TOMORROW starts here.
Introduction to IP Multicast

BRKIPM-1261

Baron Rawlins, Systems Engineer
Carolinas Enterprise Operation
Session Goal

- To provide you with a basic understanding of the concepts, mechanics and protocols used to build IP multicast networks.
- To enable you to ask the right questions, and make the correct architectural decisions in deploying and maintaining an IP Multicast enabled network.
Agenda

- Why Multicast?
- Multicast History and Fundamentals
- PIM Protocols
- Rendezvous Points
- Multicast at Layer 2
- Interdomain IP Multicast
- IPv6 Multicast
Why Multicast?
Unicast vs. Multicast Scaling

Number of Streams

Server

Router

Server

Unicast

Multicast
A Brief History of Multicast

Steven Deering, 1985, Stanford University
- Yeah, he was way ahead of his time and too clever for all of us.

A solution for Layer 2 applications in the growing Layer 3 campus network
- Think overlay broadcast domain

Broadcast Domain
- all members receive
- all members can source
- members dynamically come and go
A Brief History of Multicast

RFC966 - 1985

Multi-destination delivery is useful to several applications, including:
- distributed, replicated databases [6,9].
- distributed parallel computation, including distributed gaming [2].

All inherently many-to-many applications

No mention of one-to-many services such as Video/IPTV
A Brief History of Multicast

Overlay Broadcast Domain Requirements

- Tree building and maintenance
- Network-based source discovery
- Source route information
- Overlay mechanism – tunneling

The first solution had it all

Distance Vector Multicast Routing Protocol
RFC1075 – 1988
A Brief History of Multicast

PIM – Protocol Independent Multicast

“Independent” of which unicast routing protocol you run, But does require running one*

Uses local routing table to determine route to sources

Router-to-router protocol to build and maintain distribution trees

Source discovery handled one of two ways:
1. Flood-and-prune via PIM-DM (Dense Mode) RFC3973
2. Explicit join via RP (Rendezvous Point) and PIM-SM (Sparse Mode)
A Brief History of Multicast

Today’s dominant applications are primarily one-to-many
IPTV, Contribution video over IP, etc.
Sources are generally well known

SSM – Source Specific Multicast
RFC3569, RFC4608 – 2003

- Tree building and maintenance
- Network-based source discovery
- Source route information
- Overlay mechanism—tunneling

Very simple and the preferred solution for one-to-many applications
Multicast Application Types

- **One-to-Many (1toM)**
  - Audio/Video: Lectures, presentations, concerts, television, radio
  - Push Media: News headlines, weather updates, sports scores
  - Distribution: Web site content, executable binaries
  - Announcements: Network time, multicast session schedules, random numbers, keys, security
  - Monitoring: Stock prices, sensors

- **Many-to-Many (MtoM)**
  - Conferencing: Audio/Video conferences, whiteboards
  - Sharing Resources: Synchronized distributed databases
  - Games: Multi-player with distributed interactive simulations
  - Others: Concurrent processing, collaboration, two-way distance learning

- **Many-to-One (Mto1)**
  - Resource Discovery: Service location, device discovery
  - Data Collection: Monitoring applications, video surveillance
  - Others: Auctions, polling, jukebox, accounting

For a detailed analysis see RFC3170
Multicast Considerations

- **UDP based**: Therefore not connection-oriented
- **Best effort delivery**: Drops are to be expected; multicast applications should not expect reliable delivery of data and should be designed accordingly; reliable multicast is still an area for much research; expect to see more developments in this area; PGM, FEC, QoS
- **No congestion avoidance**: Lack of TCP windowing and “slow-start” mechanisms can result in network congestion; if possible, multicast applications should attempt to detect and avoid congestion conditions
- **Duplicates**: Some protocol mechanisms can result in the occasional generation of duplicate packets; multicast applications should be designed to expect occasional duplicate packets
- **Out of order delivery**: Some protocol mechanisms may also result in out of order delivery of packets
Multicast Components

- **End stations (hosts-to-routers)**
  - IGMP, MLD, AMT

- **Campus Multicast**
  - Switches (Layer 2 optimization)
    - IGMP snooping PIM snooping
  - Routers (multicast forwarding protocol)
    - PIM sparse mode or bidirectional PIM

- **Multicast routing across domains**
  - MBGP

- **Interdomain Multicast**
  - Multicast source discover
    - MSDP with PIM-SM
  - Source Specific Multicast
    - SSM
Multicast Fundamentals
Fundamentals - Unicast vs. Multicast Addressing

**Unicast Addressing**

- Source address: 10.1.1.1
- A unique packet addressed to each destination.

**Multicast Addressing**

- Source address: 10.1.1.1
- Same packet for each destination.
Fundamentals - Unicast vs. Multicast Addressing

How do we address one packet to different destinations?

src addr: 10.1.1.1
A unique packet addressed to each destination.

src addr: 10.1.1.1
..replicated at each node along the tree.

