Introduction to IP Multicast
BRKRST-126

Dr. Peter J. Welcher,
Chesapeake Netcraftsmen

Cisco Mid-Atlantic User's Group
Columbia MD – April 28, 2009
Washington DC – April 29, 2009

About the Speaker

- Dr. Pete Welcher
  - Cisco CCIE #1773, CCSI #94014, CCIP, CCDE written
  - Specialties: Network Design, QoS, MPLS, Wireless, Large-Scale Routing & Switching, High Availability, Management of Networks
  - Customers include large enterprises, federal agencies, hospitals, universities, cell phone provider
  - Taught many of the Cisco router / switch courses
  - Reviewer for many Cisco Press books, book proposals
  - Designed and reviewed revisions to the Cisco DESGN and ARCH courses
  - Presented lab session on MPLS VPN Configuration at Networkers 2005, 2006, 2007, BGP in 2008, BGP and CCIP: Data Center Design in 2009
- Over 140 articles, plus prior seminars, posted
  - http://www.netcraftsmen.net/welcher/

IP Multicast at Cisco Live 2008

- BRKRST-1261: Introduction to IP Multicast
- BRKRST-2261: Deploying IP Multicast
- BRKRST-2262: Multicast Security
- BRKRST-2263: Multicast Network Management
- BRKRST-3261: Advances in IP Multicast
- BRKWT-2102: IP Multicast and Multipoint Design for IP/TV Services
- TECRST-1008: Enterprise IP Multicast

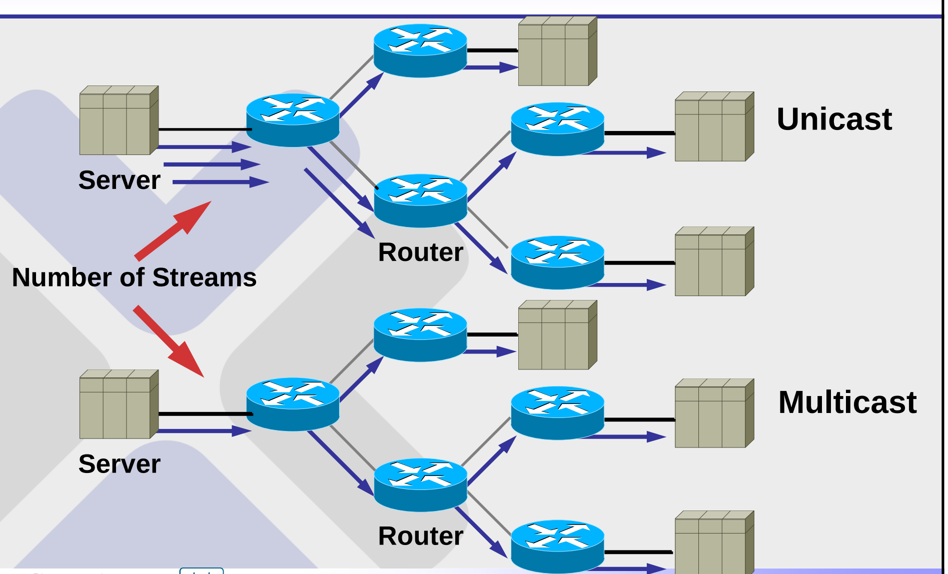
Session Goal

To Provide You with a Thorough Understanding of the Concepts, Mechanics and Protocols Used to Build IP Multicast Networks

Agenda

- Why Multicast?
- Multicast Fundamentals
- PIM Protocols
- RP Choices
- Multicast at Layer 2
- Interdomain IP Multicast
- Some Latest Additions

Unicast vs. Multicast



Multicast Uses

- Any applications with multiple receivers
  - One-to-many or many-to-many
- Live video distribution
- Collaborative groupware
- Periodic data delivery—"push" technology
  - Stock quotes, sports scores, magazines, newspapers, adverts
- Server/Website replication
- Reducing network/resource overhead
  - More than multiple point-to-point flows
- Resource discovery
- Distributed interactive simulation (DIS)
  - War games
  - Virtual reality

Unicast vs. Multicast

- TCP unicast but not multicast
  - TCP is connection-orientated protocol
  - Requires three-way handshake
  - Reliable due to sequence numbers + Ack
  - Flow control
- UDP unicast and multicast
  - Connectionless
  - Unreliable (application layer awareness)

Multicast Disadvantages

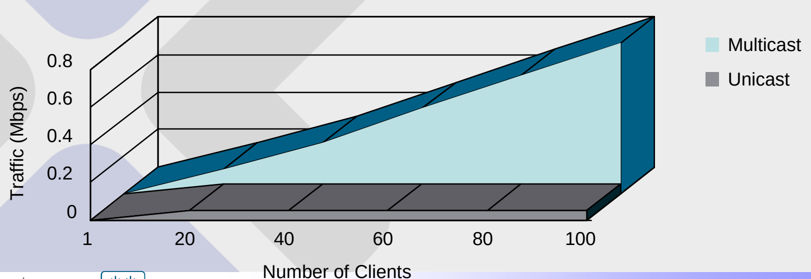
Multicast Is UDP-Based

- 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 multicast protocol mechanisms (e.g., asserts, registers, and SPT transitions) 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 Advantages

- Enhanced efficiency: **controls network traffic and reduces server and CPU loads**
- Optimized performance: **Eliminates traffic redundancy**
- Distributed applications: **Makes multipoint applications possible**

Example: Audio Streaming
All Clients Listening to the Same 8 Kbps Audio

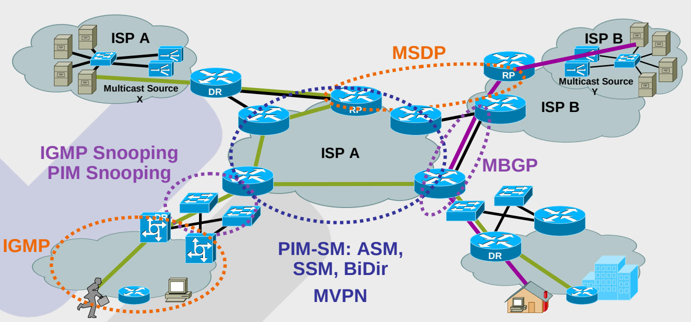


Agenda

- Why Multicast?
- Multicast Fundamentals
- PIM Protocols
- RP Choices
- Multicast at Layer 2
- Interdomain IP Multicast
- Some Latest Additions

Multicast Components

Cisco End-to-End Architecture



- End stations (hosts-to-routers)
  - IGMP

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

Chesapeake NETCRAFTSMEN

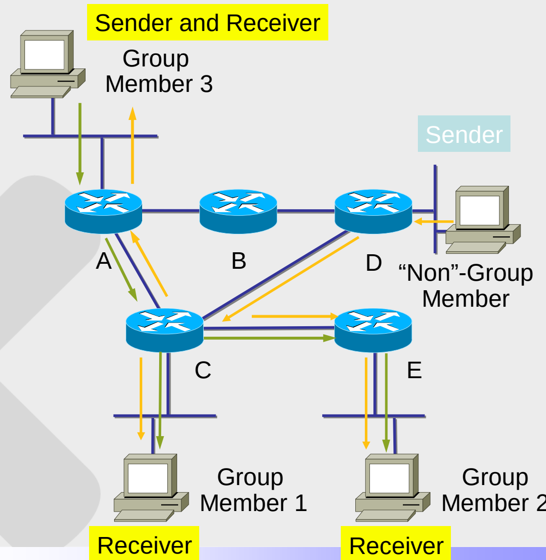
CISCO PARTNER Premier Certified

IP Multicast Group Concept

1. You must be a "member" of a group to receive its data

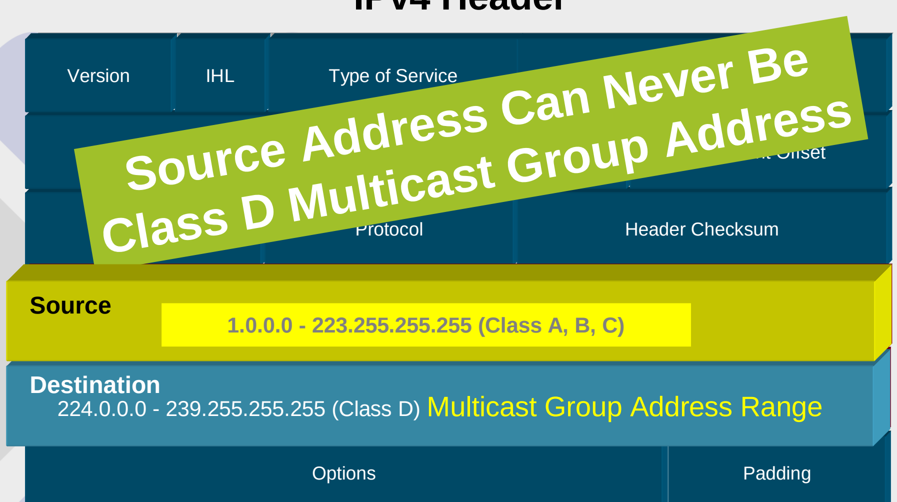
2. If you send to a group address, all members receive it

3. You do not have to be a member of a group to send to a group



Multicast Addressing

IPv4 Header



Multicast Addressing—224/4

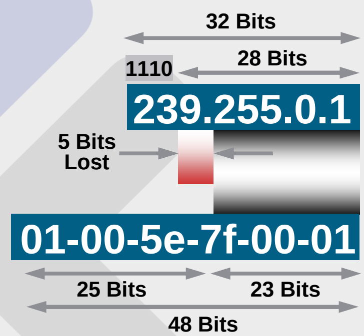
- Reserved link-local addresses
  - 224.0.0.0–224.0.0.255
  - Transmitted with TTL = 1
  - Examples
    - 224.0.0.1 All systems on this subnet
    - 224.0.0.2 All routers on this subnet
    - 224.0.0.5 OSPF routers
    - 224.0.0.13 PIMv2 routers
    - 224.0.0.22 IGMPv3
- Other reserved addresses
  - 224.0.1.0–224.0.1.255
  - Not local in scope (transmitted with TTL > 1)
  - 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
  - 239.0.0.0–239.255.255.255
  - Private address space
    - Similar to RFC1918 unicast addresses
    - Not used for global Internet traffic—scoped traffic
- SSM (Source Specific Multicast) range
  - 232.0.0.0–232.255.255.255
  - Primarily targeted for Internet-style broadcast
- GLOP (honest, it’s not an acronym)
  - 233.0.0.0–233.255.255.255
  - Provides /24 group prefix per ASN

Multicast Addressing

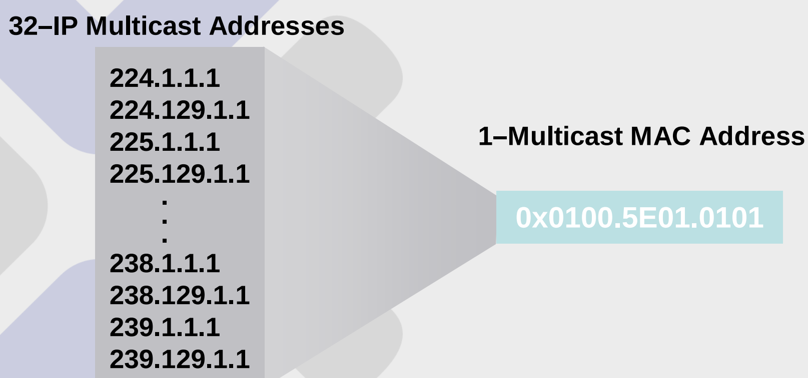
IP Multicast MAC Address Mapping



Multicast Addressing

IP Multicast MAC Address Mapping

Be Aware of the 32:1 Address Overlap



How Are Multicast Addresses Assigned?

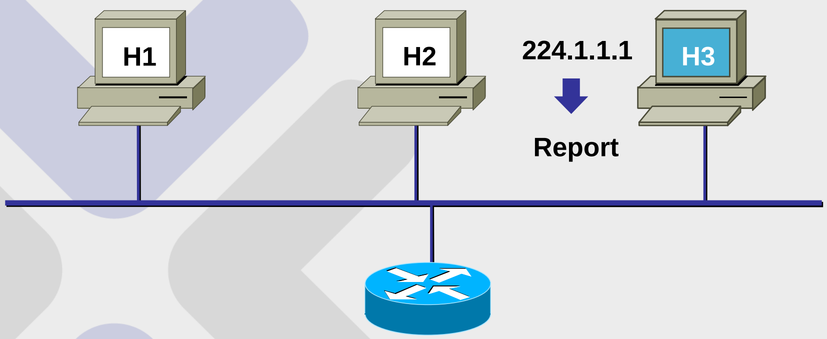
- Static global group address assignment
  - Temporary method to meet immediate needs
  - Group range: 233.0.0.0–233.255.255.255
    - Your AS number is inserted in middle two octets
    - Remaining low-order octet used for group assignment
  - Defined in RFC 2770
    - “GLOP Addressing in 233/8”
- Manual address allocation by the admin
  - Is still the most common practice

Host-Router Signaling: IGMP

- How hosts tell routers about group membership
- Routers solicit group membership from directly connected hosts
- RFC 1112 specifies version 1 of IGMP
  - Supported on Windows 95
- RFC 2236 specifies version 2 of IGMP
  - Supported on latest service pack for Windows and most UNIX systems
- RFC 3376 specifies version 3 of IGMP
  - Supported in Window XP and various UNIX systems

Host-Router Signaling: IGMP

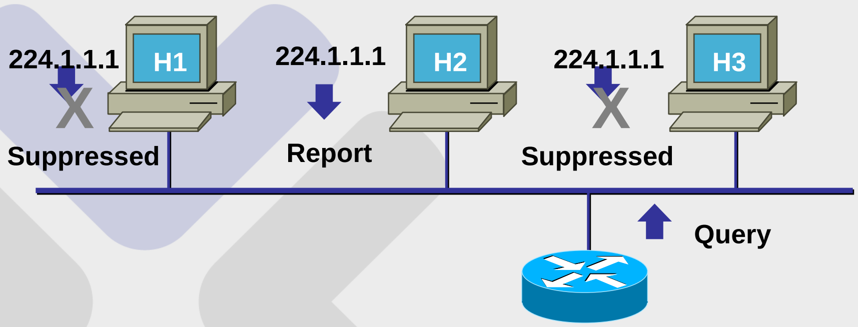
Joining a Group



- **Host sends IGMP report to join group**

Host-Router Signaling: IGMP

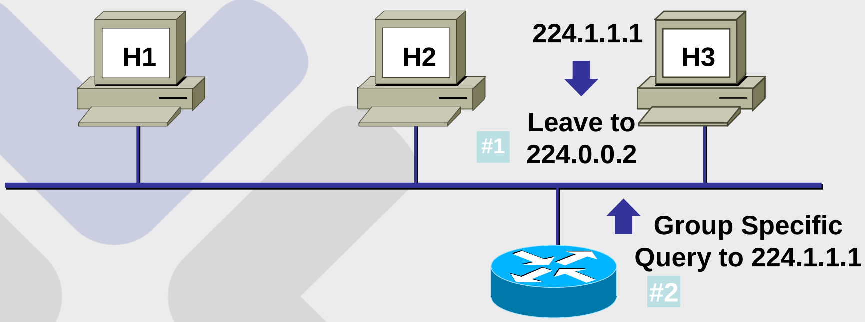
Maintaining a Group



- **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

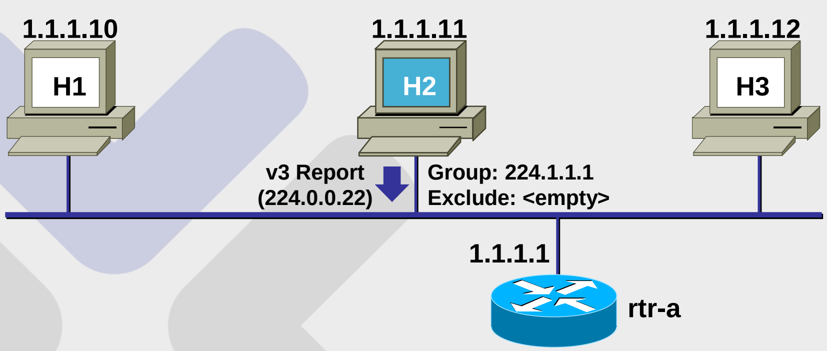
- **Adds include/exclude source lists**
- **Enables hosts to listen only to a specified subset of the hosts sending to the group**
- **Requires new 'IPMulticastListen' API**
- **New IGMPv3 stack required in the OS**
- **Apps must be rewritten to use IGMPv3 include/exclude features**

