

PIM WG  
Internet-Draft  
Intended status: Standards Track  
Expires: July 19, 2009

IJ. Wijnands  
A. Boers  
E. Rosen  
Cisco Systems, Inc.  
January 15, 2009

**The RPF Vector TLV**  
**draft-ietf-pim-rpf-vector-08**

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of [BCP 78](#) and [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 July 19, 2009.

Copyright Notice

Copyright (c) 2009 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document.

Abstract

This document describes a use of the PIM Join Attribute as defined in

[draft-ietf-pim-join-attributes](#) [[RFC5384](#)] 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

<a href="#">1.</a>	<a href="#">Specification of Requirements</a>	<a href="#">3</a>
<a href="#">2.</a>	<a href="#">Introduction</a>	<a href="#">3</a>
<a href="#">3.</a>	<a href="#">Use of the RPF Vector TLV</a>	<a href="#">4</a>
<a href="#">3.1.</a>	<a href="#">Attribute and shared tree joins</a>	<a href="#">4</a>
<a href="#">3.2.</a>	<a href="#">Attribute and Bootstrap messages</a>	<a href="#">5</a>
<a href="#">3.3.</a>	<a href="#">The Vector Attribute</a>	<a href="#">5</a>
<a href="#">3.3.1.</a>	<a href="#">Inserting a Vector Attribute in a Join</a>	<a href="#">5</a>
<a href="#">3.3.2.</a>	<a href="#">Processing a Received Vector Attribute</a>	<a href="#">5</a>
<a href="#">3.3.3.</a>	<a href="#">Vector Attribute and Asserts</a>	<a href="#">6</a>
<a href="#">3.3.4.</a>	<a href="#">Vector Attribute and Join suppression</a>	<a href="#">6</a>
<a href="#">4.</a>	<a href="#">Vector Attribute TLV Format</a>	<a href="#">7</a>
<a href="#">5.</a>	<a href="#">IANA Considerations</a>	<a href="#">7</a>
<a href="#">6.</a>	<a href="#">Security Considerations</a>	<a href="#">7</a>
<a href="#">7.</a>	<a href="#">Acknowledgments</a>	<a href="#">8</a>
<a href="#">8.</a>	<a href="#">Normative References</a>	<a href="#">8</a>
	<a href="#">Authors' Addresses</a>	<a href="#">8</a>



## **1. 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](#)].

## **2. 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 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 Sparse-Mode (PIM-SM) [[RFC4601](#)], PIM Source-Specific Mode (PIM-SSM) [[RFC4607](#)] or Bidirectional PIM (BIDIR-PIM) [[RFC5015](#)] 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 [[RFC5384](#)]. It consists of a single Vector which identifies the exit point of the network.

### 3. Use of the RPF Vector TLV

Before a router can start forwarding multicast packets, it is necessary 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.

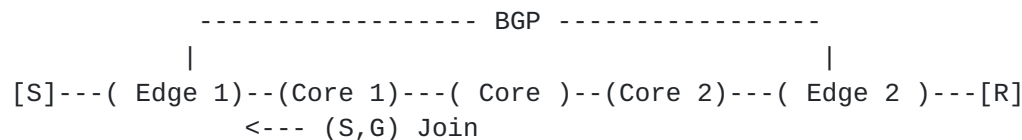


Figure 1

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 to 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 routers do their RPF (Reverse Path Forwarding) check [[RFC4601](#)] on the Attribute (IP address of Edge 1) rather than the Source, this allows the tree to be constructed.

#### 3.1. 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.

### **3.2. 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 BSR IP address is not routable in the core a solution needs to be developed for BSR.

### **3.3. The Vector Attribute**

#### **3.3.1. 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.

#### **3.3.2. 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.





### **3.3.3. Vector Attribute and Asserts**

A PIM Assert message includes 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 MUST NOT be sent in a PIM Join to an upstream neighbor which is chosen as the result of an Assert processing, if that neighbor is different than the original upstream neighbor. Reachability of the Vector is only guaranteed by the router that advertises reachability to the Vector in its IGP. If the assert winner upstream is not the real preferred next-hop, it is possible that the assert winner does not know 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 the O-list being NULL. The Vector attribute therefore MUST NOT be 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 MUST be included in the Join message.

### **3.3.4. Vector Attribute and Join suppression**

If a router receives a PIM join on the upstream LAN interface for a particular multicast state, join suppression may be applied if that PIM join is targeted to the same upstream neighbor. Which router(s) will suppress their PIM join is dependant on timing and is unpredictable. Downstream routers on a LAN MAY include different RPF vectors in the PIM joins. Therefore an upstream router on that LAN may receive and use different RPF vectors over time to reach the destination (depending on which downstream router(s) suppressed their Join). To make the upstream router behavior more predictable the RPF vector address MUST be used as additional condition to the join suppression logic. Only if the RPF vector in the PIM join matches the RPF vector in the multicast state, the suppression logic is applied. It is also possible to disable join suppression on that LAN.



#### 4. Vector Attribute TLV Format

```

0                               1                               2                               3
0 1 2 3 4 5 6 7 8 9 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      |          Value
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+

```

F bit

-----

Forward Unknown TLV. If this bit is set the TLV is forwarded regardless of whether the router understands the Type. If the TLV is known the F bit is ignored.

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.

#### 5. IANA Considerations

An new attribute type from the "PIM Join Attribute Types" registry needs to be assigned by IANA for the RPF Vector. The proposed value is 0.

#### 6. 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.



## **7. Acknowledgments**

The authors would like to thank Yakov Rekhter and Dino Farinacci for their initial ideas on this topic and Su Haiyang for the comments on the draft.

## **8. Normative References**

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [RFC4601] Fenner, B., Handley, M., Holbrook, H., and I. Kouvelas, "Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)", [RFC 4601](#), August 2006.
- [RFC4607] Holbrook, H. and B. Cain, "Source-Specific Multicast for IP", [RFC 4607](#), August 2006.
- [RFC5015] Handley, M., Kouvelas, I., Speakman, T., and L. Vicisano, "Bidirectional Protocol Independent Multicast (BIDIR-PIM)", [RFC 5015](#), October 2007.
- [RFC5384] Boers, A., Wijnands, I., and E. Rosen, "The Protocol Independent Multicast (PIM) Join Attribute Format", [RFC 5384](#), November 2008.

### Authors' Addresses

IJsbrand Wijnands  
Cisco Systems, Inc.  
De kleetlaan 6a  
Diegem 1831  
Belgium  
  
Email: [ice@cisco.com](mailto:ice@cisco.com)

Arjen Boers  
Cisco Systems, Inc.  
Avda. Diagonal, 682  
Barcelona 08034  
Spain  
  
Email: [aboers@cisco.com](mailto:aboers@cisco.com)



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

Email: [erosen@cisco.com](mailto:erosen@cisco.com)