



Advanced Network Training Multicast

Larry Mathews
Systems Engineer
lmathews@brocade.com



Training Objectives

- Session will concentrate on Multicast with emphasis on Protocol Independent Multicast Protocol (PIM)
- Multicast in a ITS environment will be emphasized
- Attendees will have a good understanding on how multicast works
- Primary focus is on technology not vendor specifics



Agenda

- Multicast Overview
 - What is a multicast
 - Why use it
 - Multicast elements
 - Multicast IPv4 addressing
- Internet Group Management Protocol (IGMP) Overview
 - Where and why is IGMP used
 - IGMP packet structure
 - IGMPv2 vs. IGMPv3
 - How does IGMP work
- Protocol Independent Multicast Protocol Overview
 - Layer 2 vs Layer 3 protocols
 - Where does PIM function
 - PIM-Dense Mode (DM) vs PIM-Sparse Mode (SM)
 - PIM-Source Specific Multicast (SSM) overview



Agenda Continued

- Protocol Independent Multicast (PIM) Protocol Operation
 - PIM components
 - Reverse Path Forwarding (RPF) – Multicast vs Unicast routing
 - PIM packet formats
- Demonstration of Router configuration and operation
 - Best practice
 - Configuration requirements
 - Demonstration of successful PIM operation
- Diagnostics and Troubleshooting
 - Multicast packet capture
 - Using command line to identify issues





Multicast Overview

What is a Multicast

- In computer networking, multicast is a one-to-many or many-to-many distribution
- M/C is group communication where information is addressed to a group of destination computers simultaneously.
- IP multicast is a technique for one-to-many communication over an IP infrastructure in a network.
- Multicast is often employed in IP applications of streaming media, such as Internet television, music on hold (MoH), video, etc.

Why use Multicast

- IP multicast means that one sender is sending data to multiple recipients, but only sending a single copy
- Multicast uses network infrastructure efficiently by requiring the source to send a packet only once, even if it needs to be delivered to a large number of receivers
- A multicast router does not need to know how to reach all other multicast devices on the network. It only needs to know about multicast paths for which it has downstream receivers
- In contrast, a unicast router needs to know how to reach all other unicast addresses in the Internet

Multicast Terms and Elements

- IP multicast group address
 - logical identifier for a group of hosts in a computer network
- Multicast distribution tree
 - tree with its root at the source and branches forming a tree through the network to the receivers
- Multicast Protocols
 - Internet Group Management Protocol (IGMP)
 - Protocol Independent Multicast Protocol (PIM)
 - Multicast Border Gateway Protocol (MBGP)
 - Multicast Source Discovery Protocol (MSDP)
 - Multicast Listener Discovery (MLD) – Used with IPv6



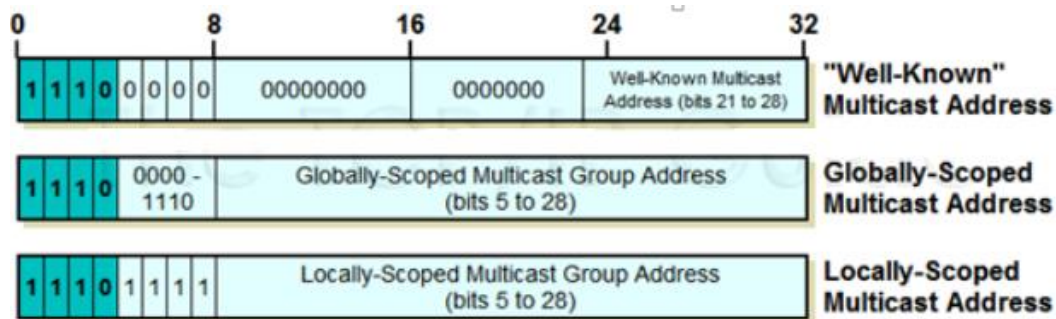
IPv4 Multicast Addressing


- IP also supports multicasting, where a source device can send to a group of devices
- Class D Multicast addresses are identified by the pattern “1110” in the first four bits, which corresponds to a first octet of 224 to 239
- The full range of multicast addresses is from 224.0.0.0 to 239.255.255.255

Range Start Address	Range End Address	Description
224.0.0.0	224.0.0.255	Reserved for special “well-known” multicast addresses.
224.0.1.0	238.255.255.255	Globally-scoped (Internet-wide) multicast addresses.
239.0.0.0	239.255.255.255	Administratively-scoped (local) multicast addresses.

<http://www.iana.org/assignments/multicast-addresses/multicast-addresses.xhtml>

Address	Purpose
224.0.0.1	All hosts on a subnet
224.0.0.2	All routers on a subnet
224.0.0.4	All DVMRP routers
224.0.0.5	All OSPF routers (DR Others)
224.0.0.6	All OSPF Designated Routers
224.0.0.7	ST Routers
224.0.0.8	ST Hosts
224.0.0.9	All RIPv2 routers
224.0.0.10	All EIGRP routers
224.0.0.11	Mobile-Agents
224.0.0.12	DHCP Server/Relay Agent
224.0.0.13	All PIM routers
224.0.0.14	RSVP Encapsulation
224.0.0.15	All CBT routers
224.0.0.18	VRRP
224.0.0.22	IGMPv3 Membership Reports
224.0.1.1	NTP





Internet Group Management Protocol (IGMP)

IGMP Overview

- IGMP is a communications protocol used by hosts and adjacent routers on IPv4 networks to establish multicast group memberships
- IGMP operates between the client computer and a local multicast router
- IGMP operates on the network layer, just the same as other network management protocols like Internet Control Message Protocol (ICMP)
- There are three versions of IGMP, as defined by Request for Comments (RFC) documents of the Internet Engineering Task Force (IETF)
 - IGMPv1 is defined by RFC 1112
 - IGMPv2 is defined by RFC 2236
 - IGMPv3 was initially defined by RFC 3376 and has been updated by RFC 4604 which defines both IGMPv3 and MLDv2
- Switches supporting IGMP snooping derive useful information by observing IGMP transactions



IGMPv1 vs IGMPv2 vs IGMPv3

— IGMPv1

- The Membership Query was always sent to 224.0.0.1 by multicast routers, and the Membership Report was always sent by stations to the group that a station wanted to join
- There was no message to announce that a station is leaving (unsubscribing) a multicast group, resulting in situations that a multicast stream was fed to a segment even after all stations have left the particular group
- Only after a timeout period the router discovered that there are no more subscribers to the group, and stopped the multicast feed

IGMPv1 vs IGMPv2 vs IGMPv3

- IGMPv2
 - Membership Query was of both types i.e. general (sent to 224.0.0.1) and group-specific (sent to a particular multicast group). The general Membership Query is used to find out all multicast groups that the stations are subscribed to. The group-specific Membership Query is used to find out if there is a subscriber for a particular group
 - Leave message to advertise that a station is unsubscribing from a multicast group, allowing the router to stop an unnecessary multicast stream feed much more quickly
 - IGMPv2 specified the way a multicast querier (a router that send Queries) is elected if there are multiple multicast routers connected to a common network. In IGMPv1, all multicast routers were expected to send Queries. The IGMPv2 stipulates that only the multicast router with the lowest IP address on the segment shall become the Querier and send Queries
 - » Other routers listen to the Replies but they do not send Queries themselves

IGMPv1 vs IGMPv2 vs IGMPv3

- IGMPv3
 - The IGMPv1/IGMPv2 does not have the capability to specify a particular sender.
 - Provided extensions to IGMP to support source-specific multicast

IGMP Version Comparison

Feature	IGMPv1	IGMPv2	IGMPv3
1 st Octet value for Query message	0x11	0x11	0x11
Group address for General query	0.0.0.0	0.0.0.0	0.0.0.0
Destination address for General query	224.0.0.1	224.0.0.1	224.0.0.1
Default Query Interval	60 sec	125 sec	125 sec
1 st octet value for Report	0x12	0x16	0x22
Group address for the report	Joining multicast group address	Joining multicast group address	Joining multicast group address and source address
Is report suppression mechanism available	Yes	Yes	No
Can max response time be configured	No , fixed at 10 sec	Yes, 0 to 25.5 sec	Yes, 0 to 53 min
Can a host send a leave group message	No	Yes	Yes
Destination address for leave group message	-	224.0.0.2	224.0.0.22
Can a Router send Group-specific query?	No	Yes	Yes
Can a Host send Source and group specific reports?	No	No	Yes
Can router end Source and Group specific Queries?	No	No	Yes
Rule for Electing a Querier?	None (depends on multicast routing protocol)	Router with the lowest IP address on the subnet	Router with the lowest IP address on the subnet
Compatible with other Versions of IGMP?	No	Yes, only with IGMP v1	Yes, with both IGMP v1 and v2

IGMP Packet Structure

IGMPv2 packet structure

+	Bits 0–7	8–15	16–31
0	Type	Max Resp Time	Checksum
32	Group Address		

IGMPv2 destination address

Message Type	Multicast Address
General Query	All hosts (224.0.0.1)
Group-Specific Query	The group being queried
Membership Report	The group being reported
Leave Group	All routers (224.0.0.2)

Where:

Type - Indicates the message type as follows: Membership Query (0x11), Membership Report (IGMPv1: 0x12, IGMPv2: 0x16, IGMPv3: 0x22), Leave Group (0x17)

Max Resp Time - Specifies the time limit for the corresponding report. The field has a resolution of 100 milliseconds, the value is taken directly. This field is meaningful only in Membership Query (0x11); in other messages it is set to 0 and ignored by the receiver.

Group Address - This is the multicast address being queried when sending a Group-Specific Query. The field is zeroed when sending a General Query

IGMPv3 Packet Structure

IGMPv3 membership query

bit offset	0-3	4	5-7	8-15	16-31
0	Type = 0x11			Max Resp Code	Checksum
32	Group Address				
64	Resv	S	QRV	QQIC	Number of Sources (N)
96	Source Address [1]				
128	Source Address [2]				
	...				
	Source Address [N]				

Where:

Max Resp Code - This field specifies the maximum time (in 1/10 second) allowed before sending a responding report. If the number is below 128, the value is used directly. If the value is 128 or more, it is interpreted as an exponent and mantissa.

Checksum - This is the 16-bit one's complement of the one's complement sum of the entire IGMP message.

Group Address - This is the multicast address being queried when sending a Group-Specific or Group-and-Source-Specific Query. The field is zeroed when sending a General Query.

Resv - This field is reserved. It should be zeroed when sent and ignored when received.

S (Suppress Router-side Processing) Flag - When this flag is set, it indicates to receiving routers that they are to suppress the normal timer updates.

QRV (Querier's Robustness Variable) - If this is non-zero, it contains the Robustness Variable value used by the sender of the Query. Routers should update their Robustness Variable to match the most recently received Query unless the value is zero.

QQIC (Querier's Query Interval Code) - This code is used to specify the Query Interval value (in seconds) used by the querier. If the number is below 128, the value is used directly. If the value is 128 or more, it is interpreted as an exponent and mantissa.

Number of Sources (N) - This field specifies the number of source addresses present in the Query. For General and Group-Specific Queries, this value is zero. For Group-and-Source-Specific Queries, this value is non-zero, but limited by the network's MTU.