12.1.1.1
11.1.1.1
13.1.1.1

How do we address one packet to different destinations?
### Fundamentals - Multicast Addressing

#### IPv4 Header

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version</td>
<td>Specifies the version of IPv4.</td>
</tr>
<tr>
<td>IHL</td>
<td>Indicates the length of the header in 32-bit words.</td>
</tr>
<tr>
<td>Type of Service</td>
<td>Specifies the type of service, such as internet control message protocol (ICMP).</td>
</tr>
<tr>
<td>Total Length</td>
<td>Indicates the total length of the IP packet in octets.</td>
</tr>
<tr>
<td>Identification</td>
<td>Helps in locating packets.</td>
</tr>
<tr>
<td>Options</td>
<td>Allows for additional information that is not part of the standard IPv4 format.</td>
</tr>
<tr>
<td>Padding</td>
<td>填补了空白，确保总长度的头大小是4的倍数。</td>
</tr>
<tr>
<td>Source Address</td>
<td>Always the unique unicast origin address of the packet – same as unicast.</td>
</tr>
<tr>
<td>Destination</td>
<td>Class A, B, C (1.0.0.0 - 223.255.255.255)</td>
</tr>
<tr>
<td>Multicast Group Address Range</td>
<td>224.0.0.0 - 239.255.255.255 (Class D)</td>
</tr>
</tbody>
</table>

Source Address Can Never Be Class D Multicast Group Address

---

© 2013 Cisco and/or its affiliates. All rights reserved. Cisco Public
“Class D” Group addresses – 224/4 (224.0.0.0 - 239.255.255.255)

Multicast Group addresses are NOT in the unicast route table.

A separate table is maintained for active multicast trees in the network.

Multicast state entries are initiated by receivers signaling their request to join a group.

Sources do not need to join, they just send.

Multicast routing protocols build and maintain the trees hop-by-hop, based on receiver membership and source reachability.

Source reachability is derived from the unicast route table.
  – Multicast relies on a dependable unicast infrastructure.
Multicast Addressing - 224/4

- **Reserved link-local addresses 224.0.0.0–224.0.0.255**
  Transmitted with TTL = 1
  Examples:
  - 224.0.0.1: All systems on this subnet
  - 224.0.0.2: All routers on this subnet
  - 224.0.0.5: OSPF routers
  - 224.0.0.13: PIMv2 routers
  - 224.0.0.22: IGMPv3

- **Other reserved addresses 224.0.1.0–224.0.1.255**
  Not local in scope (transmitted with TTL > 1) but typically bounded
  Examples:
  - 224.0.1.1: NTP (Network Time Protocol)
  - 224.0.1.32: Mtrace routers
  - 224.0.1.78: Tibco Multicast1
Multicast Addressing - 224/4

- Administratively scoped addresses
  - 239/8 (239.0.0.0 - 239.255.255.255)
  - Private address space
    Similar to RFC1918 unicast addresses
    Not used on global Internet

- GLOP
  - 233/8 (233.0.0.0 - 233.255.255.255)
  - Provides /24 group prefix per ASN

- SSM (Source Specific Multicast)
  - 232/8 (232.0.0.0 – 232.255.255.255)
  - Primarily targeted for Internet-style broadcast
How Are Multicast Addresses Assigned?

- Static global group address assignment “GLOP”
  - Temporary method to meet immediate needs
  - Group range: 233/8 (233.0.0.0–233.255.255.255)
    Your AS number is inserted in middle two octets
    Remaining low-order octet used for group assignment
  - Defined in RFC2770

- Manual address allocation by the admin
  - The most common practice
Host-Router Signaling: Internet Group Management Protocol (IGMP)

- How hosts tell routers about group membership
- Routers solicit group membership from directly connected hosts
- Three major variants in use today:
  - RFC 1112 specifies version 1
  - RFC 2236 specifies version 2
  - RFC 3376 specifies version 3
Host-Router Signaling: IGMP

Joining a Group

- Host sends IGMP report to join group
Host-Router Signaling: IGMP

Maintaining a Group

- Router sends periodic queries to 224.0.0.1
  - One member per group per subnet reports
  - Other members suppress reports
Host-Router Signaling: IGMP
Leaving a Group (IGMPv2)

- Host sends leave message to 224.0.0.2
- Router sends group-specific query to 224.1.1.1
- No IGMP report is received within ~ 3 seconds
- Group 224.1.1.1 times out
Host-Router Signaling: IGMPv3

RFC 3376 – enables SSM

- Adds include/exclude source lists
- Enables hosts to listen only to a specified subset of the sources sending to the group
- Requires ‘IPMulticastListen’ API
- IGMPv3 stack required in the OS
- Apps must be rewritten to use IGMPv3 include/ exclude features
Host-Router Signaling: IGMPv3

New Membership Report Address

- **224.0.0.22 (IGMPv3 routers)**
  - All IGMPv3 hosts send reports to this address
    Instead of the target group address as in IGMPv1/v2
  - All IGMPv3 routers listen to this address
  - Hosts do not listen or respond to this address

- **No report suppression**
  - All hosts on wire respond to queries
    Host’s complete IGMP state sent in single response
  - Response interval may be tuned over broad range
    Useful when large numbers of hosts reside on subnet
Joining member sends IGMPv3 report to 224.0.0.22 immediately upon joining.
IGMPv3 - Joining a group with specific source(s)

