Internet Engineering Task Force
INTERNET-DRAFT
draft-ietf-pim-sm-bsr-03.ps

PIM WG
Bill Fenner/AT&T
Mark Handley/ICIR
Roger Kermode/Motorola
David Thaler/Microsoft

25 February 2003
Expires: August 2003

Bootstrap Router (BSR) Mechanism for PIM Sparse Mode

Status of this Document

This document is a product of the IETF PIM WG. Comments should be addressed to the authors, or the WG's mailing list at pim@catarina.usc.edu.

Abstract

This document specifies the Bootstrap Router (BSR) mechanism for Protocol Independent Multicast - Sparse Mode (PIM-SM). BSR is one way that a PIM-SM router can learn the set of group-to-RP mappings required in order to function. The mechanism is dynamic, largely self-configuring, and robust to router failure.

Table of Contents

List of Figures

1. Introduction

Note: this document assumes familiarity with the workings of Protocol Independent Multicast - Sparse Mode, as defined in [3], and with Administratively Scoped Multicast, as described in [6].

For correct operation, every PIM Sparse-mode router within a PIM domain must be able to map a particular global-scope multicast group address to the same RP. If this is not the case then black holes may appear, where some receivers in the domain cannot receive some groups. A domain in this context is a contiguous set of routers that all implement PIM and are configured to operate within a common boundary defined by PIM Multicast Border Routers (PMBRs). PMBRs connect each PIM domain to the rest of the internet.

A PIM domain may also broken up into multiple administrative scope regions - these are regions where a border has been configured so that a range of multicast groups will not be forwarded across that border. For more information on Administratively Scoped IP Multicast, see RFC 2365. The modified criteria for admin-scoped regions are that the region is convex with respect to forwarding based on the MRIB, and that all PIM routers within the same scope region map a particular scoped group to the same RP within that region.

The PIM-SM specification does not mandate the use of a single mechanism to provide routers with the information to perform the group-to-RP mapping. This document describes the Bootstrap Router (BSR) mechanism. BSR was first defined in RFC 2362 [2], which has since been obsoleted. This document provides an updated specification of the BSR mechanism from RFC 2362, and also extends it to cope with administratively scoped region boundaries.

1.1. General Overview and Background

Every PIM-SM multicast group needs to be associated with the IP address of a Rendezvous-Point (RP). When a new multicast sender starts sending, its local Designated Router (DR) will encapsulate these data packets in a PIM Register message and send them to the RP for that multicast group. When a new multicast receiver joins, its local DR will send a PIM Join message towards the RP for that multicast group. When any PIM router sends a (*,G) Join message, it needs to know which is the next how router towards the RP for G to send the message to. Also when a PIM router is forwarding data packets using (*,G) state, it needs to know which is the correct incoming interface for packets destined for G, as it needs to reject any packets that arrive on other interfaces. Thus it is important for all the PIM routers in a domain to be able to map each multicast group to the correct RP address.

There are a number of ways that group-to-RP mapping can be done; the simplest solution is for all the routers in the domain to be configured with the same information. However, static configuration generally doesn't scale well, and also does not dynamically adapt to route around router or link failures. The mechanism specified in this document is known as the PIM BootStrap Router mechanism, or BSR for short, and provides a dynamic, adaptive mechanism to distribute group-to-RP mapping information rapidly throughout a domain.

Before we discuss the BSR mechanism itself, we should understand the rules a PIM-SM router will apply to the mapping information. Irrespective of how it obtains the mapping information, a PIM-SM router will have a mapping table containing multiple entries, each of which has the following form:

- Multicast group range, expressed as an address and prefix length.
- RP Priority.
- IP address of RP.

In general, these mapping entries may overlap in arbitrary ways; a particular multicast group may be covered by multiple mapping entries. When this is the case, the router chooses only one of the entries by applying a deterministic algorithm (specified in the PIM-SM protocol specification) so that all routers in the domain make the same choice of entry and hence apply the same group-to-RP mapping.

The BSR mechanism provides a way in which viable group-to-RP mappings can be created and distributed to all the PIM-SM routers in a domain. It is adaptive, in that if an RP becomes unreachable, this will be detected and the mapping tables will be modified so that the unreachable RP is no longer used, and the new tables will be rapidly distributed throughout the domain.

The general idea behind the BSR-mechanism is that some of the PIM routers within a PIM domain are configured to be potential RPs for the domain. These are known as candidate-RPs (C-RPs). A subset of the C-RPs will eventually be used as the actual RPs for the domain. In addition, some of the PIM routers in the domain are configured as candidate bootstrap routers (C-BSRs). One of these C-BSRs will be elected to serve as the bootstrap router (BSR) for the domain, and all the PIM routers in the domain will learn the result of this election through Bootstrap messages. The C-RPs will them report their candidacy to the elected BSR, which will choose a subset of the C-RPs to form the actual set of RPs to the used. This RP-Set will then be distributed out to all the routers in the domain through Bootstrap messages.

The mechanism is complicated slightly by the presence of administratively-scoped multicast regions within the PIM-SM domain. An admin-scope region is a convex connected set of PIM routers surrounded by an admin-scope boundary. The boundary specifies a range of multicast addresses that will not be forwarded into or out of the scoped region. This complicates BSR because we do not want a PIM router within the scoped region to use an RP outside the scoped region (or vice-versa). Thus we need to modify the basic mechanism to ensure that this doesn't happen - this is done by electing a BSR for every admin-scope region within a PIM domain, and also a global BSR for the whole PIM domain. C-RPs typically register multiple times; once to the BSR of every admin scope zone the C-RP is in. For the remainder of this overview we will ignore admin-scope regions, and concentrate on the global BSR and its role. Within each scope zone, the BSR for that zone acts in a similar manner to how the global BSR acts for the whole domain.

There are four basic phases to the BSR mechanism (although in practice all phases may by occurring simultaneously):

1. BSR election. Each Candidate-BSR originates bootstrap messages (BSMs). Every BSM contains a BSR priority field. Routers within the domain flood the BSMs throughout the domain. A C-BSR that hears about a higher-priority C-BSR than itself then suppresses its sending of further BSMs for some period of time. The single remaining C-BSR becomes the elected BSR, and its BSMs inform all the other routers in the domain that it is the elected BSR.

2. C-RP advertisement. Each Candidate-RP within a domain sends periodic Candidate-RP-Advertisement (C-RP-Adv) messages to the elected BSR. In this way, the BSR learns about possible RPs that are currently up and reachable.

