Multicast

IPv4 Multicast Overview

In This Section

- Recommended Resources
- What is Multicast?
- IPv4 Multicast Components
  - Addressing
  - Control Plane
  - Data Plane

Recommended Resources

- Books
  - Routing TCP/IP Volume II
  - Developing IP Multicast Networks
  - Interdomain Multicast Routing: Practical Juniper Networks and Cisco Systems Solutions
- Online Resources
  - Multicast Technology Documentation
  - IP Multicast SRND
  - IP Multicast Best Practices for Enterprise Customers

What is Multicast?

- Multicast is data transmission to a group of destinations simultaneously
  - I.e one to many transmission
- As opposed to…
  - Unicast – one to one transmission
  - Broadcast – one to all transmission
  - Anycast – one to nearest transmission

Why Use Multicast?

- Main goal of multicast is reduce the load on…
  - Sending server processing
  - Network bandwidth resources
  - Router forwarding processing
  - Receiving host processing

Why Not Just Use Unicast?

- Sender must generate one packet for each receiver
  - Called “head-end replication”
- Sender must know addresses of all receivers
- Routers must process packets for each receiver separately
- Bandwidth use is proportional to number of receivers

Why Not Just Use Broadcast?

- In bridged networks, broadcast packets are forwarded out all interfaces except the one received on
- All end hosts process all packets even if they don’t want them
- Result is inefficient use of resources for uninterested receivers

How Multicast Saves Resources

- Source generates single data feed for all interested recipients
- Source does not need to know who is receiving
- Routers make a single forwarding decision for all recipients
- Only one packet is replicated per interface, saving bandwidth
- Uninterested hosts do not receive packets

Multicast Disadvantages

- IP Multicast is UDP
- Connectionless transmission implies…
  - Best effort delivery
    - E.g. no acknowledgments
  - No congestion avoidance
    - E.g. no slow start
  - Possible duplicate packets
    - Usually only during network reconvergence
  - Possible out of order packets
    - Generally no resequencing in UDP applications

How Multicast Works

+ Source application sends UDP multicast traffic with “group” destination address
+ Interested receivers “join” group address by signaling routers on the LAN
+ Routers communicate to build loop free “tree” from sender to receivers
+ Portions of the network without receivers will not receive traffic for that group

Multicast Use Case Examples

- Multimedia
  - IPTV
  - Videoconferencing
  - VoIP Music on Hold
- Data distribution
  - Large scale datacenter replication
- Real-time applications
  - Stock tickers

IPv4 Multicast Components

- Multicast can be broken down into three main components
- Group Addressing
  - Layer 3 addressing
  - Layer 2 addressing
- Control Plane
  - IGMP, PIM, MSDP, MBGP
- Data Plane
  - Reverse Path Forwarding (RPF)
  - Multicast Routing Table (MRIB/MFIB)

Multicast Group Addressing

- Multicast "group" is an address agreed upon between the sender and receivers for a particular feed
  - Source sends traffic to destination address of the group
  - Receivers listen for traffic going to group address
- Traffic is always sent ***to*** a group, never ***from***
- Groups use both layer 3 and layer 2 addresses

IPv4 Multicast Addressing

- IPv4 multicast uses Class “D” Addresses
  - 224.0.0.0/4 (224.0.0.0–239.255.255.255)
- Includes reserved ranges
  - Link-local Addresses
    - 224.0.0.0/24 (224.0.0.0 - 224.0.0.255)
  - Source Specific Multicast
    - 232.0.0.0/8 (232.0.0.0 - 232.255.255.255)
  - Administratively Scoped
    - 239.0.0.0/8 (239.0.0.0 - 239.255.255.255)

Layer 2 Multicast Addressing

- IPv4 addresses map to MAC addresses to forward on the LAN
  - Allows L2 switches to forward multicast intelligently
  - E.g. don’t forward multicast as broadcast
- MAC address range is 01-00-5E-00-00-00 to 01-00-5E-7F-FF-FF
  - First 25 bits are fixed
  - Last 23 bits are mapped from IPv4 address
- Implies overlap between addresses
  - Last 23 bits must be unique to result in unique layer 2 flow