- IGMPv3 report contains desired source(s) in the include list
- Only “Included” source(s) are joined
IGMPv3—Maintaining State

- Router sends periodic queries
- All IGMPv3 members respond
- Reports contain multiple group state records
Multicast L3 Forwarding

Multicast Routing is Backwards from Unicast Routing

- Unicast routing is concerned about where the packet is going
- Multicast routing is concerned about where the packet came from
  - Initially
Unicast vs. Multicast Forwarding

Unicast Forwarding

- Destination IP address directly indicates where to forward packet
- Forwarding is hop-by-hop
  - Unicast routing table determines interface and next-hop router to forward packet
Unicast vs. Multicast Forwarding

Multicast Forwarding

- Destination IP address (Group) does not directly indicate where to forward packet
- Forwarding based on Outgoing Interface (OIF)
- Receivers must first be “connected” to the tree before traffic begins to flow

Connection messages (PIM joins) follow unicast routing table toward multicast source
Build multicast distribution trees that determine where to forward packets
Distribution trees rebuilt dynamically in case of network topology changes
Each router in the path maintains an OIF list (OIL) per tree state
Multicast State

R3#sh ip mroute
IP Multicast Routing Table
...
(2.2.2.2, 239.100.100.100), 00:00:16/00:02:48, flags: LJT
  Incoming interface: FastEthernet0/1, RPF nbr 10.23.0.2
  Outgoing interface list:
    Loopback1, Forward/Sparse, 00:00:16/00:02:43

R3#sh ip route 239.100.100.100
% Network not in table

R3#sh ip route
...
O 2.2.2.2 [110/11] via 10.23.0.2, 00:21:35, FastEthernet0/1

Multicast route entries are in (S,G) form.
Incoming interface points upstream toward the root of the tree.
Outgoing interface list is where receivers have joined downstream and where packets will be replicated and forwarded downstream.
Multicast Group addresses are NEVER in the unicast route table.
The RPF Calculation

- The multicast packets source address is checked against the unicast routing table
- This determines the interface and upstream router in the direction of the source to which PIM joins are sent
- This interface becomes the “Incoming” or RPF interface
  - A router forwards a multicast datagram only if received on the RPF interface
Reverse Path Forwarding (RPF)

RPF Calculation

- Based on source address
- Best path to source found in unicast route table
- Determines where to send join
- Joins continue towards source to build multicast tree
- Multicast data flows down tree
Reverse Path Forwarding (RPF)

RPF Calculation

- Based on source address
- Best path to source found in unicast route table
- Determines where to send join
- Joins continue towards source to build multicast tree
- Multicast data flows down tree
Reverse Path Forwarding (RPF)

RPF Calculation

- What if we have equal-cost paths?
  - We can’t use both
- Tie-breaker
  - Use highest next-hop IP address

![Diagram of network](image.png)
Reverse Path Forwarding

R3#sh ip mroute
IP Multicast Routing Table

...(2.2.2.2, 239.100.100.100), 00:00:16/00:02:48, flags: LJT
  Incoming interface: FastEthernet0/1, RPF nbr 10.23.0.2
  Outgoing interface list:
    Loopback1, Forward/Sparse, 00:00:16/00:02:43

R3#sh ip rpf 2.2.2.2
RPF information for ? (2.2.2.2)
  RPF interface: FastEthernet0/1
  RPF neighbor: ? (10.23.0.2)
  RPF route/mask: 2.2.2.2/32
  RPF type: unicast (ospf 1)
  RPF recursion count: 0
  Doing distance-preferred lookups across tables
Multicast Tree Creation

- IGMP membership reports
  - Used by hosts to indicate membership
- PIM join/prune control messages
  - Used to create/remove distribution trees
- Shortest path trees
  - Notation (S,G)
    - PIM control messages are sent toward the source
- Shared trees
  - Notation (*,G)
    - PIM control messages are sent toward RP
Multicast State

R3#sh ip mroute
IP Multicast Routing Table

... 
(*, 239.100.100.100), 00:03:10/stopped, RP 2.2.2.2, flags: SJCL
Incoming interface: FastEthernet0/1, RPF nbr 10.23.0.2
Outgoing interface list:
  Loopback1, Forward/Sparse, 00:03:10/00:02:22

Shared-tree entries are in (*,G) form.

(2.2.2.2, 239.100.100.100), 00:00:16/00:02:48, flags: LJT
Incoming interface: FastEthernet0/1, RPF nbr 10.23.0.2
Outgoing interface list:
  Loopback1, Forward/Sparse, 00:00:16/00:02:43

SPT entries are in (S,G) form.
Multicast Distribution Trees
Shortest Path or Source Distribution Tree

Notation: (S, G)
S = Source
G = Group

Source 1

Source 2

Receiver 1

Receiver 2
Multicast Distribution Trees
Shortest Path or Source Distribution Tree

Source 1

Notation: (S2, G)
S = Source
G = Group

Source 2

Receiver 1

Receiver 2
Multicast Distribution Trees

Shared Distribution Tree

Notation: \((*, G)\)