3. C-RP-Set Formation. The BSR selects a subset of the C-RPs that it has heard C-RP-Adv messages from to form the RP-Set. In general it should do this in such a way that the RP-Set is neither too large to inform all the routers in the domain about, nor too small so that load is overly concentrated on some RPs. It should also attempt to produce an RP-Set that does not change frequently.

4. RP-Set Flooding. In future bootstrap messages, the BSR includes the RP-Set information. As bootstrap messages are flooded rapidly through the domain, this ensures that the RP-Set

rapidly reaches all the routers in the domain. BSMs are originated periodically to ensure consistency after failure restoration.

In the following sections we discuss more details about BSR for global scope and for admin scoping, before specifying the protocol starting in section 2.

1.2. Overview of Bootstrap and RP Discovery for Global Scope

A small set of routers from a domain are configured as candidate bootstrap routers (C-BSRs) and, through a simple election mechanism, a single BSR is selected for that domain. A set of routers within a domain are also configured as candidate RPs (C-RPs); typically these will be the same routers that are configured as C-BSRs. Candidate RPs periodically unicast Candidate-RP-Advertisement messages (C-RP-Advs) to the BSR of that domain, advertising their willingness to be an RP. A C-RP-Adv message includes the address of the advertising C-RP, as well as an optional list of group addresses and a mask length fields, indicating the group prefix(es) for which the candidacy is advertised. The BSR then includes a set of these Candidate-RPs (the RP-Set), along with their corresponding group prefixes, in Bootstrap messages it periodically originates. Bootstrap messages are distributed hop-by-hop throughout the domain.

All the PIM routers in the domain receive and store Bootstrap messages originated by the BSR. When a DR receives an indication of local membership (typically from IGMP [4] or MLD [1]) or a data packet from a directly connected host, for a group for which it has no forwarding state, the DR uses a hash function to map the group address to one of the C-RPs from the RP-Set whose group-prefix includes the group (see [3]). The DR then sends a Join message towards that RP if the local host joined the group, or it Register-encapsulates and unicasts the data packet to the RP if the local host sent a packet to the group.

A Bootstrap message indicates liveness of the RPs included therein. If an RP is included in the message, then it is tagged as ‘up’ at the routers; RPs not included in the message are removed from the list of RPs over which the hash algorithm acts. Each router continues to use the contents of the most recently received Bootstrap message from the BSR until it accepts a new Bootstrap message.

If a PIM domain becomes partitioned, each area separated from the old BSR will elect its own BSR, which will distribute an RP-Set containing RPs that are reachable within that partition. When the partition heals, another election will occur automatically and only one of the BSRs will continue to send out Bootstrap messages. As is expected at the time of a partition or healing, some disruption in packet delivery may occur. This time will be on the order of the region’s round-trip time and the bootstrap router timeout value.

1.3. Administratively Scoped Multicast and BSR

Administratively Scoped IP Multicast, as defined in RFC 2365, permits a network provider to configure scope boundaries at multicast routers. Such a scope boundary consists of a range of multicast addresses (expressed as an address and mask) that the router will not forward across the boundary. For correct operation, such a scope zone border must be complete and convex. By this we mean that there must be no path from inside the scoped zone to outside it that does not pass through a configured scope border router, and that the multicast capable path between any arbitrary pair of multicast routers in the scope zone must remain in the zone.

For PIM-SM using BSR to function correctly with admin scoping, there must be a BSR and at least one C-RP within every admin scope region. Admin scope zone boundaries must be configured at the Zone Border Routers (ZBRs), as they need to filter PIM Join messages that might inadvertently cross the border due to error conditions. In addition, at least one C-BSR within the admin scope zone must be configured to be a C-BSR for the admin scope zone's address range.

A separate BSR election will then take place (using bootstrap messages) for every admin scope range (plus one for the global range). Admin scope ranges are identified in the bootstrap message because the group range is marked (using the "Admin Scope" bit, previously a "Reserved" bit) to indicate that this is an administrative scope range, and not just a range that a particular set of RPs are configured to handle.

Such admin scoped bootstrap message packets are flooded in the normal way, but will not be forwarded by another ZBR across the boundary for that scope zone (see Section 3.3 for the specifics of this).

We do not require that C-RPs within the scope zone be configured to know about the scope zone, as they can learn of its existence from bootstrap messages. However, we recommend that router vendors implement configuration options that allow a C-RP to be configured to be a C-RP for global scope only, for one of more admin scopes only, or for all scopes, both global and admin scoped. We also recommend that the default be that a C-RP is a C-RP for all scopes, both global and admin scoped.

Unless configured otherwise, C-RPs discover the existence of the admin scope zone and its group range from receiving a bootstrap message from the scope zone's elected BSR containing the scope zone's group-range, marked using the "Admin Scope" bit. A C-RP stores each elected BSR's address and the admin scope range contained in its bootstrap message. It separately unicasts Candidate-RP-Advertisement messages to the appropriate BSR for every admin scope range within which it is willing to serve as an RP.

All PIM routers within a PIM bootstrap domain where admin scope ranges are in use must be capable of receiving bootstrap messages and storing the winning BSR and RPset for all admin scope zones that apply. Thus PIM routers that only implement RFC 2362 (which only allows one BSR per domain) cannot be used in PIM domains where admin scope zones are configured.

2. BSR State and Timers

A PIM-SM router implementing BSR holds the following state in addition to the state needed for PIM-SM operation:

At all routers:

List of Active Scope Zones

Per Scope Zone:

Bootstrap State:

- Bootstrap Router's IP Address
- BSR Priority
- Bootstrap Timer (BST)

- List of Scope Group-Ranges for this BSR

RP Set

At a Candidate BSR:

Per Scope Zone:

- My state: One of "**Candidate-BSR**", "**Pending-BSR**", "**Elected-BSR**"

At a router that is not a Candidate BSR:

Per Scope Zone:

My state: One of "**Accept Any**", "**Accept Preferred**"

Scope-Zone Expiry Timer: SZT(Z)

Bootstrap state is described in section 3, and the RP Set is described in section 3.4.

The following timers are also required:

At the Bootstrap Router only:

Per Scope Zone (Z):

Per Candidate RP (C):

C-RP Expiry Timer: CET(C,Z)

At the C-RPs only:

C-RP Advertisement Timer: CRPT

3. Bootstrap Router Election and RP-Set Distribution

For simplicity, bootstrap messages (BSMs) are used in both the BSR election and the RP-Set distribution mechanisms.

The state-machine for bootstrap messages depends on whether or not a router has been configured to be a Candidate-BSR for a particular scope zone. The per-scope-zone state-machine for a C-BSR is given below, followed by the state-machine for a router that is not configured to be a C-BSR.