Source Address [i] - The Source Address [i] fields are a vector of n IP unicast addresses, where n is the value in the Number of Sources (N) field

Basic Multicast IGMP Process

- Host A in subnet is the multicasting source and is sending multicast data to the multicast group address
- Multicast enabled router(s) connected to subnet sends out periodic queries to all hosts located on subnet
- Host B in subnet requests group membership (Join) from its local router.
- Because Host B has joined the multicast group, its network adapter is listening for datagrams sent to the multicast group address. The remaining host in subnet has not requested group membership, so its network adapter is filtering out (dropping) traffic sent to the multicast group address
- When a host no longer wants to listen to a multicast group address then it will report to the router that it has stopped listening (Leave)

Mapping L3 M/C Address to Ethernet (L3 to L2)

- Ethernet Addresses have a 48 bit address
- Expressed in hexadecimal numbering, the first 24 bits of an Ethernet multicast address are 01:00:5e, this indicates the frame as multicast
- The next bit in the ethernet address is always 0, leaving 23 bits for the multicast address
- Because IP Multicast groups are 28 bits (1110XXXX.XXXXXXXX.XXXXXXXX.XXXXXXXX) long and there are only 23 bits available the mapping cannot be one to one, so only 23 low order bits of the multicast group ID are mapped onto the ethernet address
- With this mapping, each Ethernet Multicast address corresponds to 32 different IP Multicast Addresses (2^5)
- This means that a host in one multicast group may have to filter out multicast that are intended for other groups sharing the same ethernet address

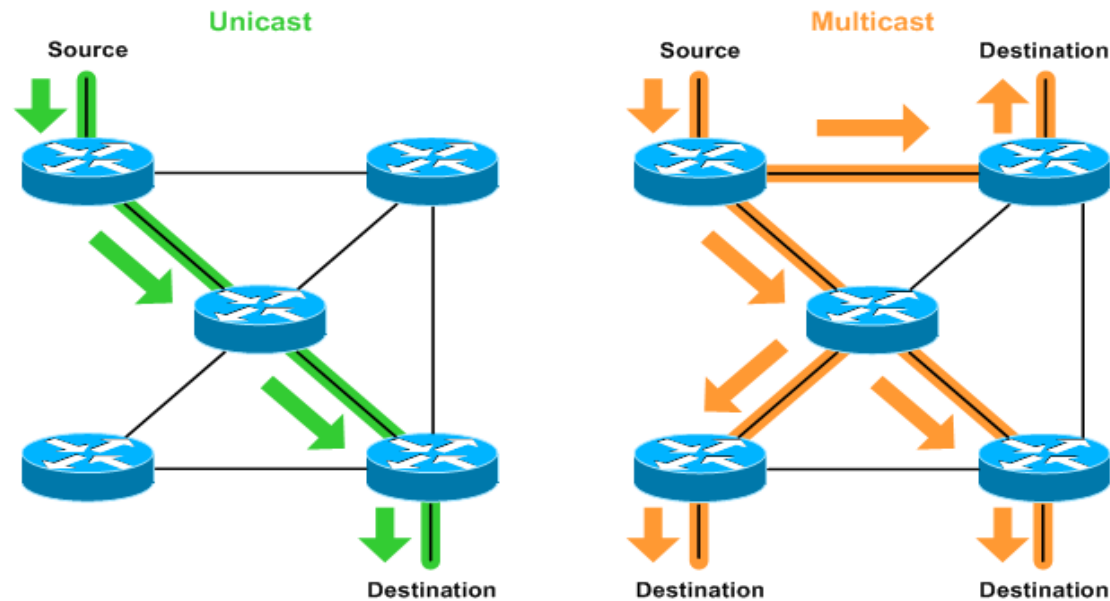




Protocol Independent Multicast (PIM)

Multicast Forwarding in Routed Network

- In multicast forwarding, each packet in a stream is transmitted once by a source and replicated by the infrastructure to efficiently reach an arbitrary number of recipients
- This introduces a number of challenges not typically of concern when dealing with strictly unicast traffic
 - The fundamental difference is that, while unicast routing protocols are designed to forward traffic down the optimal path or paths to a destination, a multicast routing protocol must forward traffic down all paths, often replicating packets out multiple interfaces on the same router



Multicast Forwarding in Routed Network

- Wherever traffic is replicated across a network containing even a single redundant path, the potential for routing loops is present
- Unicast routing protocols are well-suited to detect and avoid loops when routing to a single destination
- A multicast routing protocol is needed when paths have multiple end points
 - Protocol Independent Multicast (PIM)
 - Distance Vector Multicast Routing Protocol (DVMRP)
 - Multicast Open Shortest Path First (MOSPF)
 - Core-Based Trees (CBT)
- Multicast routing protocols, including PIM, are implemented on routers

PIM-Dense Mode (PIM-DM)

- Suitable for densely populated multicast groups, primarily in the LAN environment
- Routers initially flood multicast traffic for all groups out all multicast-enabled interfaces
- Routers which determine they have no clients interested in receiving the traffic then send prune messages up toward the source, requesting that the flow of multicast traffic downstream be pruned
- PIM-DM is straightforward to implement but generally has poor scaling properties
- Process of flooding and pruning repeats every three minutes. Because of this and other inefficiencies, PIM-DM isn't often recommended

PIM-Sparse Mode (PIM-SM)

- Initially, multicast traffic from a source isn't forwarded to group members
- When a member somewhere in the network decides it wants to receive traffic for a group, it sends a join request to its nearest router
- The join request is propagated up the multicast tree toward the source router
- Once the designated router closest to the source receives the join request, the source router begins forwarding multicast traffic for the group out the appropriate interface(s).
- PIM-SM explicitly builds unidirectional shared trees rooted at a rendezvous point (RP) per group, and optionally creates shortest-path trees per source



PIM-Source Specific Multicast (PIM-SSM)

- In traditional multicast forwarding any host can be a source for a group. Because any host can act a source, this multicast implementation is deemed Any Source Multicast (ASM)
- Source-Specific Multicast (SSM), defined in RFC 4607, extends traditional multicast to identify a set of multicast hosts not only by group address but also by source
- An SSM group, called a channel, is identified as (S,G) where S is the source address and G is the group address
 - Because an SSM channel is defined by both a source and a group address, group addresses can be re-used by multiple sources while keeping channels unique
- One of the biggest advantages SSM holds over ASM is that it does not rely on the designation of a rendezvous point (RP) to establish a multicast tree
 - Because the source of an SSM channel is always known in advance, multicast trees are efficiently built from channel hosts toward the source (based on the unicast routing topology) without the need for an RP to join a source and shared multicast tree
 - Multicast source(s) must be learned in advance via some external method (e.g. manual configuration).
- IANA has reserved for SSM the IPv4 address range 232.0.0.0/8
- Requires IGMPv3 - has the capability to specify a particular sender



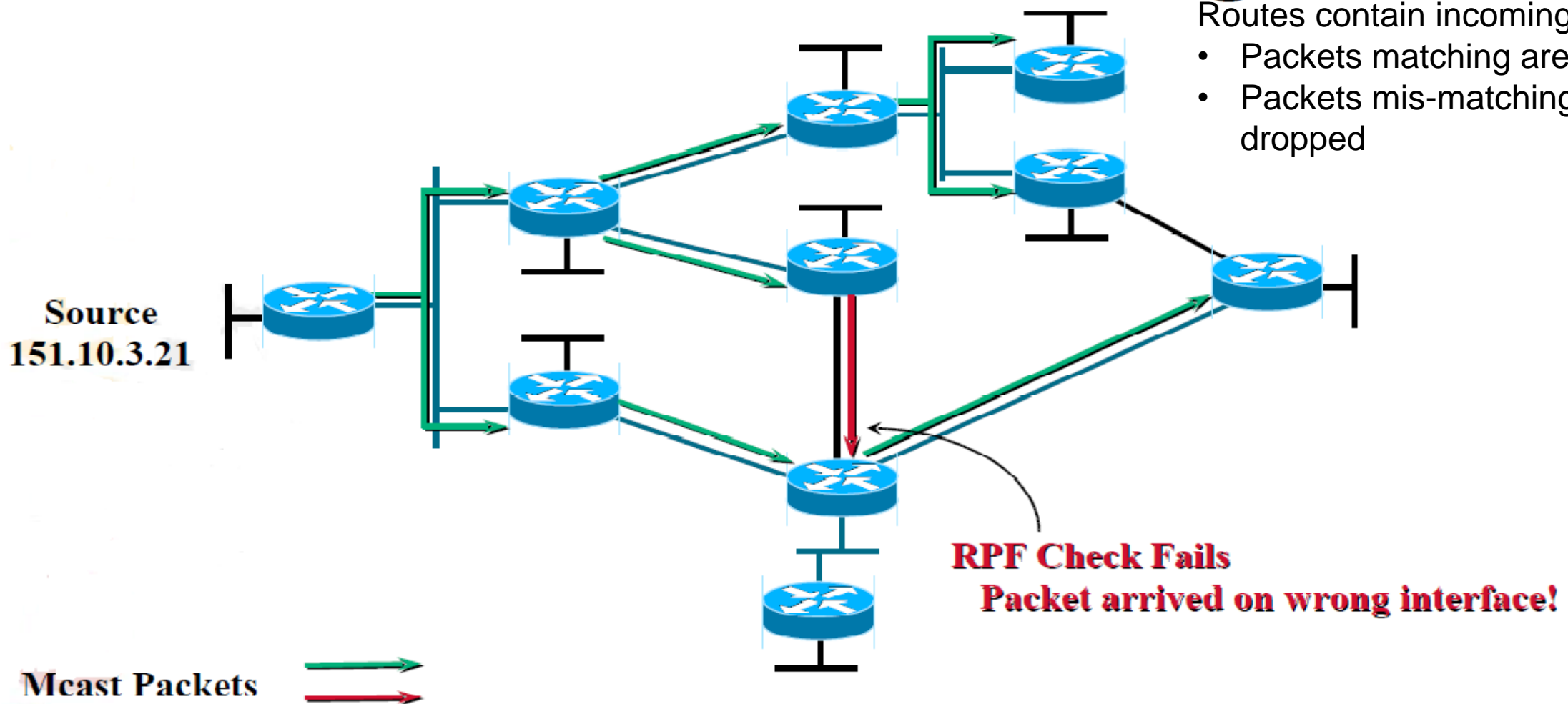
What is Reverse Path Forwarding (RPF)

- Uses unicast routes to determine path back to source
- A router forwards a multicast datagram only if received on the up stream interface to the source (i.e. it follows the distribution tree).
- The RPF Check
 - The source IP address of incoming multicast packets are checked against a unicast routing table.
 - If the datagram arrived on the interface specified in the routing table for the source address; then the RPF check succeeds.
 - Otherwise, the RPF Check fails.
- RPF checks ensures packets won't loop
- RPF checks are performed against routing table by default

RPF Check Example

RPF Checking

- Routes contain incoming interface:
- Packets matching are forwarded
 - Packets mis-matching are dropped

