Protocol Independent Multicast (PIM) in SONiC using GNS3

Nov 23, 2023

Table of Contents

Introduction

The document provides a comprehensive guide on Protocol Independent Multicast configurations in SONiC.

Intended Audience

This document is tailored for data center experts and network engineers interested in configuring **PIM-SM** in SONiC. It is designed for individuals with a solid understanding of networking principles. Whether you are a network engineer, operator, or vendor, this document aims to provide you with practical, step-by-step guidance, and best practices for deploying, configuring, and setting up BFD for SONiC devices using the GNS3 network simulation tool.

Multicast Forwarding

Multicast forwarding is the process of routing and transmitting multicast traffic within a network. Multicast forwarding enables the efficient distribution of data from one sender to multiple recipients who have expressed interest in receiving the multicast content.

In a multicast communication model, the sender (the source) sends data to a multicast group address, and only those network devices (receivers) that have joined that multicast group will receive the data. Multicast forwarding mechanisms are responsible for ensuring that the data is delivered to all interested recipients while minimising network traffic and resource usage.

There are different multicast forwarding methods and protocols, including PIM (Protocol Independent Multicast), IGMP (Internet Group Management Protocol), and various routing algorithms, which determine the path for multicast data to reach its intended recipients.

Multicast uses RPF (**Reverse Path Forwarding)** to prevent flooding/loops within the network. The RPF check uses the routing table to check against the source IP within the packet. If the packet comes in on an interface that is specified against the source of the multicast stream, the RPF check succeeds, otherwise, the RPF check fails and the packet is dropped.

Protocol Independent Multicast (PIM)

PIM, or "Protocol Independent Multicast," is a multicast routing protocol for efficient one-to-many communication in IP networks. It enables the transmission of data from one source to multiple receivers, distinguishing it from unicast (one-to-one) and broadcast (one-to-all) communication methods.

PIM is called "Protocol Independent" because it can work with various unicast routing protocols (e.g., OSPF, BGP, or RIP) without being tied to a specific one. PIM is designed to determine the best path for multicast traffic to flow through a network, making it more efficient than traditional flooding-based multicast techniques.

PIM Components

There are two key components: state and the **multicast distribution tree**.

State

Refers to the information that PIM routers maintain about the multicast distribution tree and the multicast groups they are forwarding traffic for. This state information is used to ensure the efficient routing and forwarding of multicast traffic within a network. The state information is dynamic and is updated as the network topology and multicast group membership change. PIM routers use this state information to make decisions about the forwarding of multicast traffic, such as whether to prune or forward traffic on specific interfaces to optimize the multicast delivery.

There are two primary types of state in PIM:

1. **Multicast Group State:** This type of state includes information about the multicast groups that have active sources and receivers. PIM routers keep track of which multicast groups have members and the interfaces through which multicast traffic should be forwarded to reach those group members. A multicast group state is maintained to optimize the distribution of multicast traffic to the appropriate recipients.
2. **Rendezvous Point (RP) State:** In PIM-SM (Protocol Independent Multicast - Sparse Mode), which is a common PIM variant, there is the concept of a Rendezvous Point (RP). RPs are used to facilitate the delivery of multicast traffic in a scalable manner. RPs maintain state information about which multicast groups they are responsible for and the sources that are active for those groups.

Distribution Tree

A distribution tree is a network structure or path that is used to route and deliver multicast traffic from the source (sender) to multiple receivers in a multicast group. The purpose of the distribution tree is to optimize the transmission of multicast data so that it reaches all interested recipients while minimizing network resource usage and redundancy.

There are typically two types of distribution trees in multicast routing:

1. Shared Tree (Rendezvous Point Tree - RP-Tree)

In PIM-SM (Sparse Mode), one common distribution tree is the shared tree. All receivers interested in a specific multicast group use this tree to receive data. The shared tree typically passes through a designated Rendezvous Point (RP) as shown in Figure 1, and multicast data is forwarded from the source to the RP and then distributed to the group members.