* \(=\) All Sources
G \(=\) Group

Receiver 1

Receiver 2

(C) PIM Rendezvous Point

Shared Tree
Multicast Distribution Trees

Shared Distribution Tree

Notation: \((*, G)\)

\(\ast\) = All Sources

\(G\) = Group

Source 1

Source 2

Receiver 1

Receiver 2

A

B

D (RP)

C

E

F

(RP) PIM Rendezvous Point

Shared Tree

Source Tree
Multicast Distribution Trees

- **Source or shortest path trees (S,G)**
  - Uses more memory $O(S \times G)$
  - Optimal paths from source to all receivers
  - Minimal delay

- **Shared trees (*,G)**
  - Uses less memory $O(G)$
  - May use suboptimal paths from source to all receivers
  - May introduce extra delay
PIM Protocols
Commonly deployed PIM variants

Protocol Independent Multicast – Dense Mode
  - Flood and Prune

Protocol Independent Multicast – Sparse Mode
  - Any Source Multicast (ASM), uses RP/SPT/shared tree
  - Source Specific Multicast (SSM), no RP, SPT only

Bidirectional PIM
  - Shared tree only
PIM-SM Shared Tree Join

(*, G) Join
Shared Tree

Receiver

(*, G) State Created Only
Along the Shared Tree
PIM-SM Sender Registration

- **Source**
- **RP**
- **Receiver**
- **Traffic Flow**
- **Shared Tree**
- **Source Tree**
- **(S, G) Register**
- **(S, G) Join**
- **(S, G) State Created Only Along the Source Tree**
PIM-SM Sender Registration

Traffic Flow
- Shared Tree
- Source Tree

(S, G) Register
(S, G) Register-Stop

RP Sends a Register-Stop Back to the First-Hop Router to Stop the Register Process

(S, G) Traffic Begins Arriving at the RP via the Source Tree
PIM-SM Sender Registration

Traffic Flow
Shared Tree
Source Tree

Source

Source Traffic Flows Natively Along SPT to RP
From RP, Traffic Flows Down the Shared Tree to Receivers

RP

Receiver
PIM-SM SPT Switchover

Traffic Flow
Shared Tree
Source Tree
(S, G) Join

Last-Hop Router Joins the Source Tree
Additional (S, G) State Is Created Along New Part of the Source Tree
Traffic Flow
Shared Tree
Source Tree
(S, G)RP-bit Prune

Traffic Begins Flowing Down the New Branch of the Source Tree
Additional (S, G) State Is Created Along the Shared Tree to Prune Off (S, G) Traffic
PIM-SM SPT Switchover

(S, G) Traffic Flow Is Now Pruned Off of the Shared Tree and Is Flowing to the Receiver via the Source Tree
PIM-SM SPT Switchover

(S, G) Traffic Flow Is No Longer Needed by the RP so It Prunes the Flow of (S, G) Traffic

Traffic Flow
Shared Tree
Source Tree
(S, G) Prune
PIM-SM SPT Switchover

(S, G) Traffic Flow Is Now Only Flowing to the Receiver via a Single Branch of the Source Tree
PIM-SM - Evaluation

- Effective for sparse or dense distribution of multicast receivers

**Advantages**
- Traffic only sent down “joined” branches
- Can switch to optimal source-trees for high traffic* sources dynamically
- Unicast routing protocol-independent
- Basis for interdomain multicast routing
  - When used with MBGP, MSDP and/or SSM

**Disadvantages**
- RP placement may be sub-optimal for some (*,G)

**Primary application**
- Multicast networks with sparse or dense distribution of receivers
Source Specific Multicast - SSM

Defined in RFC 3569: An Overview of Source Specific Multicast (SSM)

- Assumes a one-to-many multicast model
  - Example: video/audio broadcasts, stock market data
- Why does ASM need a shared tree?
- What if the source was already known?
  - Hosts use IGMPv3 to signal exactly which (S, G) SPT to join
  - The shared tree and RP are not necessary
  - Different sources could share the same group address and not interfere with each other
PIM-SM SSM Tree Join

Source

RP

(S,G) State Created at each router along Source Tree

Receiver

IGMPv3 (S,G) Host Report

PIM (S,G) Join

Traffic Flow

© 2013 Cisco and/or its affiliates. All rights reserved.
PIM-SM SSM Tree Join

Traffic flows natively along the source tree.
SSM - Evaluation

- **Advantages**
  - Uses a simplified subset of the PIM-SM protocol
    - Simpler network operation
  - Solves multicast address allocation problems
    - Flows differentiated by both source and group
    - Content providers can use overlapping group ranges
  - More secure and efficient

- **Disadvantages**
  - Requires IGMPv3 aware leaf routers and applications
  - Requires distribution of source information to application

- **Primary Application**
  - Applications with one source sending to many receivers
Many-to-Many State Problem

- Creates huge amounts of (S,G) state
  - State maintenance workloads skyrocket
  - Router performance begins to suffer

- Using shared trees only helps
  - Provides some (S, G) state reduction
    - Results in (S, G) state only along SPT to RP
    - Frequently still too much (S, G) state

- Need a solution that uses only (*, G) state – Bidirectional PIM
  - Defined in RFC5015
Bidirectional PIM—Overview
Bidirectional PIM—Overview