Host-Router Signaling: IGMPv3

New Membership Report Address

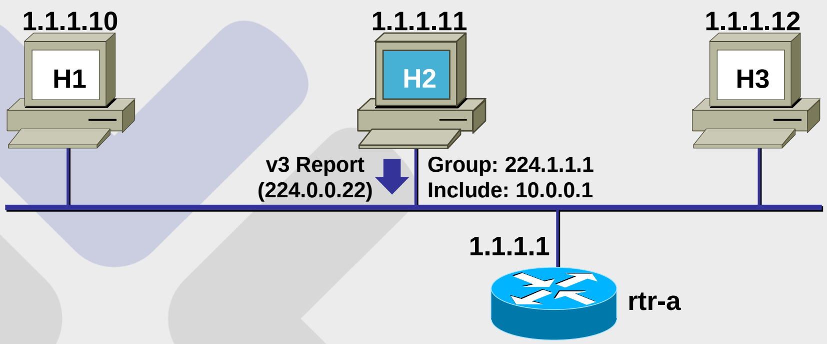
- 224.0.0.22 (IGMPv3 routers)
  - All IGMPv3 hosts send reports to this address
    - Instead of the target group address as in IGMPv1/v2
  - All IGMPv3 routers listen to this address
  - Hosts do not listen or respond to this address
- No report suppression
  - All hosts on wire respond to queries
    - 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

IGMPv3—Joining a Group



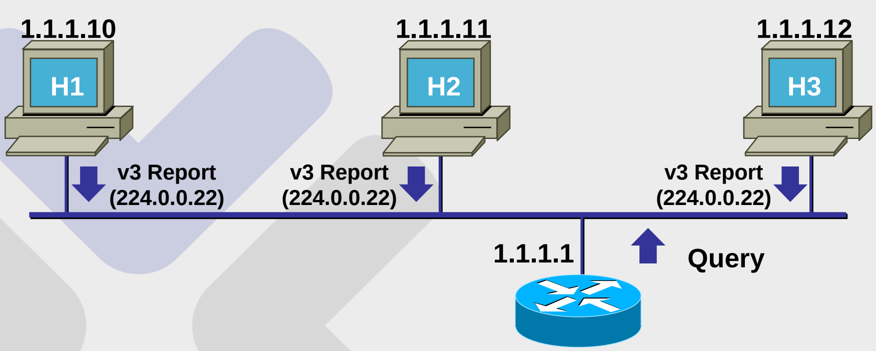
- Joining member sends IGMPv3 report to 224.0.0.22 immediately upon joining

IGMPv3—Joining 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

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) doesn't directly indicate where to forward packet**
- **Forwarding is connection-oriented**
  - **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**

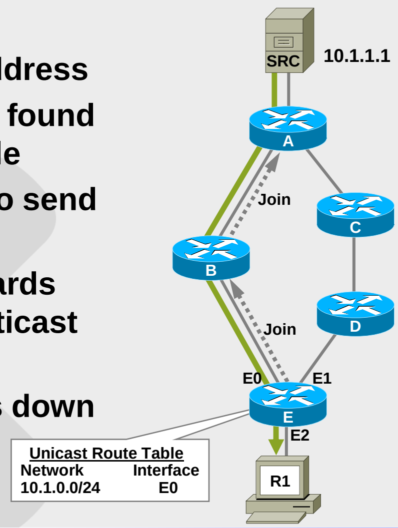
Reverse Path Forwarding (RPF)

The RPF Calculation

- **The multicast 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**

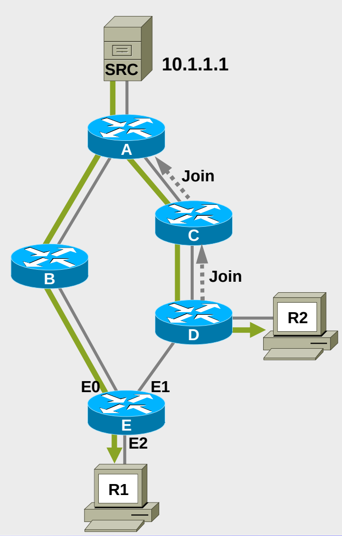
Reverse Path Forwarding (RPF)

- **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



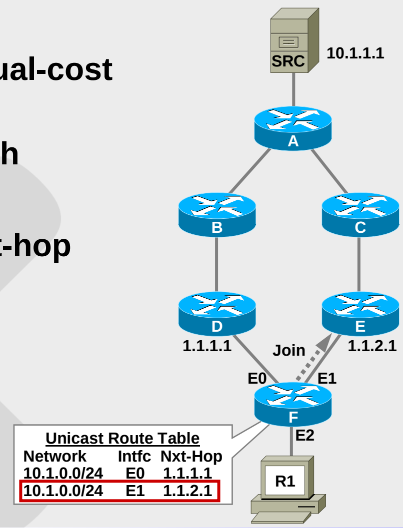
Reverse Path Forwarding (RPF)

- **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
  - Repeat for other receivers



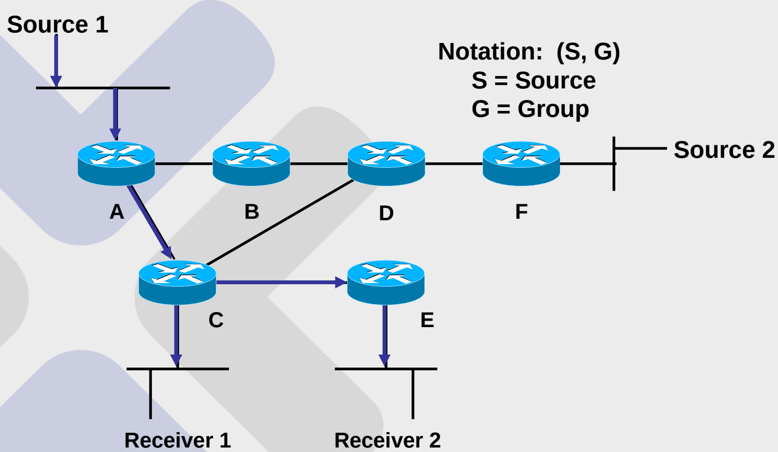
Reverse Path Forwarding (RPF)

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



Multicast Distribution Trees

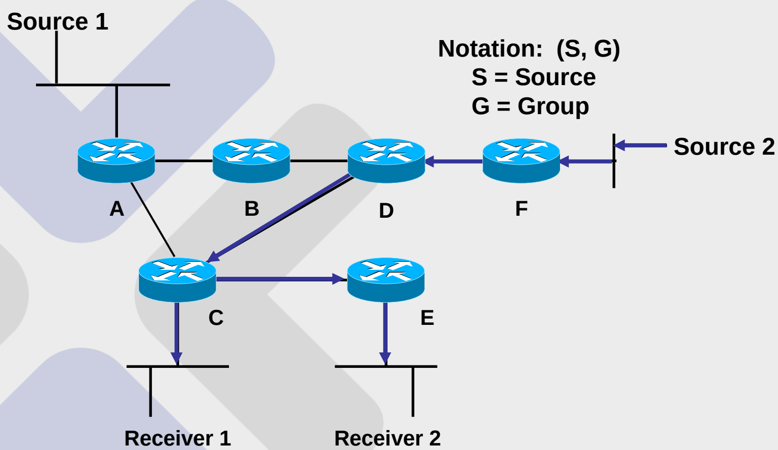
Shortest Path or Source Distribution Tree



Multicast Distribution Trees

Shortest Path or Source Distribution Tree

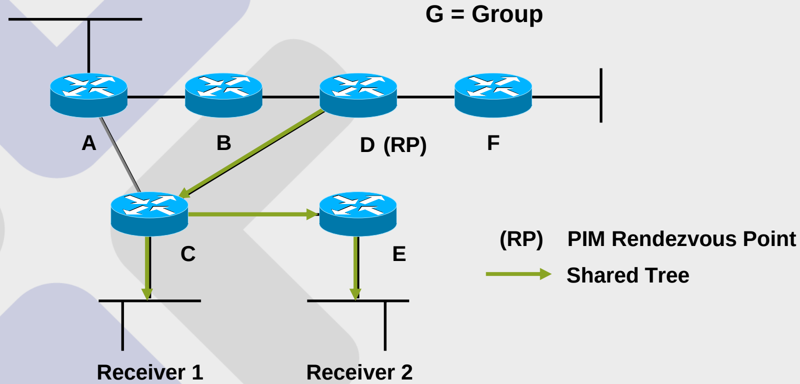
Notation: (S, G)
S = Source
G = Group



Multicast Distribution Trees

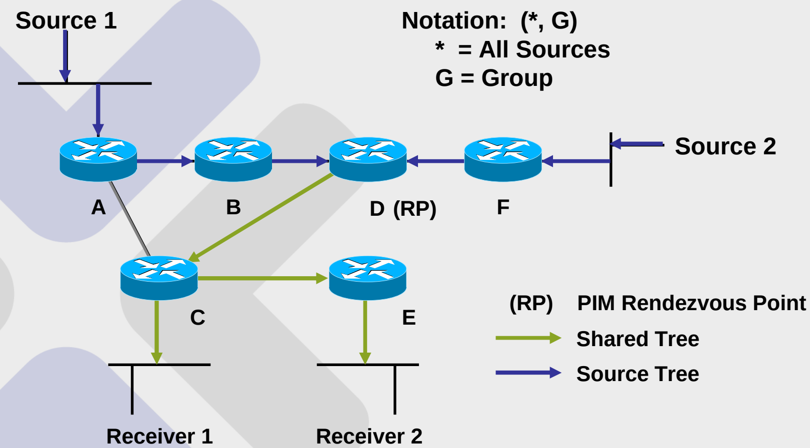
Shared Distribution Tree

Notation: (*, G)
* = All Sources
G = Group



Multicast Distribution Trees

Shared Distribution Tree



Multicast Distribution Trees

Characteristics of Distribution Trees

- **Source or shortest path trees**
  – **Uses more memory O (S x G) but you get optimal paths from source to all receivers; minimizes delay**
- **Shared trees**
  – **Uses less memory O(G) but you may get suboptimal paths from source to all receivers; may introduce extra delay**

Multicast Tree Creation

- PIM join/prune control messages
  – Used to create/remove distribution trees
- Shortest path trees
  – PIM control messages are sent toward the source
- Shared trees
  – PIM control messages are sent toward RP

Agenda

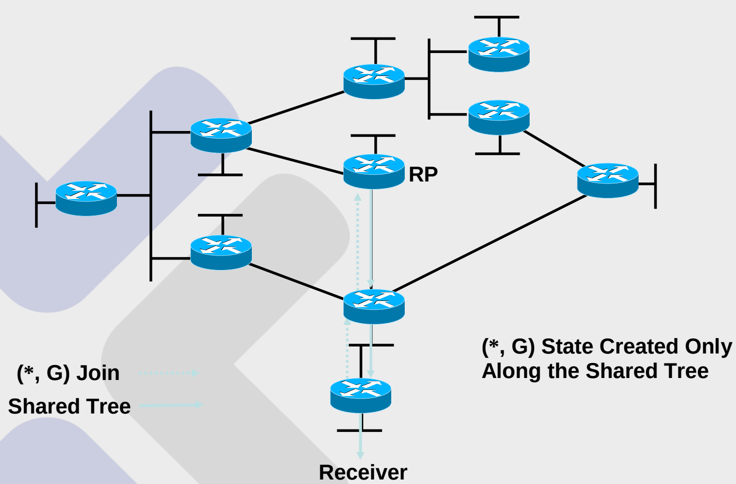
- Why Multicast?
- Multicast Fundamentals
- PIM Protocols
- RP Choices
- Multicast at Layer 2
- Interdomain IP Multicast
- Some Latest Additions

Major Deployed PIM Variants

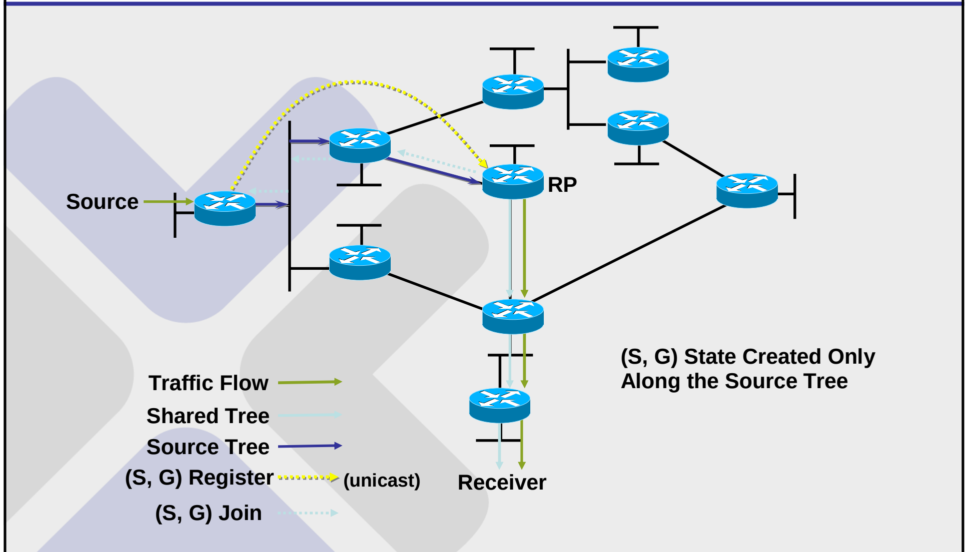
PIM-SM

- **ASM**
  - **Any Source Multicast/RP/SPT/shared tree**
- **SSM**
  - **Source Specific Multicast, no RP, SPT only**
- **BiDir**
  - **Bidirectional PIM, no SPT, shared tree only**