Per-Scope-Zone Candidate-BSR State Machine

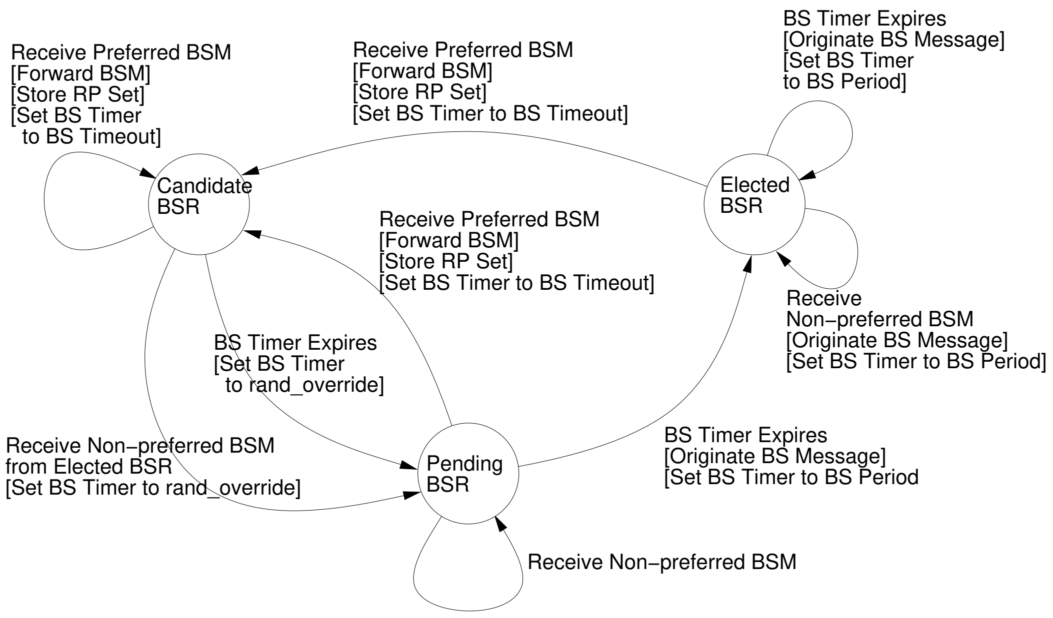

Figure 1: Per-Scope-Zone State-machine for a candidate BSR

Per-Scope-Zone State-machine for a candidate BSR in tabular form:

When in C-BSR state			
Event	Receive Preferred BSM	BS Timer Expires	Receive non-preferred BSM from Elected BSR
Action	-> C-BSR state Forward BSM; Store RP Set; Set BS Timer to BS Timeout	-> P-BSR state Set BS Timer to rand_override	-> P-BSR state Set BS Timer to rand_override

When in P-BSR state			
Event	Receive Preferred BSM	BS Timer Expires	Receive Non-preferred BSM
Action	-> C-BSR state Forward BSM; Store RP Set; Set BS Timer to BS Timeout	-> E-BSR state Originate BSM; Set BS Timer to BS Period	-> P-BSR state

When in E-BSR state			
Event	Receive Preferred BSM	BS Timer Expires	Receive Non-preferred BSM
Action	-> C-BSR state Forward BSM; Store RP Set; Set BS Timer to BS Timeout	-> E-BSR state Originate BSM; Set BS Timer to BS Period	-> E-BSR state Originate BSM; Set BS Timer to BS Period

A candidate-BSR may be in one of three states for a particular scope zone:

Candidate-BSR (C-BSR)

The router is a candidate to be the BSR for the scope zone, but currently another router is the preferred BSR.

Pending-BSR (P-BSR)

The router is a candidate to be the BSR for the scope zone. Currently no other router is the preferred BSR, but this router is not yet the BSR. For comparisons with incoming BS messages, the router treats itself as the BSR. This is a temporary state that prevents rapid thrashing of the choice of BSR during BSR election.

Elected-BSR (E-BSR)

The router is the elected bootstrap router for the scope zone and it must perform all the BSR functions.

On startup, the initial state for this configured scope zone is "Pending-BSR"; the BS Timer is initialized to the BS Timeout value.

In addition to the three states, there is one timer:

- The bootstrap timer (BS Timer) - that is used to time out old bootstrap router information, and used in the election process to terminate P-BSR state.

Per-Scope-Zone State-machine for Non-Candidate-BSR Routers

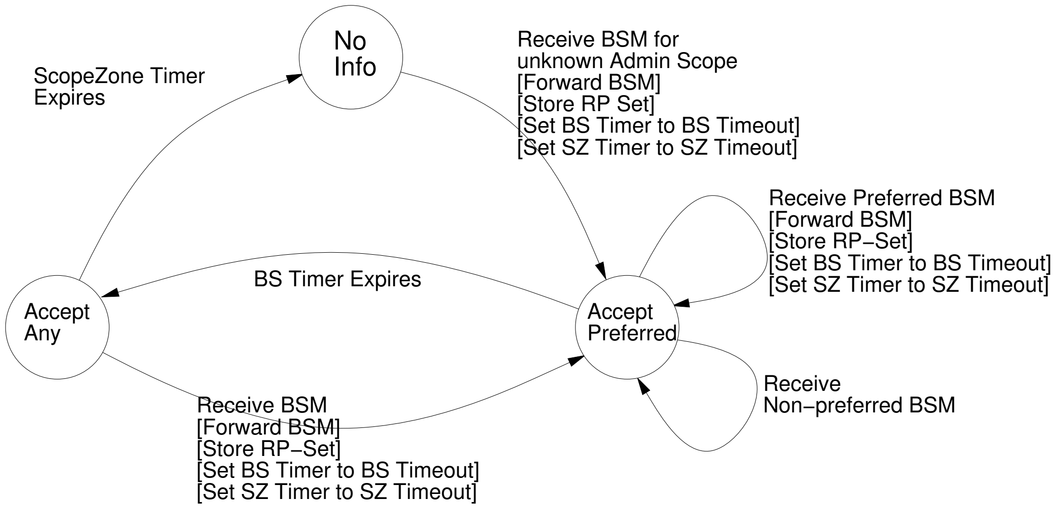

Figure 2: Per-Scope-Zone State-machine for a router not configured as C-BSR

Per-Scope-Zone State-machine for a router not configured as C-BSR in tabular form:

When in No Info state	
Event	Receive BSM for unknown Admin Scope
Action	-> AP State Forward BSM; Store RP-Set; Set BS Timer to BS Timeout; Set SZ Timer to SZ Timeout

When in Accept Any state		
Event	Receive BSM	SZ Timer Expires
Action	-> AP State Forward BSM; Store RP-Set; Set BS Timer to BS Timeout	-> No Info state cancel timers; clear state

When in Accept Preferred state			
Event	Receive Preferred BSM	BS Timer Expires	Receive Non-preferred BSM
Action	-> AP State Forward BSM; Store RP-Set; Set BS Timer to BS Timeout	-> AA State	-> AP State

A router that is not a candidate-BSR may be in one of three states:

No Info
The router has no information about this scope zone. This state does not apply if the router is configured to know about this scope zone, or for the global scope zone. When in this state, no state information is held and no timers run that refer to this scope zone.