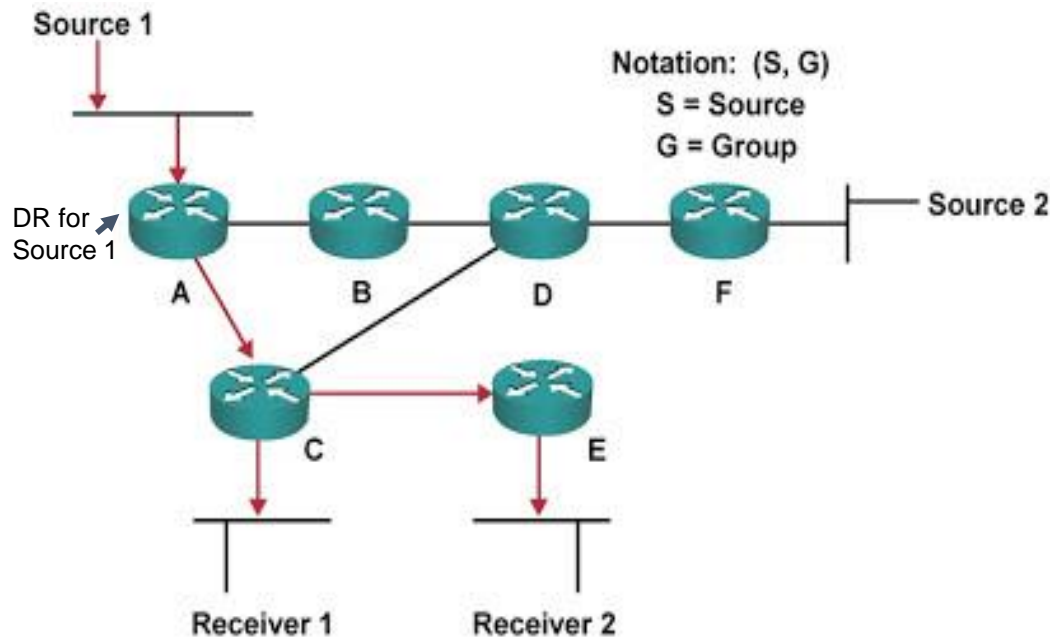


Multicast Routing Entries

- For every multicast source there must be two pieces of information:
 - the source IP address, S
 - the group address, G
- This is generally expressed as (S,G).
- Also commonly used is (*,G) - every source for a particular group.
- The router creates a table with the entries (*,G), (S,G).

Multicast Distribution Trees – Source Tree

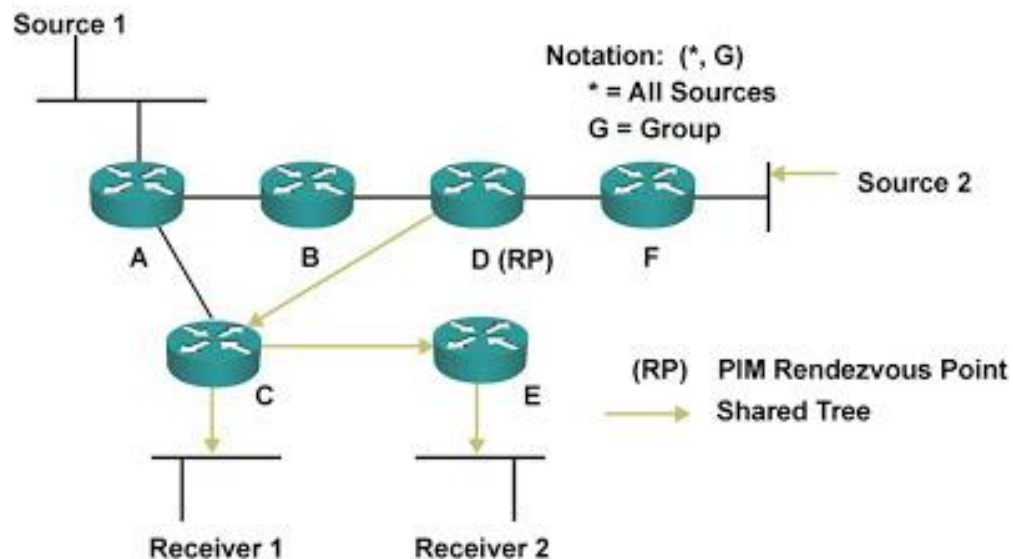
- Multicast Source will be on Top (root) of Multicast Tree - Logically, source will be the closest to the networks designated router (DR)
- Other Multicast enabled routes will be acting like branches
- PIM-Dense mode exclusively uses the Source Tree model



- Forwarding of packets based on shortest path
- The shortest path between Source 1 and Receiver 1 is via Routers A and C
- The shortest path to Receiver 2 is A, C and E
- The Tree is formed based on Source. That's why it is known as Source Tree
- Also referred to as the Shortest Path Tree (SPT)

Multicast Distribution Trees – Shared Tree

- In this topology all multicast sources Register themselves to one Central point known as RP (Rendezvous point). Multicast is managed centrally
- Host requests for any multicast stream, the DR router will send that request to mapped RP. The RP will redirect it to source
- No need to maintain (S, G) entries for every M/C router (less resources are used – scales)
- PIM-SM initially uses the shared tree model; will convert to SPT once flow is established



- Single RP in network acting as central point
- As the requests comes to Router C or Router E they will consult RP; RP will provide source info back to DR
- For example, receiver 1 wants to reach Source 1. Router C will forward that request to Router D (RP)
- The RP will inform router C source info (go to Router A)
- Router C determines there is a shorter route (then going thru Router D) and will begin using that path (SPT)
- Also referred to as RP Tree (RPT)

PIM Message Formats – Standard Header

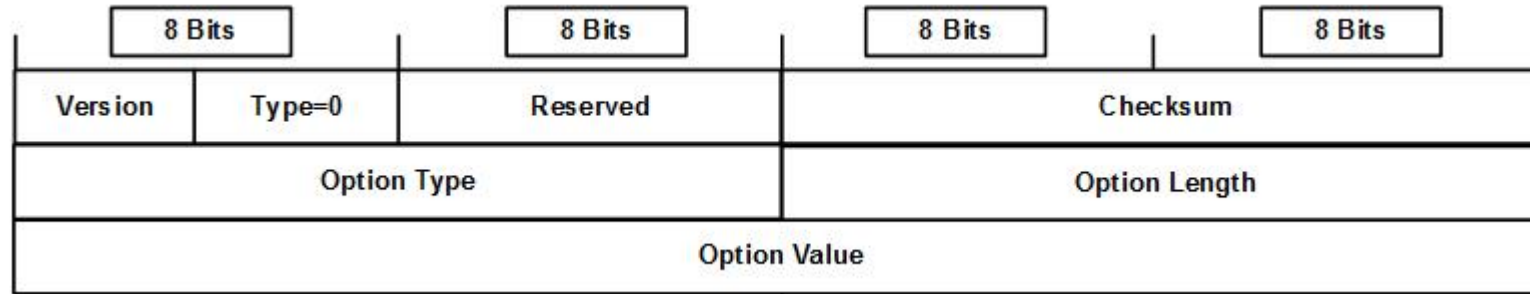


All PIM messages have a standard header

- The Version field specifies the version number, which is currently 2, but PIMv1 is still around
- The Checksum is a standard IP-style checksum, using a 16-bit one's complement of the one's complement of the PIM message, excluding the data portion of the Register message.
- The Type specifies the type of PIM message encapsulated behind the header

Type	Message
0	Hello
1	Register (used in PIM-SM only)
2	Register-Stop (used in PIM-SM only)
3	Join-Prune
4	Bootstrap (used in PIM-SM only)
5	Assert
6	Graft (used only in PIM-DM only)
7	Graft-Ack (used in PIM-DM only)
8	Candidate-RP-Advertisement (used in PIM-SM only)

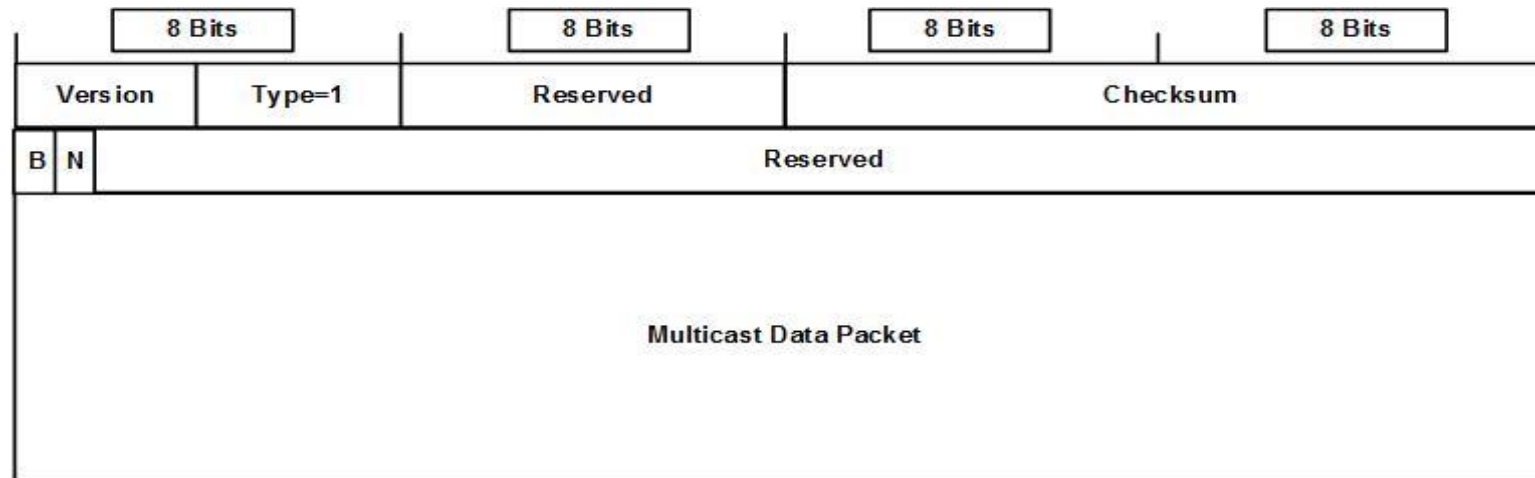
PIM Message Formats – Hello Message



The PIMv2 Hello message is used for neighbor discovery and neighbor keep-alives, which is sent every 30 seconds by default

- The Option Type field specifies the type of option in the Option Value field, where currently, only Option Type 1 is used, signifying that the field is a holdtime. The other available values of 2 to 16 are reserved.
- The Option Length field specifies the length (in bytes) of the Option Value field, which is only ever set to "1". When the Option Value field is set to "1" (specifying a holdtime), the Option Length is always "2".
- The Option Value field is a variable-length field carrying the value of whatever option is specified by the Option Type field.
- The Holdtime is the time that a router waits to hear a Hello message from a PIM neighbor before declaring the neighbor dead, which by default, is 3.5 times the Hello interval