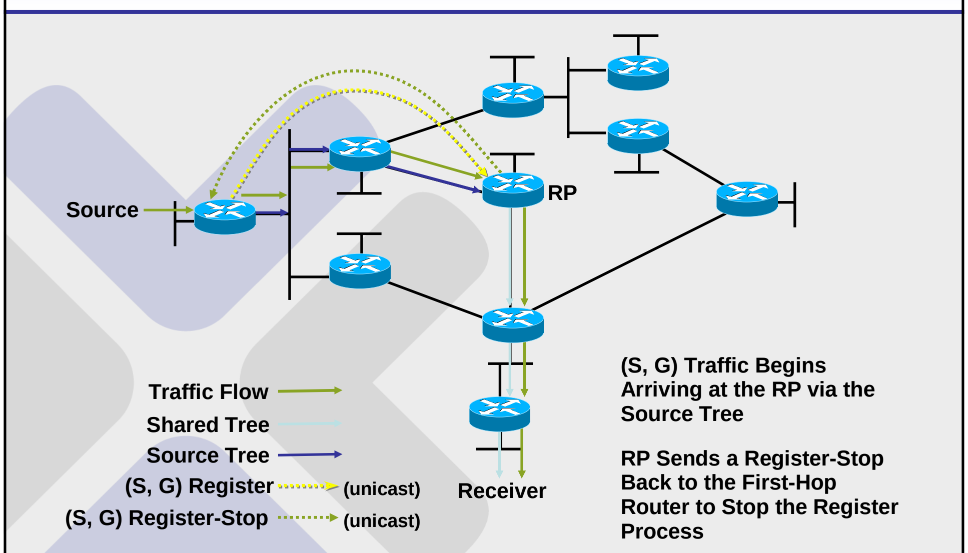
PIM-SM Shared Tree Join

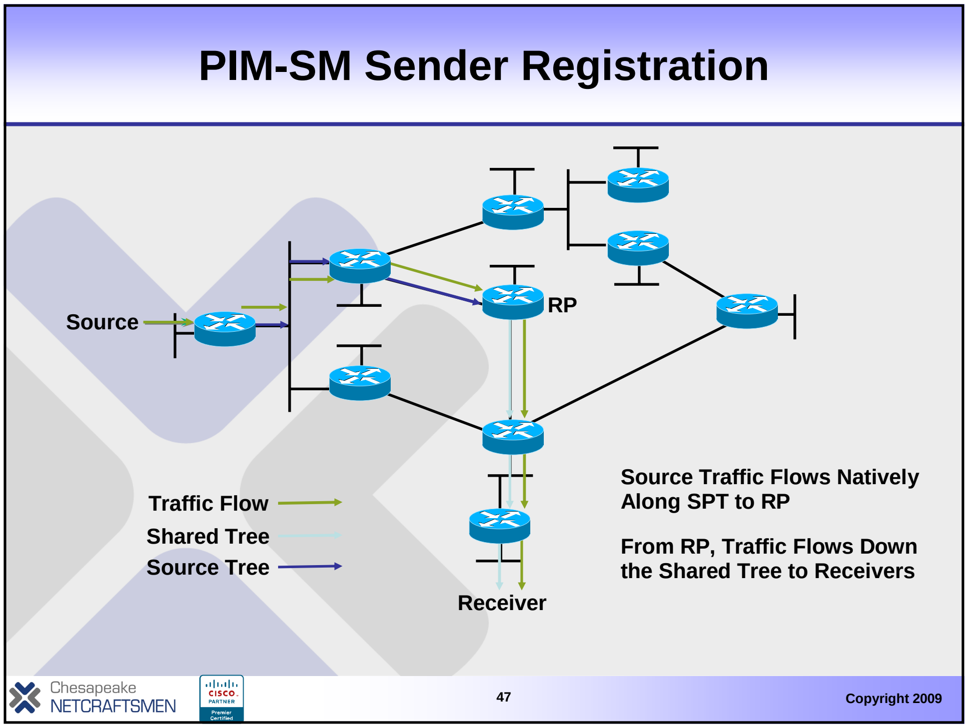


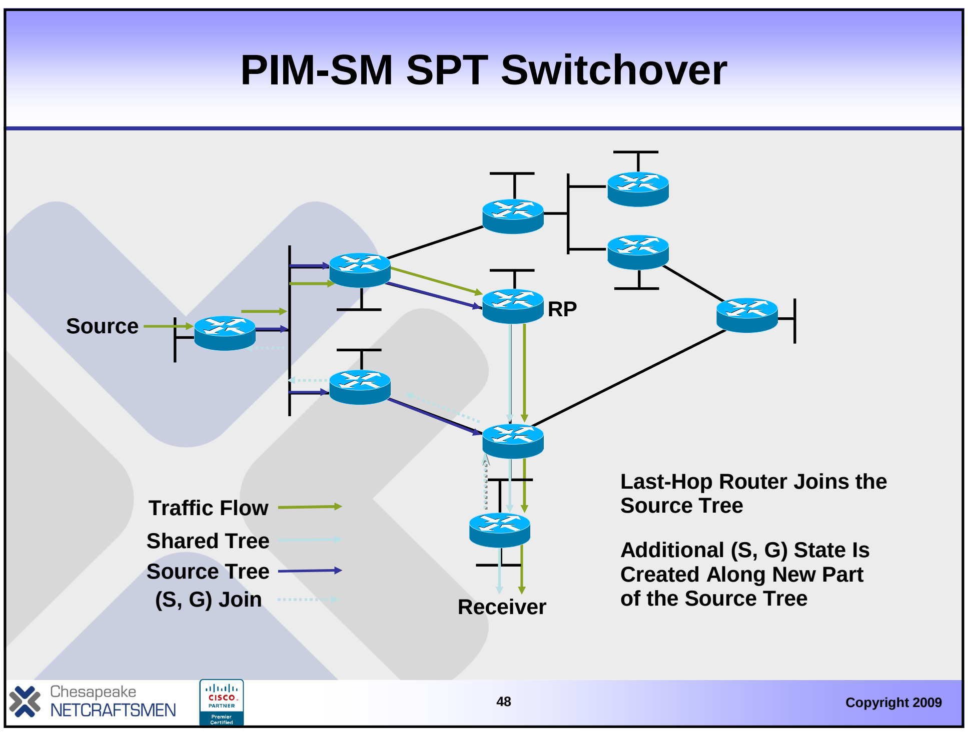
PIM-SM Sender Registration



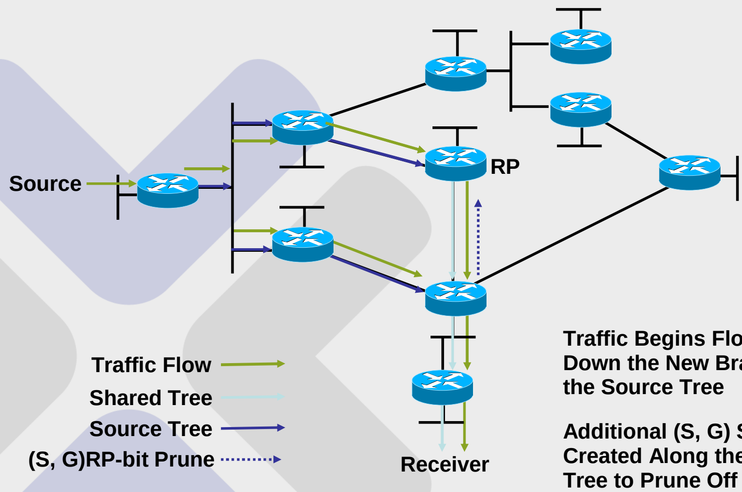
PIM-SM Sender Registration




PIM-SM Sender Registration
Source
RP
Traffic Flow
Shared Tree
Source Tree
Receiver
Source Traffic Flows Natively Along SPT to RP
From RP, Traffic Flows Down the Shared Tree to Receivers
Chesapeake NETCRAFTSMEN
47
Copyright 2009


PIM-SM SPT Switchover
Source
RP
Traffic Flow
Shared Tree
Source Tree
(S, G) Join
Receiver
Last-Hop Router Joins the Source Tree
Additional (S, G) State Is Created Along New Part of the Source Tree
Chesapeake NETCRAFTSMEN
48
Copyright 2009

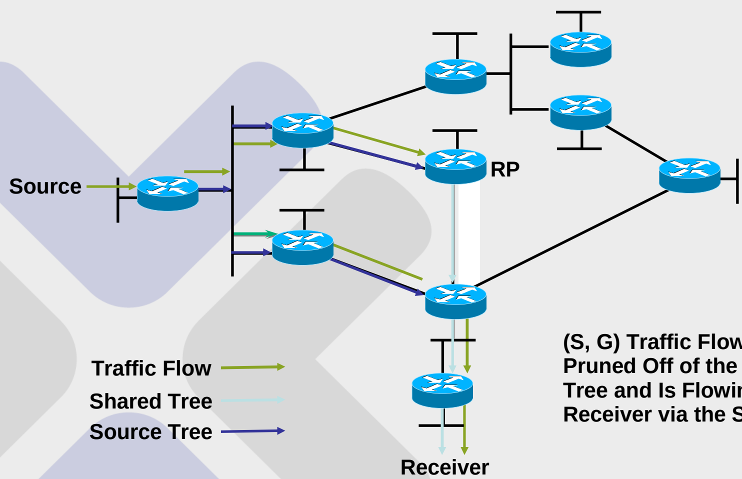
PIM-SM SPT Switchover



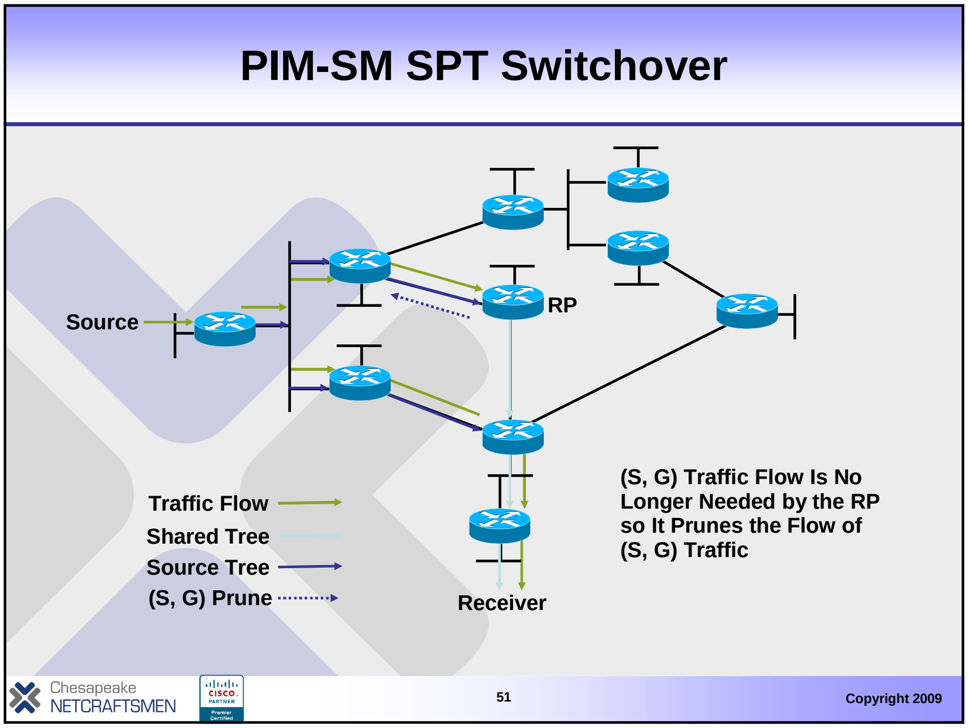
Traffic Begins Flowing Down the New Branch of the Source Tree

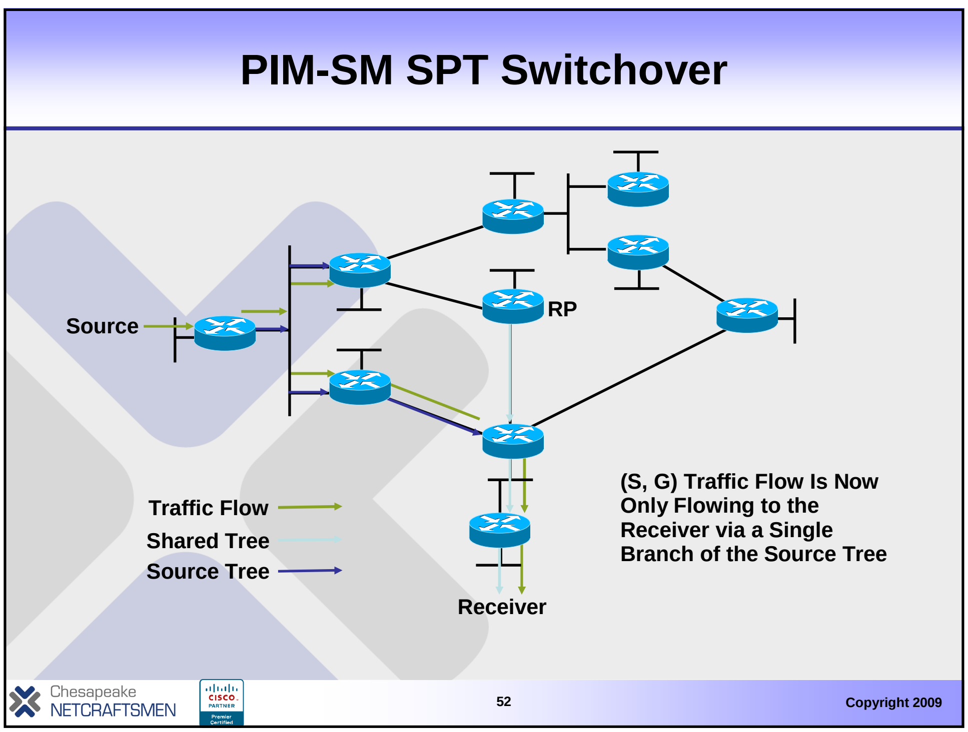
Additional (S, G) State Is Created Along the Shared Tree to Prune Off (S, G) Traffic

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
Source
RP
Receiver
Traffic Flow
Shared Tree
Source Tree
(S, G) Prune
(S, G) Traffic Flow Is No Longer Needed by the RP so It Prunes the Flow of (S, G) Traffic
Chesapeake NETCRAFTSMEN
51
Copyright 2009


PIM-SM SPT Switchover
Source
RP
Receiver
Traffic Flow
Shared Tree
Source Tree
(S, G) Traffic Flow Is Now Only Flowing to the Receiver via a Single Branch of the Source Tree
Chesapeake NETCRAFTSMEN
52
Copyright 2009

“The default behavior of PIM-SM is that routers with directly connected members will join the shortest path tree as soon as they detect a new multicast source.”

