IPv6 operator training program

IPv6 Hands-on seminar

IPv6 Multicast

＜Session Text＞

Alcatel-Lucent

Yasuo Kashimura

Today's Target

- Understanding protocol/feature's behavior used in IPv6 Multicast（PIM/MLD/MLD-snooping）.
  - Protocol sequence on router/switch.
- Configuring IPv6 multicast on router/switch, and confirm the behavior.
  - In multi-vendor environment (Alaxala/Alcatel/Cisco)

Agenda (Part1: Session)

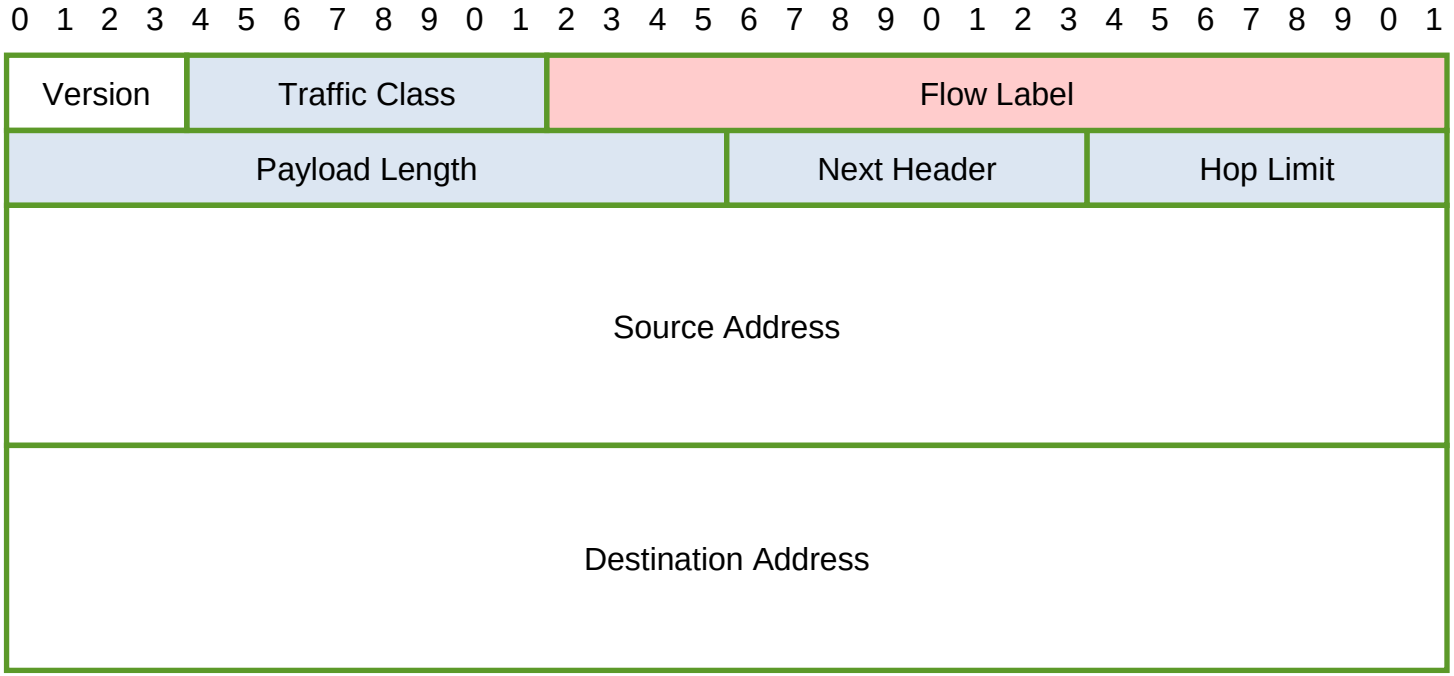
- IPv6 Basics: header format/addressing
- What is multicast ?
- MLDv1/v2
- Multicast Forwarding and PIM
  - PIM-SM (ASM: Any Source Multicast)
  - PIM-SSM (SSM: Source Specific Multicast)
- MLD Snooping
- Failure case study

Agenda (Part2: Hands-on)

1. Hands-on environment
2. Configuring IPv6 multicast and checking status
   - Enable IPv6 multicast routing
   - Enable MLD
   - Enable PIM
   - Configuring RP for PIM-ASM
   - Configuring PIM-SSM
   - Configuring MLD Snooping
3. Check the behavior
   - sending stream
   - PIM-ASM behavior
   - PIM-SSM behavior
   - MLD-Snooping behavior
   - PIM Route Failover
4. Advanced course
   - PIM DR behavior
   - Static MLD join
   - SSM mapping
5. Summary

IPv6 Basics

IPv6 Header Format



 # header length is fixed in IPv6

Newly defined field in IPv6

Field name is changed in IPv6

IPv6 Address notation

◆IPv4 address notation

#please refer also “draft-kawamura-ipv6-text-representation”

Binary digit notation(32bits)

11000000 10101000 00000000 00000001

- For every 8bits, display in decimal number and separate with “.”

192.168.0.1

◆IPv6 address notation

Binary digit notation(128bits)

0010000000000001 0000110110111000 1011111011101111 1100101011111110
0000000000000000 0000000000000000 0000000000000000 0001001000110100

- For every 16bits, display in Hex number and separate with “**:**”

2001:0db8:beef:cafe:0000:0000:0000:1234

- Leading zero in 16bits field can be abbreviated.

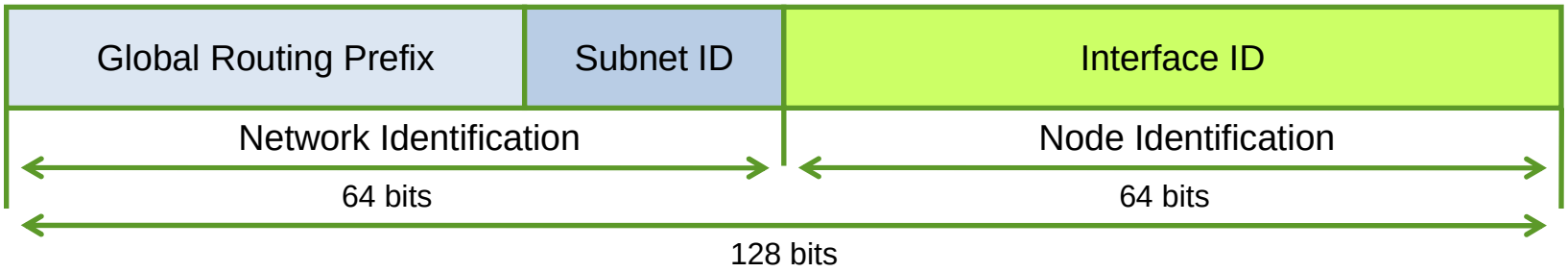
2001:db8:beef:cafe:0:0:0:1234

- Compress the zeros with “**::**”

2001:db8:beef:cafe::1234

IPv6 Address

◆IPv6 Address Architecture



●Prefix
Upper 64 bits with Global Routing Prefix + Subnet ID

◆IPv6 Address categories

●Unicast Address　one to one communication
the Address for each network interface
Global Address, Link-local Address, Unique Local Address

●Multicast Address　one(or Many) to Many communication
receiving hosts are identified by Group Address
Also used as alternative way of IPv4's broadcast as well

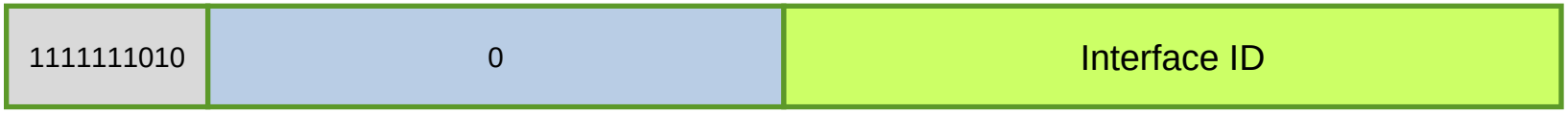
●Anycast Address　one to one of many communication
can be configured on multiple interfaces/nodes. Assigned for "feature".

Unicast Address Format

◆Global Unicast Address

- Global address (ex) 2001:db8::1

◆Link-local Unicast Address



Unique within one Link (fe80::/10)

It is used for the communication within the link.

◆Unique Local Unicast Address(ULA) [RFC4193]

Multicast Address Format

◆Multicast Address

Flag	
T Flag	0:permanently-assigned ("well-known") multicast address, assigned by IANA 1:non-permanently-assigned multicast address
P Flag	1 : Unicast-Prefix-based multicast address(RFC3306) #when P=1, T must be 1
R Flag	1 : multicast address that RP address is embedded (RFC3956) #when R=1, P/T must be 1

Scope: limit the scope of the multicast group			
0000(0)	reserved	0101(5)	site-local scope
0001(1)	interface-local scope	1000(8)	organizational-local scope
0010(2)	link-local scope	1110(E)	global scope
0100(4)	admin-local scope	1111(F)	reserved

◆ex):

PIM-SM(ASM) multicast address FF15::1234

PIM-SSM multicast address FF38::abcd

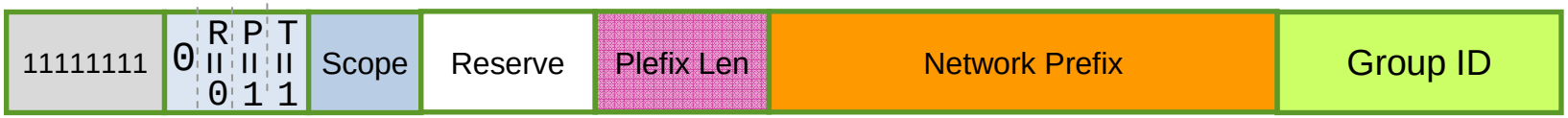
Multicast Address Format

◆**permanently-assigned multicast address: mainly used for control packets**

FF02:0:0:0:0:0:0:1	Link local All IPv6 Nodes
FF02:0:0:0:0:0:0:2	Link local All IPv6 routers
FF02:0:0:0:0:0:0:C	Link local DHCP server/relay agent
FF02:0:0:0:0:1:FFxx:xxxx	Solicited node multicast address(xx:xxxx represents lower 24bits of node's unicast/anycast address.)

◆**RFC3306 Unicast-Prefix-based multicast address**

If the user has global unicast IPv6 prefix, the user can have globally unique multicast prefix that global unicast prefix is embedded in multicast prefix.



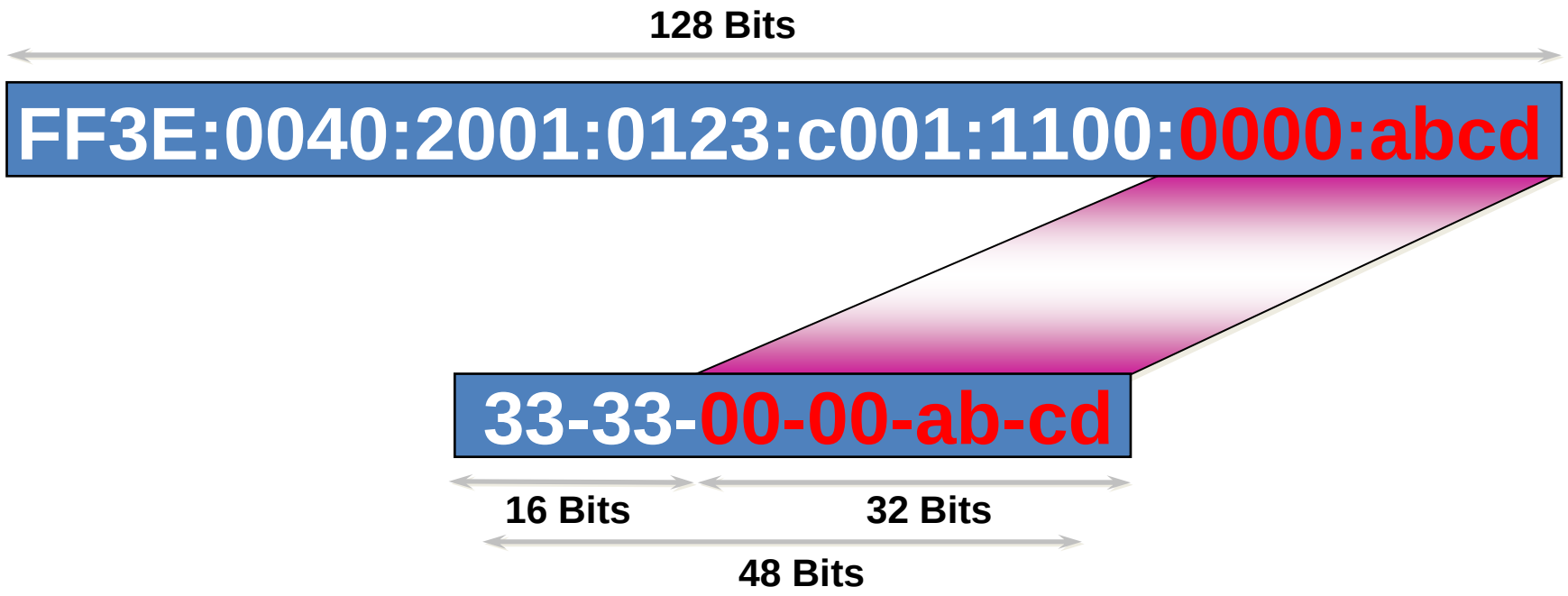
◆**ex):**

FF3E:0040:2001:0123:c001:1100:0000:abcd

Plen=64 **Unicast prefix** **groupID**

When the User has Unicast Prefix = 2001:0123:c001:1100::/64.

Multicast MAC Address for IPv6

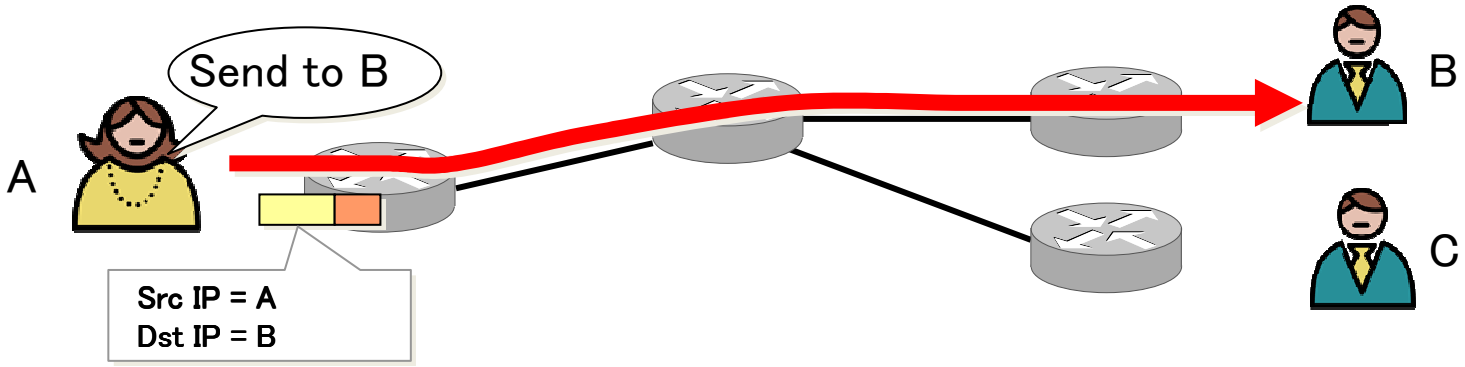


FF3E:0040:2001:0123:c001:1100:0000:abcd and
FF18:0040:2001:0234:1002:0010:0000:abcd
will have same multicast MAC address(L2-dest-address).
L2-switch may not be able to differentiate those IPv6 multicast address.