Layer 2 Multicast Address Conversion

- Conversion shortcut
  - Convert IPv4 2nd octet to binary
  - Set the first bit to 0
  - Convert to hex
  - 3rd and 4th octets convert directly to hex
- Example conversions
  - 224.0.0.1
    - 01-00-5E-00-00-01
  - 230.255.1.2
    - 01-00-5E-7F-01-02
  - 239.127.1.2
    - 01-00-5E-7F-01-02

Multicast Control Plane

- Multicast control plane used to determine…
  - Who is sending traffic and to what group(s)
  - Who is receiving traffic and for what group(s)
  - How traffic should be forwarded when it is received
    - The Multicast “Tree”
- Control plane is built with a combination of
  - Host to Router communication (IGMP)
  - Router to Router communication (PIM and MSDP)

Multicast Control Plane – IGMP

- Internet Group Management Protocol (IGMP)
  - Used for receiver to signal routers on the LAN that it wants traffic for a specific group
- Three versions
  - RFC 1112 - Host Extensions for IP Multicasting
  - RFC 2236 - Internet Group Management Protocol, Version 2
  - RFC 3337 - Internet Group Management Protocol, Version 3

IGMPv1

- Uses two message types to signal group membership
  - Host Membership Query
  - Host Membership Report
- Report used by client to “join” a group
- Query used by router to see if members of the group still exist
  - Essentially an idle timer for the group
- Legacy now, replaced by IGMPv2

IGMPv2

- Enhances IGMPv1 by adding
  - Querier election
    - If multiple routers on the segment, who sends queries?
  - Tunable timers
    - Can speed up query response timeouts
  - Group specific queries
    - Query sent to the group address instead of all multicast hosts
  - Explicit leave
    - Speeds up convergence if no other hosts are joined to that group
- Backwards compatible with IGMPv1

IGMPv3

- Used to support Source Specific Multicast (SSM)
  - IGMPv1/v2 only support group specific joins
    - (*,G) join
  - IGMPv3 supports source specific joins
    - (S,G) join
- Implies IGMPv3 receiver must already know about the sender
  - More details on this later

Next Steps From IGMP

- Router knows that host wants traffic for multicast group “G”
- How does it tell the rest of the network to deliver traffic to it for “G” ?
- Multicast “routing” protocols now take over
  - PIM
  - ***Not*** MBGP or MSDP
    - More on this later…

Multicast Control Plane – PIM

- Protocol Independent Multicast (PIM)
  - Router to router communication used to build loop-free “tree” from sender to receiver(s)
- Considered “protocol independent” because it does not advertise its own topology information
  - Implies IGP already runs in the network to build a loop-free topology
- Two versions and two “modes”
  - PIMv1 & PIMv2
  - Sparse Mode & Dense Mode

PIM Modes

- Dense Mode
  - Considered implicit join
  - All traffic unless you say you don’t want it
  - Uses Flood & Prune behavior
- Sparse Mode
  - Considered explicit join
  - No traffic unless you ask for it
  - Uses Rendezvous Point (RP) to process join requests
- PIM modes control how the tree is built, and who receives what traffic
  - More detail later…

Multicast Data Plane

- Once the tree from sender to receiver(s) is built, traffic begins to flow
- Before forwarding, Data Plane checks occur
  - Reverse Path Forwarding (RPF) check
    - Was traffic received on the correct interface?
  - Multicast Routing Table (MRIB/MFIB)
    - What interface(s) should I forward the packets out?

The RPF Check

- If PIM does not exchange its own topology, how does it know the network is loop free?
  - Multicast packet comes in, router looks at source IP address and incoming interface
  - Unicast routing table (CEF table) is checked for the reverse path back to source address
- RPF logic is…
  - If incoming multicast interface == outgoing unicast interface, RPF check passed
  - If incoming multicast interface != outgoing unicast interface, RPF check fails and packet is dropped

Why Use RPF?