PIM-SM Frequently Forgotten Fact

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**

Source Specific Multicast

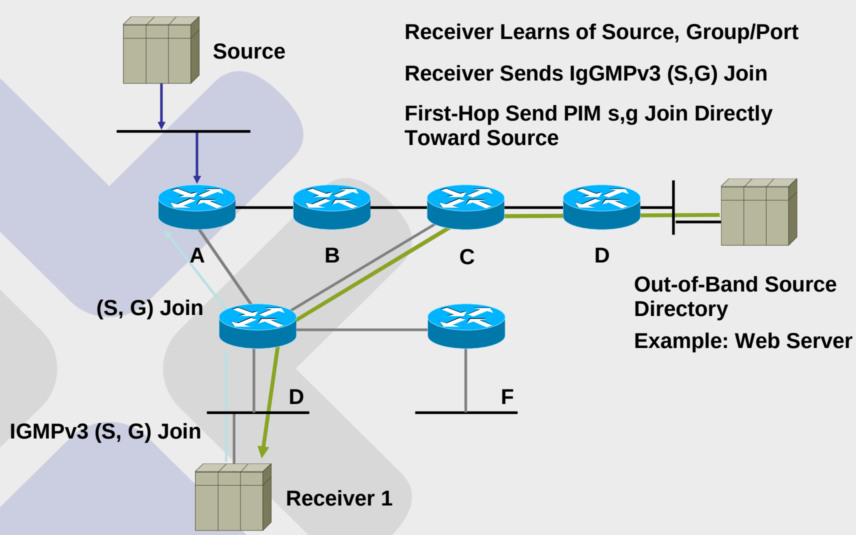
- Assume a one-to-any multicast model
  - Example: video/audio broadcasts, stock market data
- Why does ASM need a shared tree?
  - So that hosts and first hop routers can learn who the active source is for the group—source discovery
- What if this was already known?
  - Hosts could use IGMPv3 to signal exactly which (S, G) SPT to join
  - The shared tree and RP wouldn't be necessary
  - Different sources could share the same group address and not interfere with each other
- Result: Source Specific Multicast (SSM)
- RFC 3569: An Overview of Source Specific Multicast (SSM)

PIM Source Specific Mode

Receiver Learns of Source, Group/Port

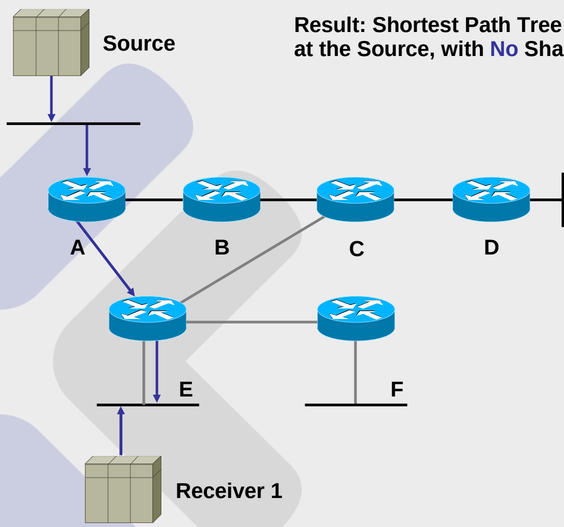
Receiver Sends IgGMPv3 (S,G) Join

First-Hop Send PIM s,g Join Directly Toward Source



PIM Source Specific Mode

Result: Shortest Path Tree Rooted at the Source, with No Shared Tree



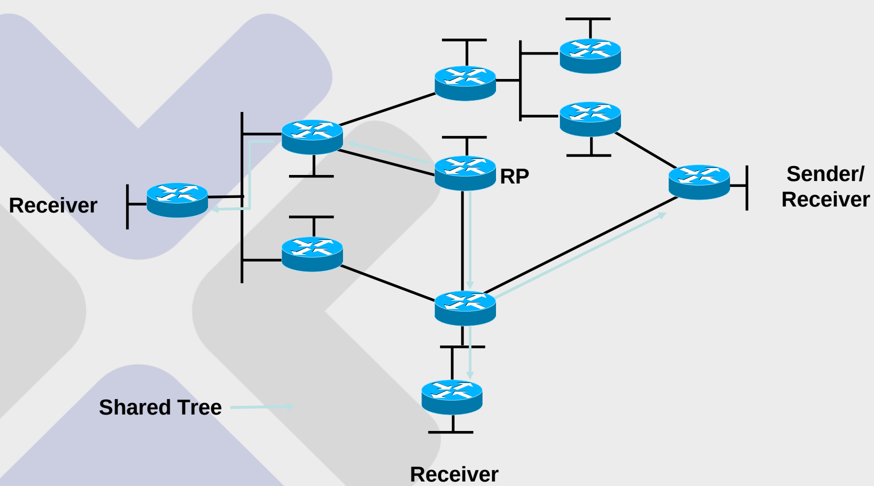
SSM—Evaluation

- Ideal for applications with one source sending to many receivers
- Uses a simplified subset of the PIM-SM protocol
  - Simpler network operation
- Solves multicast address allocation problems
  - Flows differentiated by both source and group
    - Not just by group
  - Content providers can use same group ranges
    - Since each (S,G) flow is unique
- Helps prevent certain DoS attacks
  - "Bogus" source traffic
    - Can't consume network bandwidth
    - Not received by host application

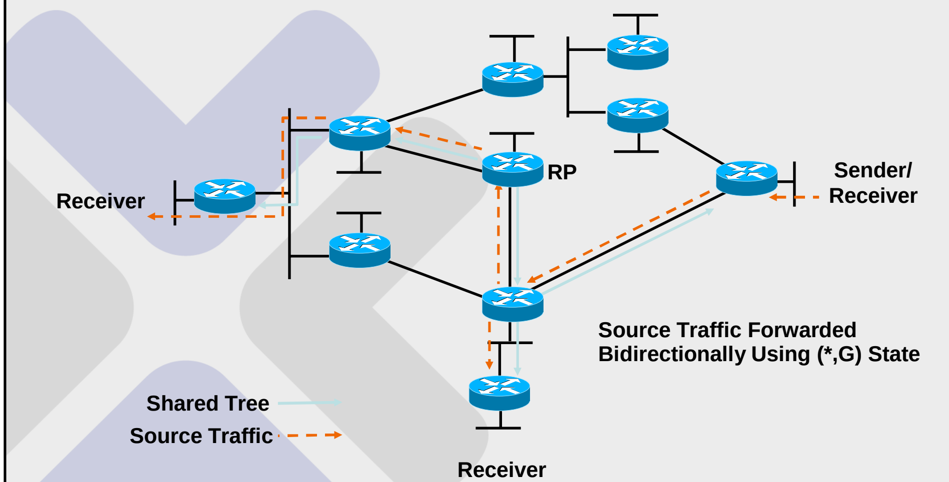
Many-to-Many State Problem

- Creates huge amounts of (S,G) state
  - State maintenance workloads skyrocket
    - High OIL fan-out makes the problem worse
  - Router performance begins to suffer
- Using shared trees only
  - 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 only uses (*, G) state

Bidirectional PIM—Overview



Bidirectional PIM—Overview



Bidir PIM—Evaluation

- Ideal for many to many applications
- 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-any applications to scale
    - Permits virtually an unlimited number of sources

Agenda

- Why Multicast?
- Multicast Fundamentals
- PIM Protocols
- RP Choices
- Multicast at Layer 2
- Interdomain IP Multicast
- Some Latest Additions

PIM-SM ASM RP Requirements

- Group to RP mapping
  - Consistent in all routers within the PIM domain
- RP redundancy requirements
  - Eliminate any single point of failure

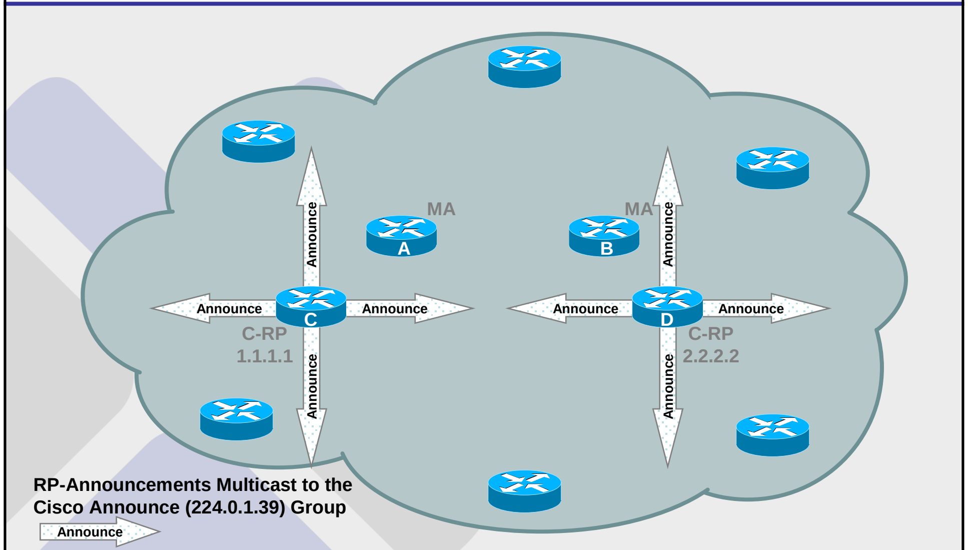
How Does the Network Know About the RP?

- Static configuration
  - Manually on every router in the PIM domain
- AutoRP
  - Originally a Cisco® solution
  - Facilitated PIM-SM early transition
- BSR
  - draft-ietf-pim-sm-bsr

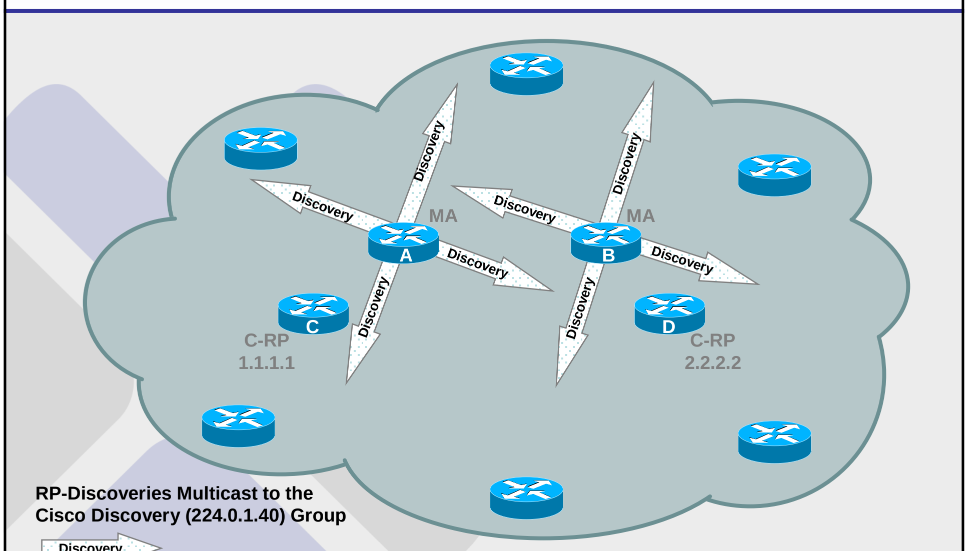
Static RPs

- Hard-configured RP address
  - When used, must be configured on every router
  - All routers must have the same RP address
  - RP failover not possible
    - Exception: if anycast RPs are used
- Command
  - `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

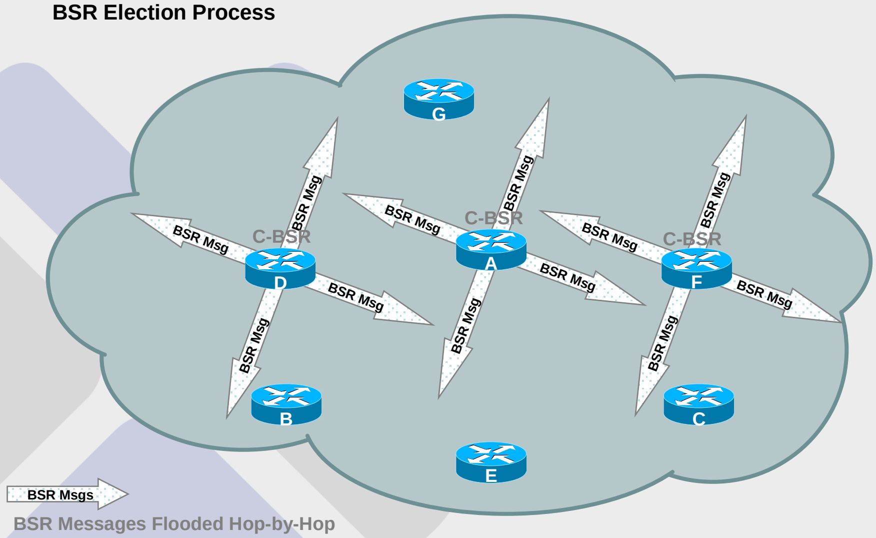


Auto-RP—From 10,000 Feet



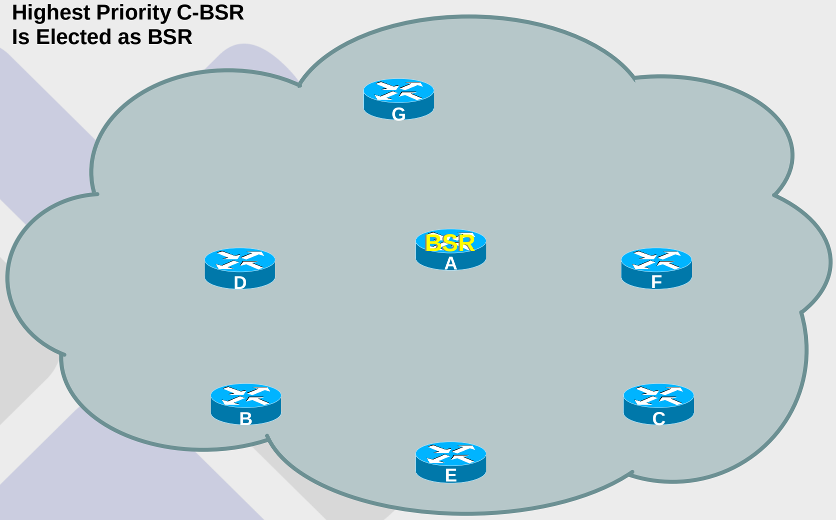
BSR—From 10,000 Feet

BSR Election Process

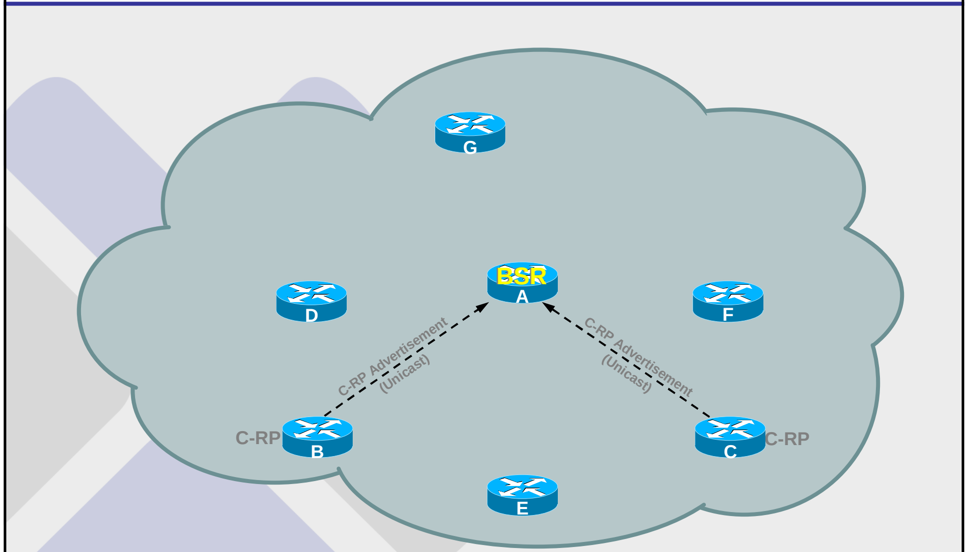


BSR—From 10,000 Feet

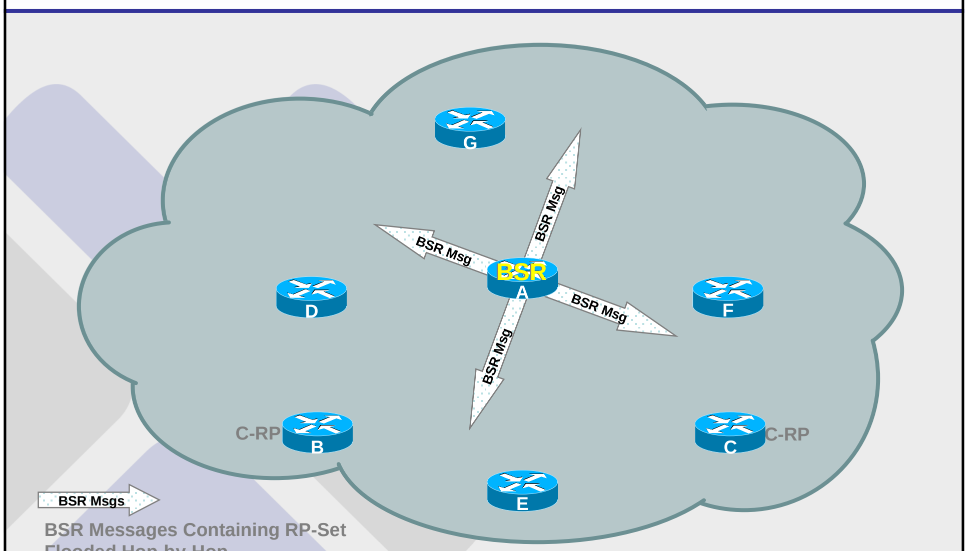
Highest Priority C-BSR
Is Elected as BSR