Accept Any (AA)
The router does not know of an active BSR, and will accept the first bootstrap message it sees as giving the new BSR's identity and the RP-Set. If the router has an RP-Set cached from an obsolete bootstrap message, it continues to use it.

Accept Preferred (AP)
The router knows the identity of the current BSR, and is using the RP-Set provided by that BSR. Only bootstrap messages from that BSR or from a C-BSR with higher weight than the current BSR will be accepted.

On startup, the initial state for this scope zone is "Accept Any" for routers that know about this scope zone, either through configuration or because the scope zone is the global scope which always exists; the SZ Timer is considered to be always running for such scope zones. For routers that do not know about a particular scope zone, the initial state is No Info; no timers exist for the scope zone.

In addition to the three states, there are two timers:

- The bootstrap timer (BS Timer) - that is used to time out old bootstrap router information.
- The scope zone timer (SZ Timer) - that is used to time out the scope zone itself if BS messages specifying this scope zone stop arriving.

Bootstrap Message Processing Checks

When a bootstrap message is received, the following initial checks must be performed:

```
if ( (DirectlyConnected(BSM.src_ip_address) == FALSE)
     OR (we have no Hello state for BSM.src_ip_address)) {
  drop the BS message silently
}
if (BSM.dst_ip_address == ALL-PIM-ROUTERS group) {
  if ( BSM.src_ip_address != RPF_neighbor(BSM.BSR_ip_address) ) {
     drop the BS message silently
  }
} else if (BSM.dst_ip_address is one of my addresses) {
  if ( (Any previous BSM for this scope has been accepted) {
     #the packet was unicast, but this wasn't
     #a quick refresh on startup
     drop the BS message silently
  }
} else {
  drop the BS message silently
}
if (the interface the message arrived on is an Admin Scope
    border for the BSM.first_group_address) {
  drop the BS message silently
}
```

Basically, the packet must have come from a directly connected neighbor for which we have active Hello state. It must have been sent to the ALL-PIM-ROUTERS group by the correct upstream router towards the BSR that originated the BS message, or the router must have no BSR state (it just restarted) and have received the BS message by unicast. In addition it must not have arrived on an interface that is a configured admin scope border for the first group address contained in the BS message.

BS State-machine Transition Events

If the bootstrap message passes the initial checks above without being discarded, then it may cause a state transition event in one of the above state-machines. For both candidate and non-candidate BSRs, the following transition events are defined:

Receive Preferred BSM
A bootstrap message is received from a BSR that has greater than or equal weight than the current BSR. In a router is in P-BSR state, then it uses its own weight as that of the current BSR.

The weighting for a BSR is the concatenation in fixed-precision unsigned arithmetic of the BSR priority field from the bootstrap message and the IP address of the BSR from the bootstrap message (with the BSR priority taking the most-significant bits and the IP address taking the least significant bits).

Receive BSM
A bootstrap message is received, regardless of BSR weight.

A non-candidate BSM also has the following transition event defined:

Receive BSM for unknown Admin Scope
As "Receive BSM", except that the admin scope zone indicated in the BSM was not previously known at this router.

BS State-machine Actions

The state-machines specify actions that include setting the BS timer to the following values:

BS Period
The periodic interval with which bootstrap messages are normally sent. The default value is 60 seconds.

BS Timeout
The interval after which bootstrap router state is timed out if no bootstrap message from that router has been heard. The default value is 2 times the BS Period plus 10 seconds, which is 130 seconds.

Randomized Override Interval
The randomized interval during which a router avoids sending a bootstrap message while it waits to see if another router has a higher bootstrap weight. This interval is to reduce control message overhead during BSR election. The following pseudocode is proposed as an efficient implementation of this "randomized" value:

```
Delay = 5 + 2 * log_2(1 + bestPriority - myPriority)
        + AddrDelay
```

where `myPriority` is the Candidate-BSR's configured priority, and `bestPriority` equals:

```
bestPriority = Max(storedPriority, myPriority)
```

and `AddrDelay` is given by the following for IPv4:

```
if ( bestPriority == myPriority) {
    AddrDelay = log_2(storedAddr - myAddr) / 16
} else {
    AddrDelay = 2 - (myAddr / 2^31)
}
```

and `AddrDelay` is given by the following for IPv6:

```
if ( bestPriority == myPriority) {
    AddrDelay = log_2(storedAddr - myAddr) / 64
} else {
    AddrDelay = 2 - (myAddr / 2^127)
}
```

where `myAddr` is the Candidate-BSR's address, `storedAddr` is the stored BSR's address, and `storedPriority` is the stored BSR's priority.

SZ Timeout
The interval after which a router will time out an Admin Scope zone that it has dynamically learned. The interval MUST be larger than the BS Timeout. The default value is ten times the BS Timeout, which is 1300 seconds.

In addition to setting the timers, the following actions may be triggered by state-changes in the state-machines:

Forward BSM
A bootstrap message that passes the Bootstrap Message Processing Checks is forwarded out of all interfaces with PIM neighbors (including the interface it is received on), except where this would cause the BSM to cross an admin-scope boundary for the scope zone indicated in the message. The source IP address of the message is the forwarding router's IP address on the interface the message is being forwarded from, the destination address is ALL-PIM-ROUTERS, and the TTL of the message is set to 1.

As an optimation, a router MAY choose not to forward a BSM out of the interface the message was received on if that interface is a point-to-point interface. On interfaces with multiple PIM neighbors, a router MUST forward an accepted BSM onto the interface that BSM was received on, but if the number of PIM neighbors on that interface is large, it MAY delay forwarding a BSM onto that interface by a small randomized interval to prevent message implosion.

Rationale: A BSM needs to be forwarded onto the interface the message was received on (in addition to the other interfaces) because the routers on a LAN may not have consistent routing information. If three routers on a LAN and A, B, and C, and at router B RPF(BSR)==A and at router C RPF(BSR)==B, then router A originally forwards the BSM onto the LAN, but router C will only accept it when router B re-forwards the message onto the LAN.

