PIM Working Group
Internet-Draft
Expires: December 13, 2008
Updates: RFC 5060

B. Joshi
Infosys Technologies Ltd.
A. Kessler
Cisco Systems, Inc.
D. McWalter
Data Connection Ltd
June 11, 2008

# PIM Group-to-RP Mapping
## draft-joshi-pim-group-rp-mapping-01.txt



Abstract

Each PIM-SM router in a PIM Domain which supports ASM maintains Group-to-RP mappings which are used to identify a RP for a specific multicast group. PIM-SM has defined an algorithm to choose a RP from the Group-to-RP mappings learned using various mechanisms. This algorithm does not allow administrator to override a specific Group-

to-RP mapping with the static Group-to-RP mapping which an administrator would want to use. This algorithm also does not consider the PIM mode and the mechanism through which a Group-to-RP mapping was learned.

This document first explains the requirements to extend the Group-to-RP mapping algorithm and then proposes the new algorithm.

Table of Contents

1. Introduction . . . . . 3
2. Terminology . . . . . 4
3. Existing algorithm . . . . . 5
4. Assumptions . . . . . 6
5. Common use cases . . . . . 7
6. Proposed algorithm . . . . . 8
7. Deprecation of MIB Objects . . . . . 10
8. Security Consideration . . . . . 11
9. IANA Consideration . . . . . 12
10. Acknowledgments . . . . . 13
11. Normative References . . . . . 13
Authors' Addresses . . . . . 14
Intellectual Property and Copyright Statements . . . . . 15

## 1. Introduction

Multiple mechanisms exist today to create and distribute Group-to-RP mappings. Each PIM-SM router may learn Group-to-RP mappings through various mechanisms.

It is critical that each router select the same 'RP' for a specific multicast group address. This is even true in the case of Anycast RP for redundancy. Routers should select the same RP address to use for a given group address. This RP address may correspond to a different physical router but it is one logical RP address and must be consistent across the PIM domain. This is usually achieved by using the same algorithm to select the RP in all the PIM routers in a domain.

PIM-SM[1] has defined an algorithm to select a 'RP' for a given multicast group address but it is not flexible enough for an administrator to apply various policies. Please refer to section 3 for more details.

PIM-STD-MIB [2] has defined an algorithm that allows administrators to override Group-to-RP mappings with static configuration. But this algorithm is not completely deterministic, because it includes an implementation-specific 'precedence' value.

Embedded-RP as defined in section-7.1 of Embedded-RP address in IPv6 Multicast address [3], mentions that to avoid loops and inconsistencies, for addresses in the range FF70::/12, the Embedded-RP mapping must be considered the longest possible match and higher priority than any other mechanism.

## 2. Terminology

In this document, the key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" are to be interpreted as described in RFC 2119. This document also uses following terms:

o PIM Mode

PIM Mode is the mode of operation a particular multicast group is used for. Wherever this term in used in this document, it refers to either Sparse Mode or BIDIR Mode.

## 3. Existing algorithm

Existing algorithm defined in PIM-SM (Section 4.7.1 in [1]) does not consider following constraints:

- It does not consider the origin of a Group-to-RP mapping and therefore will treat all of them equally.
- It does not provide the flexibility that a specific statically created Group-to-RP mapping can override any dynamically learned mappings.
- It does not provide the flexibility to give higher priority to a specific PIM mode. For example, an entry learned for PIM-BIDIR mode is treated with same priority as an entry learned for PIM-SM.

## 4. Assumptions

We have made following assumptions in defining this algorithm:

- A router MAY use hash function on Group-to-RP mappings learned through BSR mechanism [4]. This means that only a subset of Group-to-RP mappings will be available which are learned through BSR mechanism.

- A static Group-to-RP mapping entry can be configured with override-dynamic flag. If this flag is set, the static Group-to-RP mapping entry will be preferred instead of dynamically learned entries.

- Group-to-RP mappings created with the embedded RP extracted from Multicast Group addresses are special and always has the highest priority. These mappings can not be overridden by a static Group-to-RP mapping with override-dynamic flag set.

- A Group-to-RP mapping can be learned from various mechanisms. We assume that following list is in the decreasing preferences of these mechanism:

  * Embedded Group-to-RP mappings

  * Bootstrap Router Mechanism [PIM-BSR]

  * Auto-RP [Cisco]

  * Static configuration.

  * Other mapping method

- A Group-to-RP mapping learned for PIM-BIDIR mode is preferred to an entry learned for PIM-SM mode.

## 5. Common use cases

- Default static Group-to-RP mappings with dynamically learned entries

Many network operators will have a dedicated infrastructure for the standard multicast group range (224/4) and so might be using statically configured Group-to-RP mappings for this range. In this case, to support some specific applications, they might like to learn Group-to-RP mappings dynamically using either BSR or Auto-RP mechanism. In this case to select Group-to-RP mappings for these specific applications, a longer prefix match should be given preference over statically configured Group-to-RP mappings. For example 239.100.0.0/16 could be learned for a corporate communications application. Network operators may change the Group-to-RP mappings for these applications more often and would need to be learned dynamically.