- Assume that IGP has built a loop-free unicast topology
  - If RPF check passes on multicast packets, we can assume the traffic has not looped
  - If RPF check fails it’s possible a loop occurred, so traffic is dropped
- RPF is very conservative, but always loop-free

The Multicast Routing Table

- During exchange of PIM messages, routers learn where sources and receivers exist
  - Interface facing upstream towards source is the "incoming interface"
  - Downstream links to receivers are "outgoing interface list" or OIL
  - Split-horizon like behavior – link cannot be in incoming and OIL at same time
- If RPF check passes…
  - Packets flow from incoming interface to all interfaces in the OIL
- More information at…
  - Multicast Forwarding Information Base Overview
  - Verifying IPv4 Multicast Forwarding Using the MFIB

Multicast

PIM Sparse Mode

In This Section

+ PIM Sparse Mode Overview
+ PIM Sparse Mode Configuration

PIM Sparse Mode

- RFC 4601 - Protocol Independent Multicast - Sparse Mode (PIM-SM)
- Uses “pull” model or “explicit join”
  - Traffic is not flooded unless you ask for it
- Uses both Shared Trees (RPT) and Shortest Path Trees (SPT)
  - Dense mode uses only shortest path/source trees
  - More scalable than dense mode and usually the better design choice

Shared vs. Source Trees

- Multicast tree determines how traffic is routed from sender to receivers
- Source based trees
  - Uses shortest path from sender to receiver
  - Dense mode or sparse mode
- Shared trees
  - Uses shortest path from sender to Rendezvous Point (RP), then shortest path from RP to receiver
  - Sparse mode only
  - Used to eliminate flooding and pruning and make routing table more scalable

PIM Sparse Mode Operation

+ Discover PIM neighbors & elect DR
+ Discover RP
+ Tell RP about sources
+ Tell RP about receivers
+ Build shared tree from sender to receivers through RP
+ Join shortest path tree
+ Leave shared tree
+ Multicast table maintenance

Rendezvous Point Overview

- RP is used as a reference point for the root of the shared tree
- RP learns about sources through unicast PIM Register messages
  - Register tells the RP about an (S,G)
- RP learns about receivers through PIM Join messages
  - Tells the RP to add an interface to the OIL for (*,G)
- RP is used to merge the two trees together

Learning the RP's Address

- Without the RP…
  - Sources can't register
  - Joins can't be processed
- All routers must agree on the same RP address on a per-group basis
  - Registers and joins are rejected for invalid RPs
- RP address can be assigned
  - Statically
  - Dynamically
    - Auto-RP
    - BSR

PIM Register Message

- As the root of all shared trees, the RP must know about all sources
- When the first-hop router connected to sender hears traffic, a unicast Register message is sent to the RP
  - If multiple first-hop routers, only the DR registers
- If RP accepts this message, it acknowledges with Register Stop and inserts (S,G) into the table
- At this point only DR and RP know (S,G)

PIM Join Message

- As the root of all shared trees, the RP must also know about all receivers
- When a last-hop router receives an IGMP Report, a PIM Join is generated up the reverse path tree towards the RP
- All routers in the reverse path install (*,G) and forward the Join hop-by-hop to the RP
- At this point the RP and all downstream devices towards the receiver know (*,G)

Merging the Trees

- Once the RP knows about both sender and receiver…
  - RP sends a PIM Join message up reverse path to source
- All routers in the reverse path from the RP to the source install (*,G) with OIL pointing towards RP
- Once (S,G) begins to flow, the tree is built end-to-end through the RP

Joining the SPT

- The shared tree is made up of two Shortest Path Trees
  - SPT from receiver to RP
  - SPT from RP to sender
- SPT from receiver to sender may not be the same as the shared tree
  - Result is that Shared Tree is not optimal forwarding
- To fix this, last-hop router…
  - Joins SPT to source with (S,G) Join
  - Leaves the RPT by sending (*,G) Prune to RP
- Can be modified with **ip pim spt-threshold**

Routing Table Maintenance

- Like PIM Dense Mode, PIM Sparse Mode uses State Refresh to ensure that feeds do not timeout
  - (*,G) join sent to RP or up SPT to refresh the OIL
- Sparse Prune message can be used to speed up state information timeout if IGMP Leave is heard from end host

Multicast

Troubleshooting Multicast RPF Failures

In This Section

- Understanding Multicast RPF Check
- Troubleshooting RPF Failure
- Modifying RPF Check

