Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 22, 2012

IJ. Wijnands, Ed.
S. Venaas
Cisco Systems, Inc.
M. Brig
Aegis BMD Program Office

October 20, 2011

**PIM source discovery via BSR**
**draft-wijnands-pim-source-discovery-bsr-00**

Abstract

PIM Sparse-Mode use a Rendezvous Point (RP) and shared trees to forward multicast packets to Last Hop Routers (LHR). After the first packet is received by the LHR, the source of the multicast stream is learned and the Shortest Path Tree (SPT) can be joined. This draft proposes a solution to support PIM Sparse Mode (SM) without the need for PIM registers, RPs or shared trees. Multicast source information is distributed via Bootstrap Router [RFC5059] messages and flooded throughout the Multicast domain. By removing the need for RPs and shared trees, the PIM-SM procedures are simplified, improving router operations, management and making the protocol more robust.

Status of this Memo



Copyright Notice

Table of Contents

## 1. Introduction

PIM Sparse-Mode uses a Rendezvous Point (RP) and shared trees to forward multicast packets to Last Hop Routers (LHR). After the first packet is received by the LHR, the source of the multicast stream is learned and the Shortest Path Tree (SPT) can be joined. This draft proposes a solution to support PIM Sparse Mode (SM) without the need for PIM registers, RPs or shared trees. Multicast source information is distributed via Bootstrap Router [RFC5059] messages and flooded throughout the Multicast domain. By removing the need for RPs and shared trees, the PIM-SM procedures are simplified, improving router operations, management and making the protocol more robust.

BSR provides an infrastructure to advertise 'RP mappings' to all the routers in the Multicast network via Reverse Path Forwarding (RPF). This document proposes to use BSR to advertise Source Group (SG) mappings from the First Hop Router (FHR) to all the LHRs. BSR seems like a good fit since its already designed to distribute 'mappings' through out the multicast network. The requirements for distributing SG mappings seems very close to RP mappings, for that reason there seems no need to invent a new protocol.

### 1.1. Conventions used in this document

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 RFC 2119 [RFC2119].

### 1.2. Terminology

RP:  Rendezvous Point.

BSR:  Bootstrap Router.

RPF:  Reverse Path Forwarding

SPT:  Shortest Path Tree.

FHR:  First Hop Router, directly connected to the Source.

LHR:  Last Hop Router, directly connected to the receiver.

SG Mapping:  Multicast source to group mapping.

SG BSR message:  A BSR message containing a source to group mapping.

## 2. Source to Group Mappings via BSR

A Candidate Bootstrap (C-BSR) router is typically a router which is configured to be a BSR. Multiple C-BSR may be configured in the network and an election procedure is applied to select the Elected BSR (E-BSR) router. The E-BSR router is the router that is responsible for distributing the Group to RP mappings. See [RFC5059] section 3.1 for more details. In order to distribute Source to Group (SG) mappings, there is no need to elect an E-BSR router.

### 2.1. Originating SG BSR messages

Each FHR that is directly connected to an active multicast source becomes the E-BSR for that SG mapping. How a multicast router discovers the source of the multicast packet and when it considers it self the FHR follows the same procedures as the registering process described in [RFC4601]. After it is decided that a register needs to be sent, the SG is not registered via the PIM SM register procedures, but the SG mapping is distributed via a BSR message. Note, only the SG mapping is distributed in the BSR message, not the entire packet as would have been done with a PIM register. The router originating the BSR messages includes its own address in the BSR message. The BSR messages are periodically sent for as long as the multicast source is active, similar to how PIM registers are periodically sent. The timer and timeout values described in [RFC5059] apply here as well.

### 2.2. Forwarding SG BSR messages

The forwarding of BSR messages follows the same procedures as documented in [RFC5059] section 3.4 and 3.5.

### 2.3. Processing SG BSR messages

A router that receives a SG BSR messages should parse the SG BSR message and store the SG mappings with a holdtimer started with the advertised holdtime for that group. If there are directly connected receivers for that group this router should send PIM (S,G) joins for all the SG mappings advertised in the BSR message. The SG BSR mappings is kept alive for as long as the holdtimer for the source is running. Once the holdtimer expired a PIM (S,G) prune must be sent to remove itself from the tree.

## 3. The first packets and bursty sources

The PIM register procedure is designed to deliver Multicast packets to the RP in the absence of a native SPT tree from the RP to the

source. The register packets received on the RP are decapsulated and forwarded down the shared tree to the LHRs. As soon as an SPT tree is built, multicast packets would flow natively over the SPT to the RP or LHR and the register process would stop. The PIM register process bridges the gap between how long it takes to build the SPT tree to the FHR. If the packets would not be unicast encapsulated to the RP they would be dropped by the FHR until the SPT is setup. This functionality is important for applications where the first packet(s) must be received for the application to work correctly. An other reason would be for bursty sources. If the application sends out a multicast packet every 4 minutes (or longer), the SPT is torn down (typically after 3:30 minutes of inactivity) before the next packet is forwarded down the tree. This will cause no multicast packet to ever be forwarded. A well behaved application should really be able to deal with packet loss since IP is a best effort based packet delivery system. But in reality this is not always the case.

With the procedures proposed in this draft the packet(s) received by the FHR will be dropped until the LHR has learned about the source and the SPT tree is built. That means for bursty sources or applications sensitive for the delivery of the first packet this proposal would not be very applicable. This proposal is mostly useful for applications that don't have strong dependency on the first packet(s) and have a constant data rate, like video distribution for example. For applications with strong dependency on the first packet(s) we recommend using PIM Bidir [RFC5015] or SSM [RFC4607]. The protocol operations are much simpler compared to PIM SM, it will cause less churn in the network and both guarantee best effort delivery for the first packet(s).