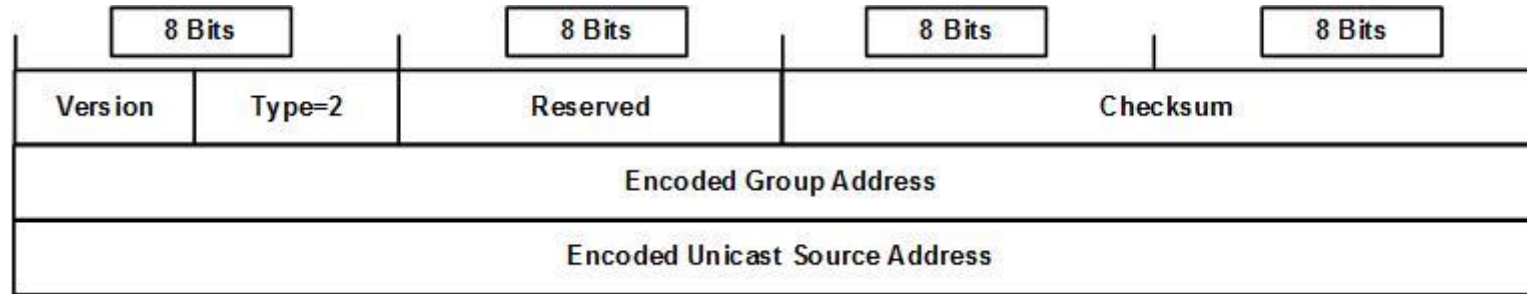
PIM Message Formats – Register Message



Register messages are only used by PIM-SM, and are unicast from the source's DR to the RP. They carry the initial multicast packets from the source, meaning that Register messages are used to tunnel multicast traffic from the source to the RP when an SPT has not yet been established from the source's DR to the RP

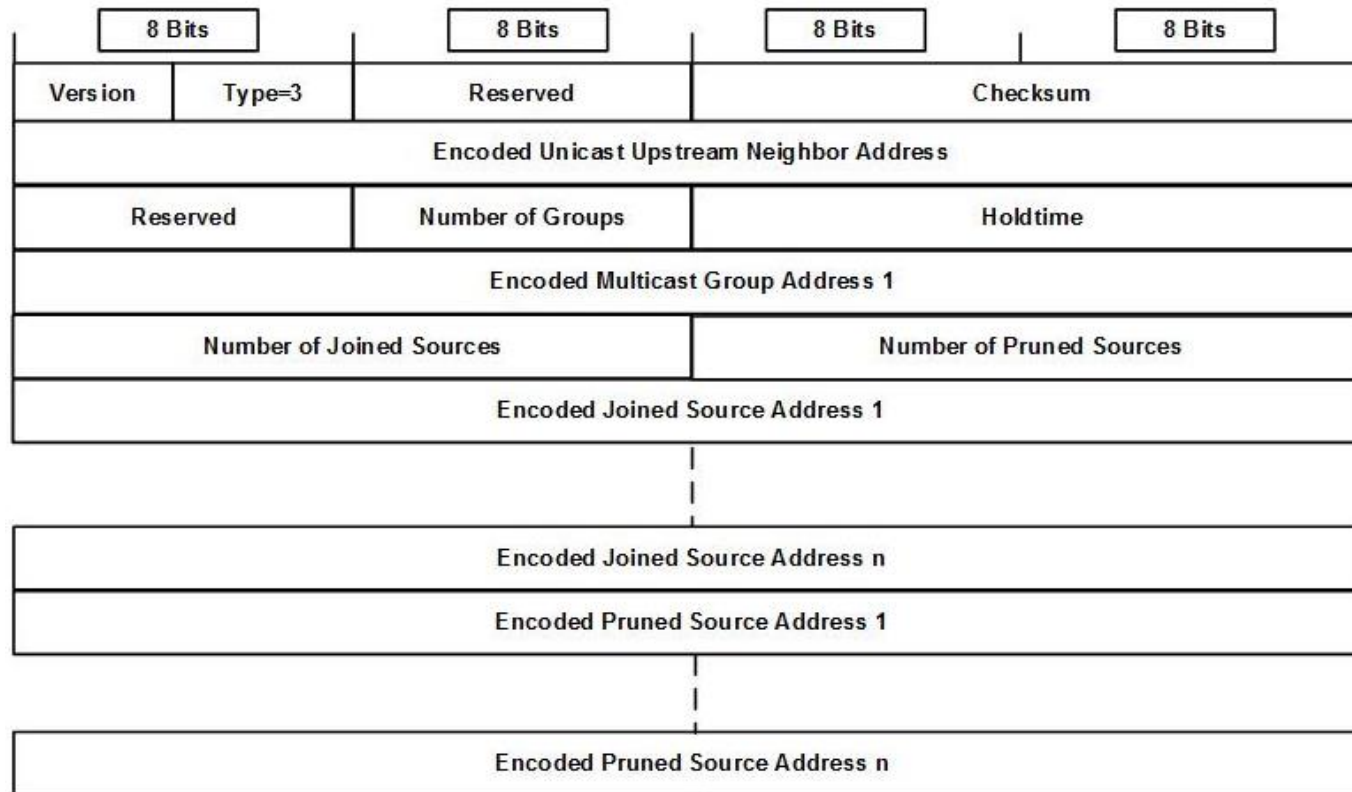
- The B bit is the Border bit, and it is set to 0 if the originator is a DR with a directly connected source. The bit is set to 1 if the source is a PIM Multicast Border Router (PMBR).
- The N bit is the Null-Register bit, and it is set to 1 when a DR probes the RP before expiring its local Register-Suppression timer.
- The Multicast Data Packet field is a single packet from the source that is being tunneled to the RP in the Register message.
- The Checksum field in Register messages is calculated only on the message header, and the data packet portion is excluded

PIM Message Formats – Register-Stop Message



- The Register Stop message is sent by an RP to a DR originating Register messages, which is used in a few different situations.
 - The first is when the RP is receiving the sourced multicast packets over the SPT and no longer needs to receive them encapsulated in Register messages.
 - The other is when there are no group members, either directly attached or over SPTs or RPTs, for the RP to forward the packets to.
- The Encoded Group Address field is the multicast group IP address for which the receiver should stop sending Register message.
- The Encoded Unicast Source Address is the IP address of the multicast source, which can also specify the wildcard source for (*, G) entries by setting the address to all zeros.

PIM Message Formats – Join/Prune Message



- Join/Prune messages are sent upstream to either RPs or sources and are used to join and prune both RPTs and SPTs.
- The message consists of a list of one or more multicast groups, where for each multicast address, there is a list of one or more source addresses.
- Together, these lists specify all (S, G) and (*, G) entries to be joined or pruned.

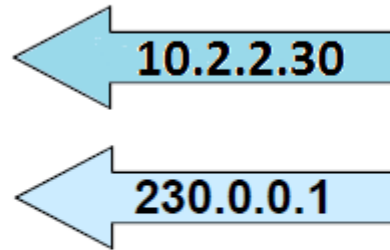
Putting it All Together – how does it work (IGMPv2)

Router adds group



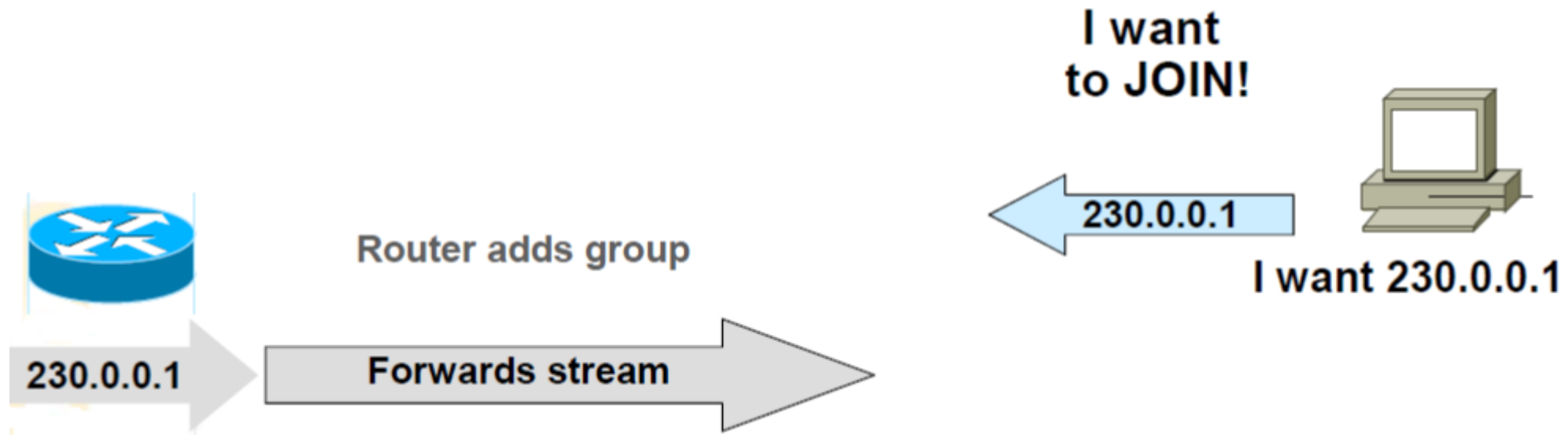
S,G(10.2.2.30, 230.0.0.1)

DR updates M/C routing table with Source and Group address (S,G) and registers it with RP as directly connect source



Source comes online and informs DR

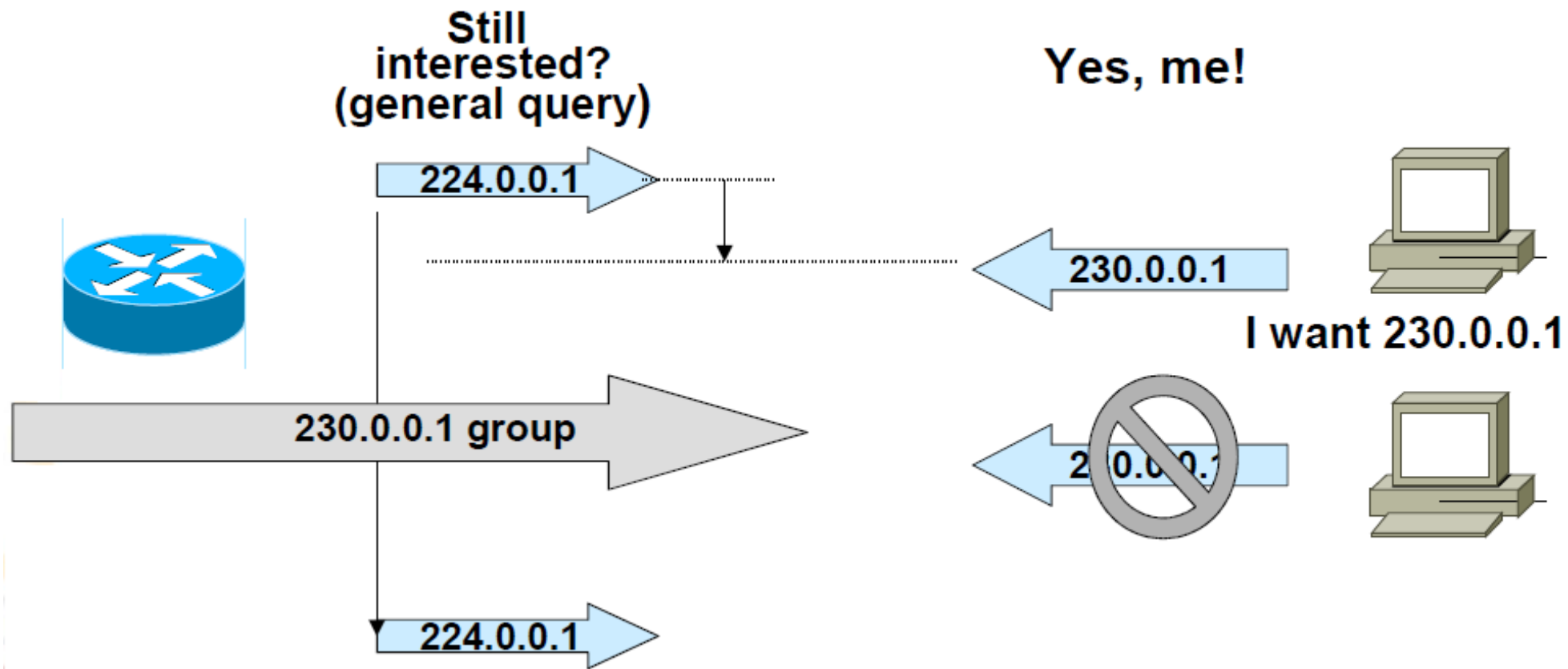
Putting it All Together – how does it work (IGMPv2)



Hosts can send unsolicited join membership messages – called reports or hosts can join by responding to periodic query from router

Router triggers group membership request to PIM.