Originate BSM
A new bootstrap message is constructed by the BSR, giving the BSR's address and BSR priority, and containing the BSR's chosen RP-Set. The message is forwarded out of all multicast-capable interfaces, except where this would cause the BSM to cross an admin-scope boundary for the scope zone indicated in the message. The IP source address of the message is the originating router's IP address on the interface the

message is being forwarded from, the destination address is ALL-PIM-ROUTERS, and the TTL of the message is set to 1.

Store RP Set

The RP-Set from the received bootstrap message is stored and used by the router to decide the RP for each group that the router has state for. Storing this RP Set may cause other state-transitions to occur in the router. The BSR's IP address and priority from the received bootstrap message are also stored to be used to decide if future bootstrap messages are preferred.

In addition to the above state-machine actions, a DR also unicasts a stored copy of the Bootstrap message to each new PIM neighbor, i.e., after the DR receives the neighbor's first Hello message, and sends a Hello message in response. It does so even if the new neighbor becomes the DR.

3.1. Sending Candidate-RP-Advertisements

Every C-RP periodically unicasts a C-RP-Adv to the BSR for that scope zone to inform the BSR of the C-RP's willingness to function as an RP. Unless configured otherwise, it does this for every Admin Scope zone for which it has state, and for the global scope zone. If the same router is the BSR for more than one scope zone, the C-RP-Adv for these scope zones MAY be combined into a single message.

If the C-RP is a ZBR for an admin scope zone, then the Admin Scope bit MUST be set in the C-RP-Adv messages it sends for that scope zone; otherwise this bit MUST NOT be set. This information is currently only used for logging purposes by the BSR, but might allow for future extensions of the protocol.

The interval for sending these messages is subject to local configuration at the C-RP, but must be smaller than the HoldTime in the C-RP-Adv.

A Candidate-RP-Advertisement carries a list of group address and group mask field pairs. This enables the C-RP router to restrict the advertisement to certain prefixes or scopes of groups. If the C-RP becomes an RP, it may enforce this scope acceptance when receiving Registers or Join/Prune messages.

The C-RP priority field determines which C-RPs get selected by the BSR to be in the RP Set. Note that a value of zero is the highest possible priority. C-RPs should by default send C-RP-Adv messages with the 'Priority' field set to 192.

When a C-RP is being shut down, it SHOULD immediately send a C-RP-Adv to the BSR for each scope for which it is currently serving as an RP; the HoldTime in this C-RP-Adv message should be zero. The BSR will then immediately time out the C-RP and generate a new BSR message removing the shut down RP from the RPset.

3.2. Creating the RP-Set at the BSR

Upon receiving a C-RP-Adv, if the router is not the elected BSR, it silently ignores the message.

If the router is the BSR, then it adds the RP address to its local pool of candidate RPs. For each C-RP, the BSR holds the following information:

IP address

The IP address of the C-RP.

Group Prefix and Mask list
The list of group prefixes and group masks from the C-RP advertisement.

HoldTime
The HoldTime from the C-RP-Adv message. This is included later in the RP-set information in the Bootstrap Message.

C-RP Expiry Timer
The C-RP-Expiry Timer is used to time out the state associated with a C-RP when the BSR fails to receive C-RP-Advertisements from it. The expiry timer is initialized to the HoldTime from the RP's C-RP-Adv, and is reset to the HoldTime whenever a C-RP-Adv is received from that C-RP.

C-RP Priority
The C-RP Priority is used to determine the subset of possible RPs to use in the RP-Set. Smaller values are deemed to be of higher priority than large ones.

When the C-RP Expiry Timer expires, the C-RP is removed from the pool of available C-RPs.

The BSR uses the pool of C-RPs to construct the RP-Set which is included in Bootstrap Messages and sent to all the routers in the PIM domain. The BSR may apply a local policy to limit the number of Candidate RPs included in the Bootstrap message. The BSR may override the prefix indicated in a C-RP-Adv unless the 'Priority' field from the C-RP-Adv is less than 128.

The Bootstrap message is subdivided into sets of {group-prefix, RP- Count, RP-addresses}. For each RP-address, the corresponding HoldTime is included in the "RP-HoldTime" field. The format of the Bootstrap message allows 'semantic fragmentation', if the length of the original Bootstrap message exceeds the packet maximum boundaries. However, we recommend against configuring a large number of routers as C-RPs, to reduce the semantic fragmentation required.

When an elected BSR is being shut down, it should immediately originate a Bootstrap message listing its current RP set, but with the BSR priority field set to the lowest priority value possible. This will cause the election of a new BSR to happen more quickly.

3.3. Forwarding Bootstrap Messages

Bootstrap messages originate at the BSR, and are hop-by-hop forwarded by intermediate routers if they pass the Bootstrap Message Processing Check. Bootstrap messages are multicast to the 'ALL-PIM-ROUTERS' group. When a BS message is forwarded, it is forwarded out of every multicast-capable interface which has PIM neighbors (excluding the one over which the message was received). The exception to this is if the interface is an administrative scope boundary for the admin scope zone indicated in the first group address in the BS message packet. The IP source address on the bootstrap message should be set to the forwarding router's IP address on the interface the message is being forwarded from. Bootstrap messages are always originated or forwarded with an IP TTL value of 1.

3.4. Receiving and Using the RP-Set

When a router receives and stores a new RP-Set, it checks if each of the RPs referred to by existing state (i.e., by (*,G), (*,*,RP), or (S,G,rpt) entries) is in the new RP-Set.

If an RP is not in the new RP-Set, that RP is considered unreachable and the hash algorithm (see PIM-SM specification) is re-performed for each group with locally active state that previously hashed to that RP. This will cause those groups to be distributed among the remaining RPs.

If the new RP-Set contains a RP that was not previously in the RP-Set, the hash value of the new RP is calculated for each group covered by the new C-RP's Group-prefix. Any group for which the new RP's hash value is greater than hash value of the group's previous RP is switched over to the new RP.

4. Message Formats

BSR messages are PIM messages, as defined in [3]. The values of the PIM message Type field for BSR messages are:

4 Bootstrap Message

8 Candidate-RP-Advertisement

In this section we use the following terms defined in the PIM-SM [3]:

- Encoded-Unicast format
- Encoded-Group format

We repeat these here to aid readability.