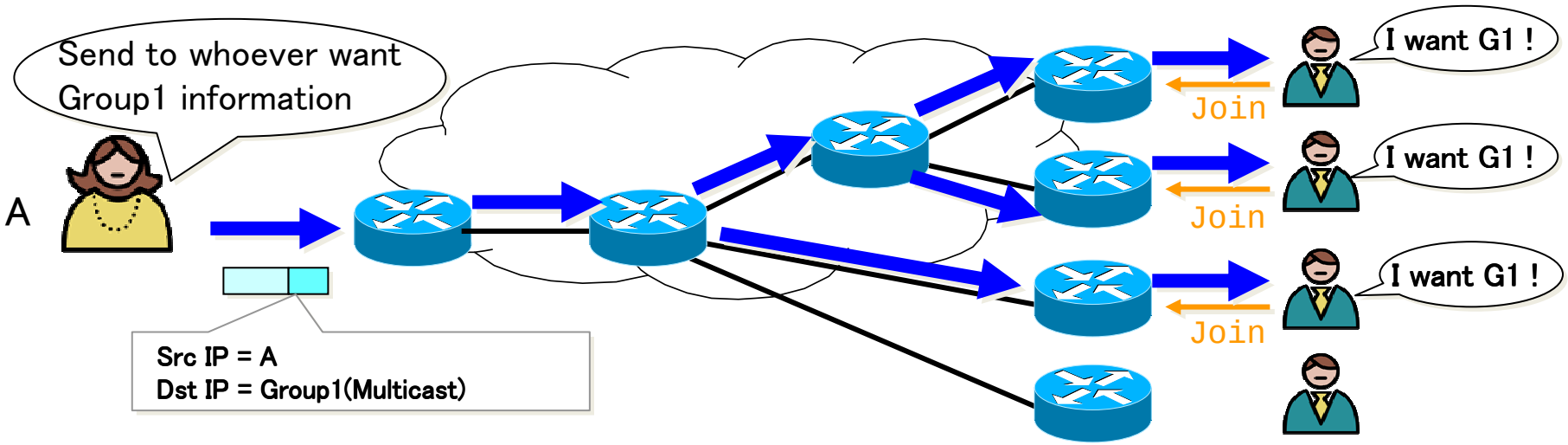
What is Multicast ?

Difference between Unicast and Multicast

◆**Unicast: one to one communication**

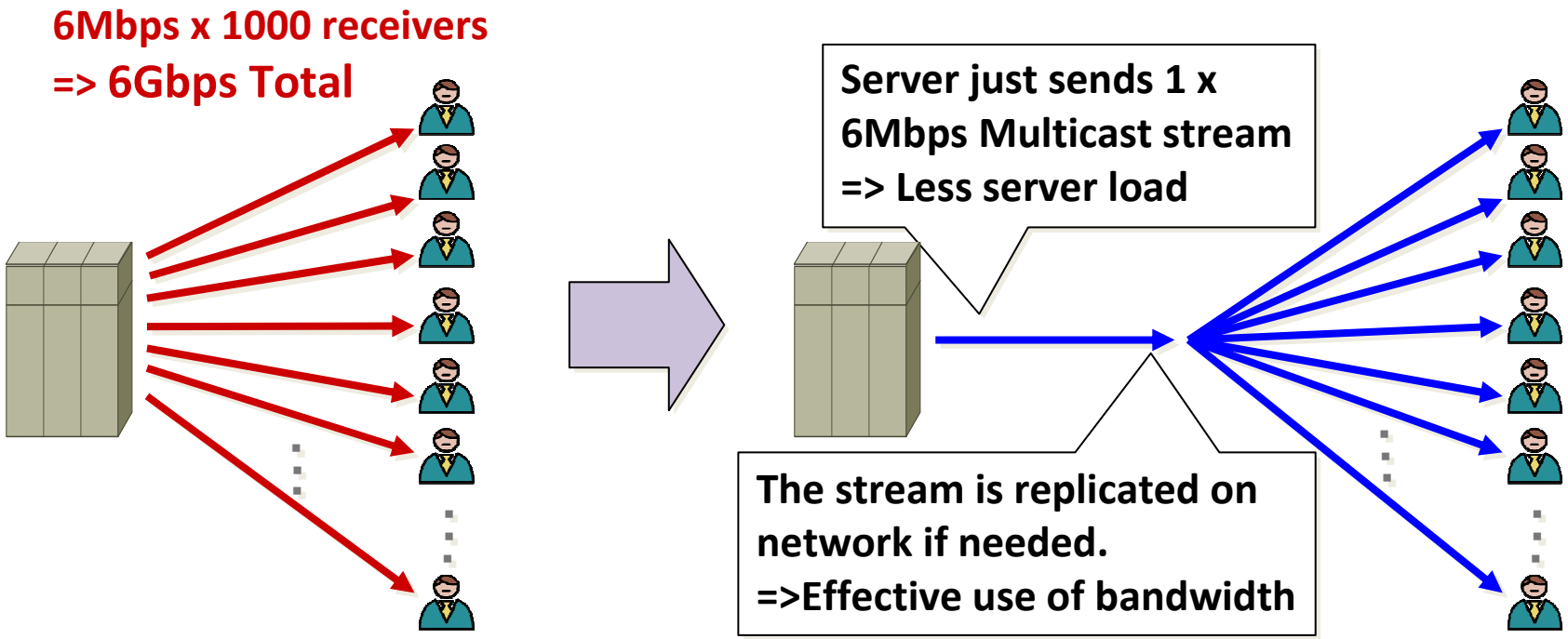


◆**Multicast: one(or Many) to Many that joined to the group**



Merit of Multicast

- ◆Reduce the server load while sending same contents to many receivers
- ◆Effective use of network bandwidth



Multicast applicability

◆Residential Broadband
IP/TV, Live Streaming
Gaming, Contents download
◆Enterprise network
E-Learning, broadcasting
Application delivery
Multipoint conference
Sensor network in Factory
◆Financial
Hoot and Holler
Financial systems(Stock exchange, etc)
◆Public
Video Surveillance for river, highway
information broadcasting(voice, disaster information)
information distribution in local town

IPv6 multicast service example

◆**IPv6 multicast broadcasting system(i-InproV6)**

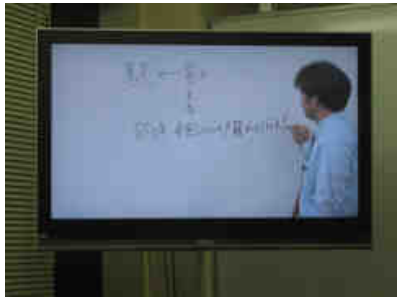
●Remote class in preparatory schools

Cost is 1/10 to compare with satellite broadcasting

- Initial cost: $several million => about $200k
- running cost: $100k/month => $10k/month

Popular teacher's class is broadcasted to all areas.
same quality in all areas, more per class profit

Remote Class



http://becare.co.jp/service/case01.html

◆**Emergency Earthquake information distribution service(OCN)**

●Emergency Earthquake information/Alert from Meteorological biz support center is distributed to users with urgency/realtime/efficiency.

Receiver Application

http://www.ntt.com/jishinsokuho/index.html

◆**contents distribution for kiosk hosts at convenience stores. (FamilyMart)**

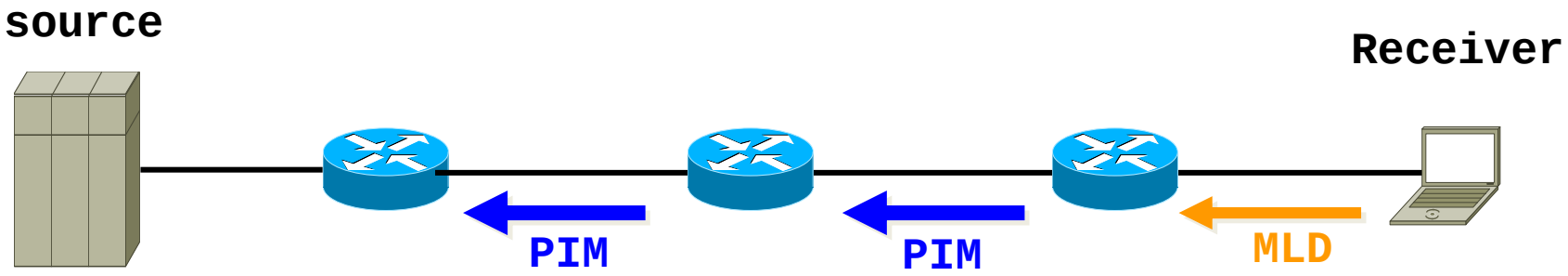
●6,000 stores are dual-stack-ed

●change from satellite to broadband with multicast

●Distributing large volumes of data by multicast like New product Add/manual for employee

Kiosk Host(Fami-port)

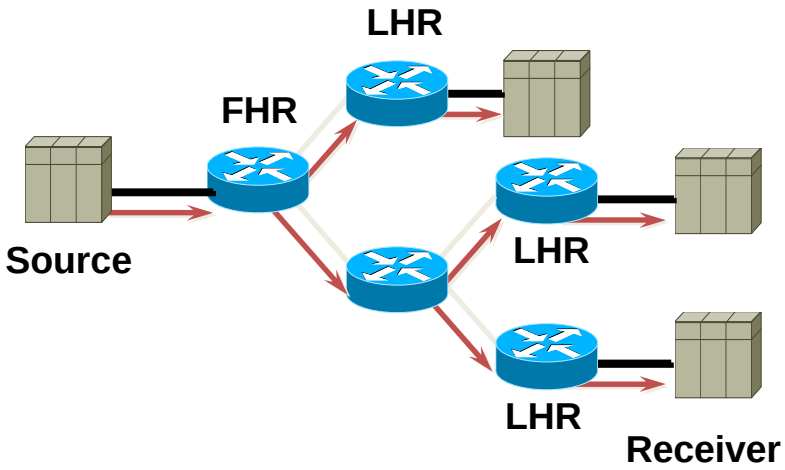
Multicast Protocols



MLD: for Signaling between Receiver and Router
PIM: for Signaling between Routers to build multicast distribution tree

Terminologies

- Source/Sender
  - The host/server sending multicast traffic
- Receiver/Listener
  - The host that receives multicast traffic
- Upstream
  - The direction traffic comes in
- Downstream
  - The direction traffic goes out



- RP（Rendezvous Point）
  - In PIM-SM, The router that “rendezvous” source and receiver info
- First Hop Router (FHR)
  - The router that source is connected
- Last Hop Router (LHR)
  - The router that receiver is connected
- (*,G)
  - *:any source, G:group address
- (S,G)
  - S:source address of multicast traffic, G:group address

Terminologies

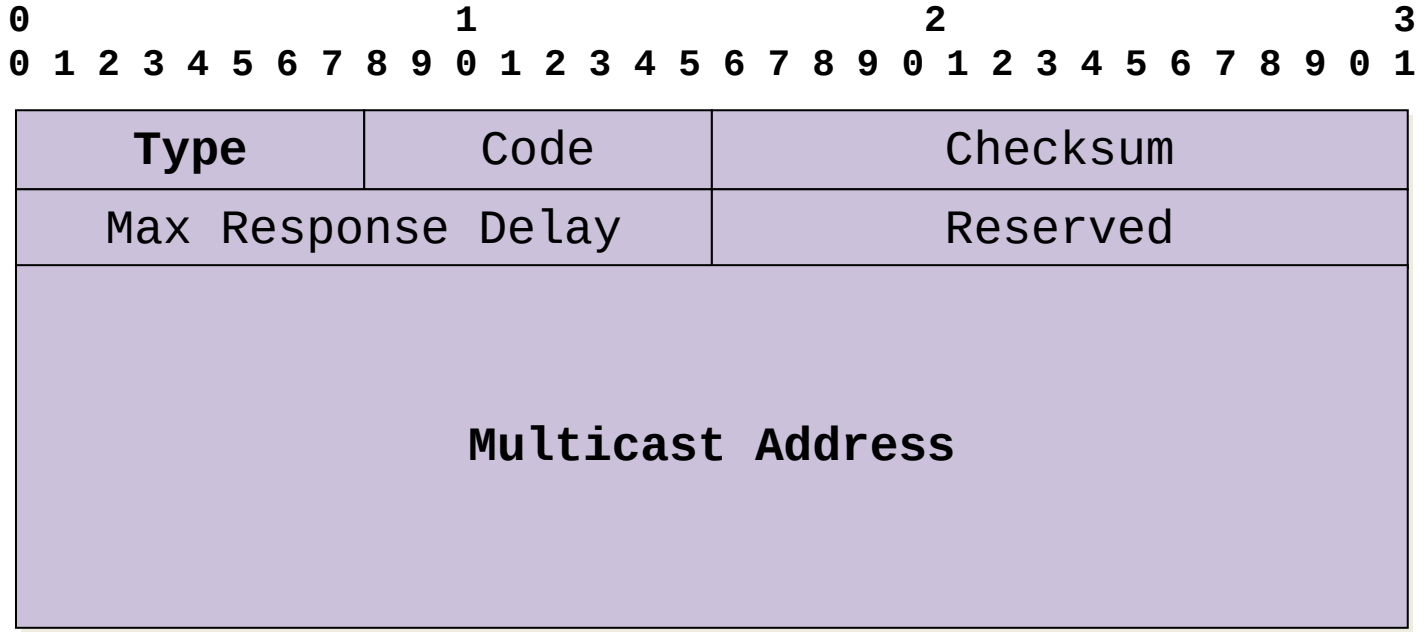
- Multicast Group Address
  - Destination address for multicast traffic
- Multicast Group
  - The group consist of sender and receiver
- MDT (Multicast Distribution Tree)
  - The tree used to distribute multicast traffic on the routers
  - Shortest Path Tree(Source Tree), Shared Tree(RP tree)
- Join
  - Joining to the multicast group to receive the traffic
- Leave(MLD)/Prune(PIM)
  - Leave from the multicast group to stop receiving the traffic
- ASM (Any Source Multicast)
  - Multicast service that only specifies group address(does not specify source address)
- SSM (Source Specific Multicast)
  - Multicast service that specifies group and source address