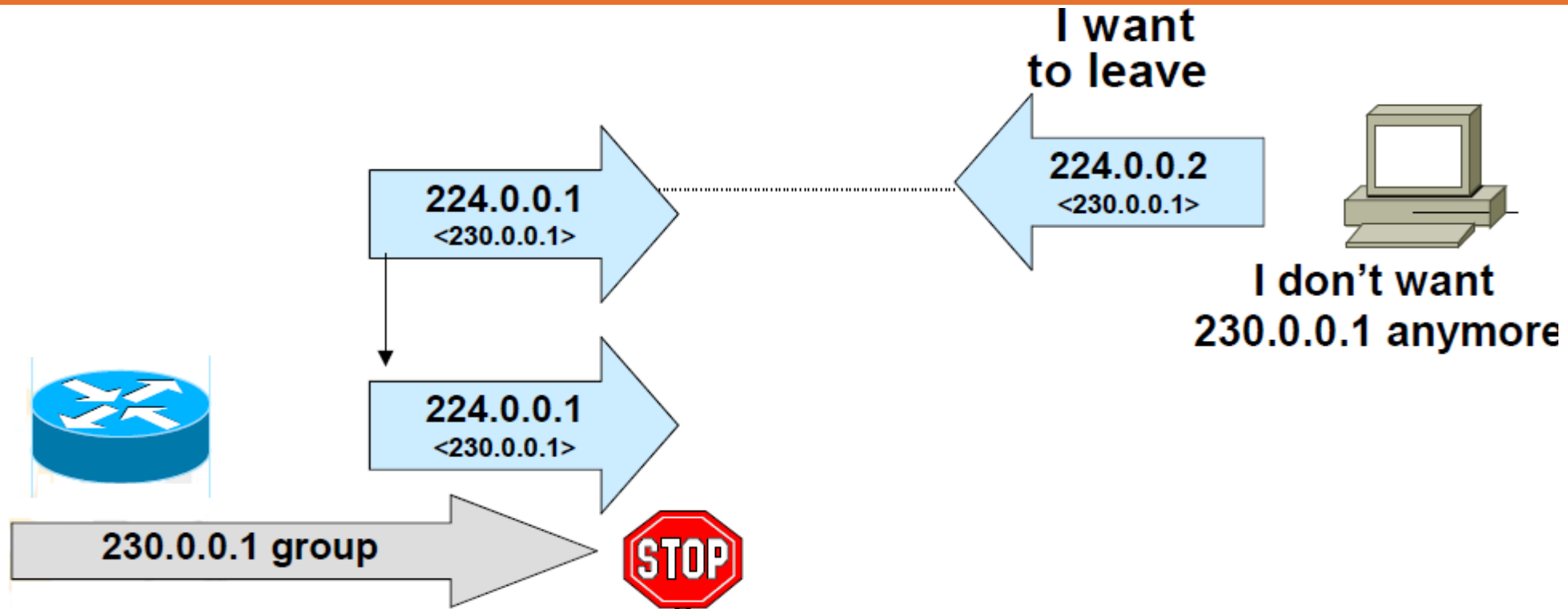
Putting it All Together – how does it work (IGMPv2)



Hosts respond to query to indicate (new or continued) interest in group(s)

After 260 sec with no response, router times out group

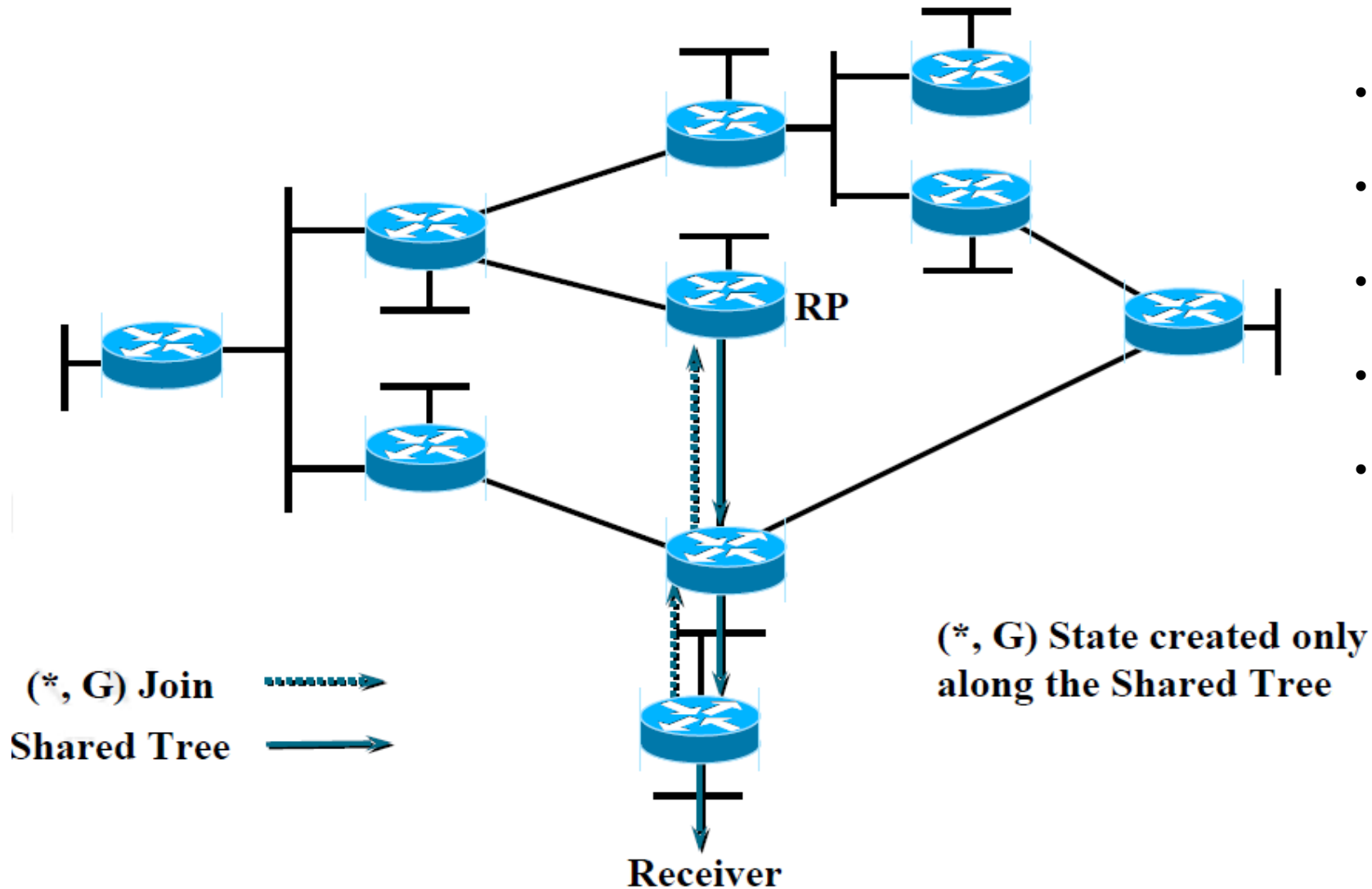
Putting it All Together – how does it work (IGMPv2)



Host sends leave message to all routers group indicating group they're leaving.

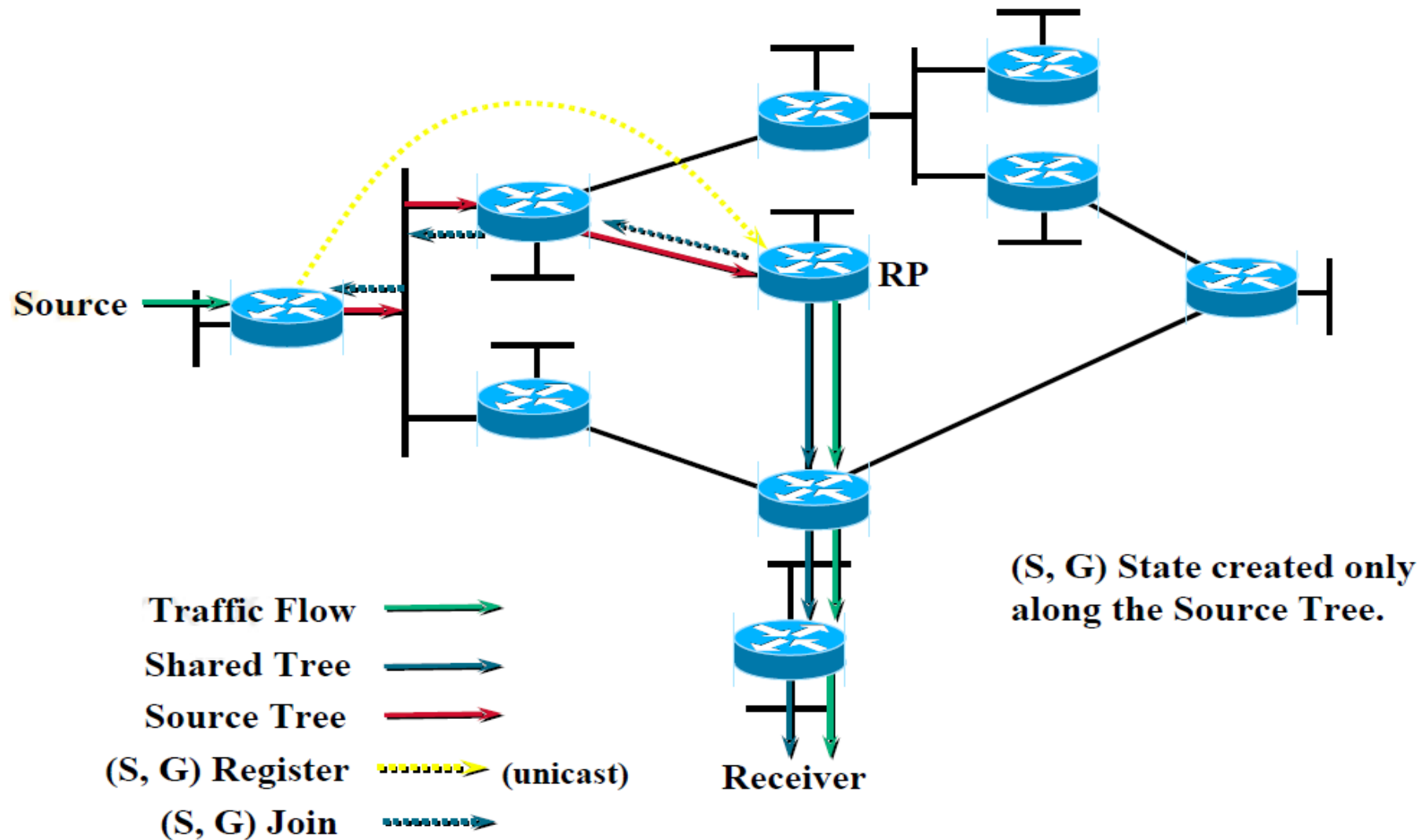
Router follows up with group-specific query message