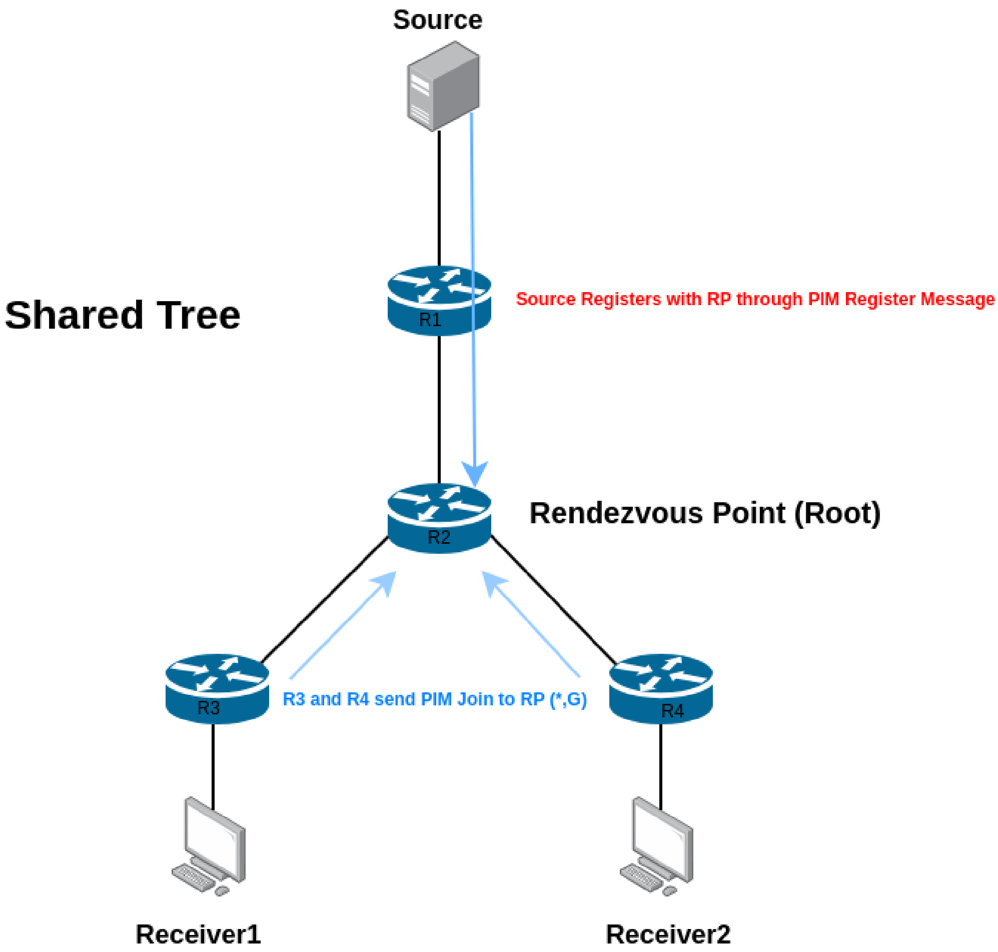


Figure 1: Shared Tree

2. Source-Specific Tree (SPT - Shortest Path Tree)

Figure 2 shows that Source-specific trees are created when PIM routers dynamically build a tree directly from the source to the receivers for a specific (S, G) pair. This is often more efficient than the shared tree because it minimizes the path length and reduces the chance of unnecessary data replication.