MLD(Multicast Listener Discovery) v1/ v2

MLDv1(RFC2710)

- Used for signaling between Multicast Listener(Receiver) and First Hop Router to inform/confirm the existence of multicast listener.
- Receiver sends “Report” to the router when join to the group.
- Receiver sends “Done” to the router when leave from the group.
- Router periodically sends MLD General Query to receiver side, and confirm the existence of listener.
- When the listener leave from the group, router sends Multicast Address Specific Query to confirm no any other listener is there.
- MLD is subset of ICMPv6 protocol

MLDv1 Packet Format



Type Field:

- **130 : Multicast Listener Query**
  - **General Query**
  - **Multicast-Address-Specific Query**
- **131 : Multicast Listener Report**
- **132 : Multicast Listener Done**

Multicast Address Field:

- **Report : Target Multicast Address**
- **General Query : zero**
- **Multicast-Address-Specific Query**
  :Target Multicast Address

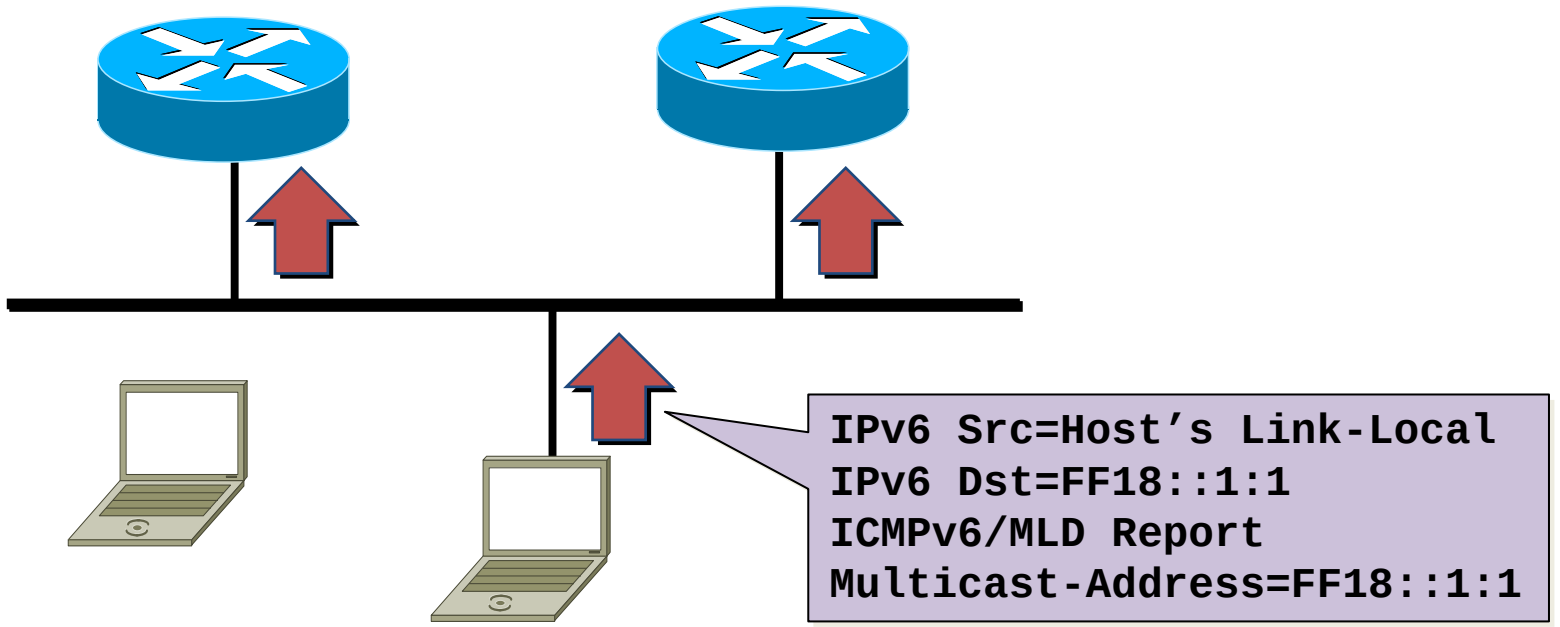
Destination address, MLDv1 timers/variable

Message Types	IPv6 destination Address
General Query	link-scope all-nodes (FF02::1)
Multicast-Address-Specific Query	Target multicast address
Report	Target multicast address
Done	link-scope all-routers (FF02::2)

Default Timer/Variable

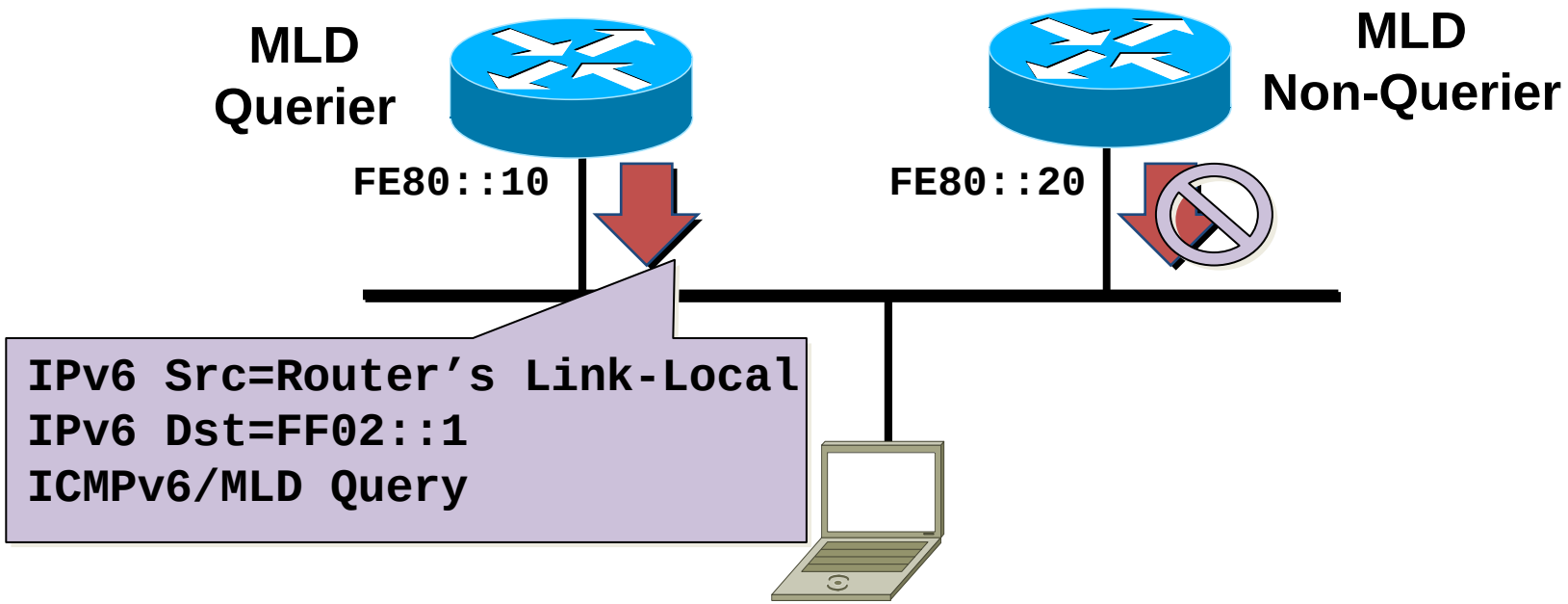
Query Interval	125 Sec
Maximum Response Delay	10000mSec
Multicast Listener Interval (expire timer on router side)	[Query Interval] * 2 + 10Sec = 260 Sec
Other querier Present interval	[Query Interval] * 2 + 5Sec = 255 Sec
Unsolicited Report Interval	10Sec
Last Listener Query Count(Robustness Variable)	2
Last Listener Query Interval	1Sec

MLDv1 Joining to the group



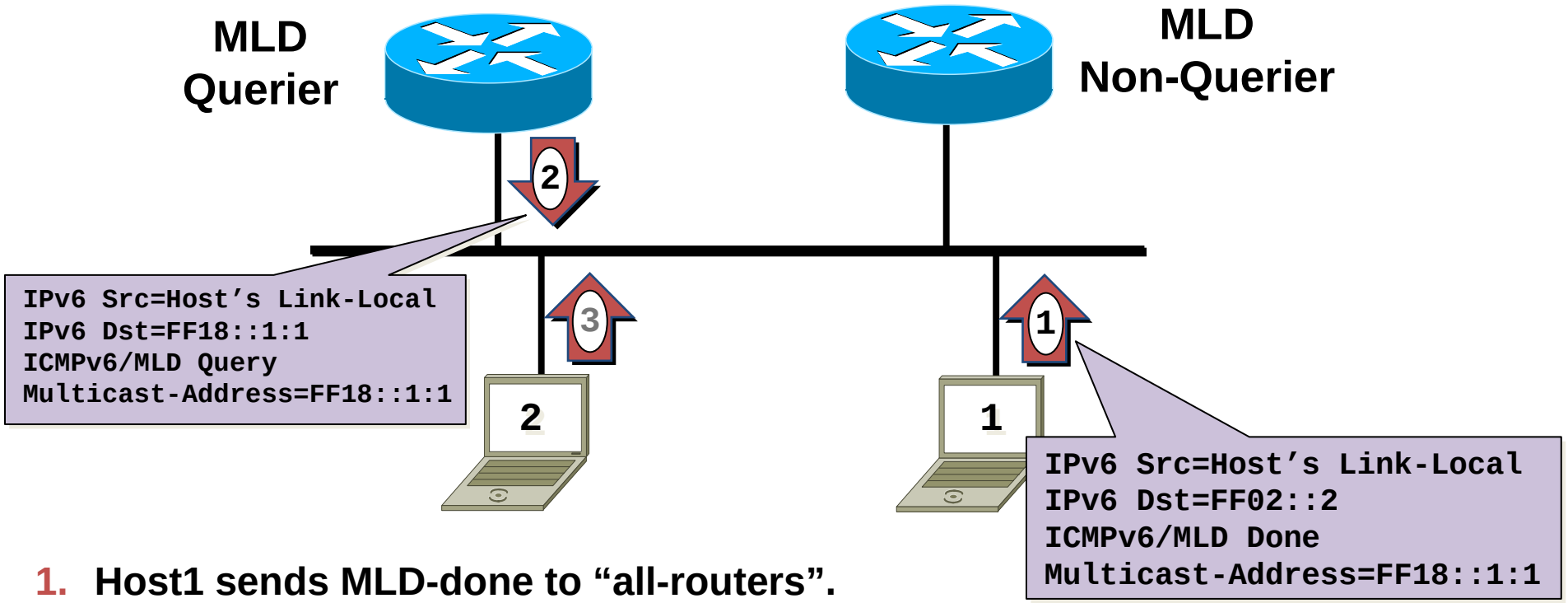
- Receiver sends MLD Report to Router for joining to the group.

MLD Querier election



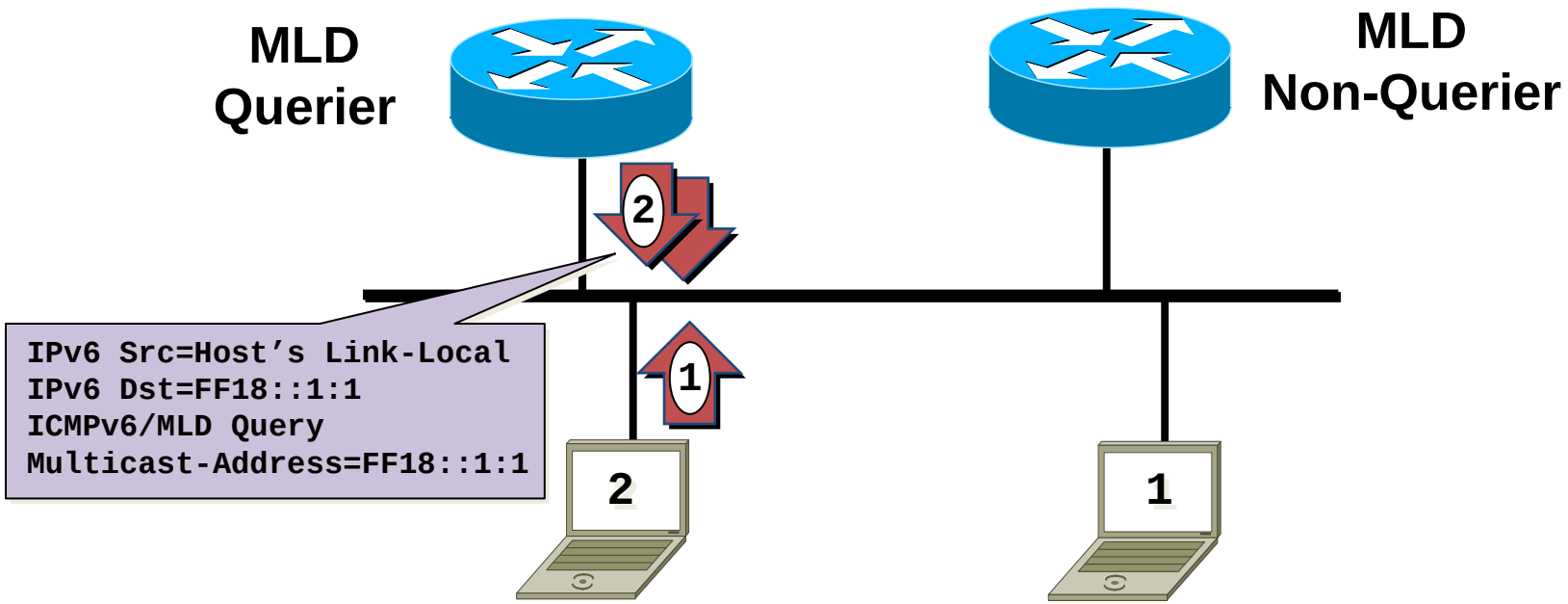
- When MLD routers become online, all routers start sending Query.
- If the router receives the query from other router that has more smaller source address, the router stop sending query(become non-querier).
- The router that has the most smallest IPv6 address become MLD querier.