BSR—From 10,000 Feet



BSR—From 10,000 Feet



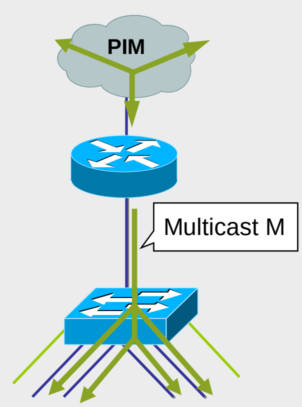
Agenda

- Why Multicast?
- Multicast Fundamentals
- PIM Protocols
- RP Choices
- Multicast at Layer 2
- Interdomain IP Multicast
- Some Latest Additions

L2 Multicast Frame Switching

Problem: Layer 2 Flooding of Multicast Frames

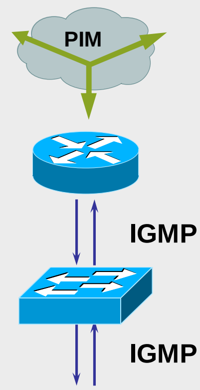
- Typical L2 switches treat multicast traffic as unknown or broadcast and must "flood" the frame to every port
- Static entries can sometimes be set to specify which ports should receive which group(s) of multicast traffic
- Dynamic configuration of these entries would cut down on user administration



L2 Multicast Frame Switching

IGMPv1–v2 Snooping

- **Switches become "IGMP"-aware**
- **IGMP packets intercepted by the NMP or by special hardware 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

Impact of IGMPv3 on ICMP Snooping

- **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—Frame Switches

IGMP Snooping

- 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!
- Shouldn't be a problem when IGMPv3 is implemented

77 Copyright 2009

Agenda

- Why Multicast?
- Multicast Fundamentals
- PIM Protocols
- RP Choices
- Multicast at Layer 2
- Interdomain IP Multicast
- Some Latest Additions

78 Copyright 2009

MBGP Overview

MBGP: Multiprotocol BGP

- Defined in RFC 2858 (extensions to BGP)
- Can carry different types of routes
  - Unicast
  - Multicast
- 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 Overview

- Separate BGP tables maintained
  - Unicast prefixes for unicast forwarding
  - Unicast prefixes for multicast RPF checking
- 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 Overview

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

Chesapeake NETCRAFTSMEN 81 Copyright 2009

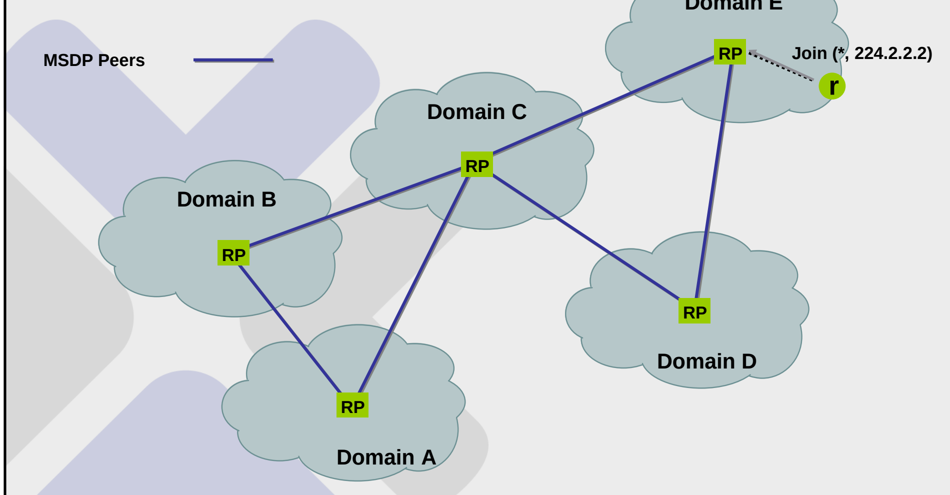
MSDP

- RFC 3618
- ASM only
  - RPs knows 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 required for interdomain ASM source discovery