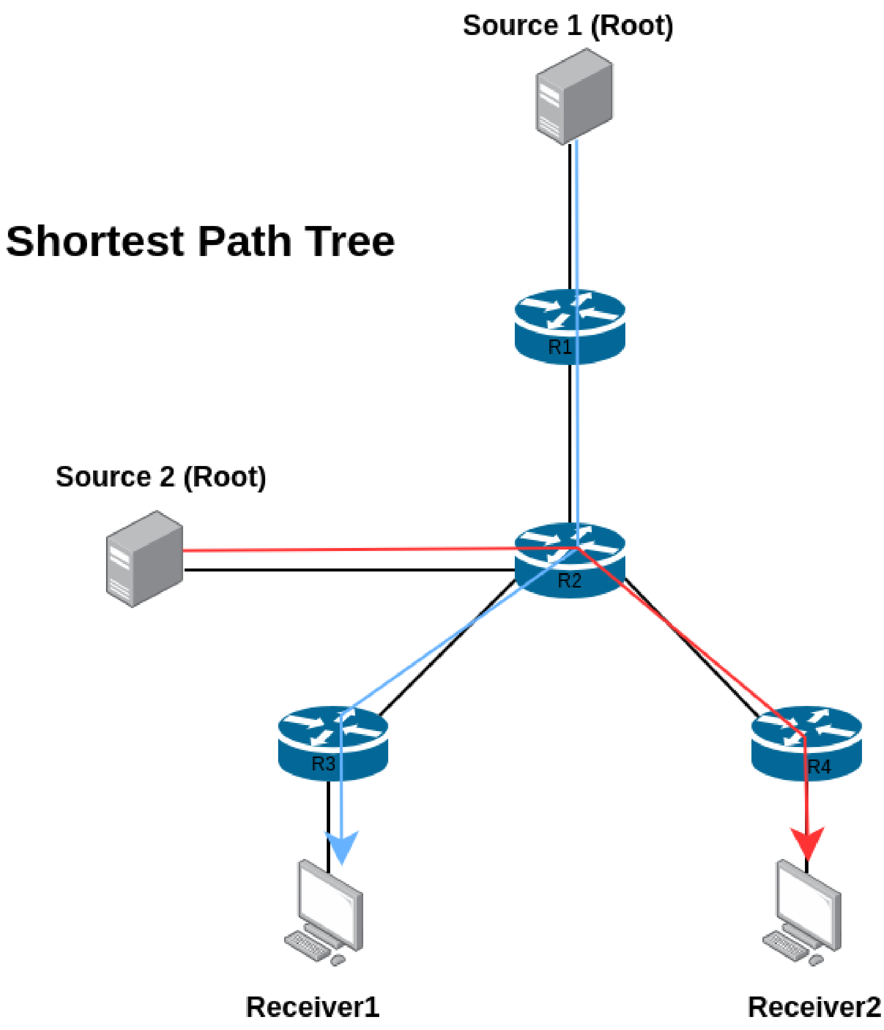


Figure 2: Shortest path Tree

The choice between using a shared tree or a source-specific tree depends on the multicast routing protocol and specific requirements of the multicast group. The distribution tree helps multicast routers determine where to forward multicast traffic and optimizes its delivery to the appropriate recipients.

PIM Multicast Protocols and Messages

Currently, there are five PIM operating modes which are:

1. PIM Dense Mode (PIM-DM)
2. PIM Sparse Mode (PIM-SM)
3. PIM Sparse Dense Mode
4. PIM Source-Specific Multicast (PIM-SSM)
5. PIM Bidirectional Mode (Bidir-PIM)

PIM Key Terms

1. **PIM Join Message** – A request sent by a router towards the RP.
2. **Internet Group Management Protocol (IGMP)** – The protocol receivers use to connect to multicast groups and begin receiving traffic from them.
3. **Prune Message** – Message sent to RP to inform other routers in the path that they can disable multicast traffic forwarding for a particular group.
4. **RPF or Reverse Path Forwarding** – PIM will only accept multicast packets on an interface that is used to reach the source. RPF failure is when multicast packets are dropped on an interface that doesn't reach the source.
5. **RPF Neighbor** – The PIM neighbor on the RPF interface.
6. **Upstream Router** – A router that receives multicast traffic from.
7. **Downstream Router** – A router that forwards multicast traffic.

PIM Messages

Message Type	Destination	PIM Protocol
Hello	All PIM routers / 224.0.0.13	PIM-SM, PIM-DM
Register	RP address / unicast	PIM-SM
Register stop	First hop router / unicast	PIM-SM
Join/prune	All PIM routers / 224.0.0.13	PIM-SM
Bootstrap	All PIM routers / 224.0.0.13	PIM-SM
Assert	All PIM routers / 224.0.0.13	PIM-SM, PIM-DM
Candidate RP advertisement	Bootstrap router (BSR) address / unicast	PIM-SM
State refresh	All PIM routers / 224.0.0.13	PIM-DM
DF election	All PIM routers / 224.0.0.13	

PIM Sparse Mode (PIM SM)

(PIM-SM) is a multicast routing protocol designed on the assumption that recipients for any particular multicast group will be sparsely distributed throughout the network. In other words, it is assumed that most subnets in the network will not want any given multicast packet. To receive multicast data, routers must explicitly tell their upstream neighbors about their interest in particular groups and sources. Routers use PIM Join and Prune messages to join and leave multicast distribution trees.

PIM-SM by default uses shared trees, which are multicast distribution trees rooted at some selected node (in PIM, this router is called the Rendezvous Point, or RP) and is used by all sources sending to the multicast group. To send to the RP, sources must encapsulate data in PIM control messages and send it by unicast to the RP. This is done by the source's Designated Router (DR), a router on the source's local network. A single DR is elected from all PIM routers on a network so that unnecessary control messages are not sent.

One of the important requirements of PIM Sparse Mode, and Bi-directional PIM, is the ability to discover the address of an RP for a multicast group using a shared tree. Various RP discovery mechanisms are used, including static configuration, Bootstrap Router, Auto-RP, Anycast RP, and Embedded RP.