MLDv1 Leave from the Group



1. **Host1 sends MLD-done to "all-routers".**
2. **MLD Querier sends Multicsat-Address-Specific-Query to target Group Address.**
3. **If Host2 still joining to the group, Host2 must send MLD report within "Last Listener Query Interval".**

MLDv1 Leave from the Group

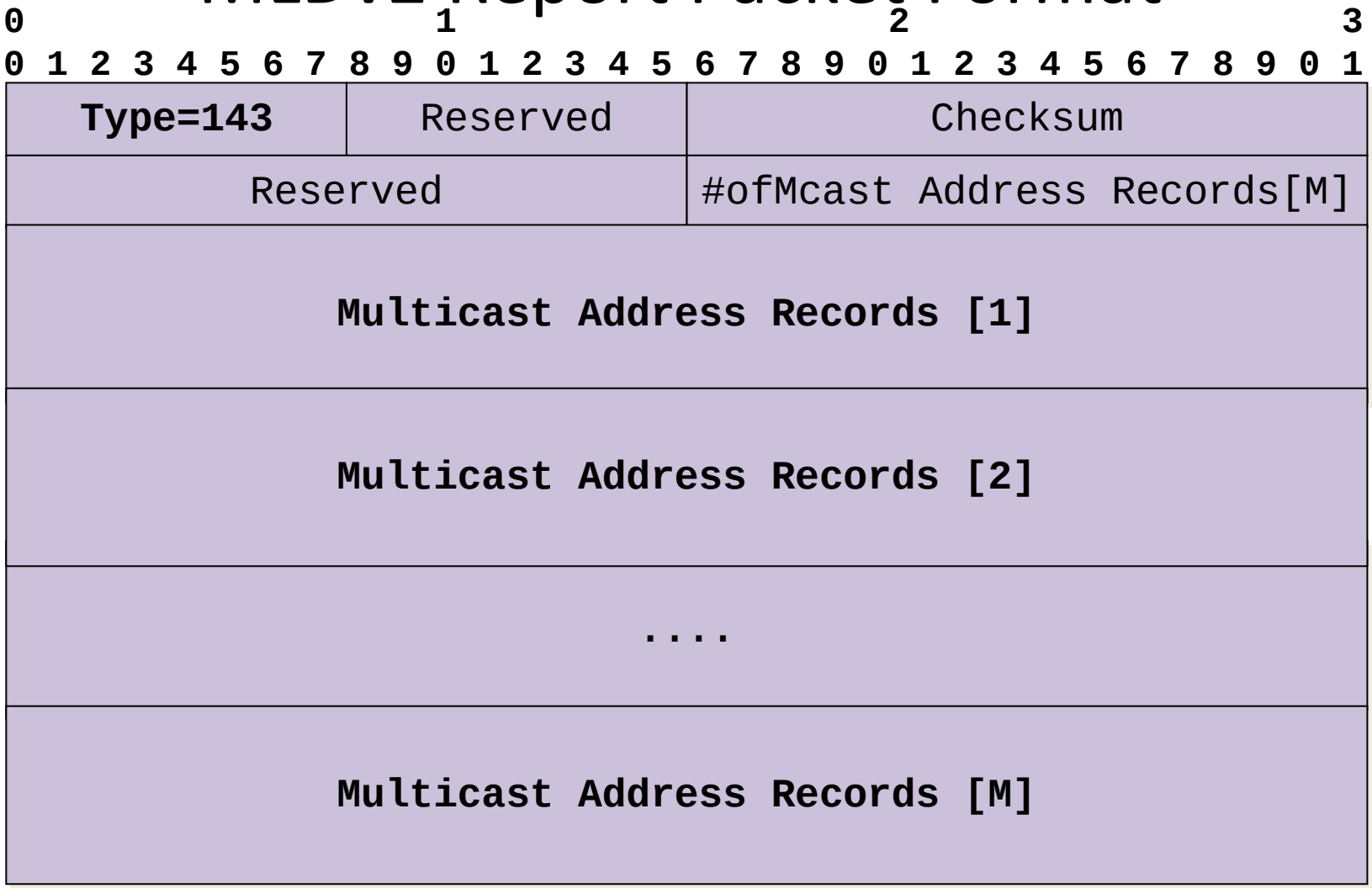


1. **Host2 sends MLD-done to "all-routers"**
2. **MLD Querier sends Multicsat-Address-Specific-Query to target Group Address. (send 2*queries with Timeout=1sec)**
3. **After timeout, routers delete MLD entry.**

MLDv2(RFC3810)

- MLDv2 enables host join/leave to Source and Group(for PIM-SSM)
- Adding Include/Exclude Source-List
- It has backward compatibility with MLDv1.
- All MLDv2 packets use same destination IPv6 address “FF02::16” in all types of messages.

MLDv2 Report Packet Format



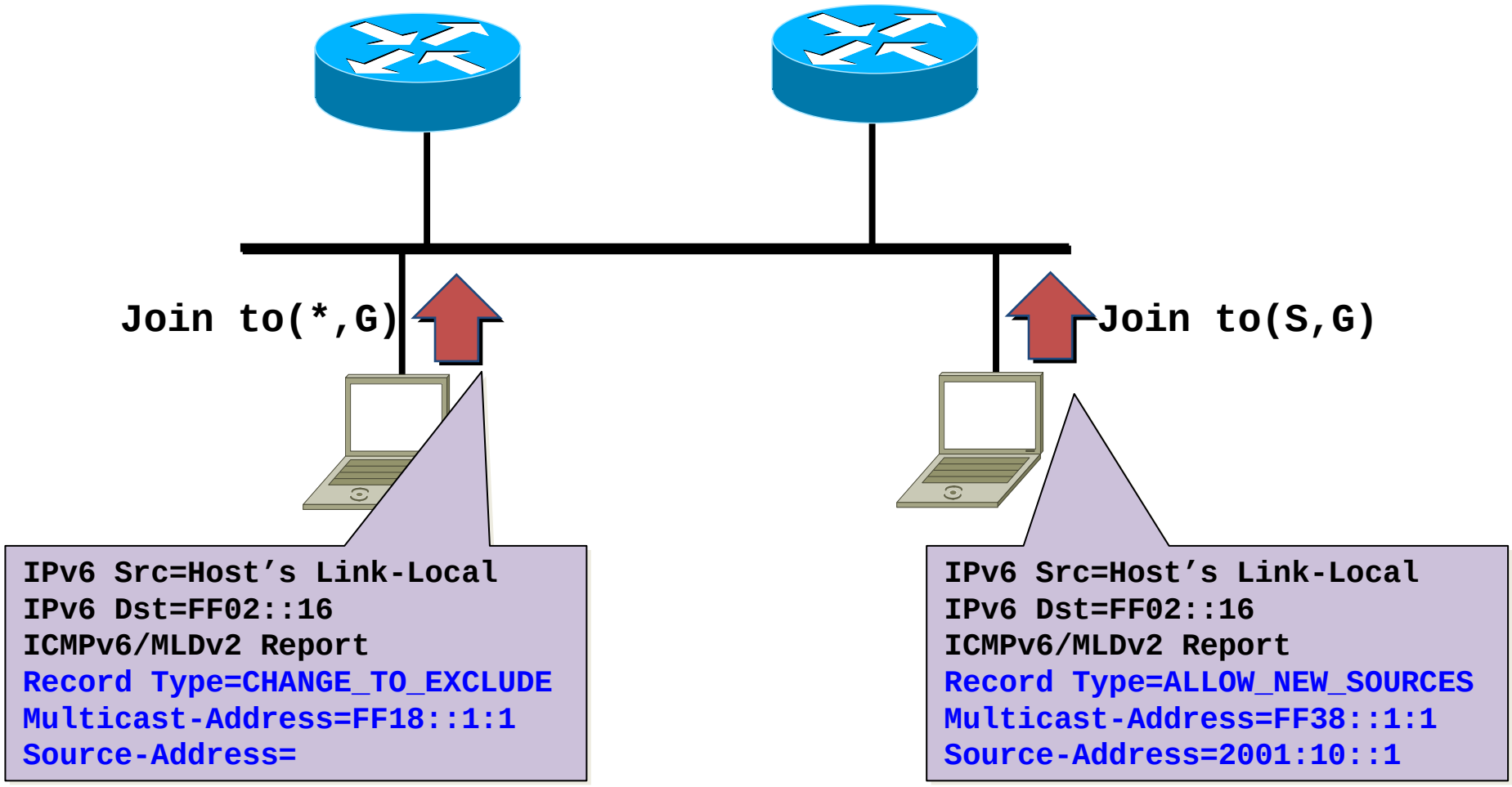
MLDv2 Multicast Address Record Format

Record Type	Aux Data Len	Number of Sources (N)
Multicast Address		
Source Address [1]		
...		
Source Address [N]		

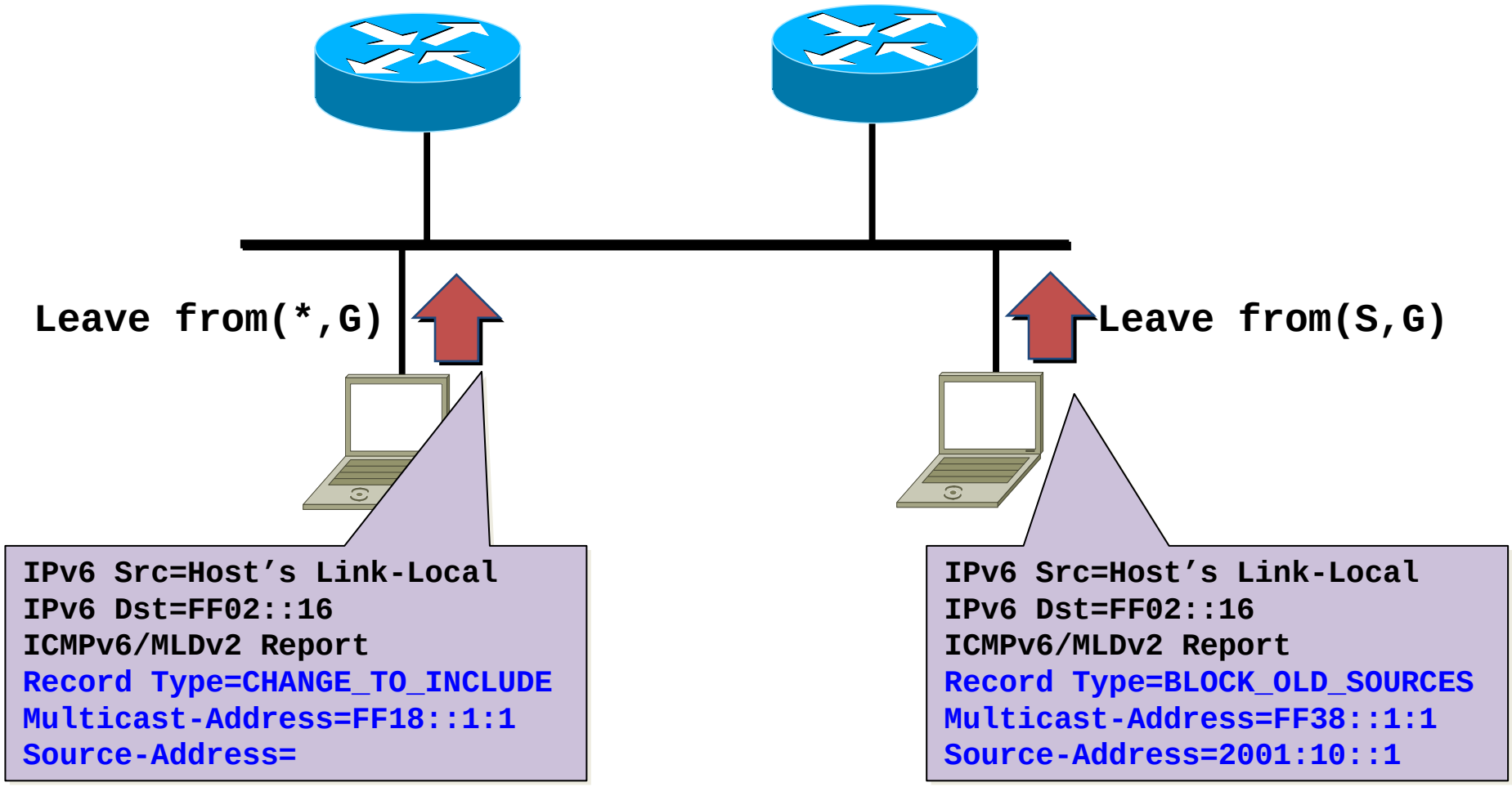
Record Type:

1. **MODE_IS_INCLUDE - IS_IN ({S},G)**
2. **MODE_IS_EXCLUDE - IS_EX ({S},G)**
3. **CHANGE_TO_INCLUDE_MODE - TO_IN ({S},G)**
4. **CHANGE_TO_EXCLUDE_MODE - TO_EX ({S},G)**
5. **ALLOW_NEW_SOURCES - ALLOW ({S},G) [join (S,G)]**
6. **BLOCK_OLD_SOURCES - BLOCK ({S},G) [Leave (S,G)]**

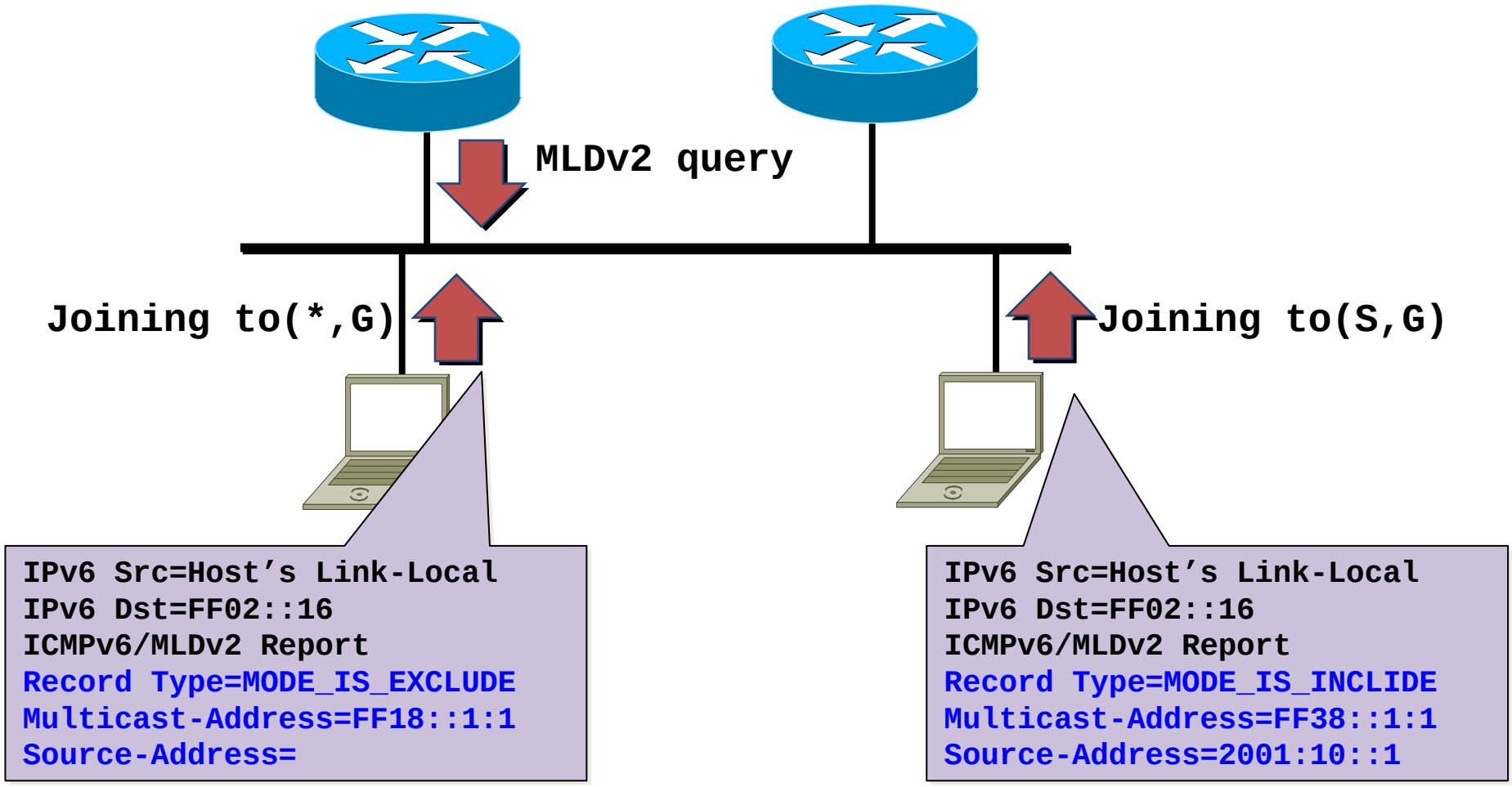
MLDv2 Join to (*,G)/(S,G)



MLDv2 Leave from (*,G)/(S,G)



MLDv2 Reply to the Query



Multicast Forwarding and PIM (Protocol Independent Multicast)

Multicast Distribution Tree(MDT)

Multicast Traffic is forwarded over the distribution tree that is built by PIM, from Upstream to Downstream.

