Anycast RP using PIM (Nexus)

TAC   **Document ID: 115011**

Contributed by Ruchir Jain, Cisco TAC Engineer.
Nov 07, 2012

Contents

Introduction

You can have redundancy for Rendezvous Point (RP) by running protocols such as auto−RP and Bootstrap. However, in case of failure their convergence is not that fast. There is a concept of Anycast RP where the same IP address (rp−address) is configured on two or more routers that you want to serve as RP. Then, advertise this IP in IGP. Other routers will select any of these routers based on best path to the rp−address. In case of a failure the convergence is the same as Interior Gateway Protocol (IGP).

With this concept a problem arises. Information needs to be synched between the different RPs because few senders and receivers might join router 1 serving as a RP, and few might join router 2 serving as a RP. These routers would not have complete information of all the sources and multicast would be broken. The solution to this problem is having a mechanism to sync information with respect to sources between all the routers which are acting as a RP. There are two protocols which can serve this purpose:

- Multicast Source Discovery Protocol (MSDP)
- PIM

MSDP has been around for awhile. Source Active messages are sent to other routers whenever a source registers to a RP. There is an enhancement to PIM which is detailed in this document. Currently, this enhancement is only available for Nexus.

Prerequisites

Requirements

Cisco recommends that you have knowledge of these topics:

- Anycast RPs
- PIM (Nexus)

Components Used

This document is not restricted to specific software and hardware versions.

Conventions

Refer to Cisco Technical Tips Conventions for more information on document conventions.

Registration Process

This is an example topology:

Sender(172.16.1.1)–––––––––––(9/3)Nexus–1(9/2)–––––––––––––––(9/2)Nexus–2

Relevant running configurations

Nexus 1 relevant configuration:

```
ip pim rp-address 10.1.1.1 group-list 224.0.0.0/4
ip pim anycast-rp 10.1.1.1 192.168.1.1
ip pim anycast-rp 10.1.1.1 192.168.2.2

interface loopback1
  ip address 192.168.1.1/32
  ip router ospf 1 area 0.0.0.0
  ip pim sparse-mode

interface loopback7
  ip address 10.1.1.1/32
  ip router ospf 1 area 0.0.0.0
  ip pim sparse-mode

interface Ethernet9/2
  ip address 10.7.7.1/24
  ip router ospf 1 area 0.0.0.0
  ip pim sparse-mode

interface Ethernet9/3
  ip address 172.16.1.2/24
  ip router ospf 1 area 0.0.0.0
  ip pim sparse-mode
```

Nexus 2 relevant configuration:

```
ip pim rp-address 10.1.1.1 group-list 224.0.0.0/4
ip pim ssm range 232.0.0.0/8
ip pim anycast-rp 10.1.1.1 192.168.1.1
ip pim anycast-rp 10.1.1.1 192.168.2.2

interface loopback1
  ip address 192.168.2.2/32
  ip router ospf 1 area 0.0.0.0
  ip pim sparse-mode

interface loopback7
  ip address 10.1.1.1/32
  ip router ospf 1 area 0.0.0.0
  ip pim sparse-mode

interface Ethernet9/2
  ip address 10.7.7.2/24
  ip router ospf 1 area 0.0.0.0
  ip pim sparse-mode
```

There are two Nexus boxes: Nexus−1 and Nexus−2. Both will be used as the RP. The RP address should be 10.1.1.1. Loopback 7 is on both Nexus boxes that have this IP configured. This loopback is then advertised in Open Shortest Path First (OSPF) so different routers in the network will either reach Nexus−1 or Nexus−2 for the RP. This depends on the best path metric.

On both Nexuses, 10.1.1.1 is defined to be the RP using this command:

ip pim rp−address 10.1.1.1 group−list 224.0.0.0/4

Now you need to define something called the RP set. This is the set of all the routers which would act as RP. You need to have a loopback on each prospective RP router, which is different than the loopback that is being used as the RP address. In this example, loopback 1 is on both Nexuses that have IP address 192.168.1.1/32 and 192.168.2.2/32, respectively. This loopback 1 is used to define the RP set. The command to do the same is:

ip pim anycast−rp *<rp−address> <ip−address−of−prospective−RP>*

The commands for both Nexuses for the setup are:

- **ip pim anycast−rp 10.1.1.1 192.168.1.1**
- **ip pim anycast−rp 10.1.1.1 192.168.2.2**

One thing to note here is that you need to define your own IP, as well as in the RP set. Therefore, both these commands need to be put on both Nexus boxes.

