acknowledgments

The authors would like to thank Yakov Rekhter and Dino Farinacci for their initial ideas on this topic.

7. References

7.1. Normative References


[I-D.ietf-pim-bidir]

[I-D.ietf-pim-join-attributes]
Boers, A., "Format for using TLVs in PIM messages",
7.2. Informative References

Authors' Addresses

IJisbrand Wijnands
Cisco Systems, Inc.
De kleetlaan 6a
Diegem 1831
Belgium

Email: ice@cisco.com

Arjen Boers
Cisco Systems, Inc.
Avda. Diagonal, 682
Barcelona 08034
Spain

Email: aboers@cisco.com

Eric Rosen
Cisco Systems, Inc.
1414 Massachusetts Avenue
Boxborough, Ma 01719

Email: erosen@cisco.com