Putting it All Together – how does it work (PIM-SM)

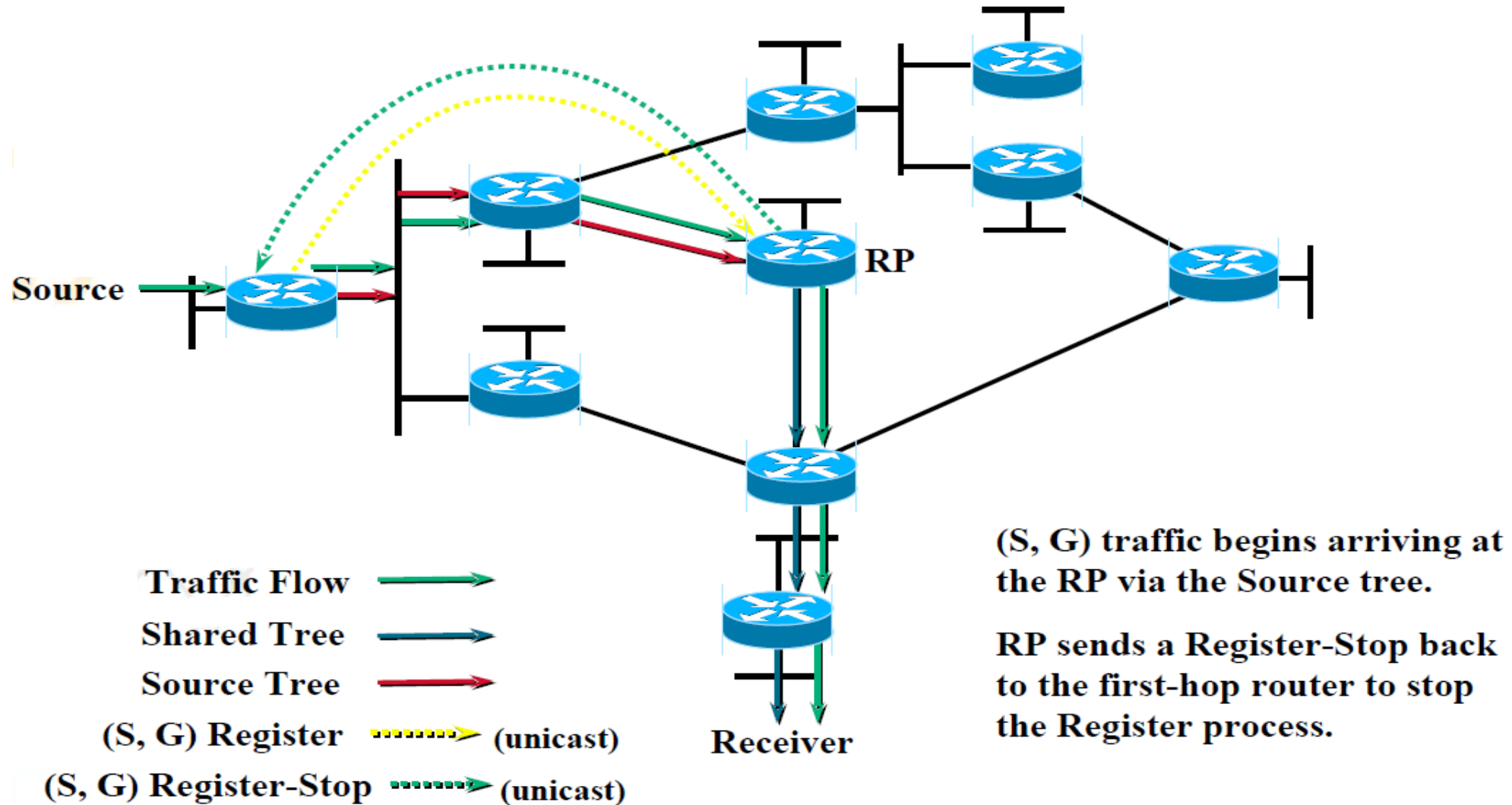


- Explicit join: assumes everyone does not want the data
- Uses unicast routing table for RPF checking
- Data and joins are forwarded to RP for initial rendezvous
- All routers in a PIM domain must have RP mapping
- Source-tree state is refreshed when data is forwarded and with Join/Prune control messages

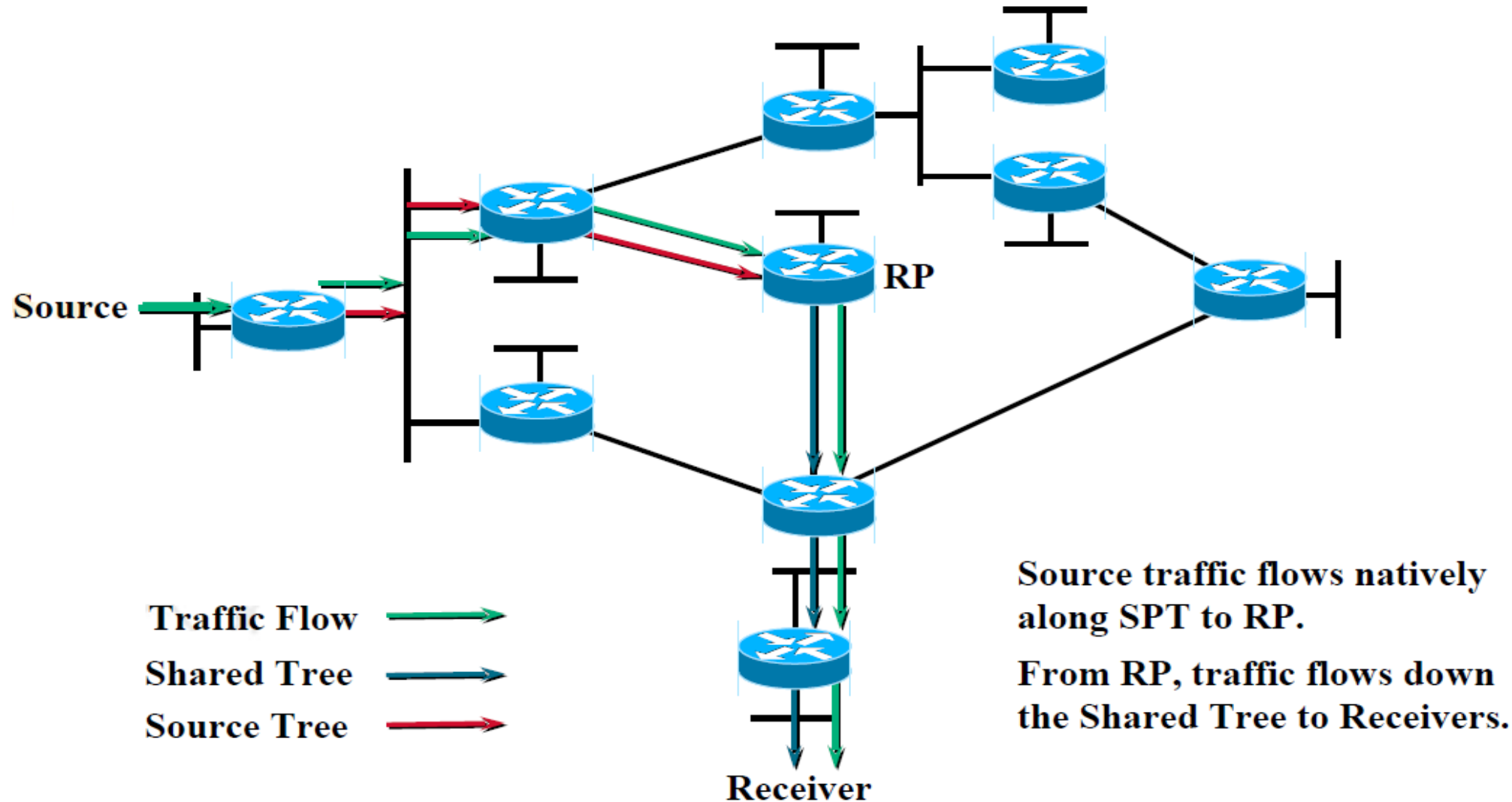
Putting it All Together – how does it work (PIM-SM)



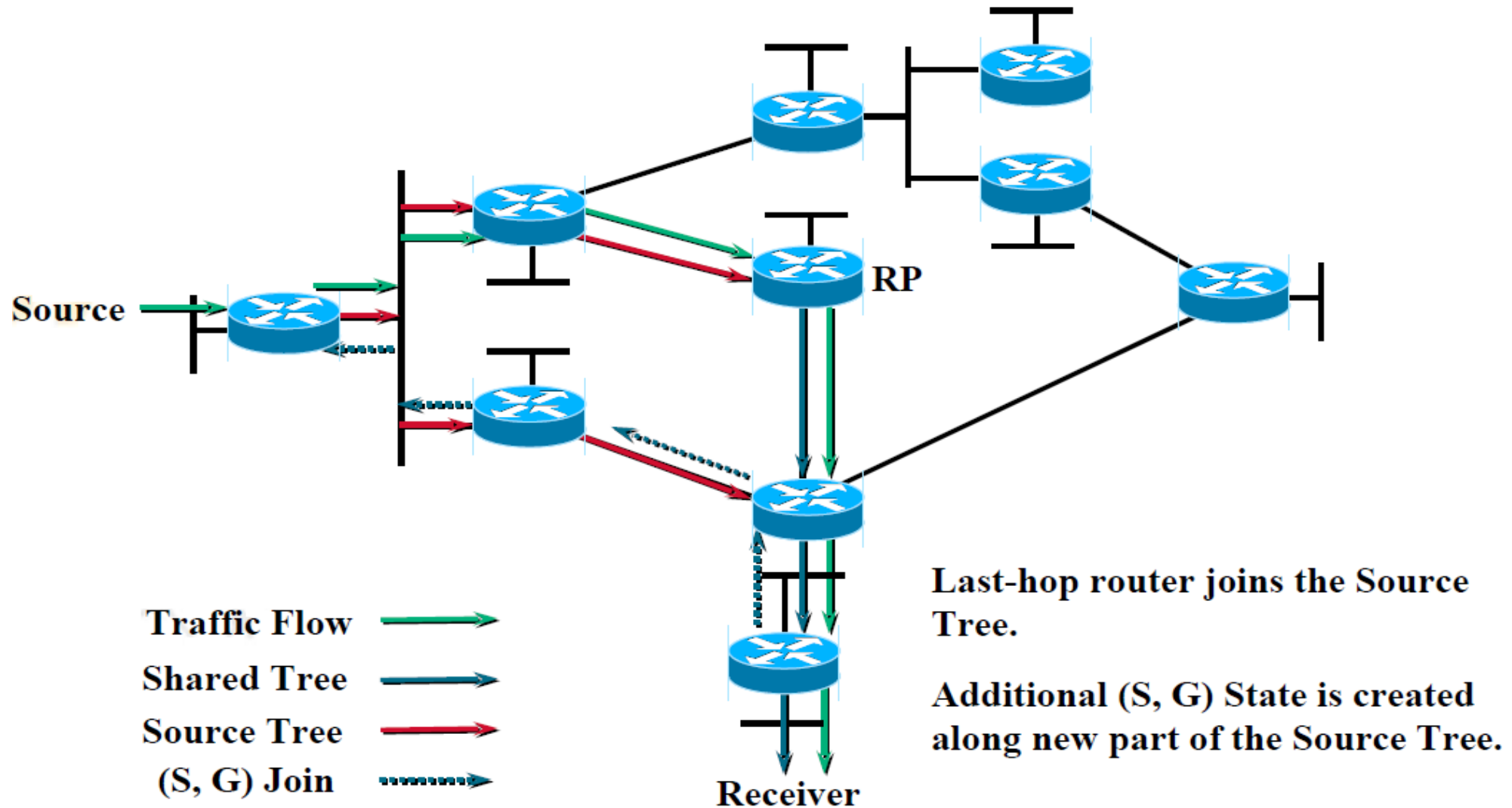
Putting it All Together – how does it work (PIM-SM)