- **Shortest Path Tree/Source Tree/(S,G) Tree**
  - **The top of Shortest Path Tree is Source**
  - **(S,G) based forwarding**
  - **shortest path from Receiver to Source**
- **Shared Tree / RP Tree / (*,G) Tree**
  - **The top of Shared Tree is RP**
  - **(*,G) based forwarding**
  - **shortest path from Receiver to RP**

Reverse Path Forwarding (RPF)

- **RPF interface(Incoming interface)**

  Each router selects the interface as upstream interface for Source/RP. Selection of the upstream interface is based on Unicast Routing information or routing information only for Multicast like BGP-mcast-addr-family/static-mroute.

- **Outgoing interface(List)**

  Downstream interface that received Join.OIF/OIL

- **RPF Neighbor**

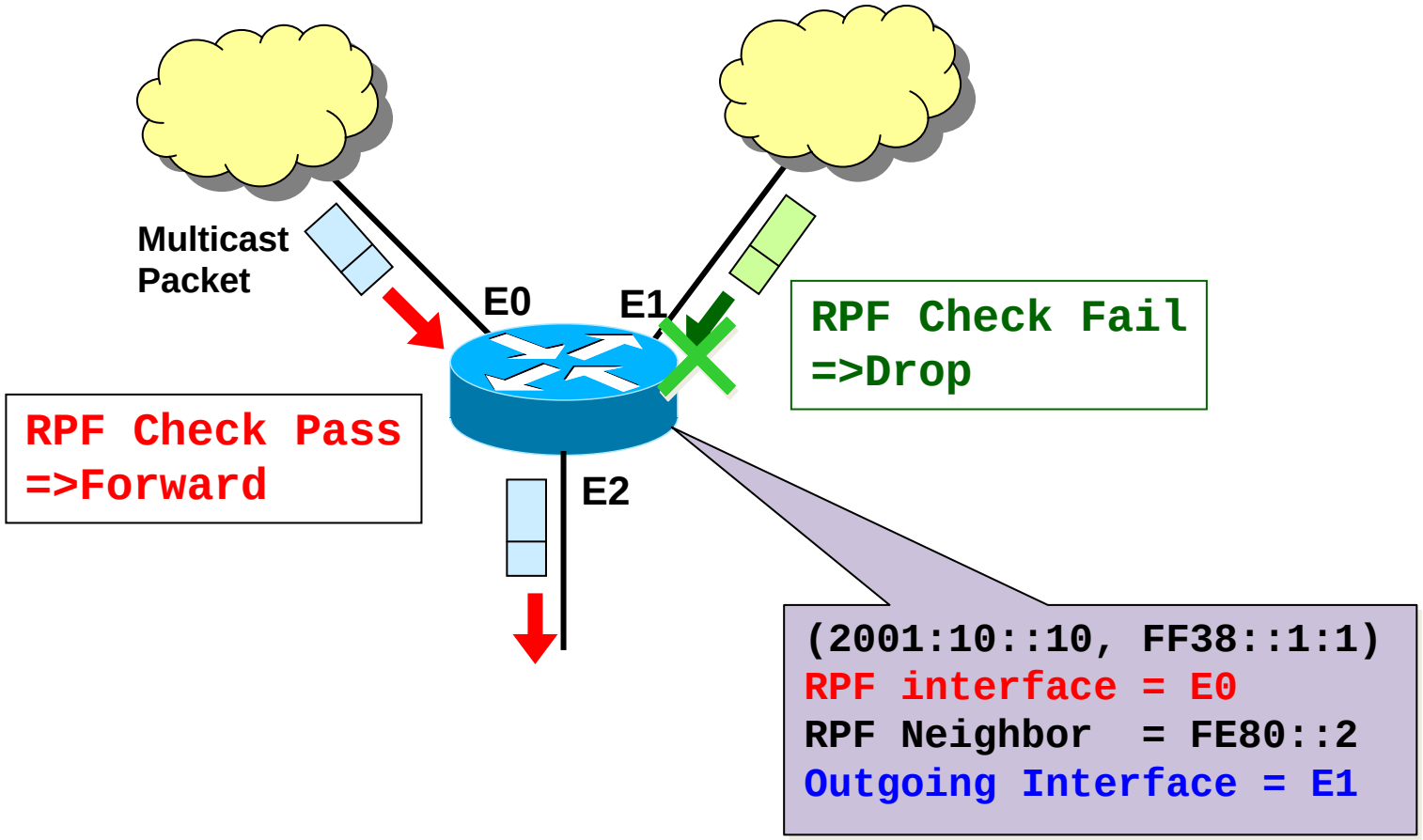
  Next hop router(address) on RPF interface(upstream side) towards source/RP. Each (*,G)/(S,G) entry will select each RPF neighbor. PIM Join/Prune must be sent out to RPF Neighbor.

- **RPF Check**

  If the multicast packet for (*,G)/(S,G) is received on RPF interface, that packet is forwarded to OIFs.

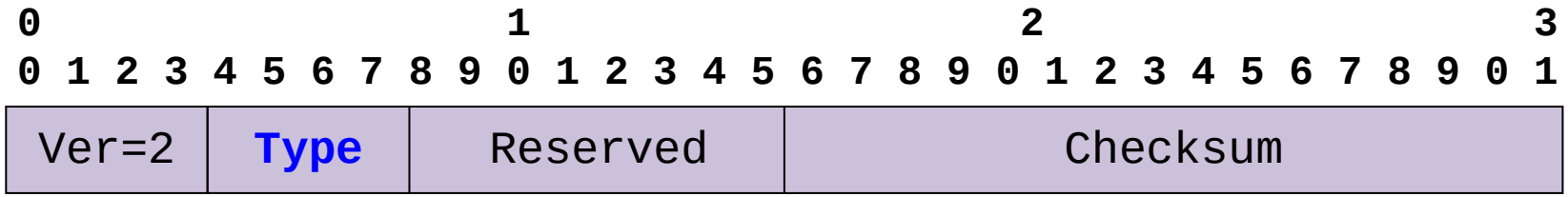
  If the multicast packet is received on non-RPF interface, that packet is discarded.

RPF Check



PIM Header Packet Format

IPv6 Source Addr = Router Link-Local address
IPv6 Destination Addr = depends of message type
IPv6 NextHeader = 103(0x67)



Message Types	IPv6 destination Address
0 = Hello	ALL-PIM-ROUTERS(FF02::D)
1 = Register	RP Address
2 = Register-Stop	First Hop Router(source of register)
3 = Join/Prune	ALL-PIM-ROUTERS(FF02::D)
4 = Bootstrap	ALL-PIM-ROUTERS(FF02::D)
5 = Assert	ALL-PIM-ROUTERS(FF02::D)
8 = Candidate-RP-Advertisement	BSR address

PIM Hello/PIM Neighbor

- PIM Routers periodically send PIM hello packet on the link, and each Router recognize other routers as PIM Neighbor.
- If RPF neighbor is recognized as PIM Neighbor, router can send PIM Join/Prune to RPF neighbor.
- When there are multiple PIM routers on same Subnet(Link), one PIM DR(Designated-Router) is selected on the link based on PIM-DR-Priority（if routers have same DR-priority, biggest address is selected as DR.)

PIM Neighbor Discovery

- The router that has most highest DR-priority is elected as DR on that link. (If DR-priority is same, most biggest address wins).
- Router periodically sends PIM Hello to "FF02::D"(All-PIM-ROUTERS)

  Default Timer Vaule

  Hello Interval = 30 Sec

  Holdtime = 30 x 3.5 = 105 sec
- When DR is Timeout, new DR is elected.
- Only DR can do the PIM Join/Prune activity by receiving MLD Join/Leave. (Non-DR can not start any PIM action)

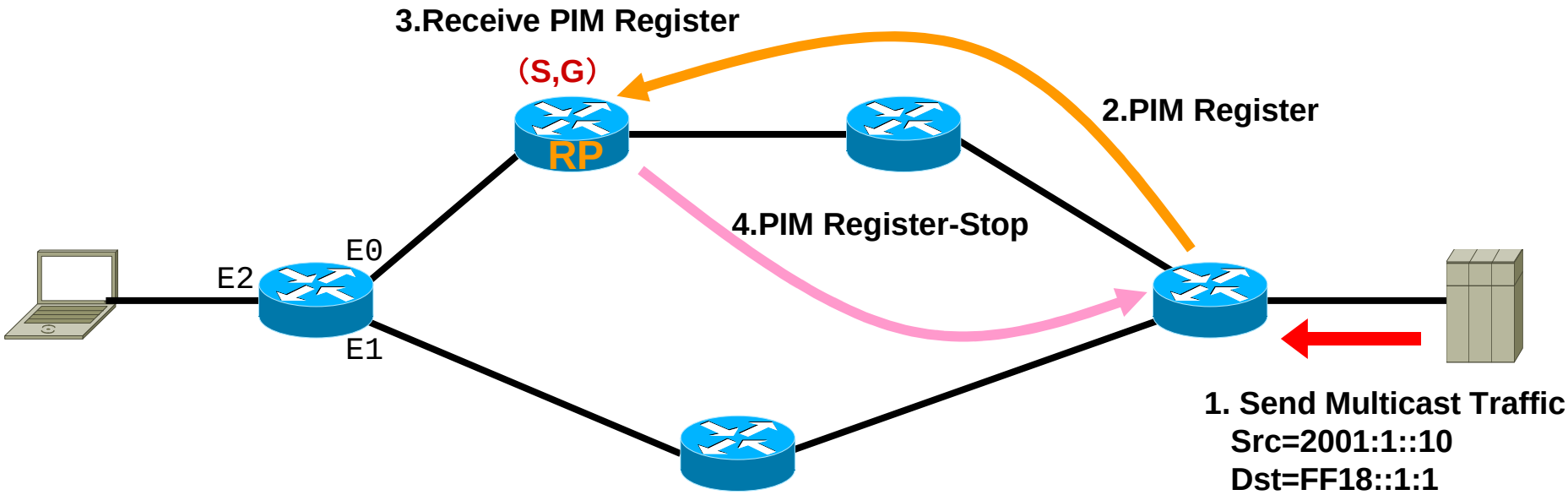
PIM SM(RFC4601)

◆Requires RP(Rendezvous Point)

◆Source information is registered on RP(by first hop router). Join request is sent toward to RP by Hop-by-Hop.RP is managing (S,G) information in that multicast domain.

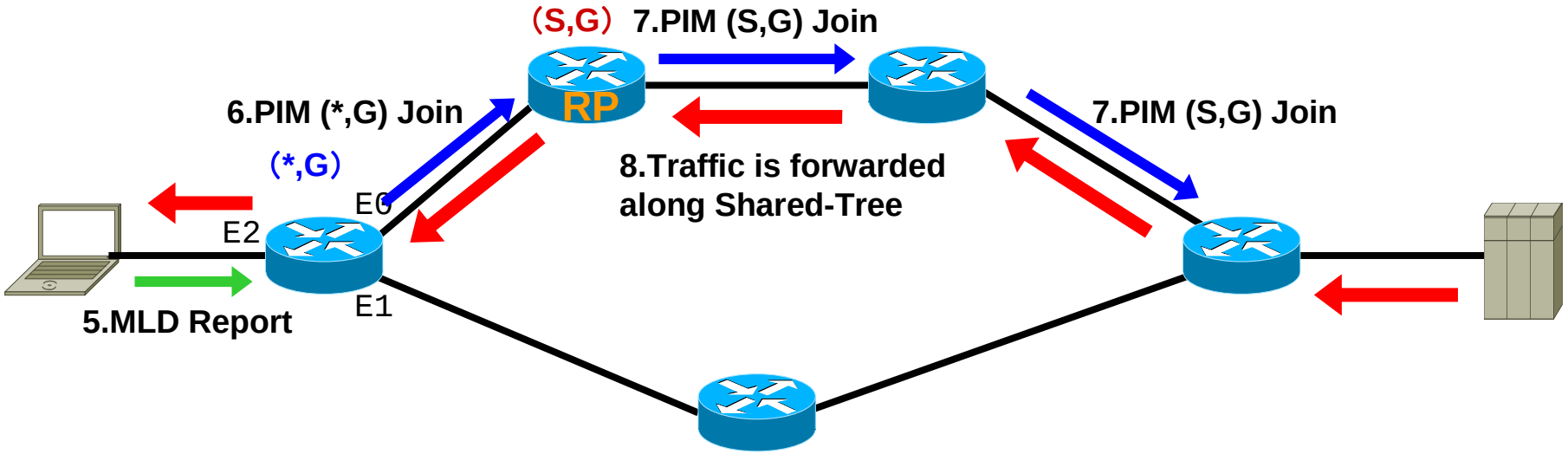
◆Effective for “one to many” or “many to many” communication.

