Network Working Group
INTERNET-DRAFT
Expiration Date: July 2003

Dino Farinacci
Procket Networks
Yiqun Cai
cisco Systems

January 15, 2003

Anycast-RP using PIM
<draft-farinacci-pim-anycast-rp-00.txt>

Status of this Memo

Copyright Notice

Abstract

This proposal allows Anycast-RP to be used inside a domain which runs PIM only. There are no other multicast protocols required to support Anycast-RP, such as MSDP, which has been used traditionally to solve this problem.

1. Introduction

Anycast-RP as described in [2] is a mechanism ISP-based backbones have used to get fast convergence when a PIM Rendezvous Point (RP) router fails. To allow receivers and sources to Rendezvous to the closest RP, the packets from a source needs to get to all RPs to find joined receivers.

This notion of receivers finding sources is the fundamental problem of source discovery which MSDP was intended to solve. However, if one would like to retain the Anycast-RP benefits from [2] with less protocol machinery, removing MSDP from the solution space is an option.

This draft extends the Register mechanism in PIM so Anycast-RP functionality can be retained without using MSDP.

2. Requirements

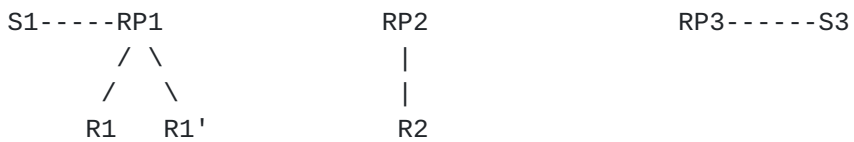
- o Each router acting as an RP MUST be configured with a loopback interface using the same (shared) IP address. This address is used to tell other routers in the PIM domain, what IP address to use for the RP address.

- o The RP address or a prefix that covers the RP address is injected into the unicast routing system inside of the domain.

- o Each RP configures all other RPs used in the Anycast-RP set. This must be consistently configured in all RPs in the set.

3. Mechanism

The following diagram illustrates a domain using 3 RPs where receivers are joining to the closest RP according to where unicast routing metrics take them and 2 sources sending packets to their respective RPs.



Assume the above scenario is completely connected where R1, R1', and R2 are receivers for a group, and S1 and S2 send to that group. Assume RP1, RP2 and RP3 are all assigned the same IP address which is used as the Anycast-RP address (let's say the IP address is RPA).

The following procedure is used when S1 starts sourcing traffic:

- o S1 sends a multicast packet.
- o The DR directly attached to S1 will form a PIM Register message to send to RP1. The IP address to use is the Anycast-RP address.
- o RP1 will receive the PIM Register message, decapsulate it, send the packet down the shared-tree to get the packet to receivers R1 and R1'.
- o RP1 is configured with RP2 and RP3's IP address. It will forward the Register message from S1's DR to both of them. RP1 will include it's own IP address as the source address for the PIM Register message.
- o RP1 sends a Register-Stop back to the DR.
- o RP1 MAY join back to the source-tree by triggering a (S1,G) Join message toward S1. However, RP1 MUST create (S1,G) state.
- o RP2 receives the Register message from RP1, decapsulates it, and also sends the packet down the shared-tree to get the packet to receiver R2.
- o RP2 sends a Register-Stop back to the RP1.
- o RP2 MAY join back to the source-tree by triggering a (S1,G) Join message toward S1. However, RP2 MUST create (S1,G) state.
- o RP3 receives the Register message from RP1, decapsulates it, but since there is no receivers joined for the group, it can discard the packet.
- o RP3 sends a Register-Stop back to the RP1.
- o RP3 creates (S1,G) state so when a receiver joins after S1 starts sending, RP3 can join quickly to the source-tree for S1.

The procedure for S3 sending follows the same as above but it is RP3 which forwards the Register originated by S3's DR to RP1 and RP2. Therefore, this example shows how sources anywhere in the domain, associated with different RPs, can reach all receivers, also associated with different RPs, in the same domain.