PIM-SM also supports the use of source-based trees, in which a separate multicast distribution tree is built for each source sending data to a multicast group. Each tree is rooted at a router adjacent to the source, and sources send data directly to the root of the tree. Source-based trees enable the use of Source-Specific Multicast (SSM), which allows hosts to specify the source from which they wish to receive data, as well as the multicast group they wish to join. With SSM, a host identifies a multicast data stream with a source and group address pair (S,G), rather than by group address alone (*,G). PIM-SM is a soft-state protocol. That is, all state is timed out a while after receiving the control message that instantiated it. To keep the state alive, all PIM Join messages are periodically retransmitted.

Below are some important terms that are used in PIM SM.

RP (Rendezvous Point)

When configuring PIM-SM on a network, at least one router must be designated as a Rendezvous Point (RP). An RP is a meeting point and is required in the shared tree to help the source information passed to the receiver. An RP performs a critical function: it establishes a common reference point from which multicast trees are grown.

FHR (First Hop Router)

The First Hop Router is the router closest to the source of the multicast traffic. It is responsible for forwarding multicast data to the RP.

LHR (Last Hop Router)

The Last Hop Router is the router closest to the receivers of the multicast traffic. It is responsible for forwarding multicast data to the end hosts.

(*,G) State

The (,G) state represents the shared tree state in PIM-SM. It indicates that a router is interested in receiving traffic for a particular multicast group (G) from any source (). This state is created toward the RP, and multicast traffic flows along the shared tree until it reaches the Last Hop Router (LHR).

(S,G) State

The (S,G) state represents the source-specific tree state in PIM-SM. It indicates that a router is interested in receiving traffic for a particular multicast group (G) from a specific source (S). This state is created toward the source, and multicast traffic flows along the source-specific tree.

PIM Dense Mode (PIM DM)

PIM Dense Mode (PIM-DM) is a multicast routing protocol designed with the opposite assumption to PIM-SM, namely that the receivers for any multicast group are distributed densely throughout the network. That is, it is assumed that most (or at least many) subnets in the network will want any given multicast packet. Multicast data is initially sent to all hosts in the network. Routers that do not have any interested hosts then send PIM Prune messages to remove themselves from the tree.

When a source first starts sending data, each router on the source's LAN receives the data and forwards it to all its PIM neighbors and to all links with directly attached receivers for the data. Each router that receives a forwarded packet also forwards it likewise, but only after checking that the packet arrived on its upstream interface. If not, the packet is dropped. This mechanism prevents forwarding loops from occurring. In this way, the data is flooded to all parts of the network.

Some routers will have no need for the data, either for directly connected receivers or for other PIM neighbors. These routers respond to receipt of the data by sending a PIM Prune message upstream, which instantiates the Prune state in the upstream router, causing it to stop forwarding the data to its downstream neighbor. In turn, this may cause the upstream router to have no need for the data, triggering it to send a Prune message to its upstream neighbor. This 'broadcast and prune' behavior means that eventually the data is only sent to those parts of the network that require it.

Eventually, the Prune state at each router will time out, and data will begin to flow back into the parts of the network that were previously pruned. This will trigger further Prune messages to be sent, and the Prune state will be instantiated once more.

PIM-DM only uses source-based trees. As a result, it does not use RPs, which makes it simpler than PIM-SM to implement and deploy. It is an efficient protocol when most receivers are interested in the multicast data, but does not scale well across larger domains in which most receivers are not interested in the data.

PIM Sparse Dense Mode

Sparse-dense mode, as the name implies, allows the interface to operate on a per-group basis in either sparse or dense mode. A group specified as dense is not mapped to an RP. Instead, data packets destined for that group are forwarded by means of PIM dense-mode rules. A group specified as sparse is mapped to an RP, and data packets are forwarded by means of PIM sparse-mode rules.

In PIM-SDM, routers use a combination of shared trees (similar to PIM-SM) and source-based trees (similar to PIM-DM) to forward multicast traffic. The choice between shared trees and source-based trees is determined by the density of group members in a particular portion of the network.

Key features of PIM Sparse-Dense Mode

Shared Trees (Sparse Mode)

Routers use shared trees (RP-based trees) to forward multicast traffic for groups with sparse membership. This is similar to PIM-SM, where an RP is used as a rendezvous point for sources and receivers.