Multicast RPF Check Review

- Reverse Path Forwarding (RPF) Check
  - PIM does not exchange its own topology
  - PIM relies on IGP for loop free path selection
  - RPF check is an extra data plane loop prevention technique
- RPF check performed on all incoming multicast packets
  - If incoming multicast interface == outgoing unicast interface, RPF check passed
  - If incoming multicast interface != outgoing unicast interface, RPF check fails and packet is dropped

RPF Check & Tree Types

- RPF check changes depending on tree type
- Shortest Path Trees (SPT)
  - Perform RPF check against source
  - Used in (S,G) trees…
    - PIM Dense Mode
    - PIM Sparse Mode from RP to Source
    - PIM Sparse Mode from Receiver to Source after SPT switchover (S-bit set)
    - PIM Source Specific Multicast (SSM)
- Shared Trees (RPT)
  - Perform RPF check against RP
  - Used in (*,G) trees…
    - PIM Sparse Mode from Receiver to RP before SPT switchover

RPF Check & Rendezvous Point

- RPF check is also performed on RP against Register messages
  - RP must have a route back to source that is being registered
  - If no route, (S,G) state cannot be created

Verifying & Troubleshooting RPF Check

- Useful commands…
  - **show ip mroute**
  - **show ip mroute count**
  - **show ip rpf**
  - **mtrace**
  - **debug ip pim**
  - **debug ip mfib pak**

Modifying the RPF Check

- RPF is based on unicast routing table
  - Implies changing unicast routing affects multicast routing
- RPF can be modified
  - Manually with ip mroute
  - Dynamically with Multicast BGP

RPF Check & PIM Join

- RPF controls how tree *must* be built, not how it *can* be built
  - PIM Join is sent out RPF interface for both (*,G) and (S,G)
  - Changing RPF results in traffic engineering for multicast

Multicast

Auto-RP

In This Section

- Auto-RP Overview
- Auto-RP Configuration

PIM Sparse Mode Review

- Traffic is not flooded unless you ask for it
- Uses RP as the root of the Shared Tree
- PIM DR hears sender and reports (S,G) to RP through PIM Register
- Last-hop router hears IGMP Join and sends (*,G) PIM Join towards the RP
- RP sends (S,G) PIM Join towards source to complete the shared tree
- Last-hop router can initiate PIM Shortest Path Tree Join and Shared Tree Prune once feed is end-to-end

Learning the RP

- Without the RP
  - Sources can’t register
  - Joins can’t be processed
- All routers must agree on the same RP address on a per-group basis
  - Registers and joins are rejected for invalid RPs
- RP address can be assigned…
  - Statically
  - Dynamically
    - Auto-RP
    - BSR

Auto-RP Overview

- Cisco proprietary method for dynamic RP advertisement
- Uses two functional roles
  - Candidate RP
    - Device(s) willing to be the RP
  - Mapping Agent
    - Chooses the RP among candidates and relays this information to the rest of the PIM domain
- Allows for redundancy of RPs

How Auto-RP Works

- Candidate RPs send announcement with group range they are willing to service
  - Uses group (S, 224.0.1.39) for announcement
- Mapping agent discovers Candidate RPs and advertises their mappings to all other routers
  - Joins (*, 224.0.1.39) to discover about Candidate RPs
  - Announces final RP advertisement with (S, 224.0.1.40)

Auto-RP Caveats

- Dynamically learned RP mappings are preferred over statically configured ones
- Auto-RP control plane messages are subject to RPF check

Auto-RP Caveats (cont.)

- Routers must join (*, 224.0.1.39) for Candidate RP and (*, 224.0.1.40) for Mapping Agent
- In PIM Sparse Mode…
  - Cannot join the Auto-RP groups without knowing where the RP is
  - Cannot know where the RP is without joining the Auto-RP groups
  - Recursive logic

Auto-RP Solutions

- Default RP assignment
  - Assign a static RP for groups 224.0.1.39 and 224.0.1.40
  - Defeats the purpose of automatic assignment
- PIM Sparse-Dense Mode
  - Dense for groups without an RP
  - Sparse for all others
- Auto-RP Listener feature
  - Dense for 224.0.1.39/224.0.1.40 only
  - Sparse for others