An other solution to address the problems described above is documented in [I-D.ietf-magma-msnip]. This proposal allows for a host to tell the FHR its willingness to act as Source for a certain Group before sending the data packets. LHRs have time to join the SPT tree before the host starts sending which would avoid packet loss. The SG mappings announced by [I-D.ietf-magma-msnip] can be advertised directly into BSR, allowing a very nice integration of both proposals. The life time of the SPT is not driven by the liveliness of Multicast data packets (which is the case with PIM SM), but by the announcements driven via [I-D.ietf-magma-msnip]. This will also prevent packet loss due to bursty sources.

## 4. Resiliency to network partitioning

In a PIM SM deployment where the network becomes partitioned, due to link or node failure, it is possible that the RP becomes unreachable to a certain part of the network. New sources that become active in

that partition will not be able to register to the RP and receivers within that partition are not able to receive the traffic.  Ideally you would want to have a candidate RP in each partition, but you never know in advance which routers will form a partitioned network. In order to be fully resilient, each router in the network may end up being a candidate RP.  This would increase the operational complexity of the network.

The solution described in this document does not suffer from that problem.  If a network becomes partitioned and new sources become active, the receivers in that partitioned will receive the BSR SG Mappings and join the source tree.  Each partition works independently of the other partition(s) and will continue to have access to sources within that partition.  As soon as the network heals, the SG Mappings are re-flooded into the other partition(s) and other receives can join to the newly learned sources.

## 5.  Fragmentation of BSR messages

[RFC5059] defines procedures to fragment BSR messages if the number of group to RP mappings is too large and the packet exceeds the MTU size.  Its important for PIM to have all the RP mappings before it applies the selection process.  Missing mappings may cause the wrong RP to be selected.  Using BSR to distribute SG mappings we don't have this problem.  There is no reason to have all the source before joining the tree.  There is no selection process applied to the SG mappings, all the known SG mappings should be joined by PIM.  For that reason there is no special fragmentation support defined for SG mappings.

## 6.  Bootstrap Source message 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
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |PIM Ver| Type  |N|  Reserved   |           Checksum            |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |             FHR Address (Encoded-Unicast format)              |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |            Group Address 1 (Encoded-Group format)             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |            Src Count           |        Src Holdtime          |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |            Src Address 1 (Encoded-Unicast format)             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |            Src Address 2 (Encoded-Unicast format)             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                               .                               |
   |                               .                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |            Src Address m (Encoded-Unicast format)             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |            Group Address 2 (Encoded-Group format)             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |            Src Count           |        Src Holdtime          |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |            Src Address 1 (Encoded-Unicast format)             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |            Src Address 2 (Encoded-Unicast format)             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                               .                               |
   |                               .                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |            Src Address m (Encoded-Unicast format)             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |            Group Address n (Encoded-Group format)             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |            Src Count           |        Src Holdtime          |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |            Src Address 1 (Encoded-Unicast format)             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |            Src Address 2 (Encoded-Unicast format)             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                               .                               |
   |                               .                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |            Src Address m (Encoded-Unicast format)             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

PIM Version:   Reserved, Checksum Described in [RFC4601].

Type:   PIM Message Type.  Value (pending IANA) for a Bootstrap Source message

[N]o-Forward bit:   When set, this bit means that the Bootstrap message fragment is not to be forwarded.

FHR Address:   The address of the FHR router for the domain.  This can be any address assigned to this router, but MUST be routable in the domain to allow successful forwarding (just like BSR address).  The format for this address is given in the Encoded-Unicast address in [RFC4601].

Group Address 1..n:   The address of the bootstrap router for the domain.  The format for this address is given in the Encoded-Unicast address in [RFC4601].

Src Count  How many unicast encoded sources address encodings follow.

Src Holdtime:   The Holdtime (in seconds) for the corresponding source(s).

Src Address:   The source address for the corresponding group range. The format for these addresses is given in the Encoded-Unicast address in [RFC4601].

## 7. Security Considerations

The security considerations are no different from what is documented in [RFC5059].

## 8. IANA considerations

This document requires the assignment of a new code point from the IANA managed registry "PIM Message Types" called "Bootstrap Source Mapping".

## 9. Acknowledgments

The authors would like to thank Arjen Boers for contributing to the initial idea and Yiqun Cai for his comments on the draft.

## 10. References

### 10.1. Normative References

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

[RFC5059] Bhaskar, N., Gall, A., Lingard, J., and S. Venaas, "Bootstrap Router (BSR) Mechanism for Protocol Independent Multicast (PIM)", RFC 5059, January 2008.

### 10.2. Informative References

[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.

[I-D.ietf-magma-msnip]
Fenner, B., "Multicast Source Notification of Interest Protocol (MSNIP)", draft-ietf-magma-msnip-05 (work in progress), March 2004.

Authors' Addresses

IJsbrand Wijnands (editor)
Cisco Systems, Inc.
De kleetlaan 6a
Diegem  1831
Belgium

Email: ice@cisco.com

Stig Venaas
Cisco Systems, Inc.
Tasman Drive
San Jose  CA  95134
USA

Email: stig@cisco.com

Michael Brig
Aegis BMD Program Office
17211 Avenue D, Suite 160
Dahlgren  VA 22448-5148
USA

Email: michael.brig@mda.mil