Source Trees (Dense Mode)

Routers use source trees (shortest path trees) to forward multicast traffic for groups with dense membership. This is similar to PIM-DM, where multicast data is initially flooded to all parts of the network, and routers prune themselves if they have no interested receivers.

Dynamic Adaptation

PIM-SDM allows routers to dynamically switch between shared trees and source trees based on the density of group membership in different parts of the network. This adaptability helps optimize multicast traffic forwarding for varying scenarios.

PIM Source-Specific Multicast (PIM-SSM)

PIM Source Specific Multicast is a multicast method in which, receivers receive multicast traffic from a specific multicast source as its name implies. In other words, for Source-Specific Multicast, source discovery responsibility shifts from multicast network to multicast receiver.

PIM-SSM is based on PIM-SM but with some differences. PIM messages and PIM timers remain unchanged for PIM Source-Specific Multicast. But this time Randevous-Point-related jobs will no longer be used. Also, there is no need for Auto-RP and Bootstrap anymore.

In PIM-SSM, multicast receivers can request traffic from a specific source within the multicast group. In other ASM (Any Source Multicast) modes, this was not possible. In ASM, any source can be accepted by the receiver as its name implies.

PIM-SSM (Source-Specific Multicast) requires IGMPv3. The relationship between multicast receivers and the source is determined by this last version of IGMP, IGMPv3. This is a requirement for Source-Specific Multicast. PIM-SSM is based on PIM Sparse Mode so, the Source-Specific Multicast mechanism uses Neighbor Discovery, DR election, and SPT Building like PIM Sparse Mode.

In this method, there is no Shared Tree, there is no Rendezvous Point (RP). Only the Shortest Path Tree is built from the receiver to the multicast source. Because, the router which receives IGMPv3 join request, knows how to reach that specific multicast source without getting help from a Randevous-Point (RP).

Bi-directional PIM (Bidir PIM)

Bi-directional PIM (BIDIR-PIM) is a third PIM protocol, based on PIM-SM. The main way BIDIR-PIM differs from PIM-SM is in the method used to send data from a source to the RP. Whereas in PIM-SM data is sent using either encapsulation or a source-based tree, in BIDIR-PIM the data flows to the RP along the shared tree, which is bi-directional - data flows in both directions along any given branch.

BIDIR-PIM's major differences from PIM-SM are as follows.

- There are no source-based trees, and in fact no (S, G) state at all. Therefore there is no option for routers to switch from a shared tree to a source-based tree, and Source-Specific Multicast is not supported.
- To avoid forwarding loops, for each RP one router on each link is elected the Designated Forwarder (DF). This is done at RP discovery time using the DF election message.
- There is no concept of a Designated Router.
- No encapsulation is used.
- The forwarding rules are much simpler than in PIM-SM, and there are no data-driven events in the control plane at all.

The main advantage of BIDIR-PIM is that it scales very well when there are many sources for each group. However, the lack of source-based trees means that traffic is forced to remain on the possibly inefficient shared tree.

Network Topology

In the configured PIM-SM topology shown in Figure 3, consisting of six routers including the First Hop Router (FHR), Rendezvous Point (RP), Last Hop Routers (LHR-1 and LHR-2), and Multicast Receivers (Receiver-1 and Receiver-2), the multicast traffic originates from a specific source. Both multicast receivers Receiver-1 and Receiver-2 are Cisco devices on which routing was disabled and used as end hosts. The FHR is responsible for initiating the multicast and sending PIM Register messages to the designated RP. The RP, upon receiving the PIM Register message, checks if there are any receivers interested in the specified multicast group. If no receivers are present, the RP will not forward the traffic. However, if there are interested receivers, the RP forwards the multicast traffic to the appropriate LHR's, and then LHR's will forward this multicast traffic to the interested receivers. The LHR's configured to receive and forward multicast traffic, and also listen for IGMP Join messages from the receivers. Receiver-1 and Receiver-2 explicitly join specific multicast groups (224.1.1.1, 224.5.5.5) and send IGMP Join messages to their respective Last Hop Routers. This explicit joining ensures that the LHR's know which receivers are interested in the multicast traffic, allowing for optimized and efficient delivery in the PIM-SM environment.