Encoded-Unicast address

An Encoded-Unicast address takes the following 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Addr Family  | Encoding Type |     Unicast Address
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...
```

Addr Family

The PIM address family of the ‘Unicast Address’ field of this address.

Values of 0-127 are as assigned by the IANA for Internet Address Families in [5]. Values 128-250 are reserved to be assigned by the IANA for PIM-specific Address Families. Values 251 though 255 are designated for private use. As there is no assignment authority for this space, collisions should be expected.

Encoding Type

The type of encoding used within a specific Address Family. The value ‘0’ is reserved for this field, and represents the native encoding of the Address Family.

Unicast Address

The unicast address as represented by the given Address Family and Encoding Type.

Encoded-Group address

Encoded-Group addresses take the following 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Addr Family  | Encoding Type |    Reserved   |Z|  Mask Len     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                Group multicast Address
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...
```

Addr Family
described above.

Encoding Type
described above.

Reserved
Transmitted as zero. Ignored upon receipt.

Admin Scope [Z]one
When set, this bit indicates that this group address range is an administratively scoped range.

Mask Len
The Mask length field is 8 bits. The value is the number of contiguous one bits left justified used as a mask which, combined with the group address, describes a range of groups. It is less than or equal to the address length in bits for the given Address Family and Encoding Type. If the message is sent for a single group then the Mask length must equal the address length in bits for the given Address Family and Encoding Type. (e.g. 32 for IPv4 native encoding and 128 for IPv6 native encoding).

Group multicast Address
Contains the group address.

4.1. Bootstrap Message Format

A bootstrap message is divided up into ‘semantic fragments’ if the original message exceeds the maximum packet size boundaries. Basically, a single bootstrap message can be sent as multiple packets (semantic fragments), so long as the fragment tags of all the packets comprising the message is the same.

If the bootstrap message contains information about more than one admin scope zone, each different scope zone MUST occupy a different semantic fragment. This allows Zone Border Routers for an admin scope zone to not forward only those fragments that should not traverse the admin scope boundary.

The format of a single ‘fragment’ is given below:

```
 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  |    Reserved   |           Checksum            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Fragment Tag          | Hash Mask len | BSR-priority  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             BSR Address (Encoded-Unicast format)              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Group Address 1 (Encoded-Group format)             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| RP Count 1    | Frag RP Cnt 1 |         Reserved              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             RP Address 1 (Encoded-Unicast format)             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          RP1 Holdtime         | RP1 Priority  |   Reserved    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             RP Address 2 (Encoded-Unicast format)             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          RP2 Holdtime         | RP2 Priority  |   Reserved    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                               .                               |
|                               .                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             RP Address m (Encoded-Unicast format)             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          RPm Holdtime         | RPm Priority  |   Reserved    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Group Address 2 (Encoded-Group format)             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                               .                               |
|                               .                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Group Address n (Encoded-Group format)             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| RP Count n    | Frag RP Cnt n |          Reserved             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             RP Address 1 (Encoded-Unicast format)             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          RP1 Holdtime         | RP1 Priority  |   Reserved    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             RP Address 2 (Encoded-Unicast format)             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          RP2 Holdtime         | RP2 Priority  |   Reserved    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                               .                               |
|                               .                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             RP Address m (Encoded-Unicast format)             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

```
|          RPm Holdtime         | RPm Priority  |   Reserved    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

PIM Version, Reserved, Checksum
Described in [3].

Type
PIM Message Type. Value is 4 for a Bootstrap Message.

Fragment Tag
A randomly generated number, acts to distinguish the fragments belonging to different Bootstrap messages; fragments belonging to same Bootstrap message carry the same 'Fragment Tag'.

Hash Mask len
The length (in bits) of the mask to use in the hash function. For IPv4 we recommend a value of 30. For IPv6 we recommend a value of 126.

BSR priority
Contains the BSR priority value of the included BSR. This field is considered as a high order byte when comparing BSR addresses. Note that for historical reasons, the highest BSR priority priority is 255 (the higher the better), whereas the highest RP Priority (see below) is 0 (the lower the better).

Unicast BSR Address
The address of the bootstrap router for the domain. The format for this address is given in the Encoded-Unicast address in [3].

Group Address 1..n
The group prefix (address and mask) with which the Candidate RPs are associated. Format described in [3]. In a fragment containing admin scope ranges, the first group address in the fragment MUST be the group range for the entire admin scope range, and this MUST have the Admin Scope bit set. This is the case even if there are no RPs in the RP set for the entire admin scope range - in this case the sub-ranges for the RP set are specified later in the fragment along with their RPs.

RP Count 1..n
The number of Candidate RP addresses included in the whole Bootstrap message for the corresponding group prefix. A router does not replace its old RP-Set for a given group prefix until/unless it receives 'RP-Count' addresses for that prefix; the addresses could be carried over several fragments. If only part of the RP-Set for a given group prefix was received, the router discards it, without updating that specific group prefix's RP-Set.

Frag RP Cnt 1..m
The number of Candidate RP addresses included in this fragment of the Bootstrap message, for the corresponding group prefix. The 'Frag RP-Cnt' field facilitates parsing of the RP-Set for a given group prefix, when carried over more than one fragment.

RP address 1..m
The address of the Candidate RPs, for the corresponding group prefix. The format for these

addresses is given in the Encoded-Unicast address in [3].

RP1..m Holdtime
The Holdtime for the corresponding RP. This field is copied from the 'Holdtime' field of the associated RP stored at the BSR.

RP1..m Priority
The 'Priority' of the corresponding RP and Encoded-Group Address. This field is copied from the 'Priority' field stored at the BSR when receiving a Candidate-RP-Advertisement. The highest priority is '0' (i.e. unlike BSR priority, the lower the value of the 'Priority' field, the better). Note that the priority is per RP per Group Address.

4.1.1. Semantic Fragmentation of BSMs

Bootstrap messages may be split over several PIM Bootstrap Message Fragment (BSMF) packets; this is known as semantic fragmentation. There are two reasons for semantic fragmentation:

- The BSM would exceed the link MTU the packet will be forwarded over.
- The BSM includes information about more than one admin scope zone.

Let us initially consider only the former case; the packet would be too large because the set of group prefixes and the RPs for each group prefix are too long to fit in one packet. The BSR will then split the BSM across several BSMF packets; each of these must be a well-formed BSMF packet in its own right.

If the BSR can split up the BSM so that different group prefixes (and their RP information) can fit in different fragments, then it should do so. If one of these BSMF packets is then lost, the state from the previous BSM for the group-prefix from the missing packet will be retained. Each fragment that does arrive will update the RP information for the group-prefixes contained in that fragment, and the new group-to-RP mapping for those can be used immediately. The information from the missing fragment will be obtained when the BSM is next transmitted. In this case, whilst the Fragment Tag must be set to the same value for all BSMFs comprising a single BSM, the tag is not actually used by routers receiving the BSM.