Source Traffic Forwarded Bidirectionally Using (*,G) State
Bidirectional PIM - Evaluation

- Ideal for many to many applications
- Advantages
  - Drastically reduces network mroute state
  - Eliminates **ALL** \((S,G)\) state in the network
    - SPTs between sources to RP eliminated
    - Source traffic flows both up and down shared tree
  - Allows many-to-many applications to scale
    - Permits virtually an unlimited number of sources
- Disadvantages
  - All traffic must flow to the root of the tree (RP)
  - Longer recovery times after path failures
Rendezvous Points
PIM-SM ASM RP Key Points

- Placement
- RP address must be known by all routers
- Group to RP mapping
  - Must be consistent in all routers within the PIM domain
- RP redundancy recommended
  - Eliminate any single point of failure
How Does the Network Learn RP Address?

RP address must be known by all routers

- **Static configuration**
  - Manually on every router in the PIM domain

- **AutoRP**
  - Originally a Cisco® solution
  - Facilitated PIM-SM early transition

- **Bootstrap Router (BSR)**
  - draft-ietf-pim-sm-bsr
Static RP Assignment

- Hard-configured RP address
  - When used, must be configured on every router
  - All routers must have the same RP address
  - RP failover not possible by default
  - Redundancy achieved via MSDP

- Command(s):
  - `ip pim rp-address <address> [group-list <acl>] [override]`
  - Optional group list specifies group range
    - Default range = 224.0.0.0/4 *(includes auto-RP groups!)*
  - Override keyword “overrides” auto-RP information
    - Default: auto-RP learned info takes precedence
Auto-RP—From 10,000 Feet

RP-Announcements Multicast to the Cisco Announce (224.0.1.39) Group
Auto-RP—From 10,000 Feet

RP-Discoveries Multicast to the Cisco Discovery (224.0.1.40) Group
BSR—From 10,000 Feet

BSR Election Process

BSR Messages Flooded Hop-by-Hop
Highest Priority C-BSR
Is Elected as BSR
BSR—From 10,000 Feet
BSR—From 10,000 Feet

BSR Messages Containing RP-Set Flooded Hop-by-Hop
Multicast at Layer 2
Multicast Addressing

IP Multicast MAC Address Mapping

- 32 Bits
- 28 Bits
- 5 Bits Lost
- 25 Bits
- 23 Bits
- 48 Bits

239.255.0.1

01-00-5e-7f-00-01
Multicast Addressing

IP Multicast MAC Address Mapping

Be Aware of the 32:1 Address Overlap

32–IP Multicast Addresses

224.1.1.1
224.129.1.1
225.1.1.1
225.129.1.1
...
238.1.1.1
238.129.1.1
239.1.1.1
239.129.1.1

1–Multicast MAC Address

01-00-5E-01-01-01
L2 Multicast Frame Switching

Problem: Layer 2 Flooding of Multicast Frames

- Typical L2 switches treat multicast traffic as unknown or broadcast and must “flood” the frame to every port.

- Static entries can sometimes be set to specify which ports should receive which group(s) of multicast traffic.

- Dynamic configuration of these entries would cut down on user administration.
L2 Multicast Frame Switching

IGMPv1–v2 Snooping

- Switches become IGMP “aware”
- IGMP packets intercepted by the NMP/CPU or by port ASICs
  
  Requires special hardware to maintain throughput
- Switch must examine contents of IGMP messages to determine which ports want what traffic
  
  IGMP membership reports
  IGMP leave messages
- Impact on low-end, Layer 2 switches
  
  Must process all Layer 2 multicast packets
  Admin load increases with multicast traffic load
  Generally results in switch meltdown
L2 Multicast Frame Switching with IGMPv3

- IGMPv3 reports sent to separate group (224.0.0.22)
  - Switches listen to just this group
  - Only IGMP traffic - no data traffic
  - Substantially reduces load on switch CPU
  - Permits low-end switches to implement IGMPv3 snooping

- No report suppression in IGMPv3
  - Enables individual member tracking

- IGMPv3 supports source-specific includes/excludes
Summary - L2 Multicast Frame Switching

- IGMP Snooping used to limit flooding to member ports
- Switches with Layer 3-aware hardware/ASICs
  - High-throughput performance maintained
  - Increases cost of switches
- Switches without Layer 3-aware hardware/ASICs
  - Suffer serious performance degradation or even meltdown!
  - Should not be a problem when IGMPv3 is implemented
Interdomain IP Multicast
MBGP - Multiprotocol BGP

- Defined in RFC 2858 (Extensions to BGP)
- Allows BGP to carry different types of routes
  - Unicast
  - Multicast
  - MPLS Layer3 VPN
- Both routes carried in same BGP session
- Does NOT propagate multicast state info
  - That’s PIM’s job
- Same path selection and validation rules
  - AS-Path, LocalPref, MED…
MBGP - Multiprotocol BGP

- Separate BGP tables maintained
  - Unicast prefixes for unicast forwarding
  - Unicast prefixes for multicast RPF checking
  - Implemented via Address Family Indicator (AFI)

- AFI = 1, Sub-AFI = 1
  - Contains unicast prefixes for unicast forwarding
  - Populated with BGP unicast NLRI

