Internet Engineering Task Force
Internet Draft
Expiration Date: November 2003

P. Savola
CSC/FUNET

B. Haberman
Caspian Networks

May 2003

Embedding the Address of RP in IPv6 Multicast Address

draft-savola-mboned-mcast-rpaddr-03.txt

Status of this Memo

Abstract

As has been noticed, there is exists a huge deployment problem with global, interdomain IPv6 multicast: Protocol Independent Multicast - Sparse Mode (PIM-SM) Rendezvous Points (RPs) have no way of communicating the information about multicast sources to other multicast domains, as there is no Multicast Source Discovery Protocol (MSDP), and the whole interdomain Any Source Multicast model is rendered unusable; SSM avoids these problems. This memo outlines a way to embed the address of the RP in the multicast address, solving the interdomain multicast problem. The problem is three-fold: specify an address format, adjust the operational procedures and configuration if necessary, and modify PIM-SM implementations of those who want to join or send to a group (Designated Routers) or

provide one (Rendezvous Points). In consequence, there would be no need for interdomain MSDP, and even intra-domain RP configuration could be simplified. This memo updatres RFC 3306.

Table of Contents

1. Introduction

As has been noticed [V6MISSUES], there exists a huge deployment problem with global, interdomain IPv6 multicast: PIM-SM [PIM-SM] RPs have no way of communicating the information about multicast sources to other multicast domains, as there is no MSDP [MSDP], and the whole interdomain Any Source Multicast model is rendered unusable; SSM [SSM] avoids these problems.

However, it has been noted that there are some problems with SSM deployment and support: it seems unlikely that SSM could be usable as the only interdomain multicast routing mechanism in the short term. This memo proposes a short-to-midterm fix to interdomain multicast routing, and provides an additional method for the RP discovery with the intra-domain case.

This memo outlines a way to embed the address of the RP in the multicast address, solving the interdomain multicast problem. The problem is three-fold: specify an address format, adjust the operational procedures and configuration if necessary, and modify PIM-SM implementations used in the multicast path as described in this memo. In consequence, there would be no need for interdomain MSDP, and even intra-domain RP configuration could be simplified.

The solution is founded upon unicast-prefix-based IPv6 multicast addressing [UNIPRFXM] and making some assumptions about IPv6 address assignment for the RPs in the PIM-SM domain.

Further, a change in how interdomain PIM-SM operates with these addresses is presented: multicast receivers' and senders' DR's join or send to (respectively) the RP embedded in the address -- not their otherwise locally configured RP (if any).

It is self-evident that one can't embed, in the general case, two 128-bit addresses in one 128-bit address. In this memo, some assumptions on how this could be done are made. If these assumptions can't be followed, either operational procedures and configuration must be slightly changed or this mechanism not be used.

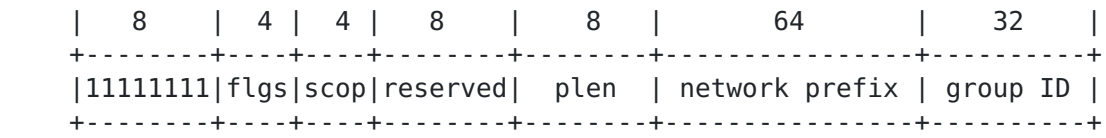
The assignment of multicast addresses is outside the scope of this document; however, the mechanisms are very probably similar to ones used with [UNIPRFXM].

This memo updates the addressing format presented in RFC 3306.

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. Unicast-Prefix-based Address Format

As described in [UNIPRFXM], the multicast address format is as follows:



Where flgs are "0011". (The first two bits are yet undefined and thus zero.)