If the list of RPs for a single group-prefix is too long to fit in a single BSMF packet, then that information must be split across multiple BSMF packets. In this case, all the BSMF packets comprising the information for that group-prefix must be received before the group-to-RP mapping in use can be modified. This is the purpose of the RP Count field - a router receiving BSMF packets from the same BSM (ie that have the same fragment tag) must wait until the BSMFs providing RP Count RPs for that group-prefix have been received before the new group-to-RP mapping can be used for that group-prefix. In a single BSMF from such a large group-prefix is lost, then that entire group-prefix will have to wait until the next BSM is originated.

Next we need to consider how a BSR would remove group-prefixes from the BSM. A router receiving a set of BSMFs cannot tell if a group-prefix is missing. If it has seen a group-prefix before, it must assume that that group-prefix still exists, and that the BSMF describing it has been lost. It should retain this information for BS Timeout seconds. Thus for a BSR to remove a group-prefix from the BSR, it should include that group-prefix, but with a RP Count of zero, and it should resend this information in each BSM for BS Timeout seconds.

Finally, we come to the case of fragments for the purpose of conveying admin scope group-prefixes. In general, the information for each admin scope range is independent of information about other admin scope ranges. As no BSMF is allowed to convey information for more than one admin scope range, then the procedure above also applies to BSMs that are fragmented due to admin scoping. However, to time out all the state for an entire admin scope zone requires waiting SZ Timeout rather than BS Timeout, as admin scope zones are not expected to come and go frequently.