PIM-SM Sequence (1)



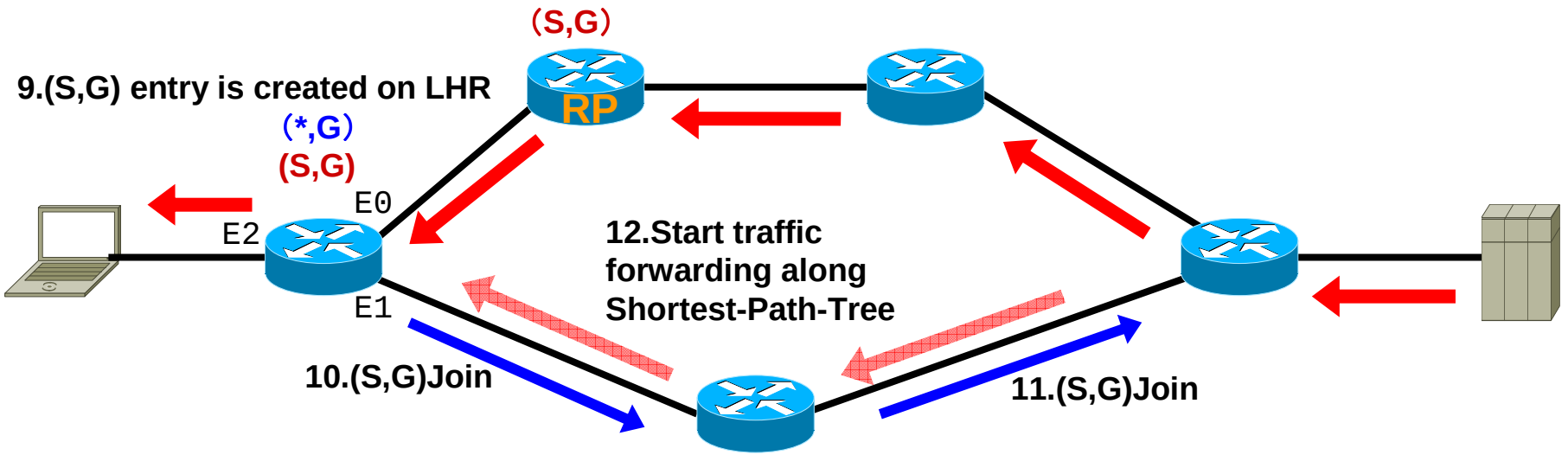
1.Source starts sending Multicast Traffic(2001:1::10,FF18::1:1).
2.First Hop Router sends PIM Register to RP in Unicast.
  Original multicast packet is encapsulated in PIM packet.
3.RP receives Register,and create (S,G) info on RP.
4.RP sends back PIM Register-Stop to First Hop Router.
   First Hop Router stop sending PIM Register.

PIM-SM Sequence (2)



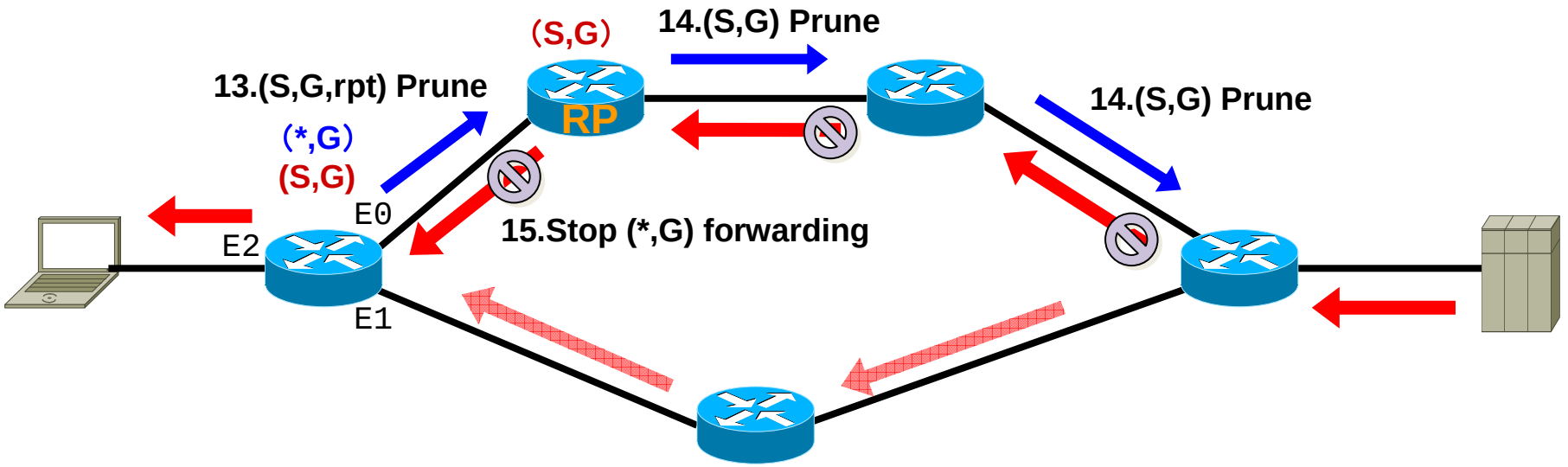
5.Receiver sends MLD Report for the group (*,FF18::1:1).
6.Last Hop Router sends PIM (*,G) Join towards RP.
#Shared-Tree(RP-Tree) is built.
7.From RP towards source, router sends PIM(S,G) Join.
8.First Hop Router receives (S,G) Join and traffic forwarding is started from Source to Receiver through RP.

PIM-SM Sequence (3)



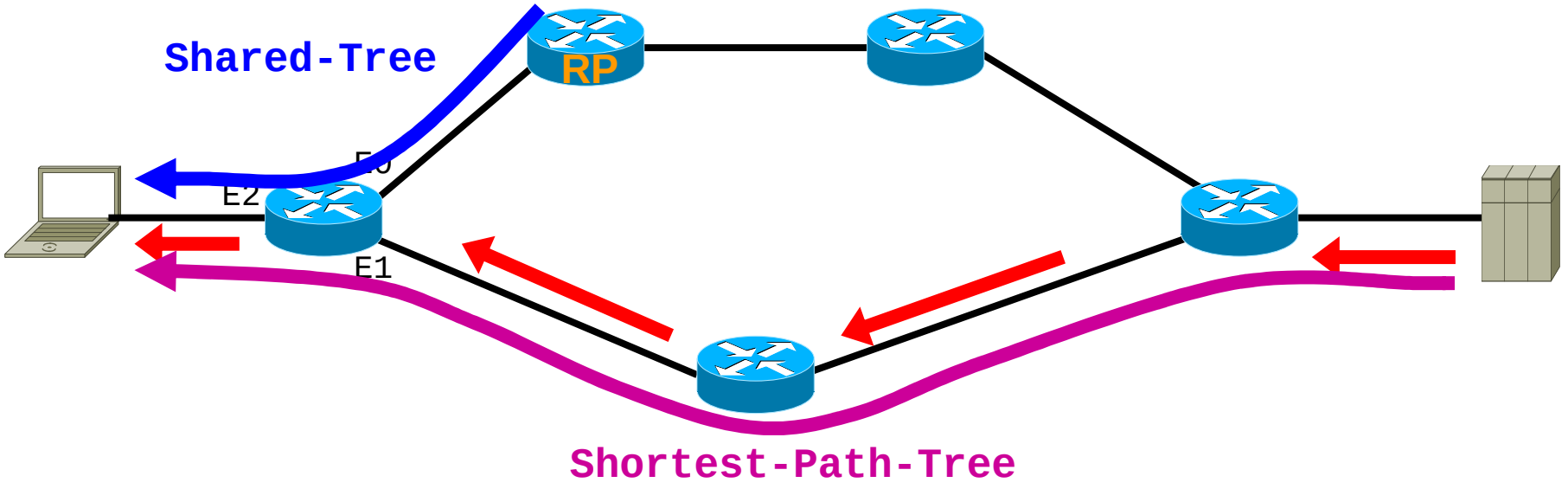
9.By receiving source traffic,(S,G)entry is created on LHR.
   In above case, (*,G) RPF-IF = E0
                  (S,G) RPF-IF = E1
10.LHR sends (S,G) Join towards RPF interface for Source.
11.Router sends (S,G) Join towards RPF interface for Source
   to build Shortest-Path-Tree.
12.Start traffic forwarding along Shortest-Path-Tree.

PIM-SM Sequence (4)



13.LHR start receiving traffic on Shortest-Path-Tree,
LHR sends (S,G,rpt)prune towards RP because the traffic via RP is not necessary any more.
(*,G)RPF-IF ≠ (S,G) RPF-IF
14.Stop Shared-Tree based forwarding.
15.(S,G) prune is sent from RP towards source, and traffic from source to RP is stopped.

PIM-SM Sequence (5)

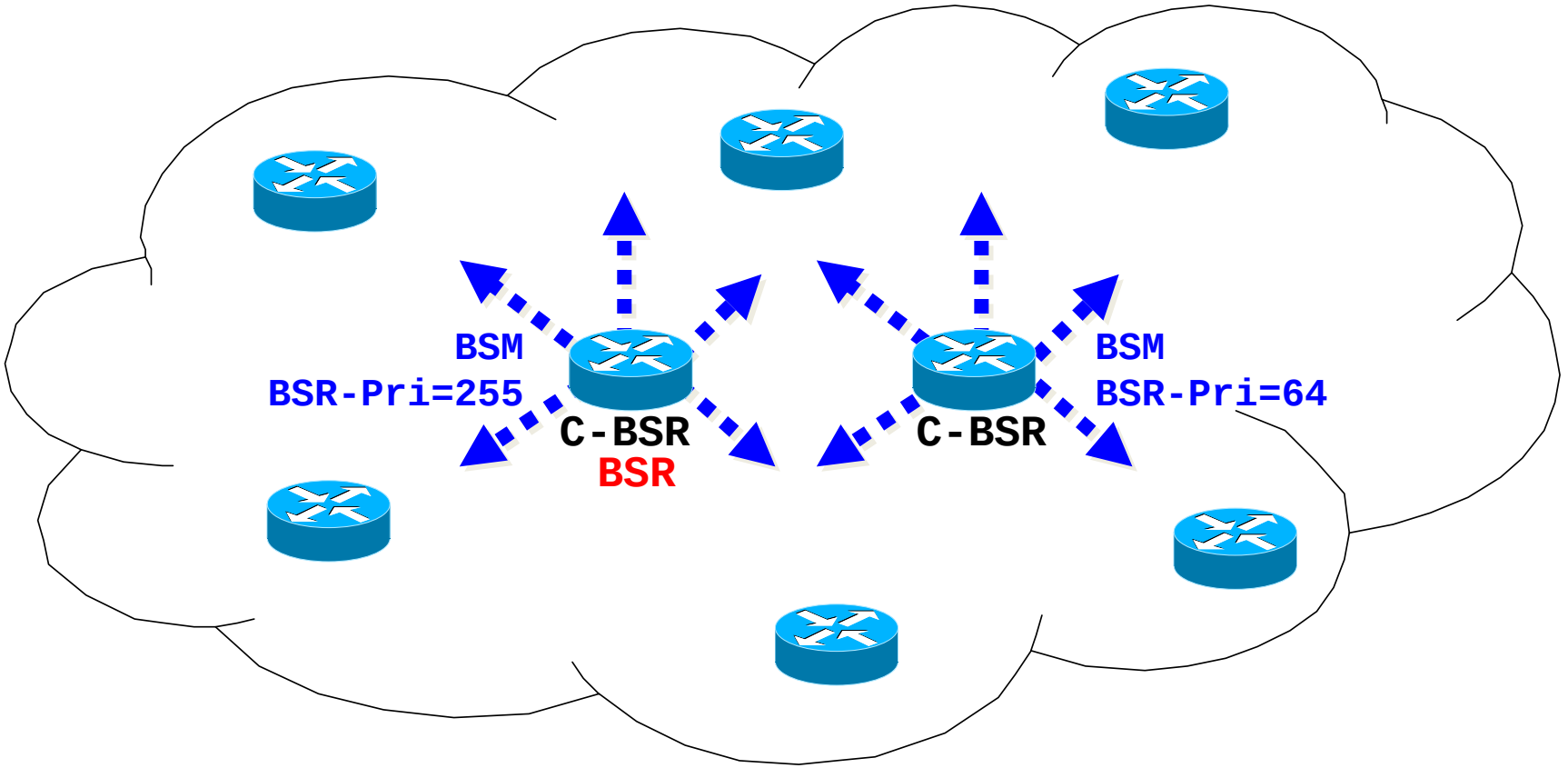


Finally traffic is forwarded only along shortest-path-tree.(SPT switchover)
After SPT is built, Shared-Tree is maintained but not used for traffic forwarding.

Defining RP

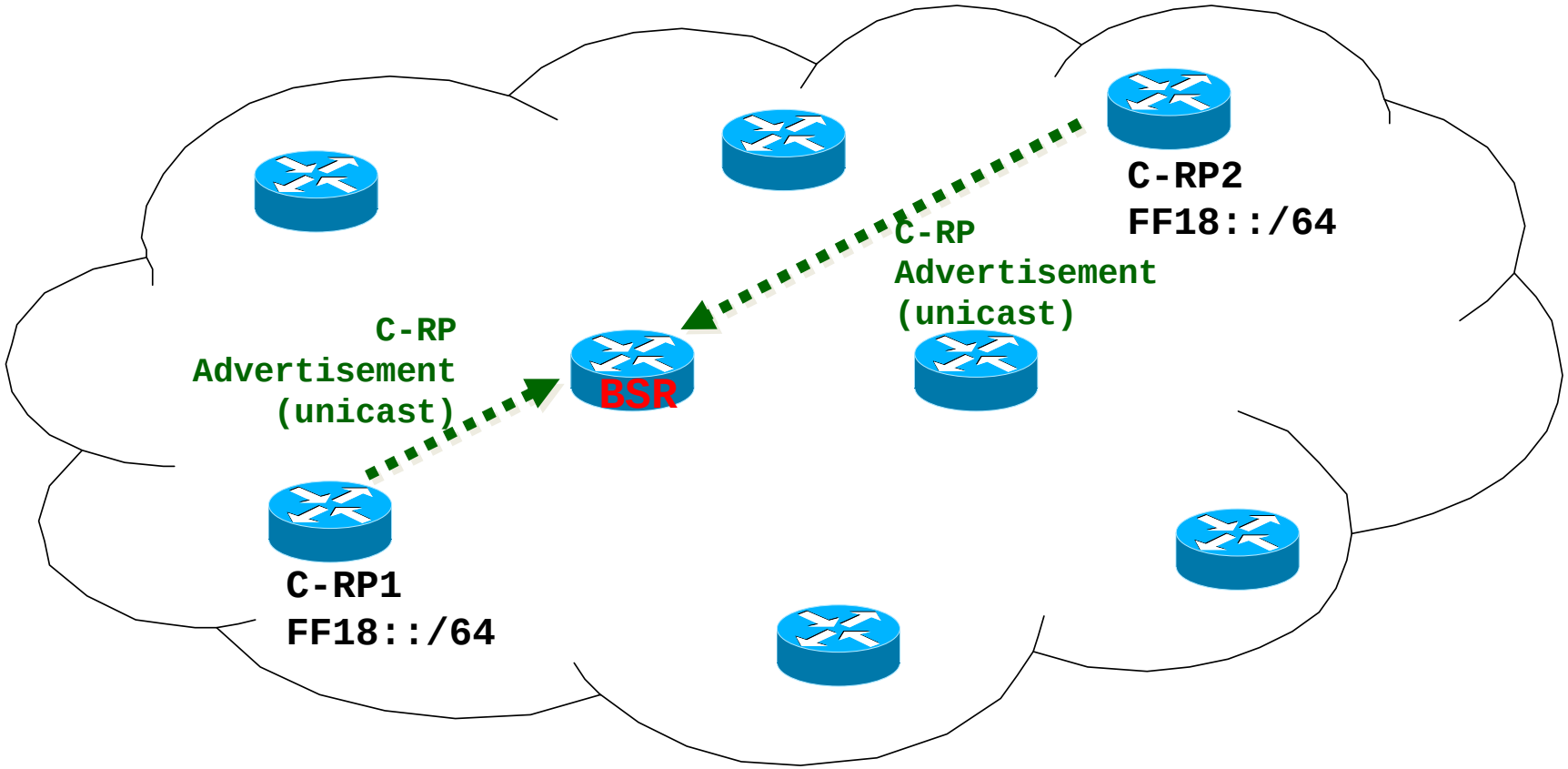
- **Static-RP: Statically configuring RP address.**
  All routers need configuration.
- **BSR(Bootstrap Router):**
  Candidate-RPs information is distributed to all routers, and RP is automatically elected from Candidate-RPs based on priority/hash.
- **Embedded-RP**
  RP address is embedded in Multicast Address.
  mainly used for PIM-SM Inter-domain connection.

