mboned Working Group P. Savola
Internet Draft CSC/FUNET
Expiration Date: September 2004

B. Haberman
Caspian Networks

March 2004

Embedding the Rendezvous Point (RP) Address in an IPv6 Multicast Address

draft-ietf-mboned-embeddedrp-02.txt

Status of this Memo

Abstract

This memo defines an address allocation policy in which the address of the Rendezvous Point (RP) is encoded in an IPv6 multicast group address. For Protocol Independent Multicast - Sparse Mode (PIM-SM), this can be seen as a specification of a group-to-RP mapping mechanism. This allows an easy deployment of scalable inter-domain multicast, and simplifies the intra-domain multicast configuration as well. This memo updates the addressing format presented in RFC 3306.

Table of Contents

1. Introduction

1.1. Background

As has been noticed [V6MISSUES], there exists a deployment problem with global, interdomain IPv6 multicast: PIM-SM [PIM-SM] RPs have no way of communicating the information about (active) multicast sources to other multicast domains, as Multicast Source Discovery Protocol (MSDP) [MSDP] has not been, on purpose, specified for IPv6. Therefore the whole interdomain Any Source Multicast model is rendered unusable; Source-Specific Multicast (SSM) [SSM] avoids these problems but is not a complete solution for several reasons.

Further, it has been noted that there are some problems with the support and deployment of mechanisms SSM would require [V6MISSUES]: it seems unlikely that SSM could be usable as the only interdomain multicast routing mechanism in the short term.

1.2. Solution

This memo describes a multicast address allocation policy in which the address of the RP is encoded in the IPv6 multicast group address, and specifies a PIM-SM group-to-RP mapping to use the encoding, leveraging and extending unicast-prefix -based addressing [RFC3306].

This mechanism not only provides a simple solution for IPv6 interdomain Any Source Multicast (ASM) but can be used as a simple solution for IPv6 intradomain ASM with scoped multicast addresses as well. It can also be used as an automatic RP discovery mechanism in those deployment scenarios which would have previously used the Bootstrap Router protocol (BSR) [BSR].

The solution consists of three elements:

- o A specification of a subrange of [RFC3306] IPv6 multicast group addresses defined by setting one previously unused bit of the Flags field to "1",

- o A specification of the mapping by which such a group address encodes the RP address that is to be used with this group, and

- o A description of operational procedures to operate ASM with PIM-SM on these IPv6 multicast groups.

Addresses in the subrange will be called embedded-RP addresses.

This scheme obviates the need for MSDP, and the routers are not required to include any multicast configuration, except when they act

as an RP.

This memo updates the addressing format presented in RFC 3306.

1.3. Assumptions and Scope

In general, a 128-bit RP address can't be embedded into a 128-bit group address with space left to carry the group identity itself. An appropriate form of encoding is thus defined by requiring that the Interface-IDs of RPs in the embedded-RP range can be assigned to be a specific value.

If these assumptions can't be followed, either operational procedures and configuration must be slightly changed or this mechanism can not be used.

The assignment of multicast addresses is outside the scope of this document; it is up to the RP and applications to ensure that group addresses are unique using some unspecified method. However, the mechanisms are very probably similar to ones used with [RFC3306].

Similarly, RP failure management methods, such as Anycast-RP, are out of scope for this document. These do not work without additional specification or deployment. This is covered briefly in Section 6.1.

1.4. Keywords

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 [RFC3306], the multicast address format is as follows:

```
|   8    |  4 |  4 |   8    |    8   |       64       |    32    |
+--------+----+----+--------+--------+----------------+----------+
|11111111|flgs|scop|reserved|  plen  | network prefix | group ID |
+--------+----+----+--------+--------+----------------+----------+
```

Where flgs are "0011". (The first two bits have been yet undefined, sent as zero and ignored on receipt.)

3. Modified Unicast-Prefix-based Address Format

This memo specifies 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" is set to 1, interpret the last low-order 4 bits of "reserved" field as signifying the RP interface ID ("RIID"), as described in this memo.