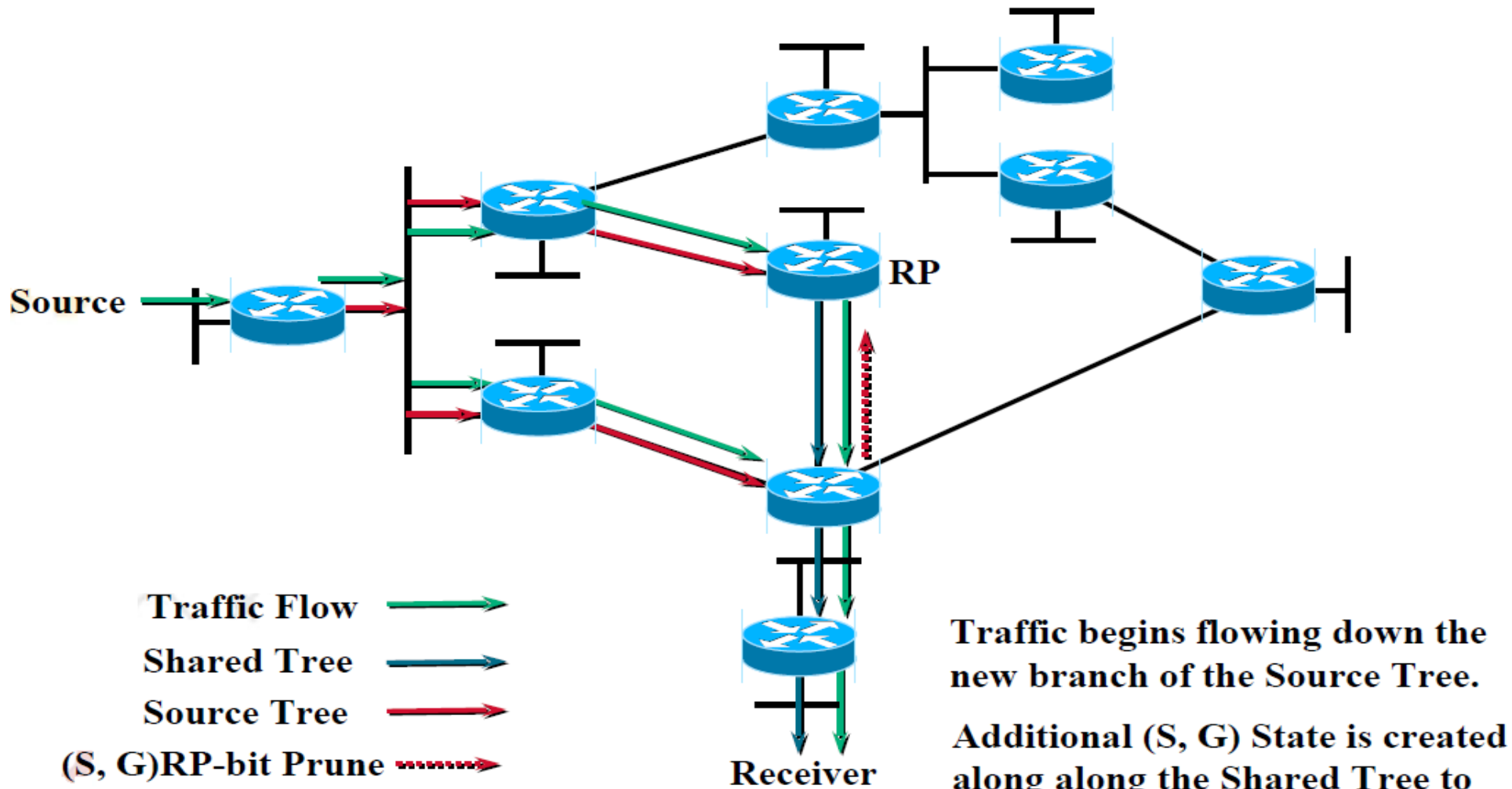
Putting it All Together – how does it work (PIM-SM)



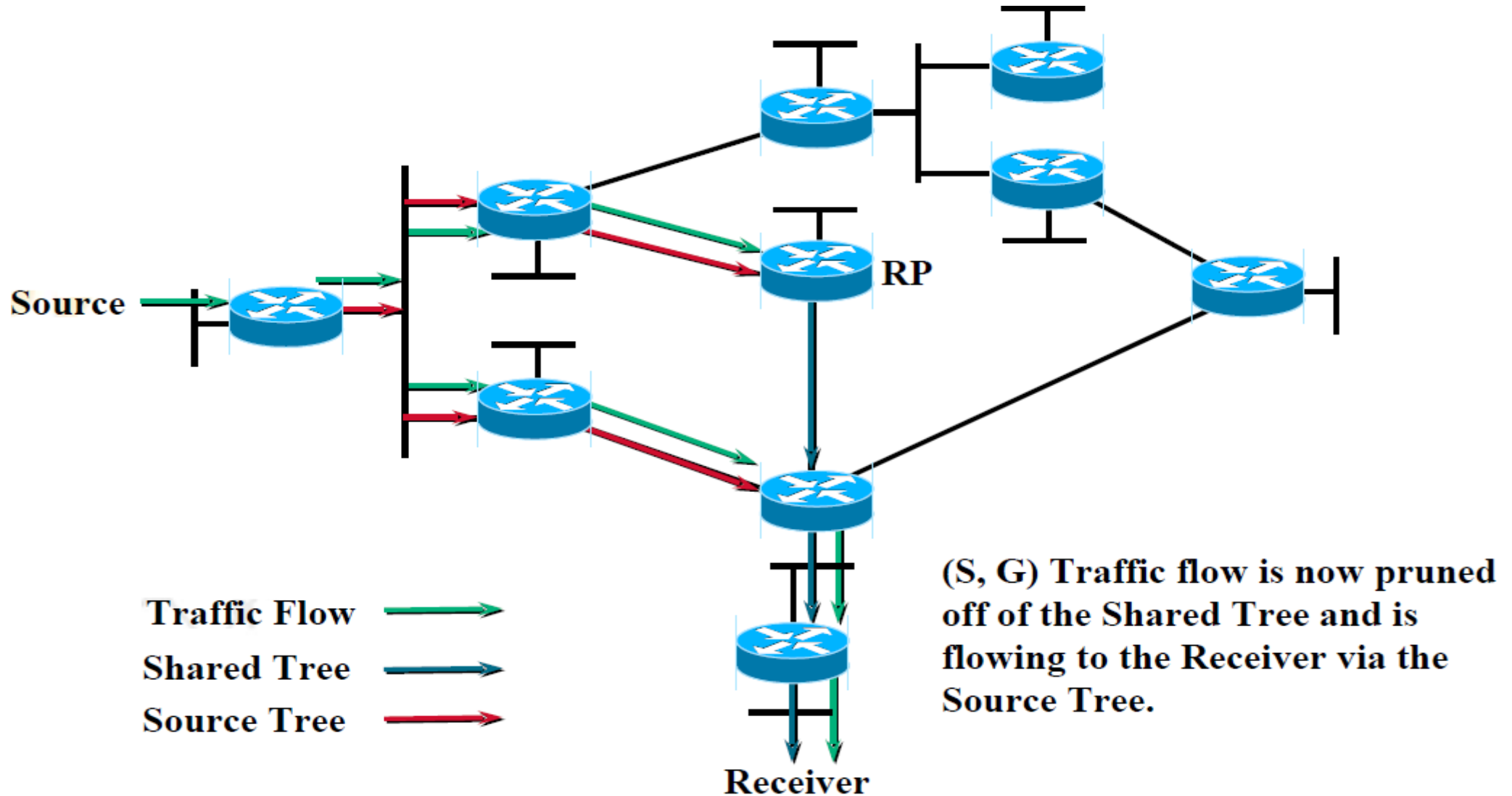
Putting it All Together – how does it work (PIM-SM)