PIM BSR



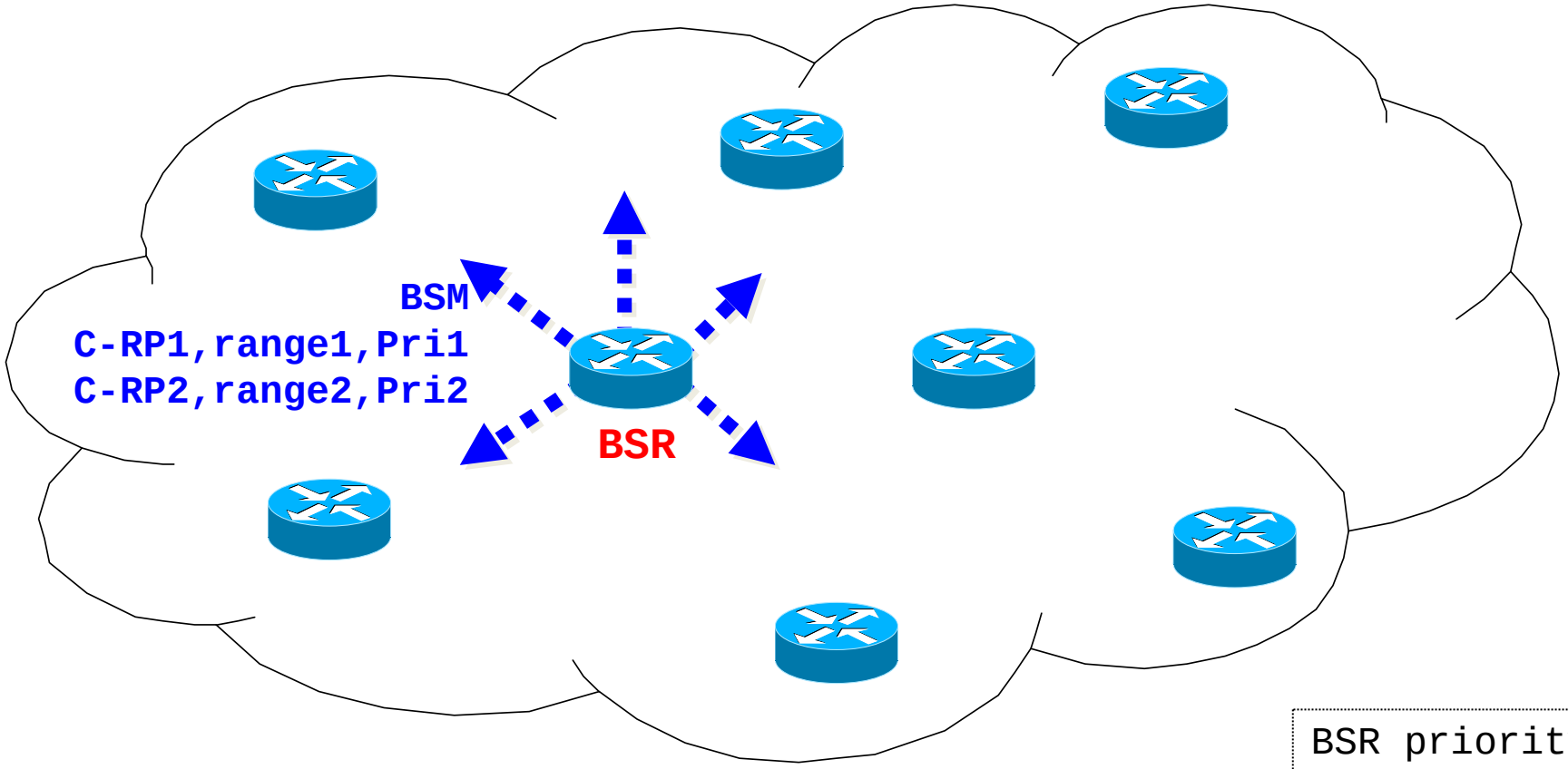
Candidate-BSR(C-BSR) floods Bootstrap Message(BSM) to all routers by Hop-by-Hop.
C-BSR that has most highest BSR Priority is elected as BSR.

PIM BSR



Candidate-RPs(C-RP) send C-RP-Advertisement to BSR in Unicast.(C-RP-Advertisement includes C-RP-Addr,Group-range, RP-Priority information.)

PIM BSR

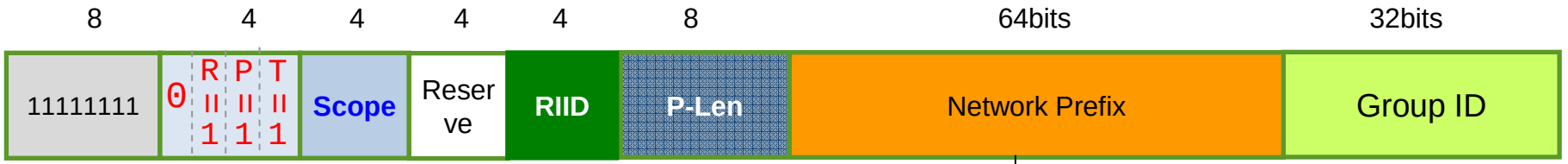


BSR floods all C-RPs information by BSM.
All other routes receive this BSM, and elect RP.

BSR priority
 255 > 0
RP priority
 0 > 255

Embedded-RP(RFC3956)

- Based on “Unicast-Prefix based Multicast Address”.
- RP address is embedded in Multicast Address.
- Router can know RP address from Multicast Address.
- Mainly used for PIM-SM inter-domain multicast.



◆example:

FF78:0140:2001:0123:c001:1100:0000:0001

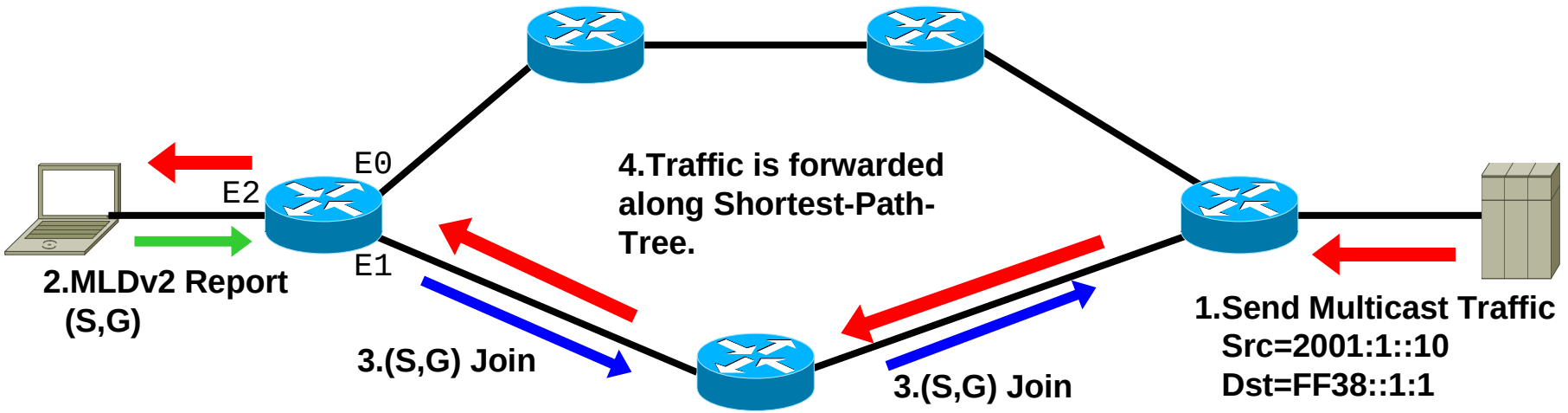
Unicast prefix　　groupID

Embedded RP Address = 2001:0123:c001:1100::1

PIM SSM (RFC4601/3569/4607)

- No RP required
- Receiver join to (S,G)
- more simpler than PIM-SM
- subset of PIM-SM

PIM-SSM Sequence



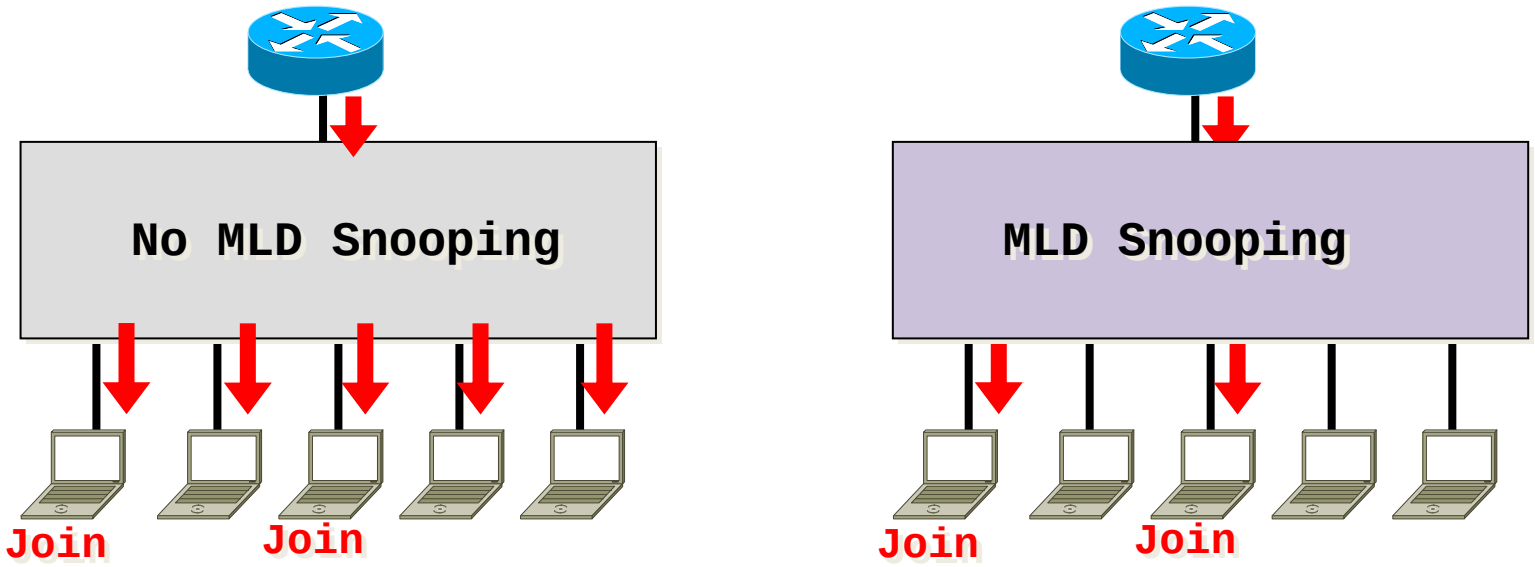
1.Source starts sending Multicast Traffic (2001:1::10,FF38::1:1).
2.Receiver send MLDv2 report for the Group/Source.
(2001:1::10,FF38::1:1)
3.Last Hop Router sends PIM (S,G)Join towards source.
Shortest-Path-Tree is built.
4.Multicast traffic is forwarded from Source to receiver along.
Shortest-Path-Tree.