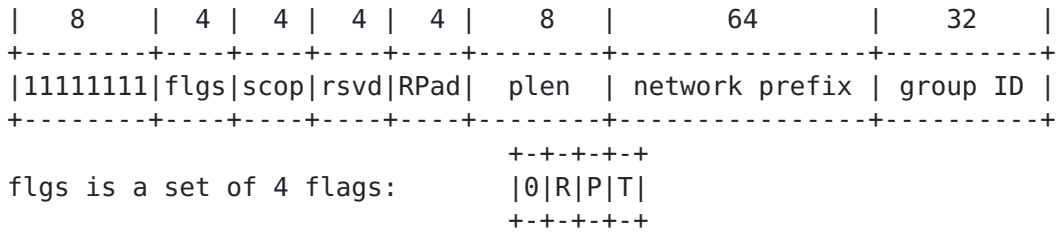
3. Modified Unicast-Prefix-based Address Format

This memo proposes a modification to the unicast-prefix-based address format:

1. If the second high-order bit in "flgs" is set to 1, the address of the RP is embedded in the multicast address, as described in this memo.

2. If the second high-order bit in "flgs" was set to 1, interpret the last low-order 4 bits of "reserved" field as signifying the RP interface ID, as described in this memo.

In consequence, the address format becomes:



R = 1 indicates a multicast address that embeds the address of the PIM-SM RP. Then P MUST BE set to 1, and consequently T MUST be set to 1, as specified in [UNIPRFXM].

In the case that R = 1, the last 4 bits of previously reserved field ("RPad") are interpreted as embedding the interface ID of the RP, as specified in this memo.

R = 0 indicates a multicast address that does not embed the address of the PIM-SM RP and follows the semantics defined in [ADDRARCH] and [UNIPRFXM]. In this context, the value of "RPad" has no meaning.

4. Embedding the Address of the RP in the Multicast Address

The address of the RP can only be embedded in unicast-prefix -based addresses, but the scheme could be extended to other forms of multicast addresses as well. Further, the mechanism cannot be combined with SSM, as SSM has no RP's.

To identify whether an address is a multicast address as specified in this memo and to be processed any further, it must satisfy all of the below:

- it MUST be a multicast address and have R, P, and T flag bits set to 1 (that is, be part of the prefix FF7::/12 or FFF::/12)

- "plen" MUST NOT be 0 (ie. not SSM)

- "plen" MUST NOT be greater than 64

The address of the RP can be obtained from a multicast address satisfying the above criteria by taking the following steps:

1. take the last 96 bits of the multicast address add 32 zero bits at the end,

2. zero the last 128-"plen" bits, and

3. replace the last 4 bits with the contents of "RPad".

One should note that there are several operational scenarios when [UNIPRFXM] statement "All non-significant bits of the network prefix field SHOULD be zero" is ignored -- and why the second step, above, is necessary. This is to allow multicast address assignments to third parties which still use your RP; see example 2 below.

"plen" higher than 64 MUST NOT be used as that would overlap with the upper bits of multicast group-id.

The implementation MUST perform at least the same address validity checks to the calculated RP address as to one received via other means (like BSR [BSR] or MSDP), to avoid e.g. the address being "::" or "::1".

One should note that the 4 bits reserved for "RPad" set the upper bound for RP's per multicast group address; not the number of RP's in a subnet, PIM-SM domain or large-scale network.

5. Examples

5.1. Example 1

The network administrator of 3FFE:FFFF::/32 wants to set up an RP for the network and all of his customers. He chooses network prefix=3FFE:FFFF and plen=32, and wants to use this addressing mechanism. The multicast addresses he will be able to use are of the form:

FF7x:y20:3FFE:FFFF:zzzz:zzzz:<group-id>

Where "x" is the multicast scope, "y" the interface ID of the RP address, and "zzzz:zzzz" will be freely assignable within the PIM-SM domain. In this case, the address of the PIM-SM RP would be:

3FFE:FFFF::y

(and "y" could be anything from 0 to F); the address 3FFE:FFFF::y/128 is added as a Loopback address and injected to the routing system.

5.2. Example 2

As above, the network administrator can also allocate multicast addresses like "FF7x:y20:3FFE:FFFF:DEAD::/80" to some of his customers within the PIM-SM domain. In this case the RP address would still be "3FFE:FFFF::y".

Note the second rule of deriving the RP address: the "plen" field in the multicast address, (hex)20 = 32, refers to the length of "network prefix" field considered when obtaining the RP address. In this case, only the first 32 bits of the network prefix field, "3FFE:FFFF" are preserved: the value of "plen" takes no stance on actual unicast/multicast prefix lengths allocated or used in the networks, here from 3FFE:FFFF:DEAD::/48.