Auto-RP With Multiple Candidates

+ For redundancy and load distribution multiple Candidate RPs can be configured
+ ACL applied on Candidate RP controls what groups they service
+ If multiple overlapping Candidate RPs, Mapping Agent chooses highest RP address

Mapping Agent Security

- Mapping Agents should be protected against false Candidate RP advertisements
- RP Announce Filter feature can permit or deny Candidate RP to be accepted

Multicast

Bootstrap Router (BSR)

In This Section

- BSR Overview
- BSR Configuration

Bootstrap Router

- RFC 5059 - Bootstrap Router (BSR) Mechanism for Protocol Independent Multicast (PIM)
  - Functionally similar to Auto-RP
- Defines two roles in BSR domain
  - RP Candidate
    - Analogous to Candidate RP in Auto-RP
    - Uses unicast PIM to advertise itself to the Bootstrap Router
  - Bootstrap Router (BSR)
    - Analogous to Mapping Agent in Auto-RP
    - Advertises RP information to other routers with multicast PIM on a hop-by-hop basis

Controlling Auto-RP and BSR Messages

- By default Auto-RP and BSR messages are sent on all PIM enabled interfaces
- For added security, these message should be filtered on network edge
  - Auto-RP via Multicast Boundary
  - BSR via BSR Border
- Filtering can also occur based on TTL
  - Called administrative scoping

Multicast

Anycast RP and MSDP

In This Section

- Anycast Overview
- Anycast RP Overview
- MSDP Overview
- Anycast & MSDP Configuration

What is Anycast?

- Anycast is one to nearest routing
  - Multiple destinations share the same address
  - Route to the closest one based on the IGP/BGP table
- Why use Anycast?
  - Poor man’s load balancing & HA

How Anycast Works

- Mirror the application data to multiple devices in the topology
  - E.g. RFC 3258 - Distributing Authoritative Name Servers via Shared Unicast Addresses
  - http://root-servers.org
- Assign each device the same duplicate IP & advertise it
  - E.g. same /32 Loopback into BGP/IGP
- Use the routing table for load balancing & HA
  - Who you route to depends on where you are physically in the topology
  - If anycast device fails, use routing convergence to find the next closest device

What is Anycast RP?

- Uses anycast load balancing to decentralize the placement of PIM Sparse Mode RPs
  - PIM Register and Join messages go to the closest RP in the topology
  - If one RP goes down, convergence is up to IGP
  - As long as one anycast RP is up, new trees can be built
  - RP failure does not necessarily affect current trees

Anycast RP Design Issues

- For anycast to work, all RPs must share the same information about senders and receivers
- What if PIM Register is sent to one anycast RP, and PIM Join is sent to another?
  - One RP knows about the sender
  - Another RP knows about the receiver
  - How can we build the tree from the RP to source if we don’t know the source?
- Multicast Source Discovery Protocol (MSDP)

What is MSDP?

- RFC 3618 - Multicast Source Discovery Protocol (MSDP)
- Used to advertise (S,G) pairs between RPs
  - Listen for PIM Register message about (S,G)
  - Tell other RPs about (S,G) through an MSDP Source Active (SA) message
  - Essentially like an inter-RP PIM Register message
- Allows PIM domains to use independent RPs
  - Originally designed for Inter-AS Multicast
  - Our use case is Anycast RP for Intra-AS Multicast

How Anycast RP Works

- Anycast RPs assign a duplicate Loopback address and advertise into IGP
- All routers point to anycast RP address
  - Could be static or dynamic assignment
- Anycast RPs are MSDP peers using a unique address
  - E.g. each device has a globally routable Loopback plus the Anycast Loopback
  - If three of more RPs, usually a mesh group
- When PIM Register is received, MSDP SA is sent to MSDP peers
  - Results in synchronization of (S,G) information
  - RP that knows about receiver can now join the (S,G) tree

Anycast RP Caveats

- Requires duplicate addresses
  - Ensure that control plane protocols do not use Anycast address as identifier
  - E.g. prevent duplicate OSPF/LDP/BGP, etc. Router-IDs
- Requires unique address to sync the application data
  - Application is hosted on the Anycast address
  - Application data sync between Anycast members needs to be routable
    - E.g. DNS zone transfers
    - E.g. MSDP TCP peering