In consequence, the address format becomes:

```
|   8    |  4 |  4 |  4 |  4 |    8   |       64       |    32    |
+--------+----+----+----+----+--------+----------------+----------+
|11111111|flgs|scop|rsvd|RIID|  plen  | network prefix | group ID |
+--------+----+----+----+----+--------+----------------+----------+
                                +-+-+-+-+
flgs is a set of 4 flags:       |0|R|P|T|
                                +-+-+-+-+
```

R = 1 indicates a multicast address that embeds the address on the RP. Then P MUST be set to 1, and consequently T MUST be set to 1, as specified in [RFC3306].

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

R = 0 indicates a multicast address that does not embed the address of the RP and follows the semantics defined in [ADDRARCH] and [RFC3306]. In this context, the value of "RIID" MUST be sent as zero and MUST be ignored on receipt.

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

The address of the RP can only be embedded in unicast-prefix -based ASM addresses.

That is, 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:

o it MUST be a multicast address and have R, P, and T flag bits set to 1 (that is, be part of the prefixes FF70::/12 or FFF0::/12),

o "plen" MUST NOT be 0 (ie. not SSM), and

o "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 two steps:

1. copy the first "plen" bits of the "network prefix" to a zeroed 128-bit address structure, and
2. replace the last 4 bits with the contents of "RIID".

These two steps could be illustrated as follows:

```
     | 20 bits | 4  | 8  |       64        |    32    |
     +---------+----+----+----------------+----------+
     |xtra bits|RIID|plen| network prefix | group ID |
     +---------+----+----+----------------+----------+
                 ||    \\  vvvvvvvvvvv
                 ||     ``====> copy plen bits of "network prefix"
                 ||       +------------+-----------------------+
                 ||       | network pre| 0000000000000000000000 |
                 ||       +------------+-----------------------+
                  \\
                   ``=================> copy RIID to the last 4 bits
                          +------------+---------------------+--+
                          | network pre| 0000000000000000000 |ID|
                          +------------+---------------------+--+
```

One should note that there are several operational scenarios (see Example 2 below) when [RFC3306] statement "all non-significant bits of the network prefix field SHOULD be zero" is ignored. This is to allow multicast group address allocations to be consistent with unicast prefixes, while the multicast addresses would still use the RP associated with the network prefix.

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

When processing an encoding to get the RP address, the multicast routers 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 for IPv4). At least fe80::/10, ::/16, and ff00::/8 MUST be excluded. This is particularly important as the information is obtained from an untrusted source, i.e., any Internet user's input.

One should note that the 4 bits reserved for "RIID" set the upper bound for RPs for the combination of scope, network prefix, and group ID -- without varying any of these, you can have 4 bits worth of different RPs. However, each of these is an IPv6 group address of its own (i.e., there can be only one RP per multicast address).

5. Examples

Four examples of multicast address allocation and resulting group-to-RP mappings are described here, to better illustrate the possibilities provided by the encoding.

5.1. Example 1

The network administrator of 2001:DB8::/32 wants to set up an RP for the network and all the customers. (S)he chooses network prefix=2001:DB8 and plen=32, and wants to use this addressing mechanism. The multicast addresses (s)he will be able to use are of the form:

```
FF7x:y20:2001:DB8: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 to anyone. In this case, the address of the RP would be:

```
2001:DB8::y
```

(and "y" could be anything from 1 to F, as 0 must not be used); the address 2001:DB8::y/128 is added on a router as a loopback address and injected to the routing system.

5.2. Example 2

As in Example 1, the network administrator can also allocate multicast addresses like "FF7x:y20:2001:DB8:DEAD::/80" to some of customers. In this case the RP address would still be "2001:DB8::y".

Note the second rule of deriving the RP address: the "plen" field in the multicast address, 0x20 = 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, "2001:DB8" are preserved: the value of "plen" takes no stance on actual unicast/multicast prefix lengths allocated or used in the networks, here from 2001:DB8:DEAD::/48.

In short, this distinction allows more flexible RP address configuration in the scenarios where it is desirable to have the

group addresses to be consistent with the unicast prefix allocations.

5.3. Example 3

In the network of Examples 1 and 2, the network admin sets up addresses for use by their customers, but an organization wants to have their own PIM-SM domain. The organization can pick multicast addresses like "FF7x:y30:2001:DB8:BEEF::/80", and then their RP address would be "2001:DB8:BEEF::y".

5.4. Example 4

In the above networks, if the administrator wants to specify the RP to be in a non-zero /64 subnet, (s)he could always use something like "FF7x:y40:2001:DB8:BEEF:FEED::/96", and then their RP address would be "2001:DB8:BEEF:FEED::y". There are still 32 bits of multicast group-id's to assign to customers and self.

6. Operational Considerations

This desction describes the major operational considerations for those deploying this mechanism.

6.1. RP Redundancy

A technique called "Anycast RP" is used within a PIM-SM domain to share an address and multicast state information between a set of RP's mainly for redundancy purposes. Typically, MSDP has been used for that as well [ANYCASTRP]. There are also other approaches, like using PIM for sharing this information [ANYPIMRP].

RP failover cannot be used with this specification without additional mechanisms or techniques such as MSDP, PIM-SM extensions, or "anycasting" (i.e., the shared-unicast model [ANYCAST]) the RP address in the IGP without state sharing (depending on the redundancy requirements, this may or may not be enough, though). However, the redundancy mechanisms are outside of the scope of this memo.

6.2. RP Deployment

As there is no need to share inter-domain state with MSDP, each DR connecting multicast sources could act as an RP without scalability concerns about setting up and maintaining MSDP sessions.

This might be particularly attractive when concerned about RP redundancy. In the case where the DR close to a major source for a group acts as the RP, a certain amount of fate-sharing properties can be obtained without using any RP failover mechanisms: if the DR goes

down, the multicast transmission may not work anymore in any case.

Along the same lines, it's may also be desirable to distribute the RP responsibilities to multiple RPs. As long as different RPs serve different groups, this is is trivial: each group could map to a different RP (or sufficiently many different RPs that the load on one RP is not a problem). However, load sharing one group faces the similar challenges as Anycast-RP.

6.3. 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 "RIID".

RIID = 0 must not be used as using it would cause ambiguity with the Subnet-Router Anycast Address [ADDRARCH].

If an administrator wishes to use an RP address that does not conform to the addressing topology but is still from the network provider's prefix (e.g., an additional loopback address assigned on a router, as described in example 1 in Section 5.1), that address can be injected into the routing system via a host route.

6.4. Use as a Substitute for BSR

With embedded-RP, use of BSR or other RP configuration mechanisms throughout the PIM domain is not necessary, as each group address specifies the RP to be used.

7. The Embedded-RP Group-to-RP Mapping Mechanism

This section specifies the group-to-RP mapping mechanism works for Embedded RP.

7.1. PIM-SM Group-to-RP Mapping

The only PIM-SM modification required is implementing this mechanism as one group-to-RP mapping method.

The implementation will have to recognize the address format and derive and use the RP address using the rules in Section 4. This information is used at least when performing Reverse Path Forwarding (RPF) lookups, when processing Join/Prune messages, or performing Register-encapsulation.

To avoid loops and inconsistancies, the group-to-RP mapping specified in this memo MUST be used for all embedded-RP groups (i.e., addresses

with prefix FF70::/12 or FFF0::/12).

It is worth noting that compared to the other group-to-RP mapping mechanisms, which can be precomputed, the embedded-RP mapping must be redone for every new IPv6 group address which would map to a different RP. For efficiency, the results may be cached in an implementation-specific manner, to avoid computation for every embedded-RP packet.