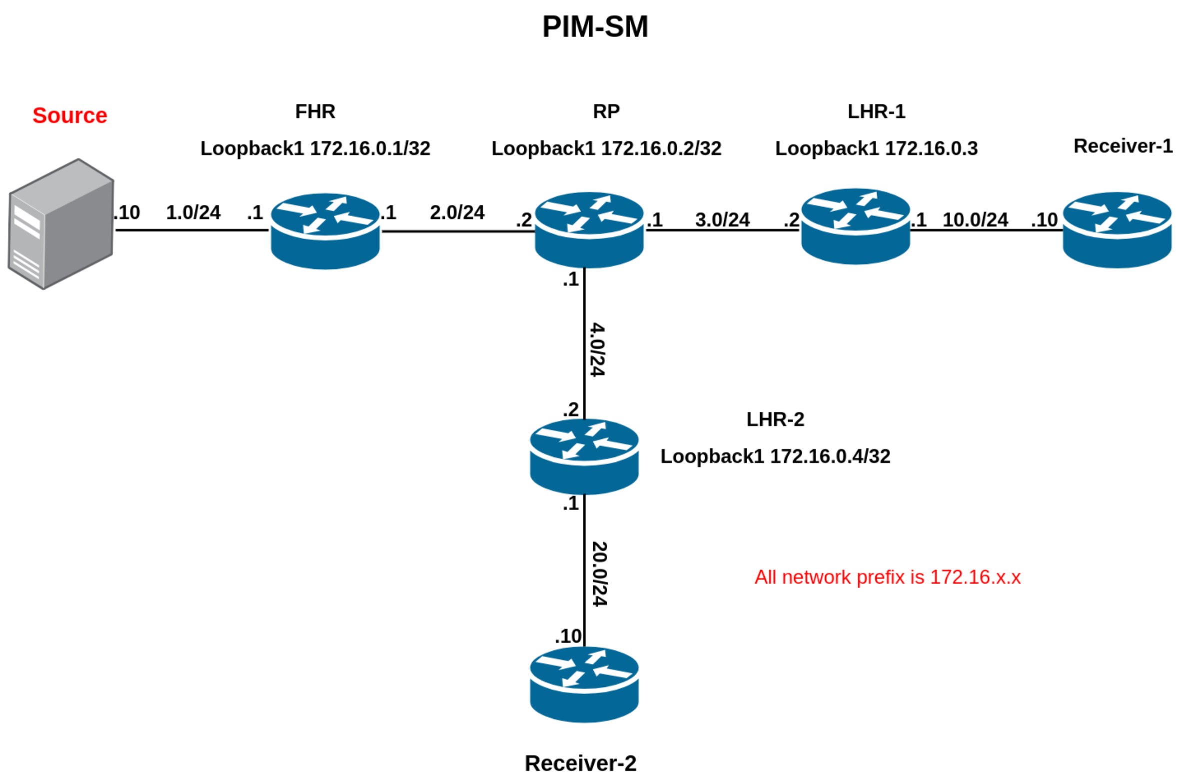


Figure 3:PIM Topology

PIM-SM Configurations

OSPF

OSPF Configurations for the FHR are shown below. All other routers in topology can be configured in a similar way.

FHR

Step 1. Configure IPs on interfaces.

```
admin@sonic:~$ sudo config interface ip rem Ethernet0 10.0.0.0/31
admin@sonic:~$ sudo config interface ip rem Ethernet4 10.0.0.2/31
admin@sonic:~$ sudo config interface ip add Ethernet0 172.16.1.1/24
admin@sonic:~$ sudo config interface ip add Ethernet4 172.16.1.10/24
admin@sonic:~$ sudo config interface ip add Lopback1 172.16.1.0/32
```

Step 2. Configure OSPF

```
sonic# configure
sonic(config)# router ospf
sonic(config-router)# network 172.16.1.0/24 area 0
sonic(config-router)# network 172.16.0.1/24 area 0
sonic(config-router)# exit
sonic(config)# exit
sonic# write
```

Step 3. Enable multicast routing on interfaces.

```
sonic# configure
sonic(config)# interface Ehernet0
sonic(config-if)# ip pim
sonic(config-if)# exit
sonic(config)# interface Ehernet4
sonic(config-if)# ip pim
sonic(config-if)# exit
sonic(config)# exit
sonic# write
```

Step 4. Configure rp address on FHR.

```
sonic# configure
sonic(config)# ip rp 172.16.0.2
sonic(config)# exit
sonic# write
```

LHR

Configurations of LHR's are the same as other routers except enabling IGMP on interfaces facing Receivers to process IGMP requests i.e. igmp join.

```
sonic# confiure
sonic(config)# interface Etherent0
sonic(config-if)# ip igmp version 3
sonic(config-if)# exit
sonic(config)# exit
sonic# write
```