Chesapeake NETCRAFTSMEN 82 Copyright 2009

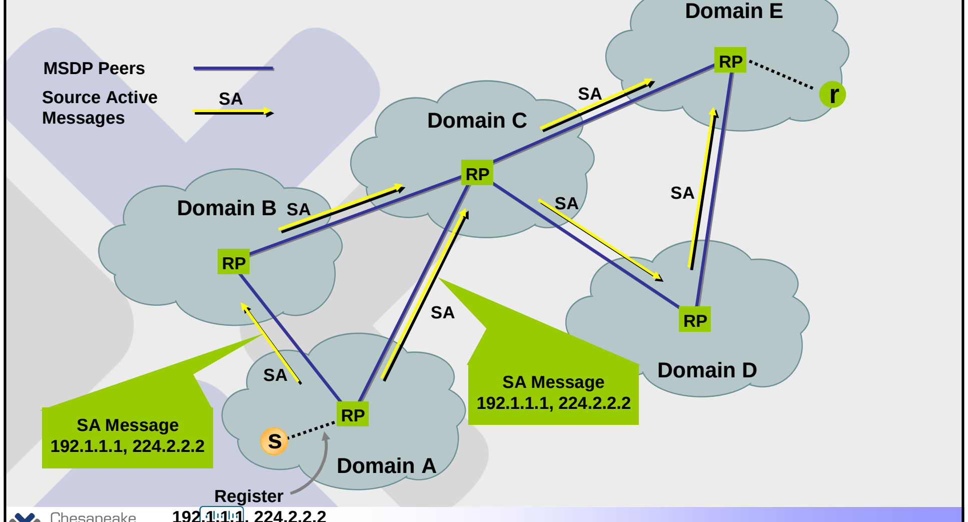
MSDP Overview

MSDP Example



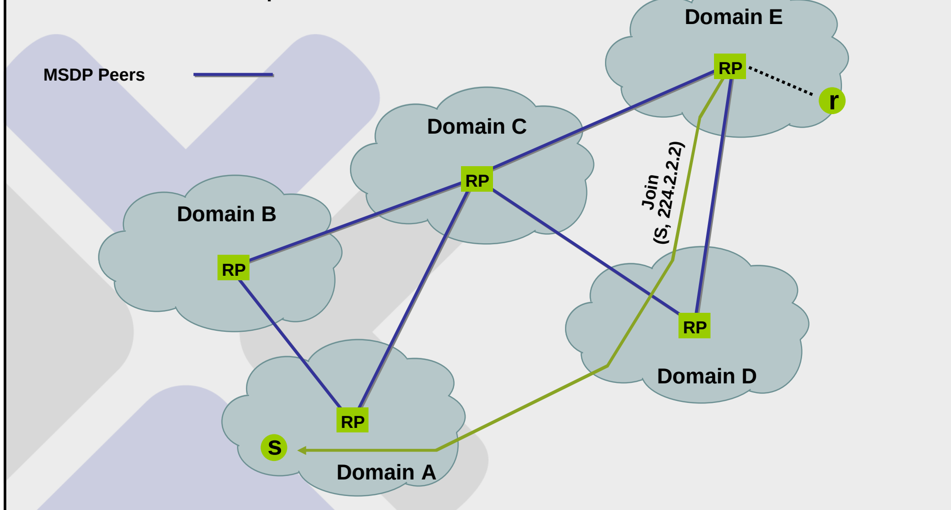
MSDP Overview

MSDP Example



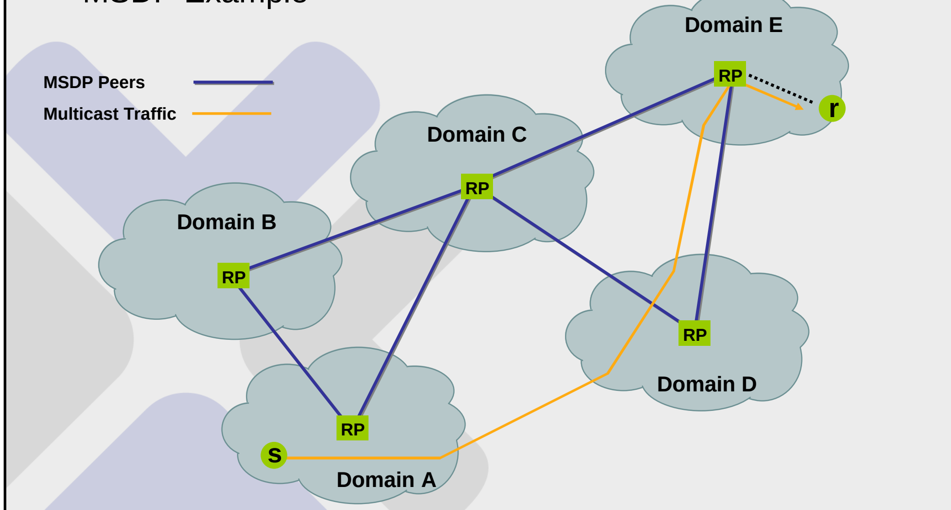
MSDP Overview

MSDP Example



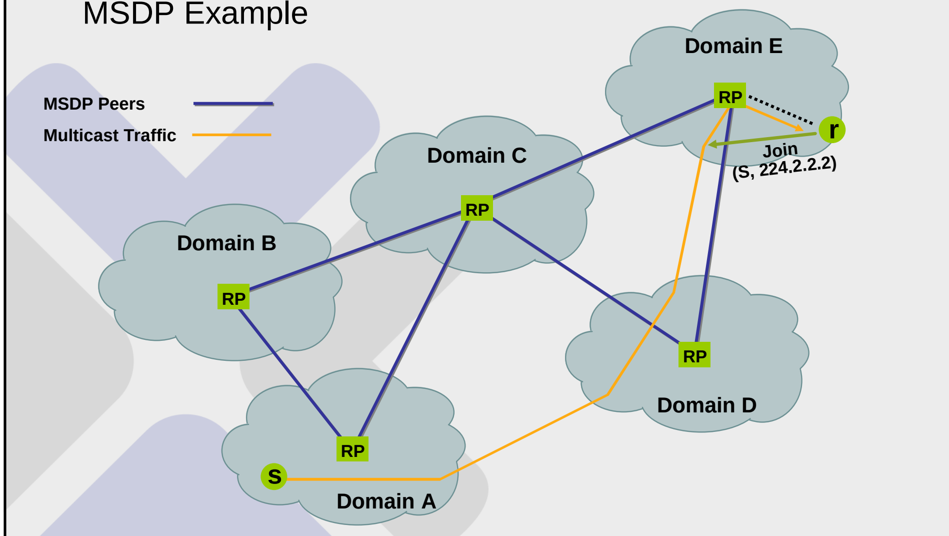
MSDP Overview

MSDP Example



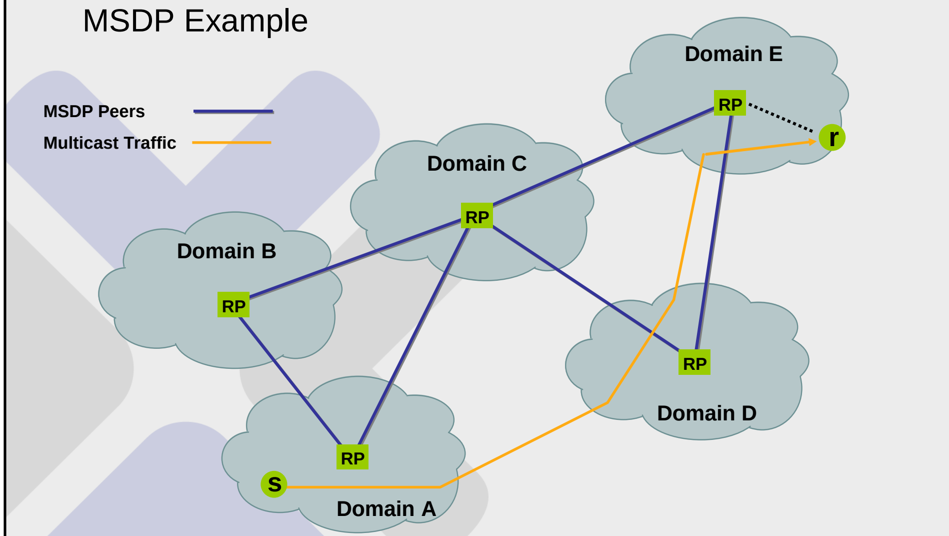
MSDP Overview

MSDP Example

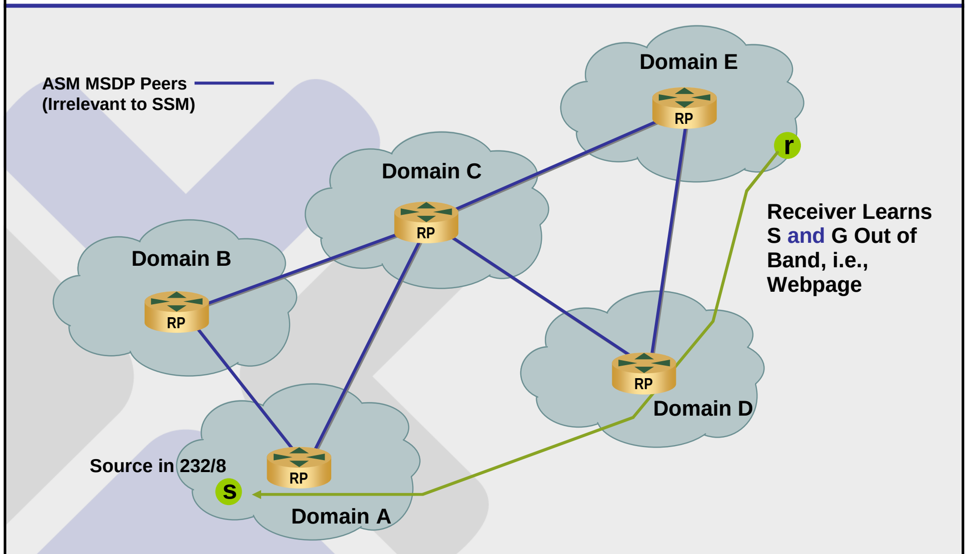


MSDP Overview

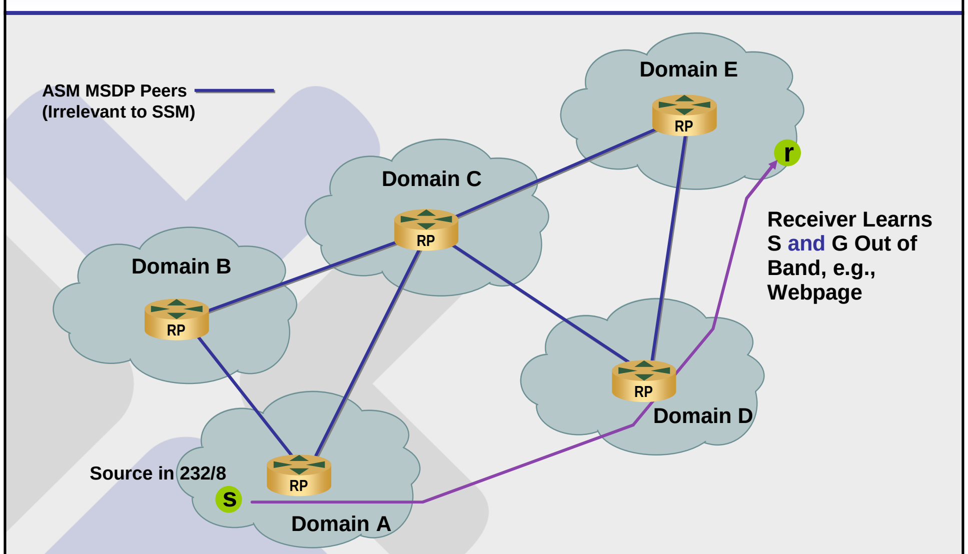
MSDP Example



MSDP wrt SSM—Unnecessary



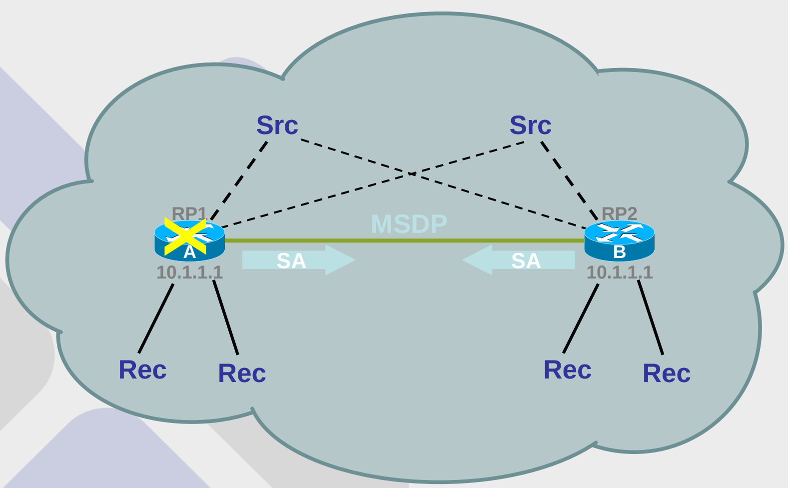
MSDP wrt SSM—Unnecessary



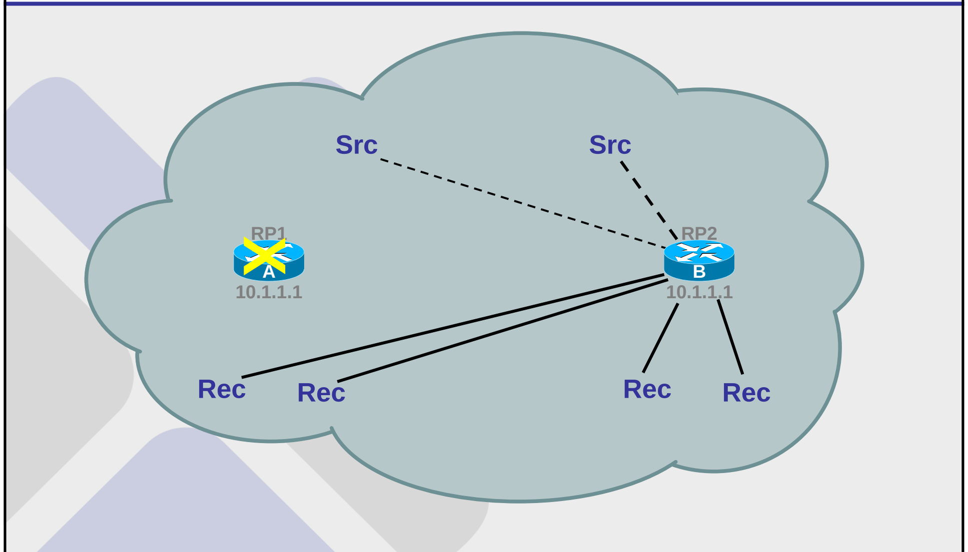
Anycast RP—Overview

- Redundant RP technique for ASM which uses MSDP for RP synchronization
- Uses single defined RP address
  - Two or more routers have same RP address
    - RP address defined as a loopback interface
    - Loopback address advertised as a host route
  - Senders and receivers join/register with closest RP
    - Closest RP determined from the unicast routing table
    - Because RP is statically defined
- MSDP session(s) run between all RPs
  - Informs RPs of sources in other parts of network
  - RPs join SPT to active sources as necessary

Anycast RP—Overview



Anycast RP—Overview

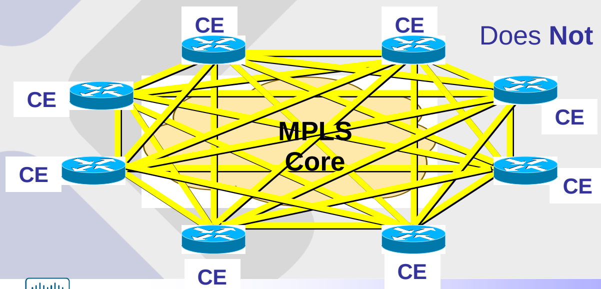


Agenda

- Why Multicast?
- Multicast Fundamentals
- PIM Protocols
- RP Choices
- Multicast at Layer 2
- Interdomain IP Multicast
- Some Recent Additions

Multicast VPN—Customer Requirement

- MPLS VPN customers want to run multicast within their VPNs
- Multicast deployment is expanding
- MPLS VPNs do not support multicast today
- Multicast options in MPLS VPNs today
  - GRE tunnels from CE to CE



Does **Not** Scale

Multicast VPN (MVPN)

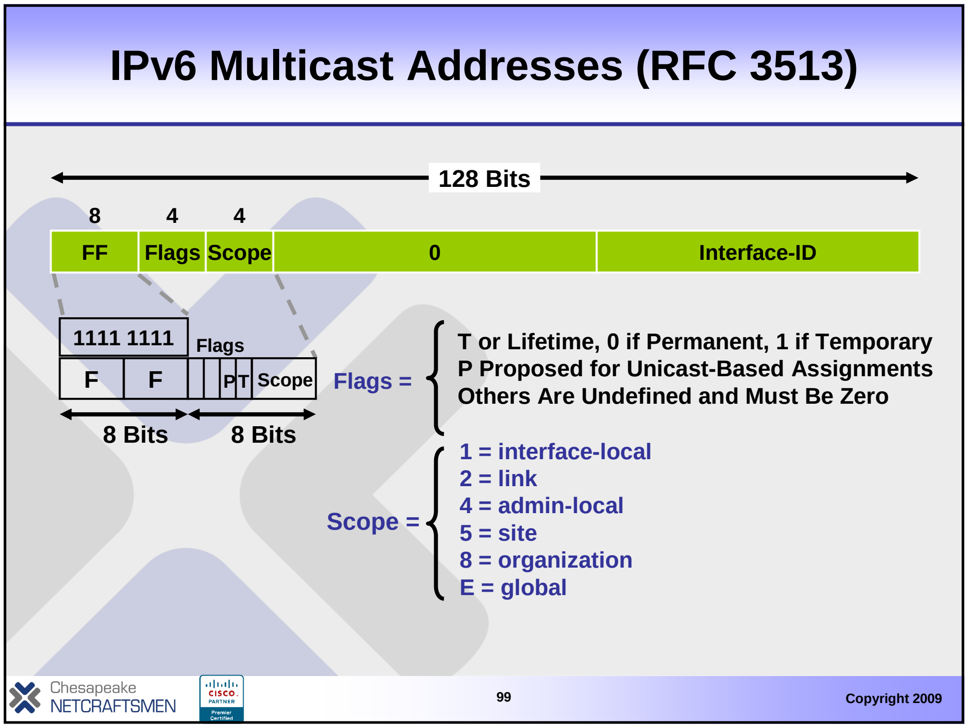
- Allows an ISP to provide its MPLS VPN customers the ability to transport their multicast traffic across MPLS packet-based core networks
- Requires IPmc enabled in the core
- MPLS may still be used to support unicast
- A scalable architecture solution for MPLS networks based on native multicast deployment in the core

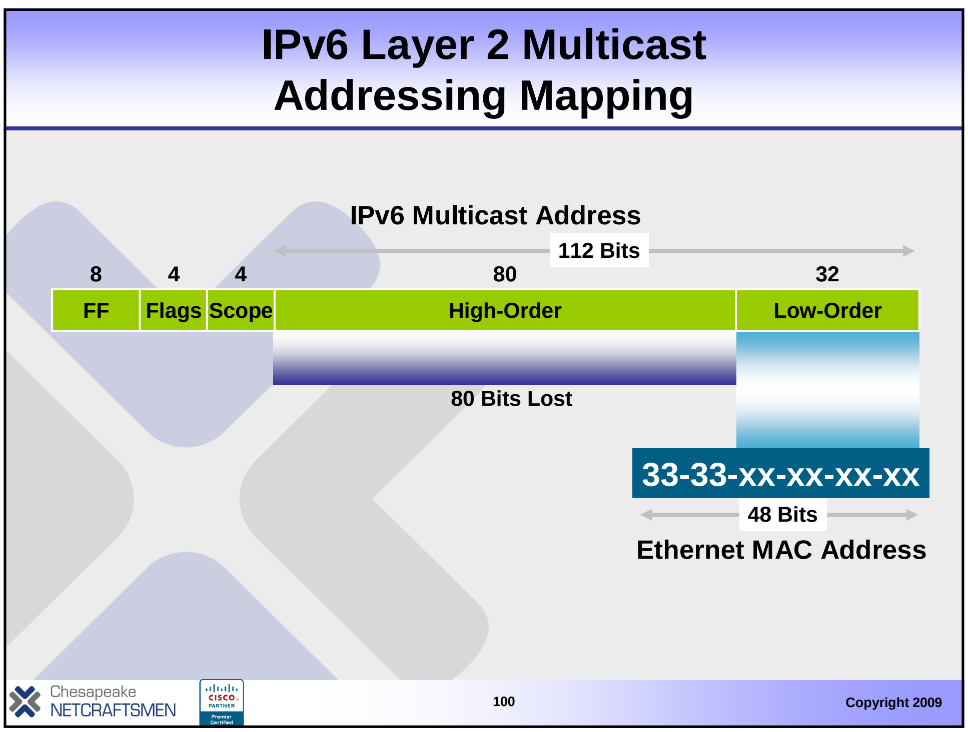
Multicast VPN (MVPN)

- Uses draft-rosen-vpn-mcast encapsulation and signaling to build MVPN Multicast VPN (MVPN)'
  - GRE encapsulation
  - PIM inside PIM
- Not universally deployed
  - Not all VPN providers offer MVPN services

IPv4 Versus IPv6 Multicast

IP Service	IPv4 Solution	IPv6 Solution
Address Range	32-Bit, Class D	128-Bit (112-Bit Group)
Routing	Protocol-Independent All IGPs and GBP4+	Protocol-Independent All IGPs and BGP4+ with v6 Mcast SAFI
Forwarding	PIM-DM, PIM-SM: ASM, SSM, BiDir	PIM-SM: ASM, SSM, BiDir
Group Management	IBMPv1, v2, v3	MLDv1, v2
Domain Control	Boundary/Border	Scope Identifier
Interdomain Source Discovery	MSDP Across Independent PIM Domains	Single RP Within Globally Shared Domains


IPv6 Multicast Addresses (RFC 3513)
128 Bits
8
4
4
FF
Flags
Scope
0
Interface-ID
1111 1111
Flags
F
F
P
T
Scope
8 Bits
8 Bits
Flags =
T or Lifetime, 0 if Permanent, 1 if Temporary
P Proposed for Unicast-Based Assignments
Others Are Undefined and Must Be Zero
Scope =
1 = interface-local
2 = link
4 = admin-local
5 = site
8 = organization
E = global
Chesapeake NETCRAFTSMEN
99
Copyright 2009


IPv6 Layer 2 Multicast Addressing Mapping
IPv6 Multicast Address
112 Bits
8
4
4
80
32
FF
Flags
Scope
High-Order
Low-Order
80 Bits Lost
33-33-xx-xx-xx-xx
48 Bits
Ethernet MAC Address
Chesapeake NETCRAFTSMEN
100
Copyright 2009

Unicast-Based Multicast Addresses

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

- 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
    - 1234:5678:9abc::/64
  - Multicast address
    - FF36:0030:1234:5678:9abc::0001

Chesapeake NETCRAFTSMEN 101

IP Routing for Multicast

- RPF-based on reachability to v6 source same as
  with v4 multicast
- RPF still protocol-independent
  - Static routes, mroutes
  - Unicast RIB: BGP, ISIS, OSPF, EIGRP, RIP, etc.
  - Multiprotocol BGP (mBGP)
    - Support for v6 mcast subaddress family
    - Provide translate function for nonsupporting peers

Chesapeake NETCRAFTSMEN 102

IPv6 Multicast Forwarding

- PIM-Sparse Mode (PIM-SM)
  - RFC4601
- PIM Source Specific Mode (SSM)
  - RFC3569 SSM overview (v6 SSM needs MLDv2)
  - Unicast, prefix-based multicast addresses ff30::/12
  - SSM range is ff3X::/32
    - Current allocation is from ff3X::/96
- PIM BiDirectional Mode (BiDir)
  - draft-ietf-pim-bidir-09.txt

103 Copyright 2009

RP Mapping Mechanisms for IPv6 PIM-SM

- Static RP assignment
- BSR
- Auto-RP—no current plans
- Embedded RP

104 Copyright 2009

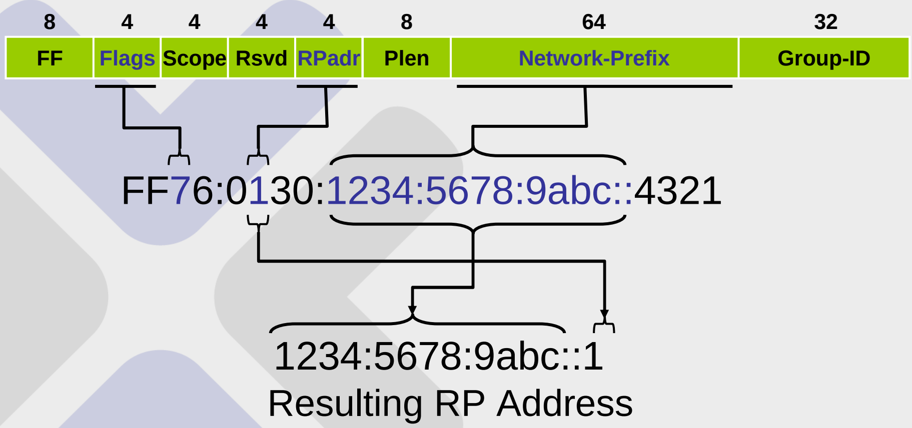
Embedded RP Addressing—RFC3956

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

- 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 → 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)

Embedded RP Addressing—Example

Multicast Address with Embedded RP Address

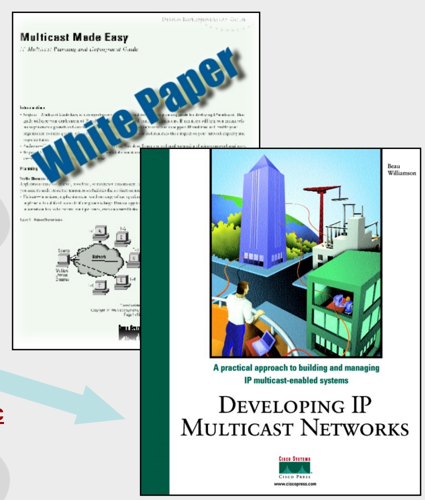


Multicast Listener Discover—MLD

- MLD is equivalent to IGMP in IPv4
- MLD messages are transported over ICMPv6
- Version number confusion
  - MLDv1 corresponds to IGMPv2
    - RFC 2710
  - MLDv2 corresponds to IGMPv3, needed for SSM
    - RFC 3810
- MLD snooping
  - draft-ietf-magma-snoop-12.txt

More Information

- White papers
- Web and mailers
- Cisco Press®



RTFB

- CCO multicast
  - http://www.cisco.com/go/ipmulticast
- Customer support mailing list
  - tac@cisco.com

RTFB = "Read the Fine Book"

Any Questions?