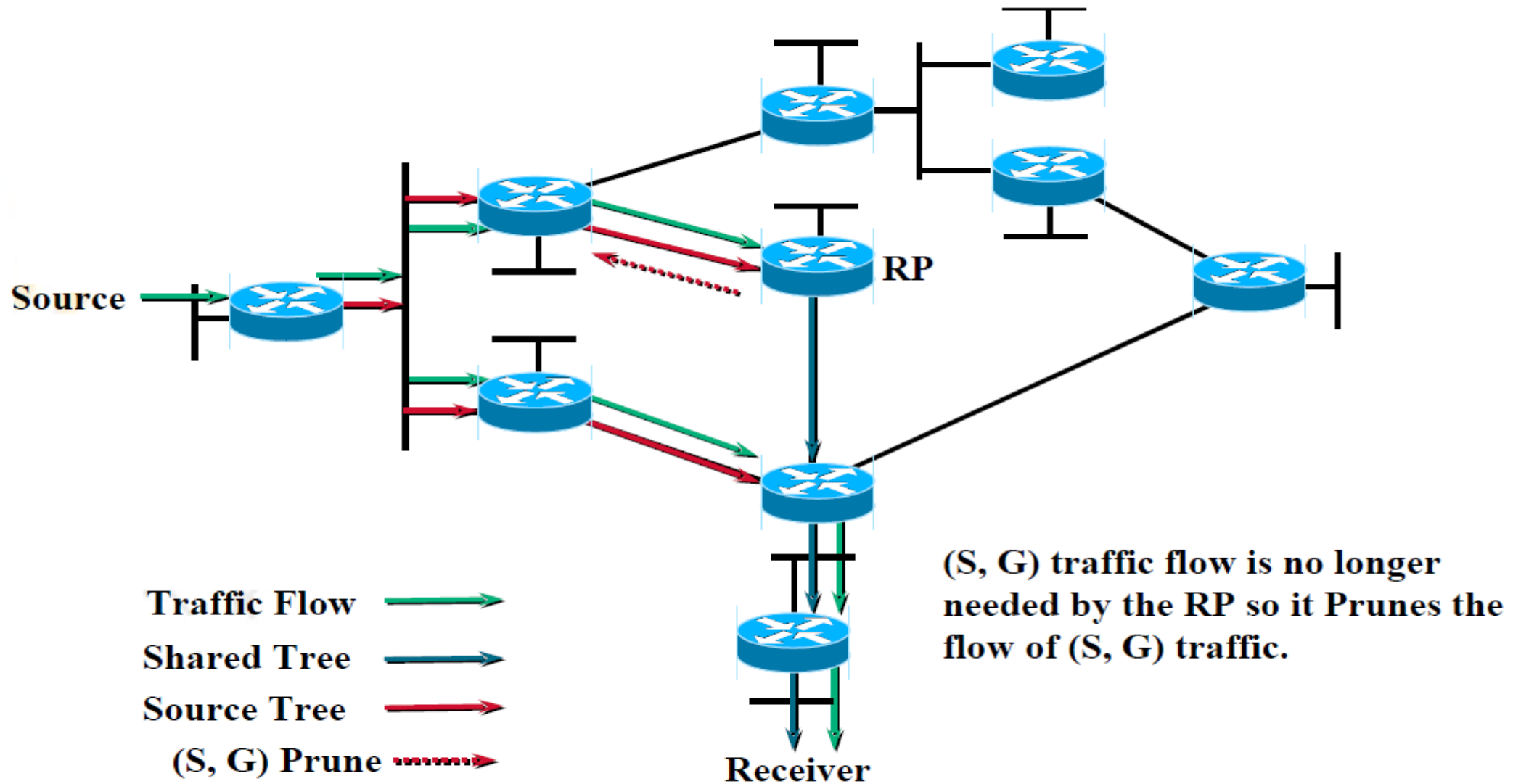
Putting it All Together – how does it work (PIM-SM)



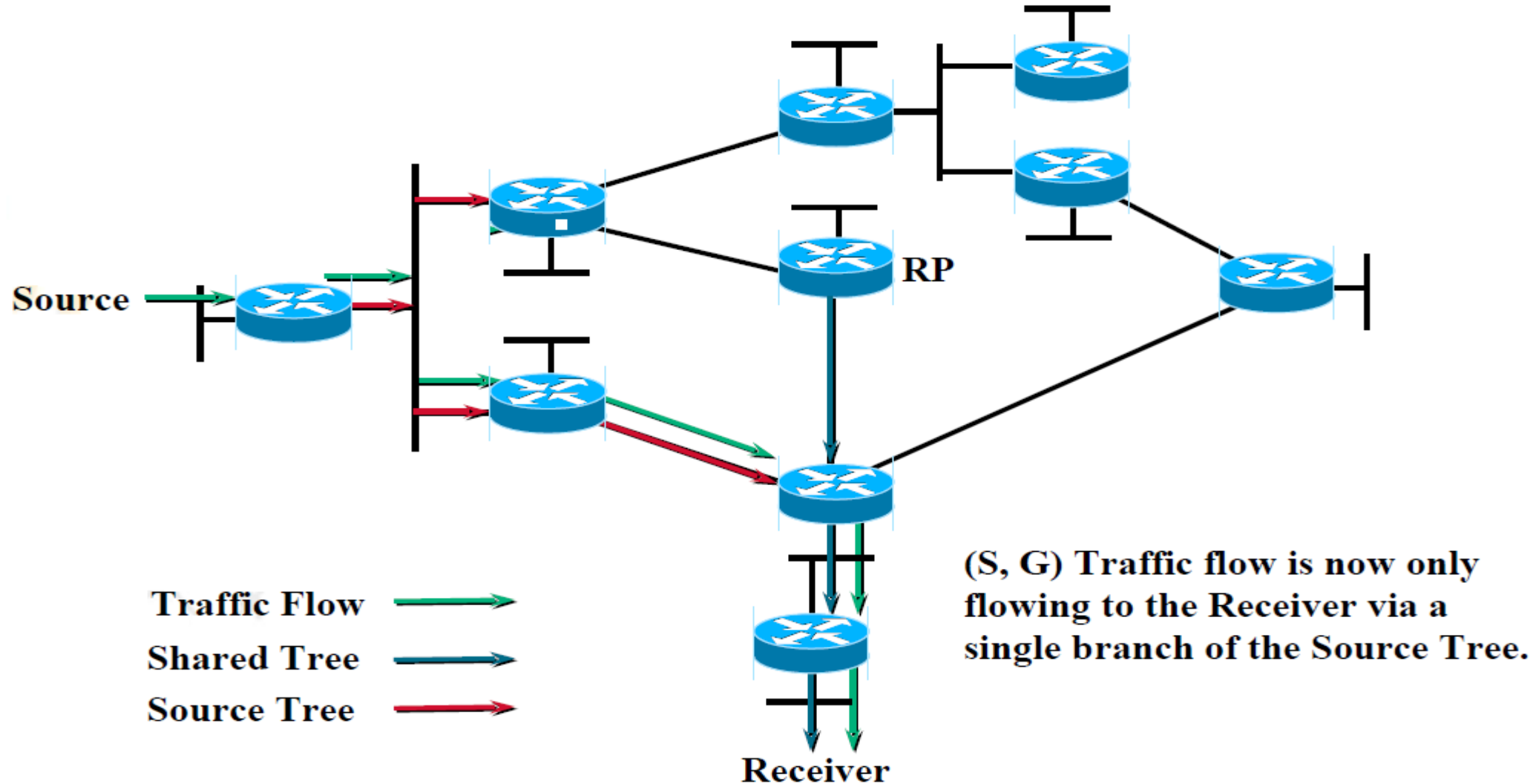
Putting it All Together – how does it work (PIM-SM)



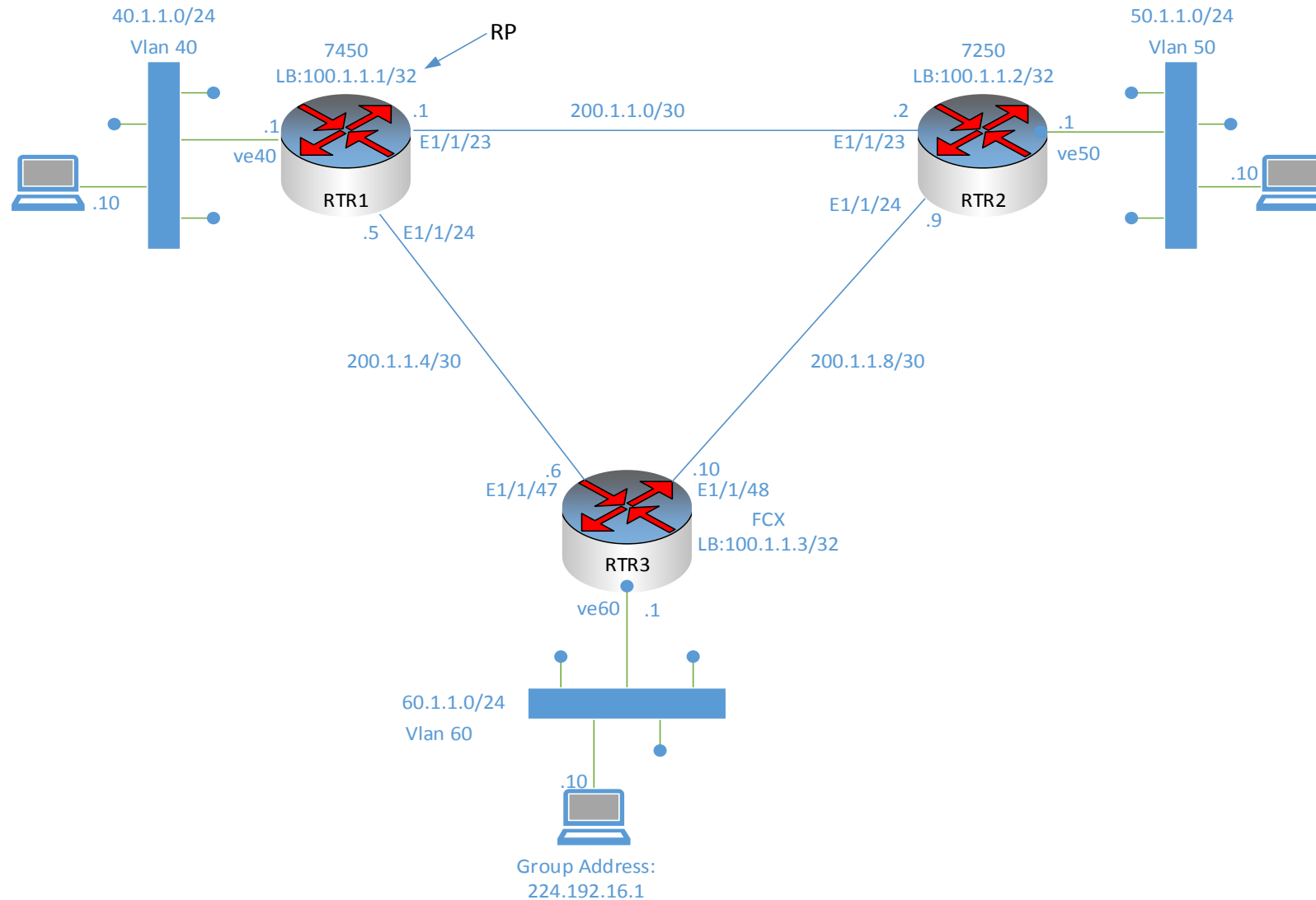
Putting it All Together – how does it work (PIM-SM)



Putting it All Together – how does it work (PIM-SM)



Multicast Demo Topology





IP Multicast Best Practices and Troubleshooting

Best Practices

- Manage group addresses
- Use IGMP Snooping on layer 2 domain devices (LAN)
- Use loopback addresses for IGP (i.e. OSPF) and PIM Global Functions
- Use point-to-point links in core
- A common layer 2 segment between routers introduces a number of unnecessary complexities and inefficiencies.
 - » For example, When a router or a link fails in a P2P environment the carrier signal is dropped and creates a triggered event that will cause immediate IGP convergence, which will be followed by IP Multicast convergence.
 - » In a switched environment, a router can fail and it will not be detected until several hello messages are missing at a layer 3 protocol level. This will increase the convergence time.

Troubleshooting Multicast

- Start near the source
- Identify the PIM-SM Designated Router
- Verify IGMP state in the Designated Router
- Look for (S,G) state in the Designated Router
- Follow the Reverse Path Forwarding (RPF) from the Designated Router back towards the source
- Verify PIM-SM has been configured on each interface along the RPF, because that determines the forwarding tree topology.
- Check (S,G) state in each router.
- Check (S,G) counters in each router.



Thank You