Once the RP set is defined, this is the output you see for RP mapping:

```
Nexus-1# show ip pim rp
PIM RP Status Information for VRF "default"
BSR disabled
Auto-RP disabled
BSR RP Candidate policy: None
BSR RP policy: None
Auto-RP Announce policy: None
Auto-RP Discovery policy: None

Anycast-RP 10.1.1.1 members:
  192.168.1.1*  192.168.2.2

RP: 10.1.1.1*, (0), uptime: 00:00:58, expires: never,
  priority: 0, RP-source: (local), group ranges:
      224.0.0.0/4


Nexus-2# show ip pim rp
PIM RP Status Information for VRF "default"
BSR disabled
Auto-RP disabled
BSR RP Candidate policy: None
BSR RP policy: None
Auto-RP Announce policy: None
Auto-RP Discovery policy: None

Anycast-RP 10.1.1.1 members:
  192.168.1.1  192.168.2.2*

RP: 10.1.1.1*, (0), uptime: 02:46:54, expires: never,
  priority: 0, RP-source: (local), group ranges:
      224.0.0.0/4
```

For example, you receive a register message on one of the routers which is part of the RP set. This router will add S,G for this source in its table. Also, the router will send a PIM register message to all the other members of RP set. The source IP of this register message would be the address of this router which is in the RP set, and destination would be address of each router in the RP set.

In this example, when source 172.16.1.1 sends multicast packet destined to 239.1.1.1 to Nexus−1, Nexus−1 first registers this source as the RP and sends a register message to Nexus−2 for source 172.16.1.1 and group 239.1.1.1. When Nexus−2 receives this register it adds S,G entry for 172.16.1.1, 239.1.1.1 in the mroute table. Now both RP Nexus−1 and Nexus−2 know about this source. The register message that was sent from Nexus−1 to Nexus−2 had source IP 192.168.1.1 and destined IP 192.168.2.2 as defined in the RP set.

Mroute entries on both Nexus boxes after the registration process:

```
Nexus-1# sh ip mroute 239.1.1.1
IP Multicast Routing Table for VRF "default"

(172.16.1.1/32, 239.1.1.1/32), uptime: 00:00:22, ip pim
  Incoming interface: Ethernet9/3, RPF nbr: 172.16.1.1, internal
  Outgoing interface list: (count: 0)


Nexus-2# sh ip mroute 239.1.1.1
IP Multicast Routing Table for VRF "default"

(172.16.1.1/32, 239.1.1.1/32), uptime: 00:00:17, pim ip
  Incoming interface: Ethernet9/2, RPF nbr: 10.7.7.1, internal
  Outgoing interface list: (count: 0)
```