Multicast

Source Specific Multicast (SSM)

In This Section

- Any Source Multicast Review
- Source Specific Multicast Overview
- Source Specific Multicast Configuration

Any Source Multicast Review

- Any Source Multicast (ASM)
  - Traditional PIM Sparse Mode with an RP
  - Receiver does not yet know who the sender is
  - Sender and Receiver are connected through the RP
  - I.e. both (S,G) trees and (*,G) trees

ASM Operational Review

- Source begins sending traffic
  - PIM DR hears application feed (S,G)
  - Unicast PIM Register is sent from DR to RP
  - RP acks DR with Register Stop
  - RP now knows about (S,G)

ASM Operational Review (cont.)

- Receiver signals group membership
  - Application sends IGMPv1/v2 Report for (*,G)
  - IGMP Querier translates to (*,G) PIM Join towards RP
  - PIM Join forwarded up RPF path to RP
  - RP now knows about receiver

ASM Operational Review (cont.)

- RP joins the (S,G)
  - RP sends (S,G) PIM Join up RPF path to source
  - Application now flows from Source to RP
- RP forwards to receiver via (*,G)
  - Receiver now gets the application flow
- SPT Switchover
  - Last hop sends PIM Join (S,G)
  - Last hop sends PIM Prune (*,G)
  - Receiver is now joined to the (S,G)

ASM Design Issues

- Receivers don’t know about senders in advance
  - RP is used to find the senders
- RP is a bottleneck in the control plane
  - RP failure means that new trees can’t be built
  - RP is at least temporarily in the data plane
- Solution?
  - Have the receiver pre-learn the source out-of-band

Source Specific Multicast Overview

- Source Specific Multicast (SSM)
  - Uses group address range 232.0.0.0/8
- Receiver knows the application source before it signals membership
  - Receiver uses IGMPv3 Report to signal (S,G) join
- RP is not needed to build the shared tree
  - Application already knows the source
  - RP not needed to build the control plane
- Result is SSM uses only (S,G) trees
  - Last hop router sends (S,G) PIM Join up RPF towards source
  - Result is each tree is SPT for (S,G)

SSM Configuration

- Enable multicast routing
  - **ip multicast-routing [distributed]**
- Define global SSM group range
  - **ip pim ssm [default|range]**
- Enable PIM Sparse at interface level
  - **ip pim sparse-mode**
- Enable IGMPv3 on links to receivers
  - **ip igmp version 3**

INE

Multicast

Bidirectional PIM (BIDIR)

In This Section

+ Bidirectional PIM Overview
+ Bidirectional PIM Configuration

What is Bidirectional PIM?

- RFC 5015 - Bidirectional Protocol Independent Multicast (BIDIR-PIM)
- Traditional Sparse Mode forms two trees
  - Unidirectional SPT from source to RP
  - Unidirectional Shared tree from RP to receivers
- Results in (*,G) and (S,G) entries in control plane
  - For many-to-many multicast applications, doesn’t scale well
- Bidirectional PIM solves this by only allowing the Shared Tree (*,G) and never a SPT (S,G)

How Bidirectional PIM Works

- Define an RP and group range as bidirectional
  - Stops formation of (S,G) for that range
- Build single (*,G) tree towards RP
  - Traffic flows upstream from source to RP
  - Traffic flows downstream from RP to receivers
- Removes PIM Register process
  - Implies that traffic from sources always flows to the RP
- Uses Designated Forwarder (DF) for loop prevention

PIM Bidir Designated Forwarder

- One DF is elected per PIM segment
  - Similar to Assert, lowest metric to the RP wins
  - Highest IP address in a tie
- Only DF can forward traffic upstream towards RP
- All other interfaces in OIL are downstream facing
- Removes the need for RPF check
  - Due to this all routers must agree on Bidir or loops can occur