5.3. Example 3

In the above network, the network admin sets up addresses as above, but an organization wants to have their own PIM-SM domain; that's reasonable. The organization can pick multicast addresses like "FF7x:y30:3FFE:FFFF:BEEF::/80", and then their RP address would be "3FFE:FFFF:BEEF::y".

5.4. Example 4

In the above networks, if the admin wants to specify the RP to be in a non-zero /64 subnet, he could always use something like "FF7x:y40:3FFE:FFFF:BEEF:FEED::/96", and then their RP address would be "3FFE:FFFF:BEEF:FEED::y". There are still 32 bits of multicast group-id's to assign to customers and self.

6. Operational Requirements

6.1. Anycast-RP

One should note that MSDP is also used, in addition to interdomain connections between RPs, in anycast-RP [ANYCASTRP] -technique, for sharing the state information between different RPs in one PIM-SM domain. However, there are other propositions, like [ANYPIMRP].

Anycast-RP mechanism is incompatible with this addressing method unless MSDP is specified and implemented. Alternatively, another method for sharing state information could be used.

Anycast-RP and other possible RP failover mechanisms are outside of the scope of this memo.

6.2. Guidelines for Assigning IPv6 Addresses to RPs

With this mechanism, the RP can be given basically any network prefix up to /64. The interface identifier will have to be manually configured to match "RPad".

If an administrator wishes to use an RP address that does not conform to the addressing topology, that address can be injected into the routing system via a host route. This RP address SHOULD be assigned out of the network's prefix in order to ensure aggregation at the border.

7. Required PIM-SM Modifications

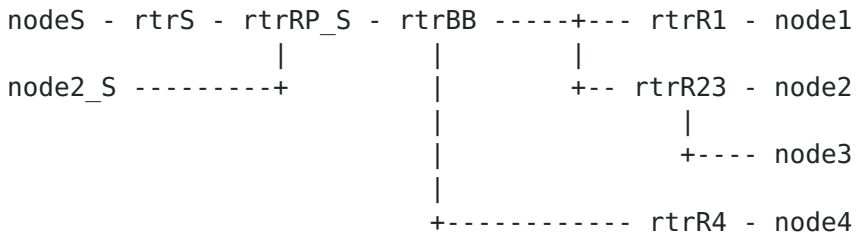
The use of multicast addresses with embedded RP addresses requires additional PIM-SM processing. Namely, a PIM-SM router will need to be able to recognize the encoding and derive the RP address from the address using the rules in section 4 and to be able to use the embedded RP, instead of its own for multicast addresses in this specified range.

The three key places where these modifications are used are the Designated Routers (DRs) on the receiver/sender networks, the backbone networks, and the RPs in the domain where the embdedded address has been derived from (see figure below).

For the foreign DR's (rtrR1, rtrR23, and rtrR4), this means sending PIM-SM Join/Prune/Register messages towards the foreign RP (rtrRP_S). Naturally, PIM-SM Register-Stop and other messages must also be allowed from the foreign RP. DR's in the local PIM-SM domain (rtrS) do the same.

For the RP (rtrRP_S), this means being able to recognize and validate PIM-SM messages which use RP-embedded addressing originated from any DR at all.

For the other routers on the path (rtrBB), this means recognizing and validating that the Join/Prune PIM-SM messages using the embedded RP addressing are on the right path towards the RP they think is in charge of the particular address.



In addition, the administration of the PIM-SM domains MAY have an option to manually override the RP selection for the embedded RP multicast addresses: the default policy SHOULD be to use the embedded RP.

The extraction of the RP information from the multicast address should be done during forwarding state creation. That is, if no state exists for the multicast address, PIM-SM must take the embedded RP information into account when creating forwarding state. Unless otherwise dictated by the administrative policy, this would result in a receiver's DR initiating a PIM-SM Join towards the foreign RP or a source's DR sending PIM-SM Register messages towards the foreign RP.

It should be noted that this approach removes the need to run inter-domain MSDP. Multicast distribution trees in foreign networks can be joined by issuing a PIM-SM Join/Prune/Register to the RP address encoded in the multicast address.