This group-to-RP mapping mechanism must be supported by the DR adjacent to the senders and any router on the path from any receiver to the RP. Further, as the switch-over to Shortest Path Tree (SPT) is also possible, it must be supported on the path between the receivers and the senders as well. It also must be supported by any router on the path from any sender to the RP -- in case the RP issues a Register-Stop and Joins the sources. So, in practice, the mechanism must be supported by all routers on any path between the RP, receivers, and senders.

7.2. Overview of the Model

This section gives a high-level, non-normative overview of how Embedded RP operates, as specified in the previous section.

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 encoded in the multicast address, irrespective of whether it is in the "local" or "remote" PIM domain.

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

1. A sender finds out a group address using an unspecified method (e.g, by contacting the administrator for group assignment or using 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 fact, all the messages go as specified in [PIM-SM] -- embedded-RP just acts as a group-to-RP mapping mechanism; instead of obtaining the address of the RP from local configuration or configuration protocols (e.g., BSR), it is derived transparently from the encoded

multicast address.

8. Scalability Analysis

Interdomain MSDP model for connecting PIM-SM domains is mostly hierarchical in configuration and deployment, but flat with regard to information distribution. The embedded-RP inter-domain model behaves as if all of the Internet was a single PIM domain, with just one RP per group. So, the inter-domain multicast becomes a flat, RP-centered topology. The scaling issues are described below.

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 (i.e., the prefix where the group address was derived from). 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 past receivers, for PIM Prunes) and the RP, in addition to (S, G) states between the receivers and senders, if SPT is used. However, the traditional ASM model also requires MSDP state to propagate everywhere in inter-domain, so the total amount of state is smaller.

The embedded-RP model is practically identical in both inter-domain and intra-domain cases to the traditional PIM-SM in intra-domain. On the other hand, PIM-SM has been deployed (in IPv4) in inter-domain using MSDP; compared to that inter-domain model, this specification simplifies the multicast routing by removing the RP for senders and receivers in foreign domains, and eliminating the MSDP information distribution.

As the address of the RP is tied to the multicast address, the RP failure management becomes more difficult, as failover or redundancy mechanisms (e.g., BSR, Anycast-RP with MSDP) cannot be used as-is. On the other hand, Anycast-RP using PIM could be used. This described briefly in Section 6.1.

The PIM-SM specification states, "Any RP address configured or learned MUST be a domain-wide reachable address". What "reachable" precisely means is not clear, even without embedded-RP. This statement cannot be proven especially with the foreign RPs as one can not even guarantee that the RP exists. Instead of configuring RPs and DRs with a manual process (configuring a non-existent RP was possible though rare), with this specification the hosts and users using multicast indirectly specify the RP themselves, lowering the

expectancy of the RP reachability. This is a relatively significant problem but not much different from the current multicast deployment: e.g., MLDv2 (S,G) joins, whether ASM or SSM, yield the same result [PIMSEC].

Being able to join/send to remote RPs raises security concerns that are considered separately, but it has an advantage too: every group has a "responsible RP" which is able to control (to some extent) who are able to send to the group.

A more extensive description and comparison of the inter-domain multicast routing models (traditional ASM with MSDP, embedded-RP, SSM) and their security properties has been described in [PIMSEC].

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 an 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. Toerless Eckert, Hugh Holbrook, and Dave Meyer provided very extensive feedback. The whole MboneD working group is also acknowledged for the continued support and comments.

10. Security Considerations

The address of the RP is encoded in the multicast address -- and thus become more visible as single points of failure. Even though this does not significantly affect the multicast routing security, it may expose the RP to other kinds of attacks. The operators are encouraged to pay special attention to securing these routers. See Section 6.1 on considerations regarding failover and Section 6.2 on placement of RPs leading to a degree of fate-sharing properties.

As any RP will have to accept PIM-SM Join/Prune/Register messages from any DR, 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 encode the address of the RP. The implementation MAY also allow manual configuration of which multicast addresses or prefixes embedding the RP could be used.

In a similar fashion, when a receiver joins to an RP, the DRs must accept similar PIM-SM messages back from RPs. However, this is not a considerable threat.