- AFI = 1, Sub-AFI = 2
  - Contains unicast prefixes for RPF checking
  - Populated with BGP multicast NLRI
MBGP - Multiprotocol BGP

- MBGP Allows Divergent Paths and Policies
- Same IP address holds dual significance
  - Unicast routing information
  - Multicast RPF information
- For same IPv4 address two different NLRI with different next-hops
- Can therefore support both congruent and incongruent topologies
MSDP – Multicast Source Discovery Protocol

- Required for interdomain source discovery
- Defined in RFC 3618
- Supports PIM-ASM only
  - RPs know about all sources in their domain
    - Sources cause a “PIM Register” to the RP
    - Tell RPs in other domains of it’s sources
      - Via MSDP SA (Source Active) messages
  - RPs know about receivers in a domain
    - Receivers cause a “(*, G) Join” to the RP
    - RP can join the source tree in the peer domain
      - Via normal PIM (S, G) joins
MSDP Overview

MSDP Peers

Domain B

Domain A

Domain C

Domain D

Domain E

Join (*, 224.2.2.2)

Receiver
MSDP Overview

MSDP Peers
Source Active Messages

Domain A
Source Register
192.1.1.1, 224.2.2.2

Domain B
SA

Domain C
RP
SA

Domain D
RP
SA

Domain E
RP
Receiver

SA Message 192.1.1.1, 224.2.2.2

SA Message 192.1.1.1, 224.2.2.2
MSDP Overview

MSDP Peers

Domain A
Source

Domain B
RP

Domain C
RP

Domain D
RP

Domain E
RP

Receiver

Join
(S, 224.2.2.2)
MSDP Overview

MSDP Peers

Multicast Traffic

Domain A

Domain B

Domain C

Domain D

Domain E

Source

Receiver

MSDP Peers

Multicast Traffic
MSDP Overview

- **MSDP Peers**
- **Multicast Traffic**

MSDP Overview Diagram:
- **Domain A**
  - Source
- **Domain B**
  - RP
- **Domain C**
  - RP
- **Domain D**
  - RP
- **Domain E**
  - RP
  - Receiver

Multicast Traffic Flow:
- **Join (S, 224.2.2.2)**

MSDP Overview

MSDP Peers
Multicast Traffic

Domain A
Source

Domain B
RP

Domain C
RP

Domain D
RP

Domain E
RP

Receiver

Multicast Traffic:
- From Source to Domain A
- From Domain A to Domain B
- From Domain B to Domain C
- From Domain C to Domain D
- From Domain D to Domain E

MSDP Peers:
- Between Domain A and Domain B
- Between Domain B and Domain C
- Between Domain C and Domain D
- Between Domain D and Domain E
MSDP with SSM – Not Needed!

ASM MSDP Peers (Irrelevant to SSM)

Multicast Traffic

Domain A

Source

Source in 232/8

Domain B

Domain C

Domain D

Domain E

Receiver Learns S and G Out of Band, i.e., Webpage

Receiver

Multicast Traffic
MSDP wrt SSM—Unnecessary

ASM MSDP Peers
(Irrelevant to SSM)

Multicast Traffic

Data flows natively along the interdomain source tree

Domain A
Source in 232/8

Domain B
RP

Domain C
RP

Domain D
RP

Domain E
RP

Receiver

Source
Anycast RP - Overview

- Redundant RP technique for ASM which uses MSDP for RP synchronization

- Uses single defined RP address in domain
  Two or more routers have same RP address
    RP address defined as a loopback interface
    Loopback address advertised as a host route via IGP
  Senders and receivers join/register with nearest RP
    Nearest RP determined via unicast routing table because RP is statically defined

- MSDP session(s) established between RPs
  Informs RPs of sources in other parts of network
  RPs join SPT to active sources as necessary
Anycast RP—Overview

[Diagram showing Anycast RP with MSDP and bidirectional SA traffic between RP1 and RP2.]

- RP1 (A) with Anycast address 10.1.1.1
- RP2 (B) with Anycast address 10.1.1.1
- MSDP connection between RP1 and RP2
- Rec (Receiver) and Src (Source) indications
Anycast RP—Overview
IPv6 Multicast
# IPv4 vs. IPv6 Multicast

<table>
<thead>
<tr>
<th>Service</th>
<th>IPv4</th>
<th>IPv6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Address Range</td>
<td>32-Bit, Class D</td>
<td>128-Bit (112-Bit Group)</td>
</tr>
<tr>
<td>Routing</td>
<td>Protocol-Independent</td>
<td>Protocol-Independent</td>
</tr>
<tr>
<td></td>
<td>All IGPs and BGP4+</td>
<td>All IGPs and BGP4+ with v6 Mcast SAFI</td>
</tr>
<tr>
<td>Forwarding</td>
<td>PIM-DM, PIM-SM: ASM, SSM, BiDir</td>
<td>PIM-SM: ASM, SSM, BiDir</td>
</tr>
<tr>
<td>Group Management</td>
<td>IGMPv1-3</td>
<td>MLDv1, v2</td>
</tr>
<tr>
<td>Domain Control</td>
<td>Boundary/Border</td>
<td>Scope Identifier</td>
</tr>
<tr>
<td>Interdomain Source Discovery</td>
<td>MSDP Across Independent PIM Domains</td>
<td>Single RP Within Globally Shared Domains</td>
</tr>
</tbody>
</table>
IPv6 Multicast Addresses - RFC 4291