4. Observations and Guidelines about this Proposal

- o An RP will forward a Register only if the Register is received

from an IP address not in the Anycast-RP list (i.e. the Register came from a DR and not another RP).

- Each DR that PIM registers for a source will send the message to it's closest RP address. Therefore there are no changes to the DR logic.
- Packets flow to all receivers no matter what RP they have joined to.
- The source gets Registered to a single RP by the DR, it's the responsibility of the RP, the DR selects, to get the packet to all other RPs in the Anycast-RP set.
- Logic is changed only in the RPs. The logic change is for forwarding Register messages. Register-Stop processing is unchanged. However, an implementation MAY suppress sending Register-Stop messages in response to a Register received from an RP.
- The rate-limiting of Register and Register-Stop messages are done end-to-end. That is from DR -> RP1 -> {RP2 and RP3}. There is no need for specific rate-limiting logic between the RPs.
- When topology changes occur, the existing source-tree adjusts as it does today according to [1]. The existing shared-trees, as well, adjust as it does today according to [1].
- Physical RP changes are as fast as unicast route convergence. Retaining the benefit of [2].
- An RP that doesn't support this draft can be mixed with RPs that do support this draft. However, the non-supporter RPs should not have sources registering to it but may have receivers joined to it.
- If Null Registers are sent (Registers with an IP header and no IP payload), they MUST be replicated to all of the RPs in the Anycast-RP set so that source state remains alive for active sources.
- The number of RPs in the Anycast-RP set should remain small so the amount of non-native replication is kept to a minimum.

5. Possible Configuration Language

A possible set of commands to be used could be:

```
ip pim anycast-rp <anycast-rp-addr> <rp-addr>
```

Where:

<anycast-rp-addr> describes the Anycast-RP set for the RP which is assigned to the group range. This IP address is the address that first-hop and last-hop PIM routers use to register and join to.

<rp-addr> describes the IP address where Register messages are forwarded to. This IP address is any address assigned to the RP router not including the <anycast-rp-addr>.

Example:

From the illustration above, the configuration commands would be:

```
ip pim anycast-rp RPA RP1
ip pim anycast-rp RPA RP2
ip pim anycast-rp RPA RP3
```

Comment:

It may be useful to include the local router's IP address in the command set so the above lines can be cut-and-pasted or scripted into all the RPs in the Anycast-RP set.

But the implementation would have to be aware of it's own address and not inadvertently send a Register to itself.

6. Interaction with MSDP running in an Anycast-PIM Router

The objective of this Anycast-PIM proposal is to remove the dependence on using MSDP. This can be achieved by removing MSDP peering between the Anycast RPs. However, to advertise internal sources to routers outside of a PIM routing domain, MSDP may still be required.

In this capacity, an Anycast-RP that receives a Register message should process it as described in this specification as well as how to process a Register message as described in [2].

A source should be considered internal to the domain, when it is discovered by an Anycast-RP through a received Register message. Regardless, if the Register message was sent by a DR, another Anycast-RP member, or the router itself.

It is recommended that an Anycast-RP set defined in this specification should not be mixed with MSDP peering amoung the members. In some cases, the source discovery will work but it may not

be obvious to the implementations what sources are local to the domain and which are not. This may affect external MSDP advertisement of internal sources.

7. Acknowledgments

The authors would like to thank Yiqun Cai and Dino Farinacci for prototyping this draft in the cisco IOS implementation and Procket implementation, respectively.

The authors would like to thank John Zwiebel for doing interoperability testing of the two prototype implementations.

And finally, the authors would like to thank Greg Shepard for comments on the draft.

8. Author Information

Dino Farinacci
Procket Networks
dino@procket.com

Yiqun Cai
cisco Systems
ycai@cisco.com

9. References

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

[2] Kim, Meyer, Kilmer, Farinacci, "Anycast RP mechanism using PIM and MSDP", Internet Draft draft-ietf-mboned-anycast-rp-08.txt, May 2001.