Receiver

Initiate the IGMP join process by having Receiver-1 and Receiver-2 send IGMP join messages to their corresponding LHR's, expressing their interest in joining a specific multicast group. This ensures that the LHR's know the receivers' multicast group preferences and can appropriately forward the multicast traffic to the intended recipients in the PIM-SM configuration.

```
sonic# confiure
sonic(config)# interface Etherent0
sonic(config-if)# ip igmp join 224.1.1.1
sonic(config-if)# exit
sonic(config)# exit
sonic# write
```

Rendezvous Point

Step 1. Enable multicast routing on interfaces.

```
sonic# configure
sonic(config)# interface Ehernet0
sonic(config-if)# ip pim
sonic(config-if)# exit
sonic(config)# interface Ehernet4
sonic(config-if)# ip pim
sonic(config-if)# exit
sonic(config)# interface Ehernet8
sonic(config-if)# ip pim
sonic(config-if)# exit
sonic(config)# exit
sonic# write
```

Step 2. Configure rp address on RP.

```
sonic# configure
sonic(config)# ip rp 172.16.0.2
sonic(config)# exit
sonic# write
```

Results

In the network topology shown in Figure 3, OSPF was configured across all devices within Area 0 to ensure efficient communication and routing. PIM-SM (Protocol Independent Multicast - Sparse Mode) was configured, emphasizing a multicast approach for optimized resource utilization. A specific device served as the RP (Rendezvous Point), managing multicast group memberships. The experiment incorporated a multicast flow with one device as the source, generating traffic for distribution, and two strategically placed receivers explicitly joining the 224.1.1.1 multicast group. Two switches functioned as Last Hop Routers (LHRs) near the receivers, and one as the First Hop Router (FHR) near the source, allowing examination of multicast traffic traversal.

- First, OSPF neighborship among all switches was configured, confirming that they communicate and share routing information effectively. re

```
sonic# show ip ospf
                   OSPF Routing Process, Router ID: 172.16.4.1
                   Supports only single TOS (TOS0) routes
                   This implementation conforms to RFC2328
                   RFC1583Compatibility flag is disabled
                   OpaqueCapability flag is disabled
                   Initial SPF scheduling delay 0 millisec(s)
                   Minimum hold time between consecutive SPFs 50
                  millisec(s)
                   Maximum hold time between consecutive SPFs 5000
                  millisec(s)
                   Hold time multiplier is currently 1
                   SPF algorithm last executed 2508480d00h00m ago
                   Last SPF duration 528 usecs
                   SPF timer is inactive
                   LSA minimum interval 5000 msecs
                   LSA minimum arrival 1000 msecs
                   Write Multiplier set to 20
                   Refresh timer 10 secs
```

- Check multicast interfaces. These interfaces are capable of sending and receiving multicast traffic. Multicast routing should be enabled on all interfaces in the network topology.

```
sonic# show ip pim interfaces
Interface          Address              ifi Vif   PktsIn PktsOut     BytesIn
BytesOut
Ethernet0          172.16.20.1       32    1         0        0           0
0
Ethernet2          172.16.4.2        34    2         0        0           0
0
pimreg             0.0.0.0           74    0         0        0           0
0
```

- Check rp-info on each router. The parameter " I am RP" shows whether this router is RP or not. If it says "yes" than this router is RP and if it says "no" than this router is not RP.

```
sonic# show ip pim rp-info
RP address       group/prefix-list   OIF                 I am RP     Source
172.16.2.2       224.0.0.0/4         Ethernet0           yes         Static
```