- For a copy of the presentation, email me at pjw@netcraftsmen.net
- References: see web article I will post at
  http://www.netcraftsmen.net/welcher/papers/index.htm
- About Chesapeake Netcraftsmen:
  - Cisco Premier Partner
  - Cisco Customer Satisfaction Excellence rating
  - We wrote the original version of the Express Foundations courses required for VAR Premier Partner status (and took and passed the tests), and the recent major CCDA/CCDP refresh
  - Cisco Advanced Specializations:
    - Advanced Unified Communications (and IP Telephony)
    - Advanced Wireless
    - Advanced Security
    - Advanced Routing & Switching
    - Advanced Data Center Networking Infrastructure
  - We have deep expertise in Routing and Switching (several R&S and two double CCIE's)
  - We do network / security / net mgmt / unified communications Design and Assessment
  - Expertise and experience in many other areas as well

Chesapeake NETCRAFTSMEN 109 Copyright 2009

Chesapeake NETCRAFTSMEN 110 Copyright 2009

So You Still Want More?

- Internet IP multicast
- AMT—Automatic Multicast Tunneling

Internet IP Multicast

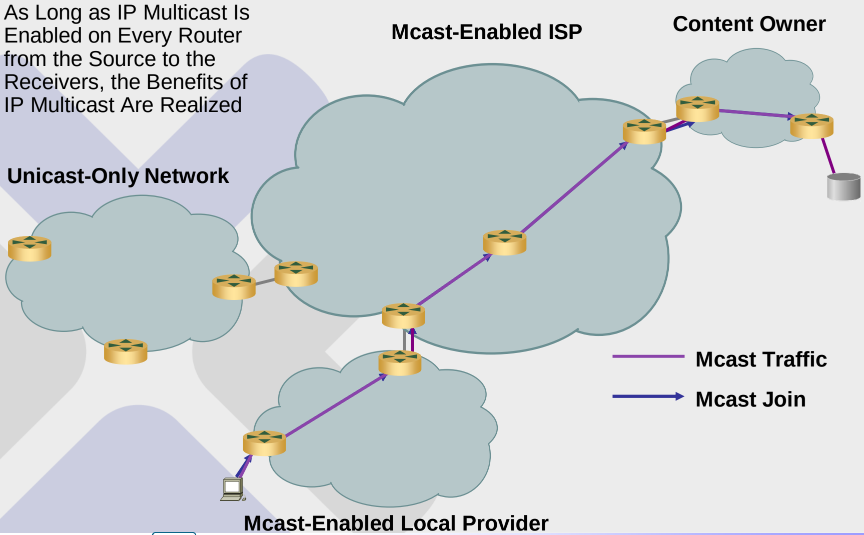
- We can build multicast distribution trees.
  - PIM
- We can RPF on interdomain sources
  - MBGP
- We can find interdomain active ASM sources
  - MSDP
- So interdomain IP Multicast is in every home, right?

Internet IP Multicast

- What worked?
- What didn't work?
- What's being done to fix it?

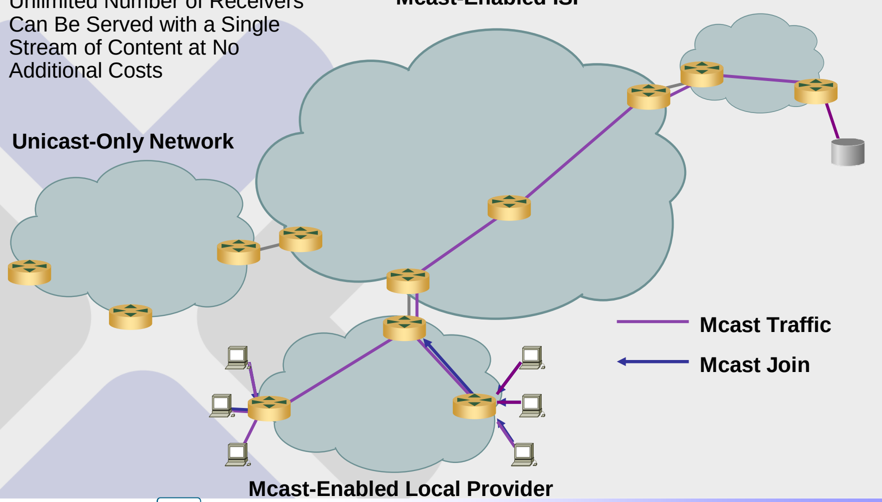
What Worked?

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



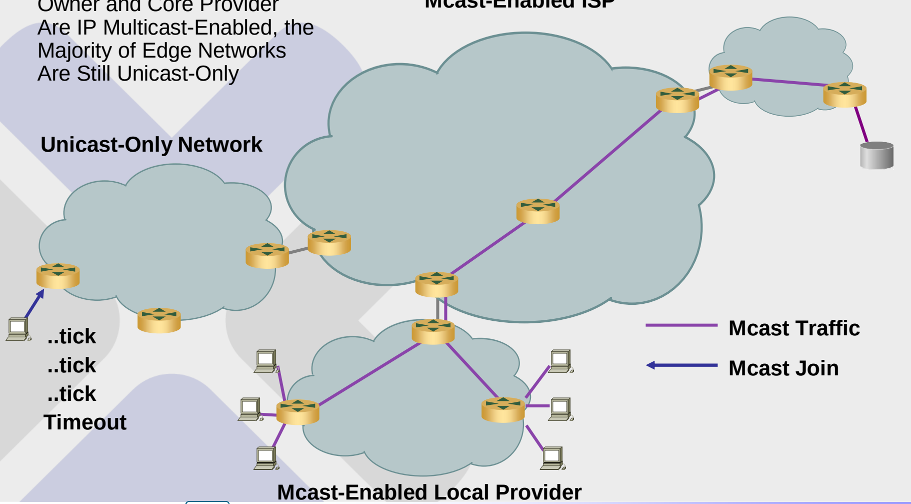
What Worked?

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

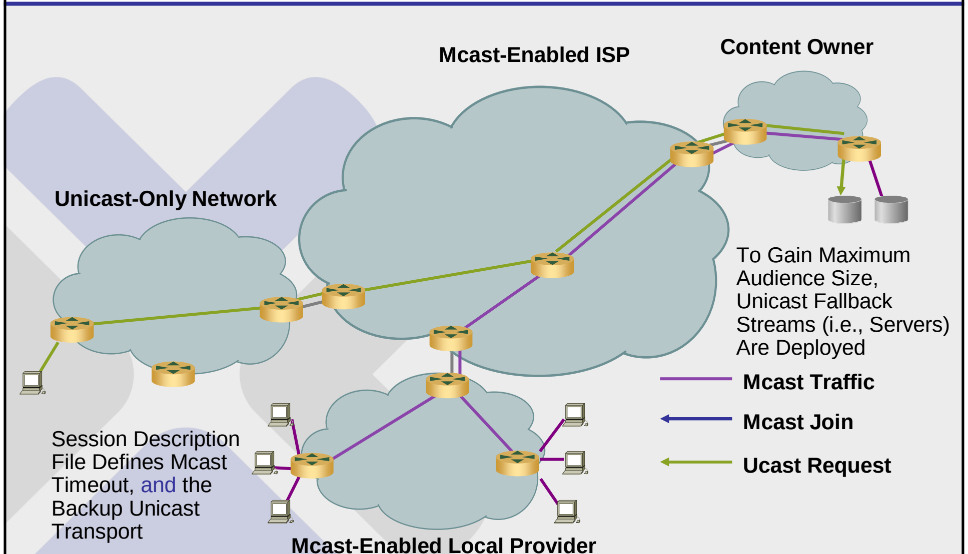


What Didn't?

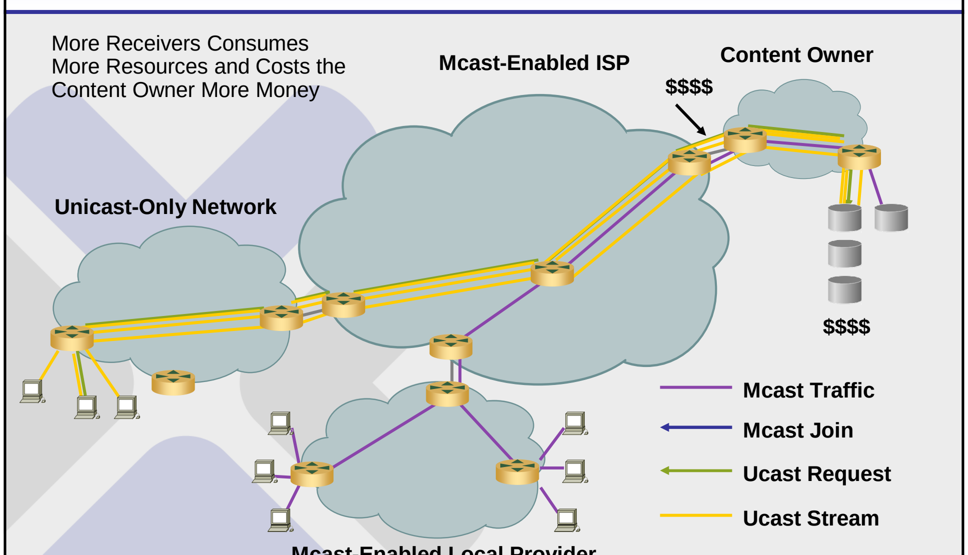
Even Though the Content Owner and Core Provider Are IP Multicast-Enabled, the Majority of Edge Networks Are Still Unicast-Only



What Didn't?



What Didn't?

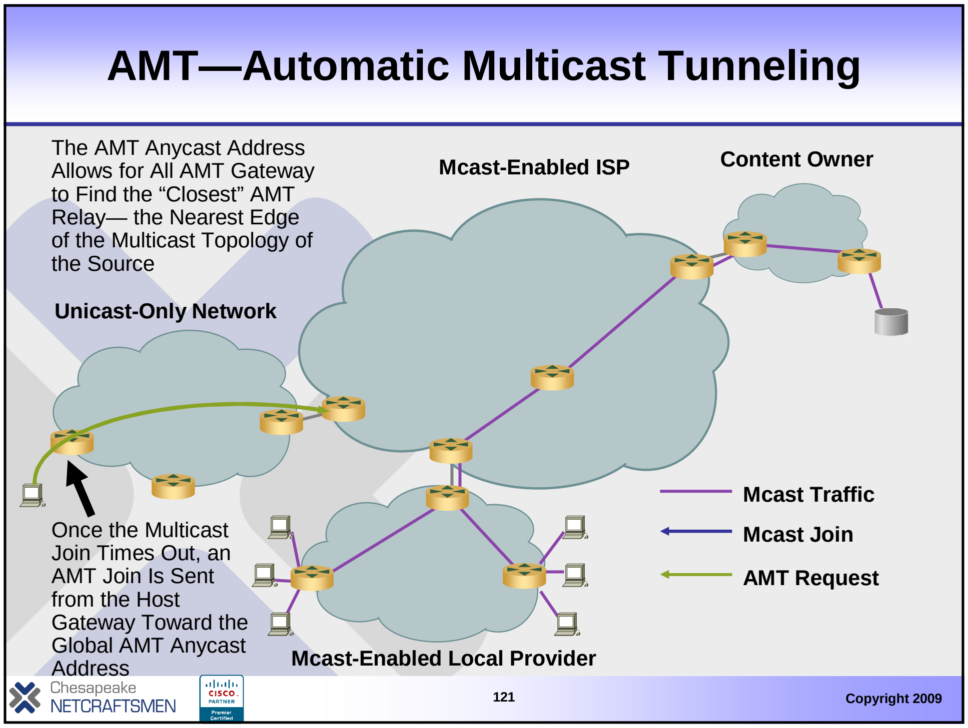


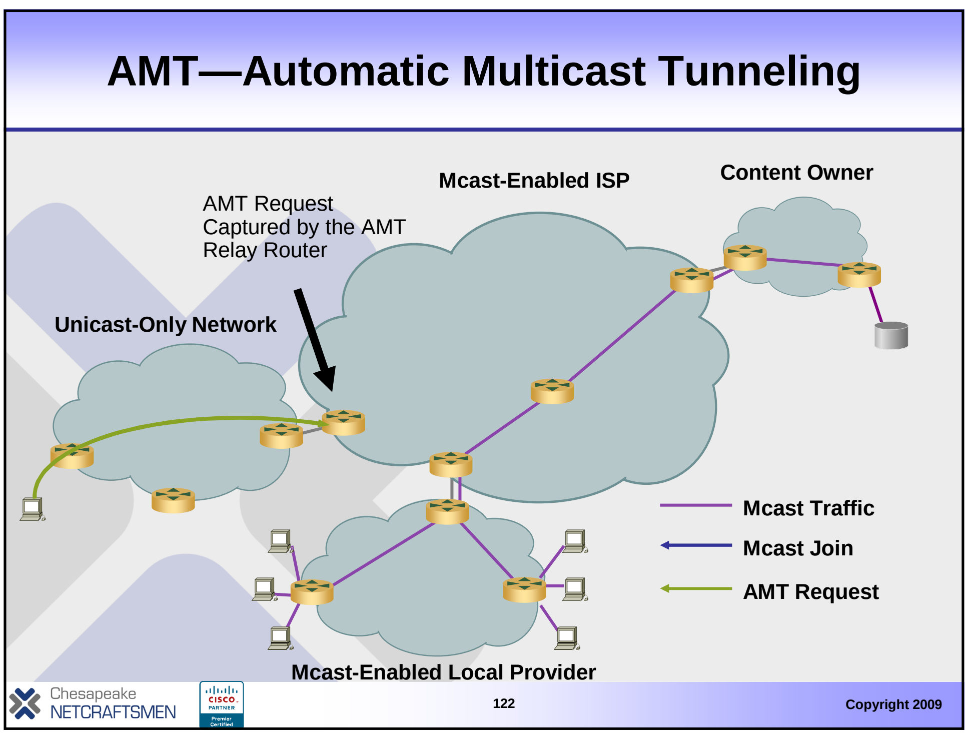
What's Wrong?