4.2. Candidate-RP-Advertisement Format

Candidate-RP-Advertisements are periodically unicast from the C-RPs to the BSR.

```
 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  |   Reserved    |           Checksum            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Prefix Cnt    |   Priority    |           Holdtime            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             RP Address (Encoded-Unicast format)               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Group Address 1 (Encoded-Group format)             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                               .                               |
|                               .                               |
|                               .                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Group Address n (Encoded-Group format)             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

PIM Version, Reserved, Checksum
Described in [3].

Type
PIM Message Type. Value is 8 for a Candidate-RP-Advertisement Message.

Prefix Cnt
The number of encoded group addresses included in the message; indicating the group prefixes for which the C-RP is advertising. A Prefix Cnt of ‘0’ implies all multicast groups, e.g. for IPv4 a prefix of 224.0.0.0 with mask length of 4. If the C-RP is not configured with Group-prefix information, the C-RP puts a default value of ‘0’ in this field.

Priority
The ‘Priority’ of the included RP, for the corresponding Encoded-Group Address (if any). highest priority is ‘0’ (i.e. the lower the value of the ‘Priority’ field, the higher the priority). This field is stored at the BSR upon receipt along with the RP address and corresponding Encoded-Group Address.

Holdtime
The amount of time the advertisement is valid. This field allows advertisements to be aged out.

RP Address
The address of the interface to advertise as a Candidate RP. The format for this address is given in the Encoded-Unicast address in [3].

Group Address-1..n
The group prefixes for which the C-RP is advertising. Format described in Encoded-Group-Address in [3].

5. Default Values for Timers

Timer Name: Bootstrap Timer (BST)

Value Name	Value	Explanation
BS Period	Default: 60 secs	Period between bootstrap messages
BS Timeout	2 * BS_Period + 10 seconds	Period after last BS message before BSR is timed out and election begins
rand_override	rand(0, 5.0 secs)	Suppression period in BSR election to prevent thrashing

Timer Name: C-RP Expiry Timer (CET(R))

Value Name	Value	Explanation
C-RP Timeout	from message	Hold time from C-RP-Adv message

C-RP Advertisement messages are sent periodically with period C-RP-Adv-Period. C-RP-Adv-Period defaults to 60 seconds. The holdtime to be specified in a C-RP-Adv message should be set to (2.5 * C-RP-Adv-Period).

Timer Name: C-RP Advertisement Timer (CRPT)

Value Name	Value	Explanation
C-RP-Adv-Period	Default: 60 seconds	Period with which periodic C-RP Advertisements are sent to BSR

Timer Name: Scope Zone Expiry Timer (SZT(Z))

Value Name Value Explanation		
SZ Timeout	1300 seconds	Interval after which a scope zone will be timed out if the state is not refreshed

6. Security Considerations

6.1. Possible Threats

Threats affecting the PIM BSR mechanism are primarily of two forms: denial of service attacks, and traffic diversion attacks. An attacker that subverts the BSR mechanism can prevent multicast traffic from reaching the intended recipients, can divert multicast traffic to a place where they can monitor it, and can potentially flood third parties with traffic.

Traffic can be prevented from reaching the intended recipients by one of two mechanisms:

- Subverting a BSM, and specifying RPs that won't actually forward traffic.
- Registering with the BSR as a C-RP, and then not forwarding traffic.

Traffic can be diverted to a place where it can be monitored by both of the above mechanisms; in this case the RPs would forward the traffic, but are located so as to aid monitoring or man-in-the-middle attacks on the multicast traffic.

A third party can be flooded by either of the above two mechanisms by specifying the third party as the RP, and register-encapsulated traffic will then be forwarded to them.

6.2. Limiting Third-Party DoS Attacks

The third party DoS attack above can be greatly reduced if PIM routers acting as DR do not continue to forward Register traffic to the RP in the presence of ICMP Protocol Unreachable or ICMP Host Unreachable responses. If a PIM router sending Register packets to an RP receives one of these responses to a data packet it has sent, it should rate-limit the transmission of future Register packets to that RP for a short period of time.

As this does not affect interoperability, the precise details are left to the implementor to decide. However we note that a router implementing such rate limiting must only do so if the ICMP packet correctly echoes part of a Register packet that was sent to the RP. If this check were not made, then simply sending ICMP Unreachable packets to the DR with the source address of the RP spoofed would be sufficient to cause a denial-of-service attack on the multicast traffic originating from that DR.

6.3. BS Message Security

If a legitimate PIM router is compromised, there is little any security mechanism can do to prevent that router subverting PIM traffic in that domain. However we recommend that implementors provide a mechanism whereby a PIM router using the BSR mechanisms can be configured with the IP addresses of valid BSR routers, and that any BS Message from any other BSR should then be dropped and logged as a security issue. We also recommend that this not be enabled by default, as it makes the initial configuration of a PIM domain problematic - it is the sort of feature that might be enabled once the configuration of a domain has stabilized.

The primary security requirement for BSR (as for PIM) is that it is possible to prevent hosts that are not legitimate PIM routers, either within or outside the domain, from subverting the BSR mechanism.

The Bootstrap Message Processing Checks prevent a router from accepting a BS message from outside of the PIM Domain, as the source address on BS Messages must be an immediate PIM neighbor. There is however a small window of time after a reboot where a PIM router will accept a bad BS Message unicast from an immediate neighbor, and it might be possible to unicast a BS Message to a router during this interval from outside the domain, using the spoofed source address of a neighbor. This can be prevented if PMBRs perform source-address filtering to prevent packets entering the PIM domain with IP source addresses that are infrastructure addresses in the PIM domain.

The principle threat to BS Message security comes from hosts within the PIM domain that attempt to subvert the BSR mechanism. They may be able to do this by sending PIM messages to their local router, or by unicasting a BS message to another PIM router during the brief interval after it has restarted.

All BS Messages SHOULD carry the Router Alert IP option. If a PIM router receives a BS Message that does not carry the router alert option, it SHOULD drop it (a configuration option should also be provided to disable this check on a per-interface basic for backward compatibility with older PIM routers). The Router Alert option allows a PIM router to perform checks on unicast packets it would otherwise blindly forward. All PIM routers should check that packets with Router Alert that are not destined for the router itself are not PIM BootStrap messages. Any such packets should be dropped and logged as a possible security issue - it is never acceptable for a PIM BS message to travel multiple IP hops.

Most hosts that are likely to attempt to subvert PIM BSR are likely to be located on leaf subnets. We recommend that implementors provide a configuration option that specifies an interface is a leaf subnet, and that no PIM packets are accepted on such interfaces.

On multi-access subnets with multiple PIM routers and hosts that are not trusted, we recommend that IPsec AH is used to protect communication between PIM routers, and that such routers are configured to drop and log communication attempts from any host that do not pass the authentication check. When all the PIM routers are under the same administrative control, this authentication may use a configured shared secret. The securing of interactions between PIM neighbors is discussed in more detail in the Security Considerations section of [3], and so we do not discuss the details further here. The same security mechanisms than can be used to secure PIM Join, Prune and Assert messages should also be used to secure BS messages.

6.4. C-RP-Advertisement Security

Even if it is not possible to subvert BS Messages, an attacker might be able to perform most of the same attacks by simply sending C-RP-Adv messages to the BSR specifying the attacker's choice of RPs. Thus it is necessary to control the sending of C-RP-Adv messages in essentially the same ways that we control BS Messages. However, C-RP-Adv messages are unicast and normally travel multiple hops, so controlling them is a little harder.

We specify that C-RP-Adv messages SHOULD also carry the Router Alert IP option, and that the BSR SHOULD by default drop and log C-RP-Adv messages that do not carry this option. Setting Router Alert on these packets is practical because the rate of C-RP-Adv messages should be very low, so the extra load on routers forwarding these packets will be insignificant. All PIM routers forwarding such a packet are then capable of checking whether the packet came from a valid neighbor. On interfaces that are configured to be leaf subnets, all C-RP-Adv messages should be dropped. On multi-access subnets with multiple PIM routers and hosts that are not trusted, the router can at least check that the source MAC address is that of a valid PIM neighbor. PMBRs should ensure that no C-RP-Adv messages enter the domain from an external neighbor.

For true security, we recommend that all C-RPs are configured to use IPsec authentication. The authentication process for a C-RP-Adv message between a C-RP and the BSR is identical to the authentication process for PIM Register messages between a DR and the relevant RP, except that there will normally be fewer C-RPs in a domain than there are DRs, so key management is a little simpler. We do not describe the details of this process further here, but refer to the Security Considerations section of [3]. Note that the use of cryptographic security for C-RP-Advs does not remove the need for the non-cryptographic mechanisms, as explained below.

6.5. Denial of Service using IPsec

An additional concern is that of Denial-of-Service attacks caused by sending high volumes of BS Messages or C-RP-Adv messages with invalid IPsec authentication information. It is possible that these messages could overwhelm the CPU resources of the recipient.

The non-cryptographic security mechanisms above prevent unicast BS messages from traveling multiple hops, and constrain who can originate such messages. However, it is obviously important that PIM Messages that are required to have Router Alert checked are checked for this option before the IPsec AH is checked. Thus the remaining vulnerability primarily exists for hosts on multi-access subnets containing more than one PIM router. A PIM router receiving PIM packets with Router Alert set from such a subnet should already be checking that the source MAC address is that of a valid PIM neighbor, but this is hardly strong security. In addition, we recommend that rate-limiting mechanisms can be configured, to be applied to the forwarding of unicast PIM packets containing Router Alert options. The rate-limiter MUST independently rate-limit different types of PIM packets - for example a flood of C-RP-Adv messages MUST NOT cause a rate limiter to drop low-rate BS Messages. Such a rate-limiter might itself be used to cause a denial of service attack by causing valid packets to be dropped, but in practice this is more likely to constrain bad PIM Messages close to their origin. In addition, the rate limiter will prevent attacks on PIM from affecting other activity on the destination router, such as unicast routing.

7. Authors' Addresses

Bill Fenner
AT&T Labs - Research
75 Willow Road
Menlo Park, CA 94025
fenner@research.att.com

Mark Handley
ICIR/ICSI
1947 Center St, Suite 600
Berkeley, CA 94708
mjh@icir.org

Roger Kermode
Motorola Australian Research Centre
Locked Bag 5028
Botany NSW 1455,
Australia
Roger.Kermode@motorola.com

David Thaler
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052
dthaler@Microsoft.com

8. References

[1] S. Deering , W. Fenner , B. Haberman, "Multicast Listener Discovery (MLD) for IPv6", RFC 2710, Oct 1999.

[2] D. Estrin et al., "Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification", RFC 2362, June 1998 (now obsolete).

[3] W. Fenner, M. Handley, H. Holbrook, I. Kouvelas, "Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)", Internet Draft draft-ietf-pim-sm-v2-new-05.ps

[4] W. Fenner, "Internet Group Management Protocol, Version 2", RFC 2236, Nov 1997.

[5] IANA, "Address Family Numbers", linked from http://www.iana.org/numbers.html

[6] D. Meyer, "Administratively Scoped IP Multicast", RFC 2365, Jul 1998.

9. Acknowledgments

PIM-SM was designed over many years by a large group of people, including ideas from Deborah Estrin, Dino Farinacci, Ahmed Helmy, Steve Deering, Van Jacobson, C. Liu, Puneet Sharma, Liming Wei, Tom Pusateri, Tony Ballardie, Scott Brim, Jon Crowcroft, Paul Francis, Joel Halpern, Horst Hodel, Polly Huang, Stephen Ostrowski, Lixia Zhang, Girish Chandranmenon, Pavlin Radoslavov, John Zwiebel, Isidor Kouvelas and Hugh Holbrook. This BSR specification draws heavily on text from RFC 2362.