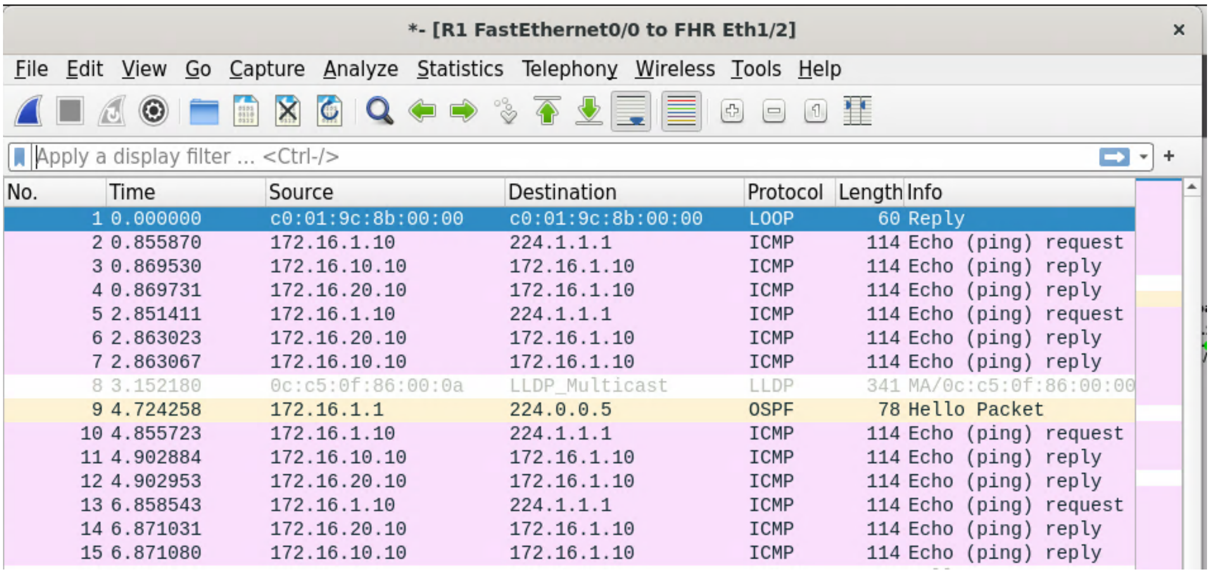
- Check multicast routes. Initially, when there is no receiver for a specific multicast group there will be no multicast route in the multicast routing table. But when there is a receiver then it will show multicast routes.

```
sonic# show ip mroute
IP Multicast Routing Table
Flags: S - Sparse, C - Connected, P - Pruned
       R - SGRpt Pruned, F - Register flag, T - SPT-bit set

Source          Group           Flags     Proto  Input            Output
TTL  Uptime
*               224.1.1.1       SC        IGMP   Ethernet2        Ethernet0
1    00:02:03
*               224.2.2.2       SC        IGMP   Ethernet2        Ethernet0
1    00:02:05
```

- Ping was initiated from the multicast traffic source with 224.1.1.1 multicast group as destination. Both receivers **Receiver-1** and **Receiver-2** joined the 224.1.1.1 multicast group and sent replies to these ICMP packets.

```
Source#ping 224.1.1.1 repeat 15

Type escape sequence to abort.
Sending 15, 100-byte ICMP Echos to 224.1.1.1, timeout is 2 seconds:
.
Reply to request 1 from 172.16.20.10, 1068 ms
Reply to request 1 from 172.16.10.10, 1760 ms
Reply to request 2 from 172.16.20.10, 24 ms
Reply to request 2 from 172.16.10.10, 28 ms
Reply to request 3 from 172.16.10.10, 40 ms
Reply to request 3 from 172.16.20.10, 40 ms
Reply to request 4 from 172.16.10.10, 24 ms
Reply to request 4 from 172.16.20.10, 24 ms
Reply to request 5 from 172.16.20.10, 24 ms
Reply to request 5 from 172.16.10.10, 24 ms
Reply to request 6 from 172.16.10.10, 40 ms
Reply to request 6 from 172.16.20.10, 40 ms
Reply to request 7 from 172.16.10.10, 200 ms
Reply to request 7 from 172.16.20.10, 200 ms
Reply to request 8 from 172.16.10.10, 36 ms
Reply to request 8 from 172.16.20.10, 36 ms
Reply to request 9 from 172.16.10.10, 104 ms
Reply to request 9 from 172.16.20.10, 104 ms
```

- Wireshark results. It shows that multicast source is sending traffic to 224.1.1.1 multicast group and there are two receivers **Receiver-1** and **Receiver-2** that are interested in receiving the traffic for that specific multicast group.



References

1. https://www.packetcoders.io/what-is-pim-protocol-independent-multicast/
2. https://www.metaswitch.com/knowledge-center/reference/what-is-protocol-independent-multicast-pim
3. https://www.juniper.net/documentation/us/en/software/junos/multicast/topics/concept/multicast-pim-overview.html
4. https://study-ccnp.com/understanding-protocol-independent-multicast-pim/
5. https://packetlife.net/blog/2008/oct/20/pim-sm-source-versus-shared-trees/
6. https://www.networkurge.com/2017/10/multicast-distribution-trees.html
7. https://docs.frrouting.org/en/latest/pim.html
8. https://networklessons.com/multicast/source-specific-multicast-ssm
9. https://ipcisco.com/lesson/pim-ssm-source-specific-multicast/