128 Bits

FF Flags Scope Group-ID

1111 1111 Flags Scope

Flags = {T or Lifetime, 0 if Permanent, 1 if Temporary
P for Unicast-based Assignments
R for Embedded RP
Others Are Undefined and Must Be Zero
1 = interface-local
2 = link
4 = admin-local
5 = site
8 = organization
E = global
0, 3, F = reserved}

Scope = {1 = interface-local
2 = link
4 = admin-local
5 = site
8 = organization
E = global
0, 3, F = reserved}

Note: other scopes (6, 7, 9-D) are unassigned but can be used
IPv6 Layer 2 Multicast

- Addressing Mapping

IPv6 Multicast Address

<table>
<thead>
<tr>
<th>FF</th>
<th>Flags</th>
<th>Scope</th>
<th>High-Order</th>
<th>Low-Order</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>80 Bits Lost</td>
<td>48 Bits</td>
</tr>
</tbody>
</table>

80 Bits Lost

33-33-xx-xx-xx-xx

Ethernet MAC Address
# Unicast-Based Multicast Addresses

<table>
<thead>
<tr>
<th>8</th>
<th>4</th>
<th>4</th>
<th>8</th>
<th>8</th>
<th>64</th>
<th>32</th>
</tr>
</thead>
<tbody>
<tr>
<td>FF</td>
<td>Flags</td>
<td>Scope</td>
<td>Rsvd</td>
<td>Plen</td>
<td>Network-Prefix</td>
<td>Group-ID</td>
</tr>
</tbody>
</table>

- Defined in RFC 3306 Unicast-based Multicast Addresses
  - Similar to IPv4 GLOP addressing
  - Solves IPv6 global address allocation problem
  - Flags = 00PT
    - \(P = 1, T = 1\) → Unicast-based multicast address

- Example
  - Content provider’s unicast prefix
    - 1234:5678:9::/48
  - Multicast address
    - FF3x:0030:1234:5678:0009::0001
IPv6 Routing for Multicast

- RPF-based on reachability to IPv6 source
  - Same as with IPv4 multicast
- RPF still protocol-independent
  - Static routes, mroutes
  - Unicast routing table: BGP, ISIS, OSPF, EIGRP, RIP, etc.
  - Multiprotocol BGP (mBGP) via Sub-AFI
IPv6 Multicast Forwarding

- PIM-Sparse Mode (PIM-SM)
  - RFC 4601
- PIM Source Specific Mode (SSM)
  - RFC 3569 SSM overview (v6 SSM needs MLDv2)
  - Unicast, prefix-based multicast addresses FF30::/12
  - SSM range is FF3X::/96
- PIM Bi-Directional Mode (BiDir)
  - RFC 5015
RP Mapping Mechanisms for IPv6

- Static RP assignment
- BSR
- Auto-RP: no plans at this time
- Embedded RP
Embedded RP Addressing

- Defined in RFC 3956
- Proposed new multicast address type
  - Uses unicast-based multicast addresses (RFC 3306)
- RP address is embedded in multicast address
- Flag bits = 0RPT
  - $R = 1$, $P = 1$, $T = 1 \rightarrow$ Embedded RP address
- Network-Prefix::RPadr = RP address
- For each unicast prefix you own, you now also own:
  - 16 RPs for each of the 16 multicast scopes (256 total) with $2^{32}$ multicast groups assigned to each RP ($2^{40}$ total)

<table>
<thead>
<tr>
<th>FF</th>
<th>Flags</th>
<th>Scope</th>
<th>Rsrvd</th>
<th>RPadr</th>
<th>Plen</th>
<th>Network-Prefix</th>
<th>Group-ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>8</td>
<td>64</td>
<td>32</td>
</tr>
</tbody>
</table>
Embedded RP Addressing - Example

FF  Flags  Scope  Rsvd  RPadr  Plen  Network-Prefix  Group-ID
   8     4      4     4     8     64          32

Resulting RP Address

FF76:0130:1234:5678:9abc::4321

1234:5678:9abc::1
Multicast Listener Discover - MLD

- MLD is equivalent to IGMP in IPv4
- MLD messages are transported over ICMPv6
- Version number confusion
  - MLDv1 corresponds to IGMPv2
    Defined in RFC 2710
  - MLDv2 corresponds to IGMPv3, needed for SSM
    Defined in RFC 3810
- MLD snooping
  - Considerations in RFC 4541
Recap

- Why Multicast?
- Multicast History and Fundamentals
- PIM Protocols
- Rendezvous Points
- Multicast at Layer 2
- Interdomain IP Multicast
- IPv6 Multicast
Q&A
Complete Your Online Session Evaluation

- Give us your feedback and you could win fabulous prizes. Winners announced daily.
- Receive 20 Cisco Daily Challenge points for each session evaluation you complete.
- Complete your session evaluation online now through either the mobile app or internet kiosk stations.