Packet dump of such register message taken on Nexus−2:

```
Ethernet II, Src: d8:67:d9:04:5f:41 (d8:67:d9:04:5f:41), Dst: d8:67:d9:02:3b:41
(d8:67:d9:02:3b:41)
    Destination: d8:67:d9:02:3b:41 (d8:67:d9:02:3b:41)
        Address: d8:67:d9:02:3b:41 (d8:67:d9:02:3b:41)
        .... ...0 .... .... .... .... = IG bit: Individual address (unicast)
        .... ..0. .... .... .... .... = LG bit: Globally unique address (factory
 default)
    Source: d8:67:d9:04:5f:41 (d8:67:d9:04:5f:41)
        Address: d8:67:d9:04:5f:41 (d8:67:d9:04:5f:41)
        .... ...0 .... .... .... .... = IG bit: Individual address (unicast)
        .... ..0. .... .... .... .... = LG bit: Globally unique address (factory
 default)
    Type: IP (0x0800)
Internet Protocol, Src: 192.168.1.1 (192.168.1.1), Dst: 192.168.2.2 (192.168.2.2)
    Version: 4
    Header length: 20 bytes
    Differentiated Services Field: 0x00 (DSCP 0x00: Default; ECN: 0x00)
        0000 00.. = Differentiated Services Codepoint: Default (0x00)
        .... ..0. = ECN-Capable Transport (ECT): 0
        .... ...0 = ECN-CE: 0
    Total Length: 128
    Identification: 0x1b27 (6951)
    Flags: 0x00
        0.. = Reserved bit: Not Set
        .0. = Don't fragment: Not Set
        ..0 = More fragments: Not Set
    Fragment offset: 0
    Time to live: 254
    Protocol: PIM (0x67)
    Header checksum: 0x9aea [correct]
        [Good: True]
        [Bad : False]
    Source: 192.168.1.1 (192.168.1.1)
```

```
              Destination: 192.168.2.2 (192.168.2.2)
       Protocol Independent Multicast
           Version: 2
           Type: Register (1)
           Checksum: 0xdeff [correct]
           PIM parameters
              Flags: 0x00000000
                  0... .... .... .... .... .... .... .... = Not border
                  .0.. .... .... .... .... .... .... .... = Not Null-Register
              Internet Protocol, Src: 172.16.1.1 (172.16.1.1), Dst: 239.1.1.1 (239.1.1.1)
                  Version: 4
                  Header length: 20 bytes
                  Differentiated Services Field: 0x00 (DSCP 0x00: Default; ECN: 0x00)
                      0000 00.. = Differentiated Services Codepoint: Default (0x00)
                      .... ..0. = ECN-Capable Transport (ECT): 0
                      .... ...0 = ECN-CE: 0
                  Total Length: 100
                  Identification: 0x0000 (0)
                  Flags: 0x00
                      0.. = Reserved bit: Not Set
                      .0. = Don't fragment: Not Set
                      ..0 = More fragments: Not Set
                  Fragment offset: 0
                  Time to live: 254
                  Protocol: ICMP (0x01)
                  Header checksum: 0xa294 [correct]
                      [Good: True]
                      [Bad : False]
                  Source: 172.16.1.1 (172.16.1.1)
                  Destination: 239.1.1.1 (239.1.1.1)
              Internet Control Message Protocol
                  Type: 8 (Echo (ping) request)
                  Code: 0 ()
                  Checksum: 0x9484 [correct]
                  Identifier: 0x0000
                  Sequence number: 0 (0x0000)
                  Data (72 bytes)

    0000  00 00 00 00 00 71 e9 54 ab cd ab cd ab cd ab cd   .....q.T........
    0010  ab cd ab cd ab cd ab cd ab cd ab cd ab cd ab cd   ................
    0020  ab cd ab cd ab cd ab cd ab cd ab cd ab cd ab cd   ................
    0030  ab cd ab cd ab cd ab cd ab cd ab cd ab cd ab cd   ................
    0040  ab cd ab cd ab cd ab cd                           ........
                      Data: 000000000071E954ABCDABCDABCDABCDABCDABCDABCDABCD...
                      [Length: 72]
```

You can use this filter to capture PIM packets on inband of the Nexus:

ethanalyzer local interface inband capture−filter "ip proto 103" limit−captured−frames 0 write logflash: pim.pcap

This will start capturing PIM packets indefinitely until cntrl+c is pressed. The packets are displayed on the screen as well as would be written to the pim.pcap file in logflash. You can also include the IP address to capture packets from a particular PIM neighbor ("ip proto 103 and host <ip_address>").

Relevant debugs

These are the two debugs that are helpful to check the registration process on Nexus:

- **debug ip pim data−register send**
- **debug ip pim data−register receive**

Nexus–1

```
2012 Sep 20 12:25:52.537472 pim: [6405] (default-base) Received Register from 172.16.1.2 fd
2012 Sep 20 12:25:52.537656 pim: [6405] (default-base) Create route for (172.16.1.1/32, 235
2012 Sep 20 12:25:52.537716 pim: [6405] (default-base) Add route (172.16.1.1/32, 239.1.1.1
2012 Sep 20 12:25:52.537745 pim: [6405] (default-base) RP 10.1.1.1 is an Anycast-RP
2012 Sep 20 12:25:52.537771 pim: [6405] (default-base) Forward Register to Anycast-RP membe
2012 Sep 20 12:25:52.537825 pim: [6405] (default-base) Send Register-Stop to 172.16.1.2 for
2012 Sep 20 12:25:52.538591 pim: [6405] (default-base) Received Register-Stop from 10.1.1.
```

Nexus–2

```
2012 Sep 20 12:26:36.367862 pim: [7189] (default-base) Received Register from 192.168.1.1
2012 Sep 20 12:26:36.368093 pim: [7189] (default-base) Create route for (172.16.1.1/32, 235
2012 Sep 20 12:26:36.368175 pim: [7189] (default-base) Add route (172.16.1.1/32, 239.1.1.1
2012 Sep 20 12:26:36.368223 pim: [7189] (default-base) RP 10.1.1.1 is an Anycast-RP
2012 Sep 20 12:26:36.368240 pim: [7189] (default-base) Register received from Anycast-RP me
```

This RFC provides more information on the PIM registration process: http://tools.ietf.org/rfc/rfc4610.txt

Also, refer to Configuring PIM and PIM6 for more information.

Related Information

- **Technical Support & Documentation – Cisco Systems**

Updated: Nov 07, 2012   Document ID: 115011