Also, the addressing model described here could be used to replace or augment the intra-domain Bootstrap Router mechanism (BSR), as the RP-mappings can be communicated by the multicast address assignment.

7.1. Overview of the Model

The steps when a receiver wishes to join a group are:

1. A receiver finds out a group address from some means (e.g. SDR or a web page).

2. The receiver issues an MLD Report, joining the group.
3. The receiver's DR will initiate the PIM-SM Join process towards the RP embedded in the multicast address.

The steps when a sender wishes to send to a group are:

1. A sender finds out a group address from some means, whether in an existing group (e.g. SDR, web page) or in a new group (e.g. a call to the administrator for group assignment, use of a multicast address assignment protocol).
2. The sender sends to the group.
3. The sender's DR will send the packets unicast-encapsulated in PIM-SM Register-messages to the RP address encoded in the multicast address (in the special case that DR is the RP, such sending is only conceptual).

In both cases, the messages then go on as specified in [PIM-SM] and other specifications (e.g. Register-Stop and/or SPT Join); there is no difference in them except for the fact that the RP address is derived from the multicast address.

Sometimes, some information, using conventional mechanisms, about another RP exists in the PIM-SM domain. The embedded RP SHOULD be used by default, but there MAY be an option to switch the preference. This is because especially when performing PIM-SM forwarding in the transit networks, the routers must have the same notion of the RP, or else the messages may be dropped.

8. Scalability/Usability Analysis

Interdomain MSDP model for connecting PIM-SM domains is mostly hierarchical. The "embedded RP address" changes this to a mostly flat, sender-centered, full-mesh virtual topology.

This may or may not cause some effects; it may or may not be desirable. At the very least, it makes many things much more robust as the number of third parties is minimized. A good scalability analysis is needed.

In some cases (especially if e.g. every home user is employing site-local multicast), some degree of hierarchy would be highly desirable, for scalability (e.g. take the advantage of shared multicast state) and administrative point-of-view.

Being able to join/send to remote RP's has security considerations that are considered below, but it has an advantage too: every group has a "home RP" which is able to control (to some extent) who are able to send to the group.

One should note that the model presented here simplifies the PIM-SM multicast routing model slightly by removing the RP for senders and receivers in foreign domains. One scalability consideration should be noted: previously foreign sources sent the unicast-encapsulated data to their local RP, now they do so to the foreign RP responsible for the specific group. This is especially important with large multicast groups where there are a lot of heavy senders -- particularly if implementations do not handle unicast-decapsulation well.

This model increases the amount of Internet-wide multicast state slightly: the backbone routers might end up with (*, G) and (S, G, rpt) state between receivers and the RP, in addition to (S, G) states between the receivers and senders. Certainly, the amount of inter-domain multicast traffic between sources and the embedded-RP will increase compared to the ASM model with MSDP; however, the domain responsible for the RP is expected to be able to handle this.

As the address of the RP is tied to the multicast address, in the case of RP failure, PIM-SM BSR mechanisms cannot pick a new RP; the failover mechanisms, if used, for backup RP's are different, and typically would depend on sharing one address. The failover techniques are outside of the scope of this memo.

9. Acknowledgements

Jerome Durand commented on an early draft of this memo. Marshall Eubanks noted an issue regarding short plen values. Tom Pusateri noted problems with earlier SPT-join approach. Rami Lehtonen pointed out issues with the scope of SA-state and provided extensive commentary. Nidhi Bhaskar gave the draft a thorough review. The whole MboneD working group is also acknowledged for the continued support and comments.

10. Security Considerations

The address of the PIM-SM RP is embedded in the multicast address. RPs may be a good target for Denial of Service attacks -- as they are a single point of failure (excluding failover techniques) for a group. In this way, the target would be clearly visible. However, it could be argued that if interdomain multicast was to be made work e.g. with MSDP, the address would have to be visible anyway (through via other channels, which may be more easily securable).