Maximize your Cisco Live experience with your free Cisco Live 365 account. Download session PDFs, view sessions on-demand and participate in live activities throughout the year. Click the Enter Cisco Live 365 button in your Cisco Live portal to log in.
Internet IP Multicast

- We can build multicast distribution trees.
  - PIM

- We can RPF on interdomain sources
  - MBGP

- We no longer need (or want) network-based source discovery
  - SSM

- So interdomain IP Multicast is in every home, right?
What’s Missing?

- Without an overlay mechanism, Multicast in the Internet is an all or nothing solution
  - Each receiver must be on an IP multicast-enabled path
  - Many core networks have IP multicast-enabled, but few edge networks accept multicast transit traffic
  - Deering had tunneling in the original solution
- Even Mcast-aware content owners are forced to provide unicast streams to gain audience size
- Unicast cannot scale dynamically for live content
  - Splitters/caches just distribute the problem
    Still has a cost per user
AMT - Automatic Multicast Tunneling

- Automatic IP multicast without explicit tunnels
  
  
  Allow multicast content distribution to extend to unicast-only connected receivers
  
  Bring the flat scaling properties of multicast to the Internet

- Provide the benefits of multicast wherever multicast is deployed
  
  Let the networks with multicast support benefit from their deployment

- Works seamlessly with existing applications
  
  No OS kernel changes
Native Interdomain Multicast (SSM)

As Long as IP Multicast Is Enabled on Every Router from the Source to the Receivers, the Benefits of IP Multicast Are Realized
Native Interdomain Multicast (SSM)

The Benefits Being an Unlimited Number of Receivers Can Be Served with a Single Stream of Content at No Additional Costs

Unicast-Only Network

Mcast-Enabled ISP

Content Owner

Mcast-Enabled Local Provider

Mcast Traffic

Mcast Join

The Benefits Being an Unlimited Number of Receivers Can Be Served with a Single Stream of Content at No Additional Costs
AMT—Automatic Multicast Tunneling

The AMT Anycast Address Allows for All AMT Gateway to Find the “Closest” AMT Relay—the Nearest Edge of the Multicast Topology of the Source

Unicast-Only Network

Once the Multicast Join Times Out, an AMT Join Is Sent from the Host Gateway Toward the Global AMT Anycast Address
AMT—Automatic Multicast Tunneling

- **Unicast-Only Network**
- **AMT Request Captured by the AMT Relay Router**
- **Mcast-Enabled ISP**
- **Content Owner**
- **Mcast-Enabled Local Provider**

- **Mcast Traffic**
- **Mcast Join**
- **AMT Request**
AMT—Automatic Multicast Tunneling

(S,G) Is Learned from the AMT Join Message, Then (S,G) PIM Join Is Sent Toward the Source

(Mcast-Enabled ISP)

(Mcast-Enabled Local Provider)

(Content Owner)

Unicast-Only Network

Mcast Traffic

Mcast Join

AMT Request
AMT—Automatic Multicast Tunneling

AMT Relay Replicates Stream on Behalf of Downstream AMT Receiver, Adding a Unicast Header Destined to the Receiver

Unicast-Only Network

Mcast-Enabled ISP

Content Owner

Mcast-Enabled Local Provider

- Mcast Traffic
- Mcast Join
- AMT Request
- Ucast Stream

AMT Relay Replicates Stream on Behalf of Downstream AMT Receiver, Adding a Unicast Header Destined to the Receiver
AMT—Automatic Multicast Tunneling

Additional Receivers Are Served by the AMT Relays; the Benefits of IP Multicast Are Retained by the Content Owner and All Enabled Networks in the Path

Unicast-Only Network  Mcast-Enabled ISP  Content Owner

Mcast-Enabled Local Provider

Mcast Traffic
Mcast Join
AMT Request
Ucast Stream
AMT—Automatic Multicast Tunneling

Creates an Expanding Radius of Incentive to Deploy Multicast

Unicast-Only Network

Mcast-Enabled ISP

Content Owner

Enables Multicast Content to a Large (Global) Audience

Mcast-Enabled Local Provider

Mcast Traffic

Mcast Join

AMT Request

Ucast Stream

Enables Multicast Content to a Large (Global) Audience

Enables Multicast Content to a Large (Global) Audience

Enables Multicast Content to a Large (Global) Audience

Enables Multicast Content to a Large (Global) Audience

Enables Multicast Content to a Large (Global) Audience
AMT—Automatic Multicast Tunneling

Creates an Expanding Radius of Incentive to Deploy Multicast

Mcast-Enabled ISP

Content Owner

Enables Multicast Content to a Large (Global) Audience

Unicast-Only Network

Mcast-Enabled Local Provider

Mcast Traffic

Mcast Join

AMT Request

Ucast Stream
AMT—Automatic Multicast Tunneling

- Creates an Expanding Radius of Incentive to Deploy Multicast
- Enables Multicast Content to a Large (Global) Audience

Mcast-Enabled ISP

Mcast-Enabled Local Provider

Content Owner

Mcast Traffic
Mcast Join
AMT Request
Ucast Stream