One should observe that the embedded-RP threat model is actually rather similar to SSM; both mechanisms significantly reduce the threats at the sender side. On the receiver side, the threats are somewhat comparable, as an attacker could do an MLDv2 (S,G) join towards a non-existent source, which the local RP could not block based on the MSDP information.

The implementation MUST perform at least the same address validity checks to the embedded-RP address as to one received via other means; at least fe80::/10, ::/16, and ff00::/8 should be excluded. This is particularly important as the information is derived from the untrusted source (i.e., any user in the Internet), not from the local configuration.

A more extensive description and comparison of the inter-domain multicast routing models (traditional ASM with MSDP, embedded-RP, SSM) and their security properties has been done separately in [PIMSEC].

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.

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

11.2. Informative References

[ANYCAST] Hagino, J., Ettikan, K., "An analysis of IPv6 anycast", work-in-progress, draft-ietf-ipngwg-ipv6-anycast-analysis-02.txt, June 2003.

[ANYCASTRP] Kim, D. et al, "Anycast RP mechanism using PIM and MSDP", RFC 3446, January 2003.

[ANYPIMRP] Farinacci, D., Cai, Y., "Anycast-RP using PIM", work-in-progress, draft-ietf-pim-anycast-rp-00.txt, November 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 Source Discovery Protocol (MSDP)", RFC 3618, October 2003.

[PIMSEC]   Savola, P., Lehtonen, R., Meyer, D., "PIM-SM Multicast Routing Security Issues and Enhancements", work-in-progress, draft-savola-mboned-mroutesec-00.txt, January 2004.

[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-09.txt, February 2004.

[SSM]   Holbrook, H. et al, "Source-Specific Multicast for IP", work-in-progress, draft-ietf-ssm-arch-04.txt, October 2003.

[V6MISSUES] Savola, P., "IPv6 Multicast Deployment Issues", work-in-progress, draft-savola-v6ops-multicast-issues-03.txt, February 2004.

Authors' Addresses

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

Brian Haberman
Caspian Networks
One Park Drive, Suite 300
Research Triangle Park, NC 27709
EMail: brian@innovationslab.net
Phone: +1-919-949-4828

A. Discussion about Design Tradeoffs

It has been argued that instead of allowing the operator to specify RIID, the value could be pre-determined (e.g., "1"). However, this has not been adopted, as this eliminates address assignment flexibility from the operator.

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.

The embedded-RP addressing could be used to convey other information (other than RP address) as well, for example, what should be the RPT

threshold for PIM-SM. These could be, whether feasible or not, encoded in the RP address somehow, or in the multicast group address. In any case, such modifications are beyond the scope of this memo.

For the cases where the RPs do not exist or are unreachable, or too much state is being generated to reach in a resource exhaustion DoS attack, some forms of rate-limiting or other mechanisms could be deployed to mitigate the threats while trying not to disturb the legitimate usage. However, as the threats are generic, they are considered out of scope and discussed separately in [PIMSEC].

The mechanism is not usable with Bidirectional PIM without protocol extensions, as pre-computing the Designated Forwarder is not possible.

B. Changes

[[RFC-Editor: please remove before publication]]

B.1 Changes since -01

- o Lots of editorial cleanups and some reorganization, without technical changes.
- o Remove the specification that RIID=0 SHOULD NOT be accepted, but state that they "must not" be used (implementation vs. operational wording).
- o Specify that the RP address MUST NOT be of prefixes fe80::/10, ::/16, or ff00::/8.

B.2 Changes since -00

- o Lots of editorial cleanups, or cleanups without techinical changes.
- o Reinforce the notion of Embedded RP just being a group-to-RP mapping mechanism (causing substantive rewriting in section 7); highlight the fact that precomputing the group-to-RP mapping is not possible.
- o Add (a bit) more text on RP redundancy and deployment tradeoffs wrt. RPs becoming SPoF.
- o Clarify the usability/scalability issues in section 8.
- o Clarify the security issues in Sections 8, Security Considerations and Appendix A, mainly by referring to a separate document.
- o Add a MUST that embedded-RP mappings must be honored by implementations.

Intellectual Property Statement

The IETF takes no position regarding the validity or scope of any intellectual property or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; neither does it represent that it has made any effort to identify any such rights. Information on the IETF's procedures with respect to rights in standards-track and standards-related documentation can be found in BCP-11. Copies of claims of rights made available for publication and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementors or users of this specification can be obtained from the IETF Secretariat.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights which may cover technology that may be required to practice this standard. Please address the information to the IETF Executive Director.

Full Copyright Statement

Acknowledgement

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