- 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
- Even Mcast-aware content owners are forced to provide unicast streams to gain audience size
- Unicast will never scale for streaming content
  - Splitters/caches just distribute the problem
    - Still has a cost per user
  - As receiver BW increases, problem gets worse
  - Creates a nonfunctional business model
  - Will never bring rich content to IP

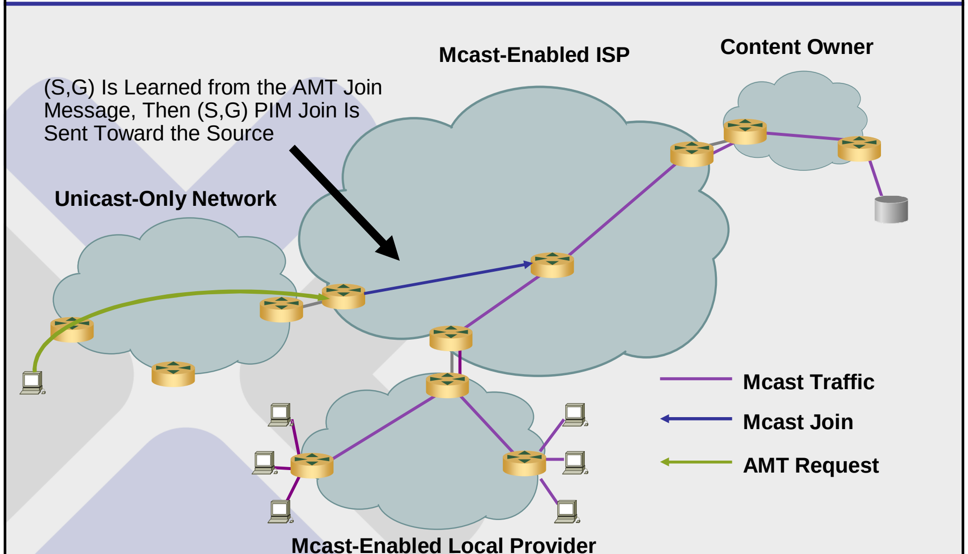
AMT—Automatic Multicast Tunneling

- Automatic IP multicast without explicit tunnels
  - http://www.ietf.org/internet-drafts/draft-ietf-mboned-auto-multicast-X.txt
- 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 which have deployed multicast benefit from their deployment
- Work seamlessly with existing applications
  - No OS kernel changes

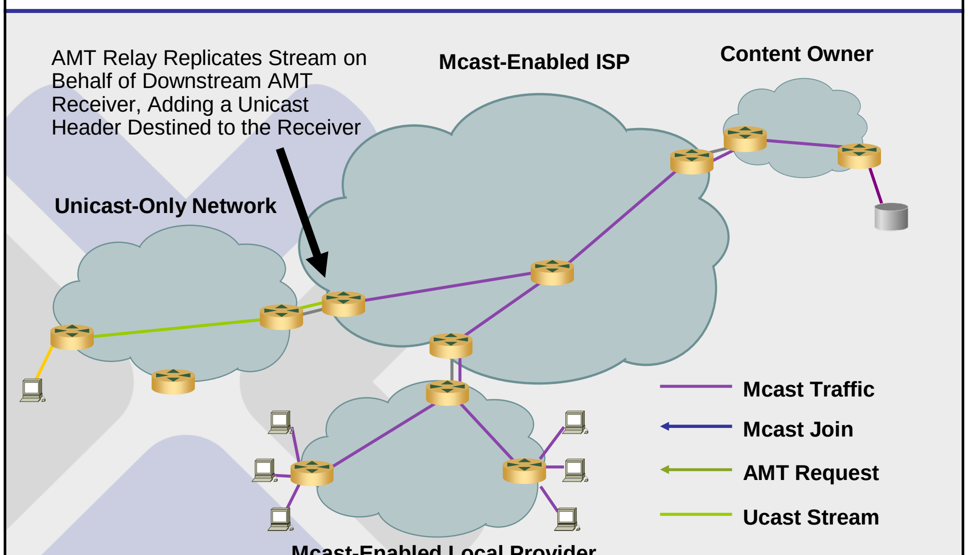

AMT—Automatic Multicast Tunneling
The AMT Anycast Address Allows for All AMT Gateway to Find the "Closest" AMT Relay— the Nearest Edge of the Multicast Topology of the Source
Mcast-Enabled ISP
Content Owner
Unicast-Only Network
Once the Multicast Join Times Out, an AMT Join Is Sent from the Host Gateway Toward the Global AMT Anycast Address
Mcast Traffic
Mcast Join
AMT Request
Mcast-Enabled Local Provider
Chesapeake NETCRAFTSMEN
121
Copyright 2009

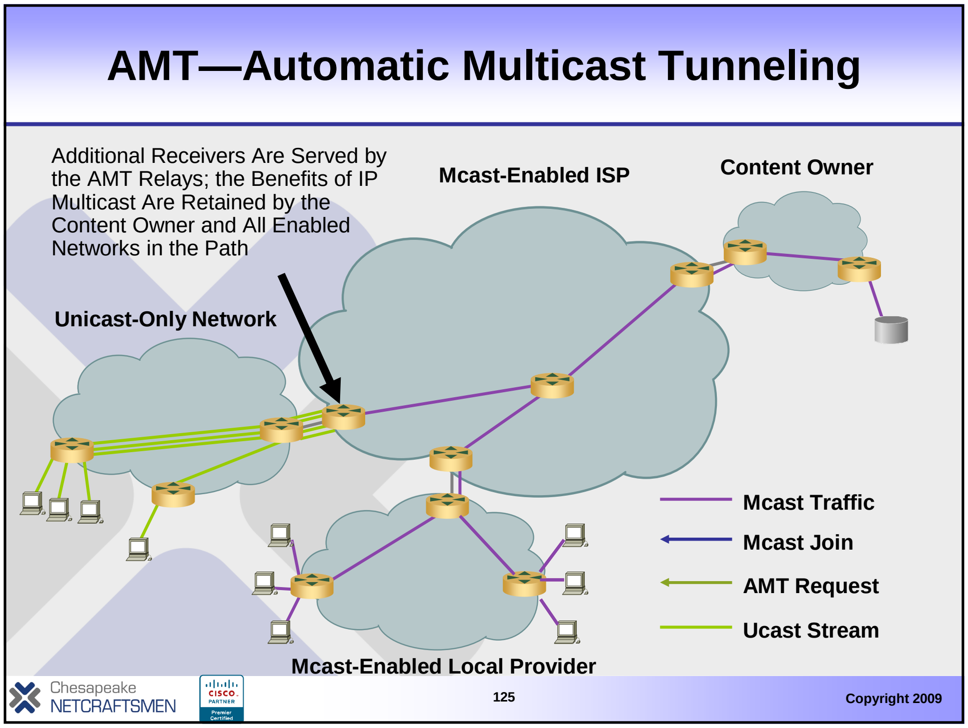

AMT—Automatic Multicast Tunneling
AMT Request Captured by the AMT Relay Router
Mcast-Enabled ISP
Content Owner
Unicast-Only Network
Mcast Traffic
Mcast Join
AMT Request
Mcast-Enabled Local Provider
Chesapeake NETCRAFTSMEN
122
Copyright 2009

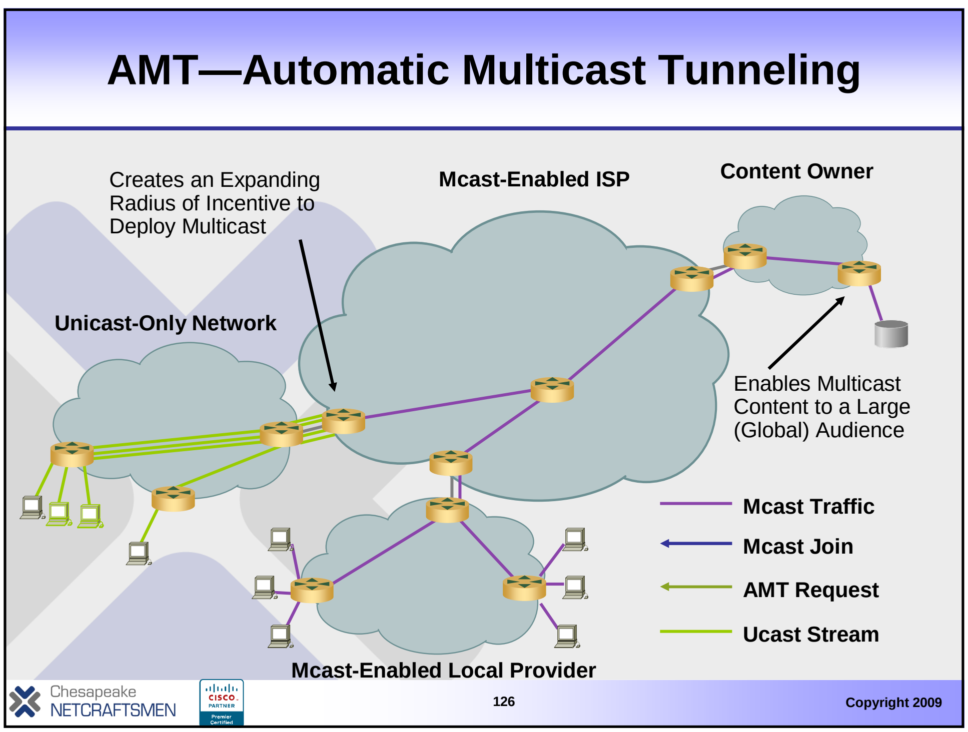
AMT—Automatic Multicast Tunneling

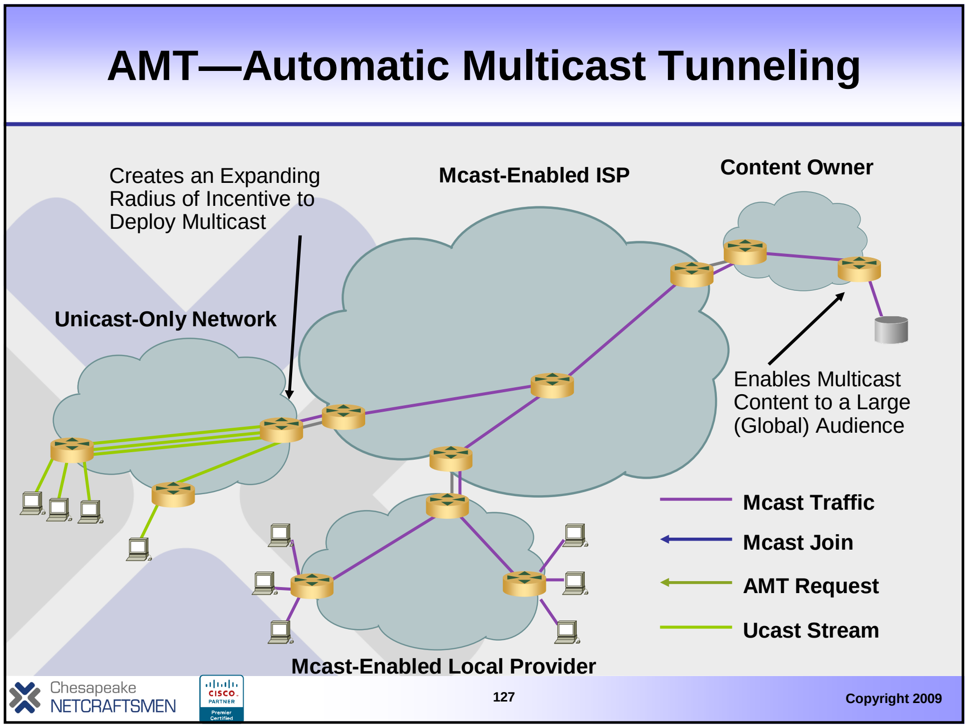


AMT—Automatic Multicast Tunneling

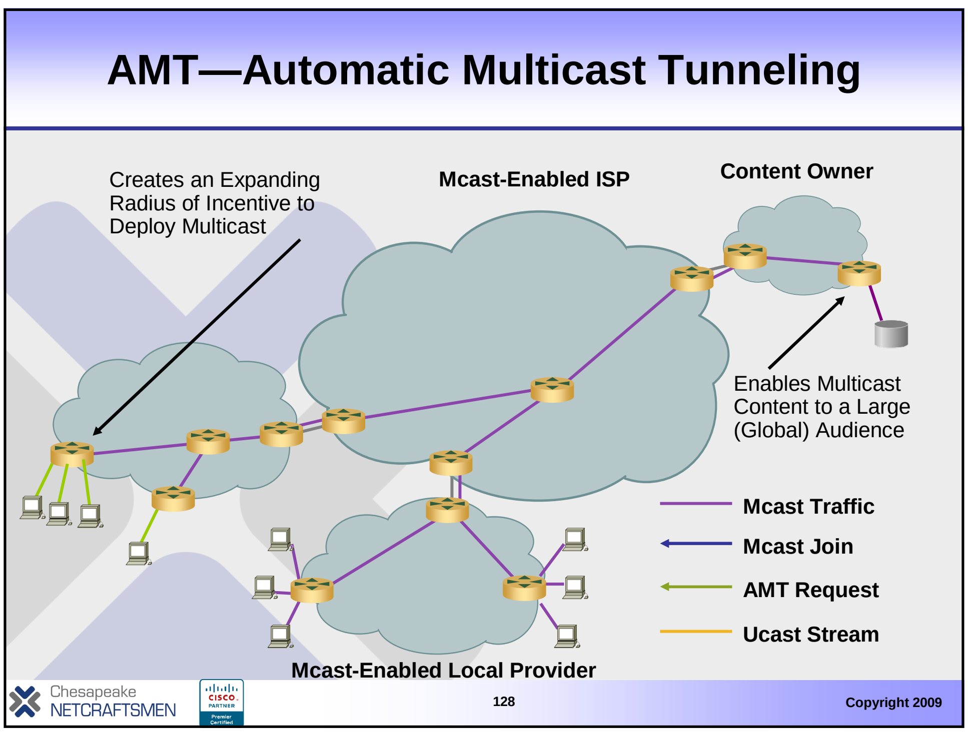



AMT—Automatic Multicast Tunneling
Additional Receivers Are Served by the AMT Relays; the Benefits of IP Multicast Are Retained by the Content Owner and All Enabled Networks in the Path
Mcast-Enabled ISP
Content Owner
Unicast-Only Network
Mcast Traffic
Mcast Join
AMT Request
Ucast Stream
Mcast-Enabled Local Provider
125
Copyright 2009


AMT—Automatic Multicast Tunneling
Creates an Expanding Radius of Incentive to Deploy Multicast
Mcast-Enabled ISP
Content Owner
Unicast-Only Network
Enables Multicast Content to a Large (Global) Audience
Mcast Traffic
Mcast Join
AMT Request
Ucast Stream
Mcast-Enabled Local Provider
126
Copyright 2009



AMT—Automatic Multicast Tunneling
Creates an Expanding Radius of Incentive to Deploy Multicast
Mcast-Enabled ISP
Content Owner
Unicast-Only Network
Enables Multicast Content to a Large (Global) Audience
Mcast Traffic
Mcast Join
AMT Request
Ucast Stream
Mcast-Enabled Local Provider
Chesapeake NETCRAFTSMEN
127
Copyright 2009



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
Mcast Traffic
Mcast Join
AMT Request
Ucast Stream
Mcast-Enabled Local Provider
Chesapeake NETCRAFTSMEN
128
Copyright 2009

Internet IP Multicast

- Will Internet IP multicast have a future?
- P2P solutions working toward over the top video solution today without end-to-end multicast
- Maybe that was just a dream… ;-)
- IP multicast deployment growing rapidly to provide edge-network content to the home
- Over the top video may use AMT, P2P, and/or could develop through cooperation with edge providers—
  but it's coming

CISCO