- Static Group-to-RP mappings with override-dynamic flag

Many Network operators would like to statically configure one or multiple Group-to-RP mappings and would always want to ignore any dynamically learned mappings through either BSR, AutoRP or embedded RP for these group prefixes. This is accomplished by providing a 'override-dynamic' flag for Group-to-RP mapping configuration. When this flag is enabled for a static Group-to-RP mapping, it will have the highest precedence and would always be use for the specified group prefix. For example: 224.1.0.0/16 is configured with override-dynamic flag enabled and uses RP address RP1. If the router learns the more specific group prefix 224.1.1.0/24 which uses RP2 through BSR, it will choose the RP1 for any group falling under 224.1.0.0/16 range.

- Migration situations

Network operators occasionally go through a migration due to an acquisition or a change in their network design. In order to facilitate this migration there is a needs to have a deterministic behavior of Group-to-RP mapping selection for entries learned using BSR and AutoRP mechanism. This will help in avoiding any unforeseen interoperability issues between different vendor's network elements.

- More use cases

By no means, the above list is complete. Please drop a mail to 'authors' if you see any other use case for this.

## 6. Proposed algorithm

We propose following algorithm here which addresses the above mentioned shortcomings in the existing mechanism:

1. If the Multicast Group Address being looked up contains an embedded RP, RP address extracted from the Group address is selected as Group-to-RP mapping.

2. From the set of all Group-to-RP mapping entries, the subset whose group prefix contains the multicast group that is being looked up, are selected.

3. If there are no entries available, then the Group-to-RP mapping is undefined.

4. If there are multiple entries available, a subset of those Group-to-RP mapping is selected that are learned using 'static' configuration and are configured with 'override-dynamic' flag.

   * If there is only one entry available then that is selected as Group-to-RP mapping.

   * If there are multiple entries available, we continue with the algorithm with this smaller set of Group-to-RP Mappings

   * If there are no static entries with 'override-dynamic' flag set then we continue with the original subset of Group-to-RP Mappings from step 2.

5. A longest prefix match is performed on the subset of Group-to-RP Mappings.

   * If there is only one entry available then that is selected as Group-to-RP mapping.

   * If there are multiple entries available, we continue with the algorithm with this smaller set of Group-to-RP Mappings

6. From the remaining set of Group-to-RP Mappings we select the subset of entries based on the preference for the PIM modes which they are assigned. A Group-to-RP mapping entry with PIM Mode 'BIDIR' will be preferred to an entry with PIM Mode 'PIM-SM'

   * If there is only one entry available then that is selected as Group-to-RP mapping.

* If there are multiple entries available, we continue with the algorithm with this smaller set of Group-to-RP Mappings

7. From the remaining set of Group-to-RP Mappings we select the subset of the entries based on the origin. Origin preference will be 'bsr', 'auto-rp', 'static' and 'other'.

   * If there is only one entry available then that is selected as Group-to-RP mapping.

   * If there are multiple entries available, we continue with the algorithm with this smaller set of Group-to-RP Mappings

8. From the remaining set of Group-to-RP Mappings we run PIM hash function as suggested by PIM-SM [1].

   * If there is only one entry available then that is selected as Group-to-RP mapping.

   * If there are multiple entries available, we continue with the algorithm with this smaller set of Group-to-RP Mappings

9. From the remaining set of Group-to-RP Mappings we will select the RP with the highest IP address. This will serve as a final tiebreaker.

## 7. Deprecation of MIB Objects

Group-to-RP mapping algorithm defined in PIM-STD-MIB [2] does not specify the usage of 'pimGroupMappingPrecedence' and 'pimStaticRPPrecedence' objects in 'pimGroupMappingTable' table clearly. With the newly proposed algorithm in this document, these MIB objects would not be required. So we propose to deprecate these MIB objects from PIM-STD-MIB.

## 8. Security Consideration

This document does not suggest any protocol specific functionality so there is no security related consideration.

## 9. IANA Consideration

This draft does not create any namespace for IANA to manage.

## 10. Acknowledgments

This draft is created based on the discussion occurred during the PIM-STD-MIB [2] work. Many thanks to Stig Vennas for providing useful comments during that discussion.

## 11. Normative References

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

[2] Sivaramu, R., Lingard, J., McWalter, D., Joshi, B., and A. Kessler, "Protocol Independent Multicast MIB", RFC 5060, January 2008.

[3] Savola, P. and B. Haberman, "Embedding the Rendezvous Point (RP) Address in an IPv6 Multicast Address", RFC 3956, November 2004.

[4] Bhaskar, N., Gall, A., Lingard, J., and S. Venaas, "Bootstrap Router (BSR) Mechanism for Protocol Independent Multicast (PIM)", RFC 5059, January 2008.

## Authors' Addresses

Bharat Joshi
Infosys Technologies Ltd.
44 Electronics City, Hosur Road
Bangalore 560 100
India

Email: bharat_joshi@infosys.com
URI: http://www.infosys.com/

Andy Kessler
Cisco Systems, Inc.
425 E. Tasman Drive
San Jose, CA 95134
USA

Email: kessler@cisco.com
URI: http://www.cisco.com/

David McWalter
Data Connection Ltd
100 Church Street
Enfield EN2 6BQ
UK

Email: dmcw@dataconnection.com

## Intellectual Property Statement



## Disclaimer of Validity



## Copyright Statement



## Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.