Multicast

Multicast BGP

In This Section

- Multicast BGP Overview
- Multicast BGP Configuration

Inter-AS Multicast Routing

- Inter-AS Multicast uses MSDP and PIM
  - MSDP to advertise sources
  - PIM to build the tree and do RPF check
- For multicast transit over the Internet all hops must run multicast
  - What if the RPF check for a multicast source is via a unicast only peer?
  - Multicast BGP solves this by separating unicast RPF and multicast RPF

Multicast BGP Overview

- Multicast BGP advertises source networks for purpose of RPF check
  - Does not replace PIM
- Multicast BGP preferred over unicast protocols for multicast RPF check
  - Like a static multicast route, but dynamic
- Doesn’t require a separate routing protocol, only BGP extensions
  - RFC 4760 - Multiprotocol Extensions for BGP 4

How Multicast BGP Works

- BGP peers negotiate Multicast Address Family during capabilities exchange
- Peers advertise NLRI under Multicast Address Family
- Network statement just like unicast
- All normal BGP rules apply
- When multicast traffic is received, MBGP learned routes are preferred over unicast

INE

Multicast

Layer 2 Multicast

In This Section

- Layer 2 Multicast Addressing
- IGMP Snooping
- IGMP Profiles

Ethernet Multicasting

- Ethernet supports Layer 2 Multicasting natively
  - Multicast bit in 48-bit address: lowest bit in the first byte
  - e.g. 01-00-CC-CC-CC-CC
- Multicast addresses used for various purposes
  - E.g. CDP, L2 protocol tunneling, etc
  - Allocated by IEEE

Layer 2 Multicast Addressing

- IPv4 addresses map to MAC addresses to forward on the LAN
  - Allows L2 switches to forward multicast intelligently
  - E.g. don’t forward multicast as broadcast
- MAC address range is 01-00-5E-00-00-00 to 01-00-5E-7F-FF-FF
  - First 25 bits are fixed
  - Last 23 bits are mapped from IPv4 address
- Implies overlap between addresses
  - Last 23 bits must be unique to result in unique layer 2 flow

Layer 2 Multicast Address Conversion

- Conversion shortcut
  - Convert IPv4 2nd octet to binary
  - Set the first bit to 0
  - Convert to hex
  - 3rd and 4th octets convert directly to hex
- Example conversions
  - 224.0.0.1
    - 01-00-5E-00-00-01
  - 230.255.1.2
    - 01-00-5E-7F-01-02
  - 239.127.1.2
    - 01-00-5E-7F-01-02

Multicast in Switched Environment

- Switches treat unknown unicast and multicast destinations like broadcast
  - Implies multicast traffic is flooded to all ports in the VLAN
  - Ideal behavior is to flood only ports that have receivers
- How can the switch know where receivers are?
  - IGMP Snooping

IGMP Snooping

- Switch listens for IGMP Reports/Leaves
  - L2 devices inspect L3 frames
  - Extracts group address reported
  - Prunes unneeded Multicast
- Multicast routers have to be discovered
  - Routers need to process all multicast traffic
  - Switch listens to PIM messages
  - i.e. PIM Snooping

IGMP Snooping Commands

+ Two commands to enable
  + **ip igmp snooping**
  + **ip igmp snooping vlan XX**
+ Statically join a port to a group
  + **ip igmp snooping vlan <N> static <IP> interface <Int>**

IGMP Snooping and STP

- STP Topology change may signal receiver moving…
  - After a TCN event switch floods all multicast groups on all ports
  - **ip igmp snooping tcn flood query count <count>**
  - The above command means: flood until <count> query intervals have expired
- Disabling flooding during TCN
  - **no ip igmp snooping tcn flood**

IGMP Profiles

- IGMP access-group applies only on Layer 3 interfaces
- IGMP Profile allows IGMP access-control at Layer 2
  - Profiles are either in permit or deny mode
  - Permit mode allows specified groups and blocks all others
  - Deny mode blocks specified groups and allows all other

IGMP Profile Example

```
ip igmp profile 1
 permit
 range 239.0.0.0
 range 237.0.0.0 238.0.0.0
!
interface FastEthernet 0/1
 ip igmp profile 1
```

IGMP Throttling

- Limits amount of groups joined on interface
  - **ip igmp max-groups NN**
  - **ip igmp max-groups action {deny|replace}**
- New groups are either denied or replace old ones