As any RP will have to accept PIM-SM Join/Prune/Register messages from any DR's, this might cause a potential DoS attack scenario. However, this can be mitigated by the fact that the RP can discard all such messages for all multicast addresses that do not embed the

address of the RP, and if deemed important, the implementation could also allow manual configuration of which multicast addresses or prefixes embedding the RP could be used; however, at least with addresses, this would increase the need for coordination between multicast sources and administration.

In a similar fashion, DR's must accept similar PIM-SM messages back from the foreign RP's for which a receiver in DR's network has joined.

One consequence of the usage model is that it allows Internet-wide multicast state creation (from receiver(s) in another domain to the RP in another domain) compared to the domain wide state creation in the MSDP model.

RPs may become a bit more single points of failure as anycast-RP mechanism is not (at least immediately) available. This can be partially mitigated by the fact that some other forms of failover are still possible, and there should be less need to store state as with MSDP.

The implementation MUST perform at least the same address validity checks to the embedded RP address as to one received via other means (like BSR or MSDP), to avoid the address being e.g. "::" or "::1".

11. References

11.1. Normative References

[ADDRARCH] Hinden, R., Deering, S., "IP Version 6 Addressing Architecture", RFC3513, April 2003.

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

[UNIPRFXM] Haberman, B., Thaler, D., "Unicast-Prefix-based IPv6 Multicast Addresses", RFC3306, August 2002.

11.2. Informative References

[ANYCASTRP] Kim, D. et al, "Anycast RP mechanism using PIM and MSDP", work-in-progress, draft-ietf-mboned-anycast-rp-08.txt, May 2001.

[ANYPIMRP] Farinacci, D., Cai, Y., "Anycast-RP using PIM", work-in-progress, draft-farinacci-pim-anycast-rp-00.txt, January 2003.

[BSR] Fenner, B., et al., "Bootstrap Router (BSR) Mechanism for PIM Sparse Mode", work-in-progress, draft-ietf-pim-sm-bsr-03.txt, February 2003.

[MSDP] Meyer, D., Fenner, B, (Eds.), "Multicast Sourc Discovery Protocol (MSDP)", work-in-progress, draft-ietf-msdp-spec-16.txt, May 2003.

[PIM-SM] Fenner, B. et al, "Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised), work-in-progress, draft-ietf-pim-sm-v2-new-06.txt, December 2002.

[SSM] Holbrook, H. et al, "Source-Specific Multicast for IP", work-in-progress, draft-ietf-ssm-arch-03.txt, May 2003.

[V6MISSUES] Savola, P., "IPv6 Multicast Deployment Issues", work-in-progress, draft-savola-v6ops-multicast-issues-01.txt, November 2002.

Authors' Addresses

Pekka Savola
CSC/FUNET
Espoo, Finland
EMail: psavola@funet.fi

Brian Haberman
Caspian Networks
One Park Drive
Suite 400
Research Triangle Park, NC 27709
EMail: bkhabs@nc.rr.com
Phone: +1-919-949-4828

A. Open Issues/Discussion

The initial thought was to use only SPT join from local RP/DR to foreign RP, rather than a full PIM Join to foreign RP. However, this turned out to be problematic, as this kind of SPT joins where disregarded because the path had not been set up before sending them. A full join to foreign PIM domain is a much clearer approach.

One could argue that there can be more RPs than the 4-bit "RPad" allows for, especially if anycast-RP cannot be used. In that light, extending "RPad" to take full advantage of whole 8 bits would seem reasonable. However, this would use up all of the reserved bits, and

leave no room for future flexibility. In case of large number of RPs, an operational workaround could be to split the PIM domain: for example, using two /33's instead of one /32 would gain another 16 RP addresses.

Some hierarchy (e.g. two-level, "ISP/customer") for RPs could possibly be added if necessary, but that would be torturing one 128 bits even more.

One particular case, whether in the backbone or in the sender's domain, is where the regular PIM-SM RP would be X, and the embedded RP address would be Y. This could e.g. be a result of a default all-multicast-to-one-RP group mapping, or a local policy decision. The embedded RP SHOULD be used by default, but there MAY be an option to change this preference.

Values 64 < "plen" < 96 would overlap with upper bits of the multicast group-id; due to this restriction, "plen" must not exceed 64 bits. This is in line with RFC 3306.