MLD Snooping

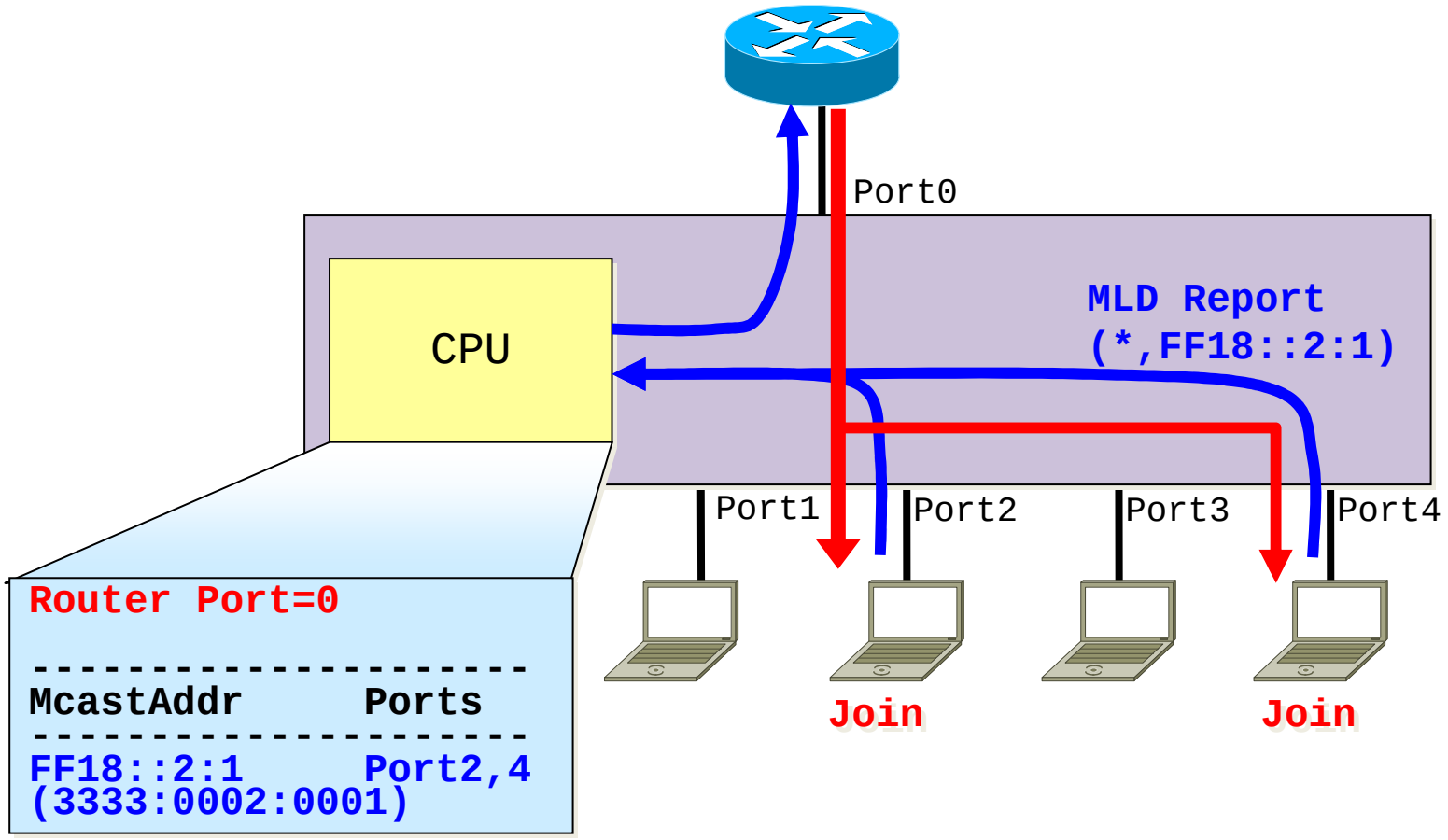
Necessity of MLD Snooping

◆ On Non-MLD-aware L2-swtich, all multicast traffic is flooded to all ports in the vlan.

◆ With MLD-snooping enabled, multicast traffic is forwarded only to the ports that receiver joined.

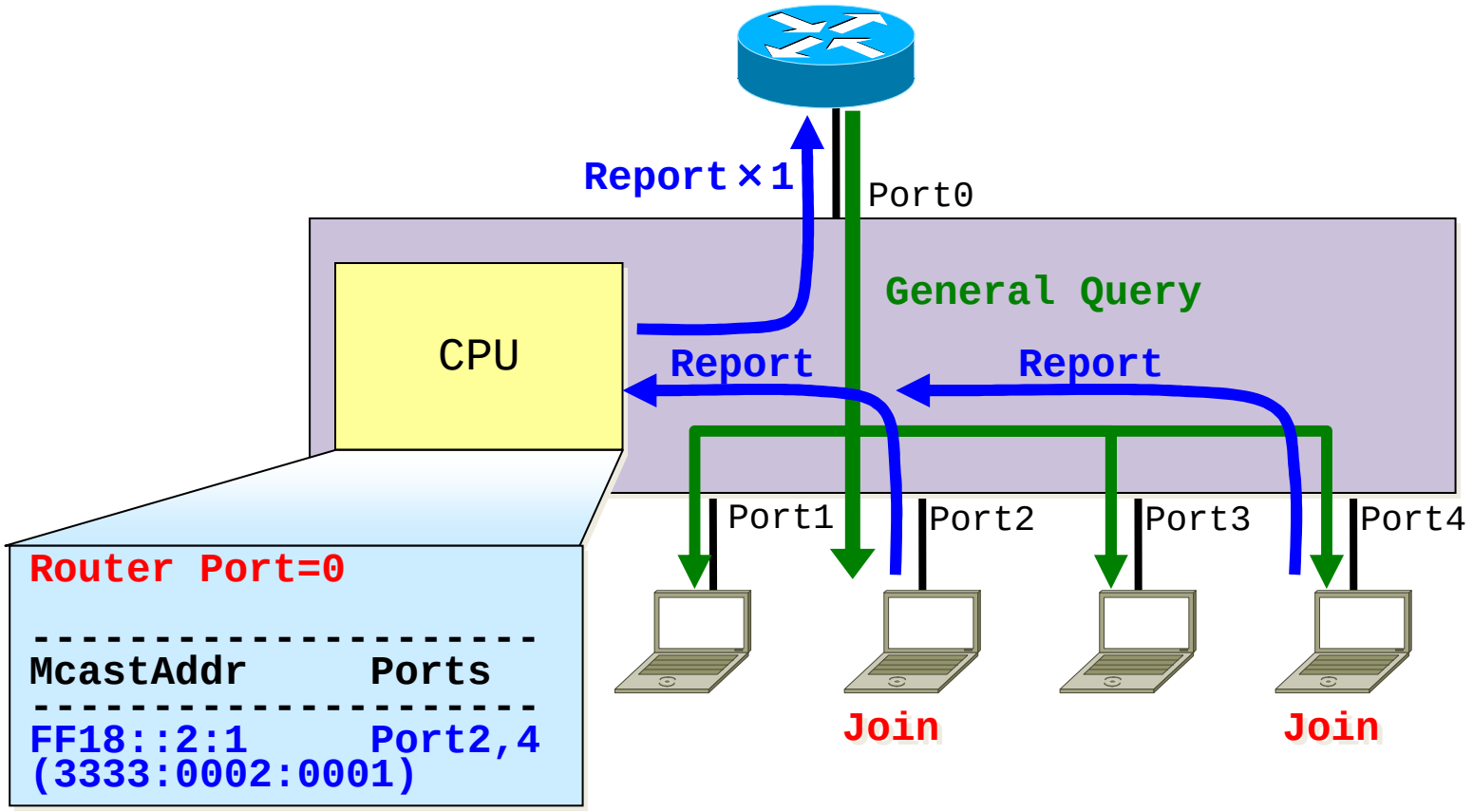


MLD Snooping: Join to the Group



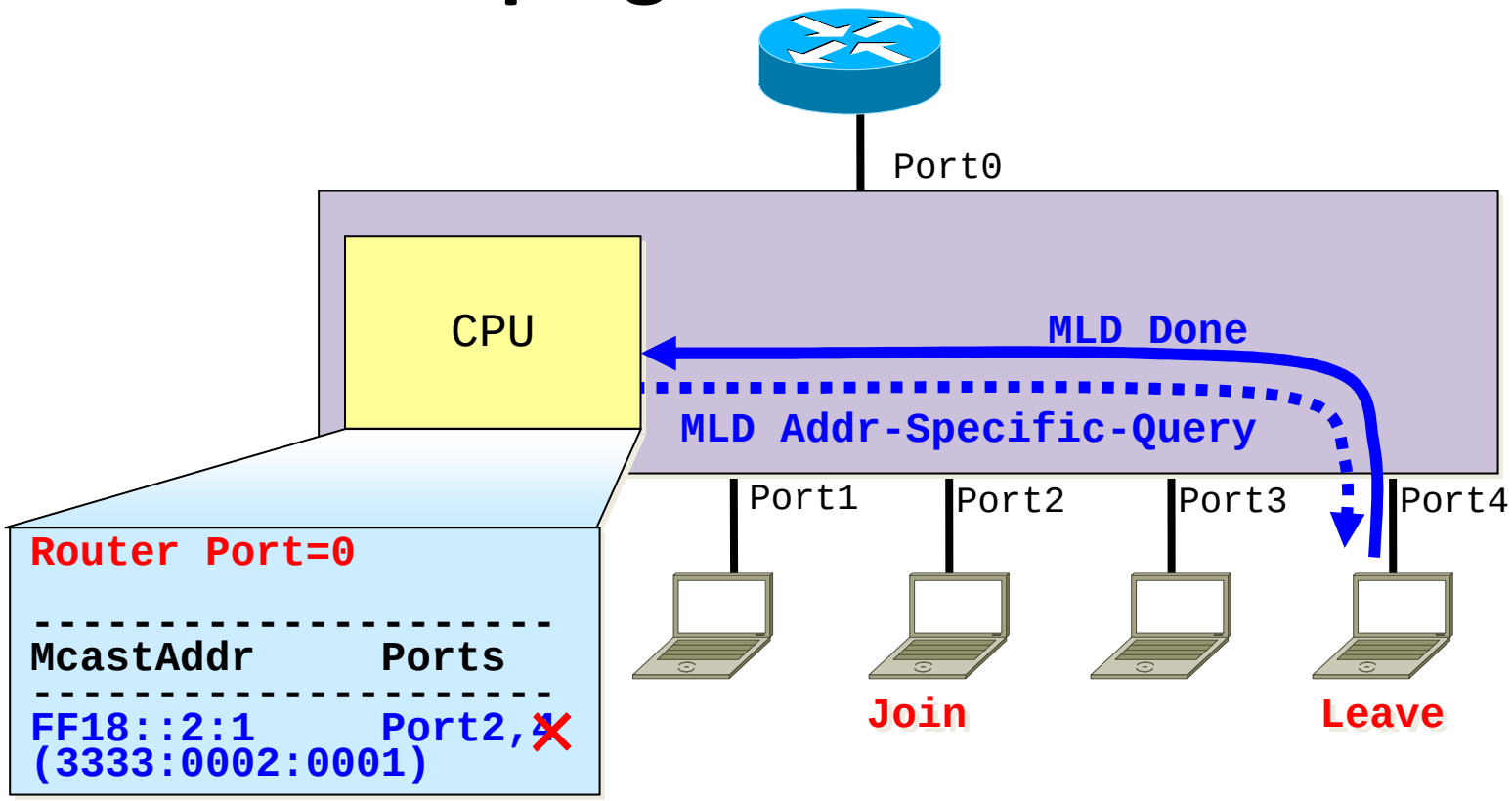
The switch snoops MLD Packet, and create L2 forwarding table based on the request of MLD packet. After that,MLD packet is forwarded to the router.

MLD Snooping: Maintaining MLD-snooping Entry



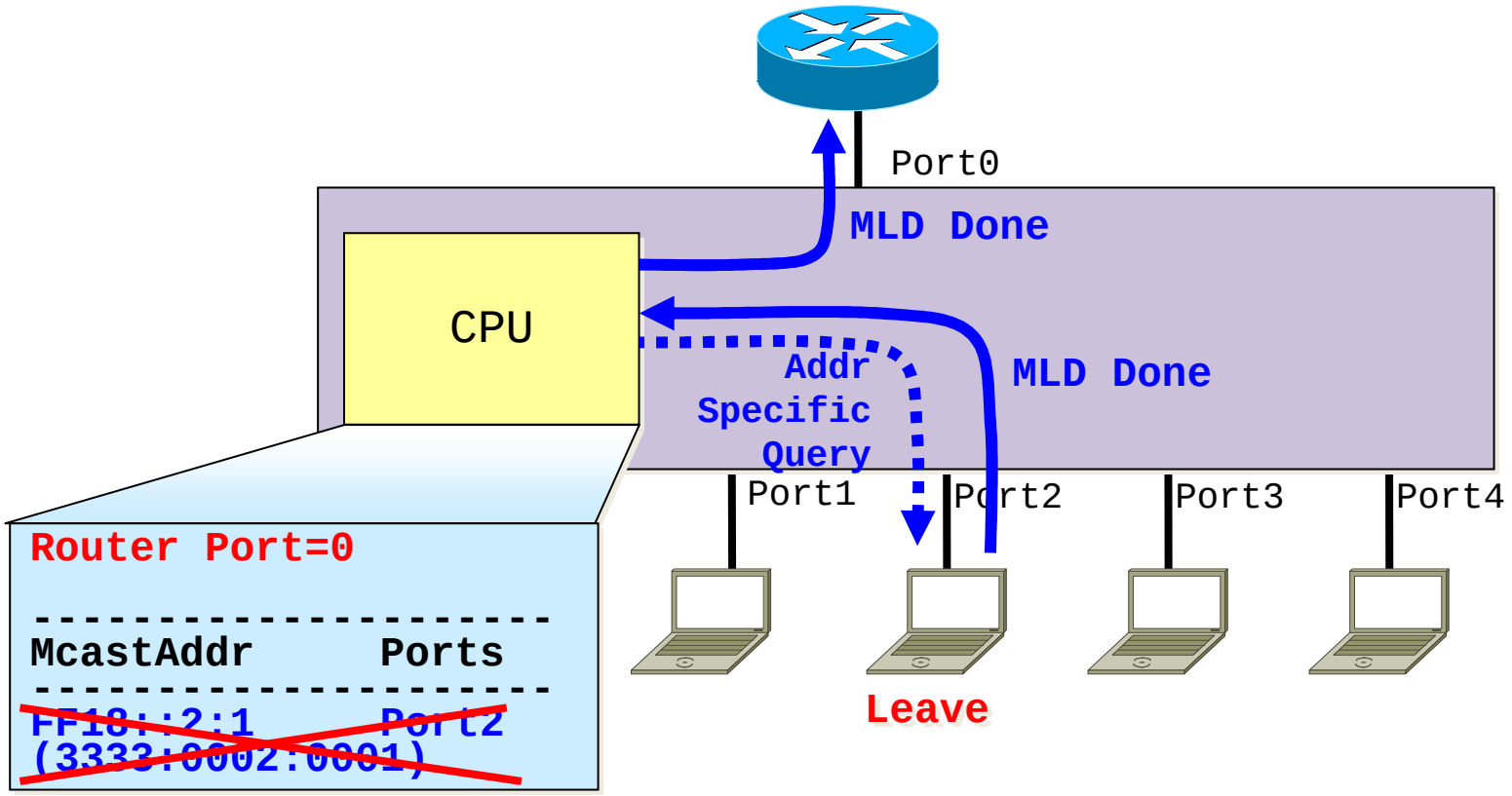
MLD General Query from the router is flooded to all ports.
Joined hosts send back MLD report and each MLD snooping entry is maintained.only one MLD report is forwarded to the router.

MLD Snooping: Leave from the Group (1)



When one host leaves from the group,switch sends multicast-address-specific-query to that port, and after timeout of the query, switch deletes that port from MLD snooping entry. When still there is any other host joining to that group, switch does not send MLD done to the router.

MLD Snooping: Leave from the Group



When Last listener leave from the group, switch sends multicast-Addr-Specifig-Query to that port.
After the query timeout, switch deletes MLD entry for that group and send MLD done to the router.

MLD Snooping: others

◆ **Router Port identification**:

MLD Snooping does work correctly only when Router Port exists or there is MLD Querier on the vlan. Router Port can be identified automatically by receiving MLD Query/PIM Hello packet, or need static configuration.(depends on switch's implementation.)

◆ **The role of Router Port**:

Receiver side:**MLD packet is forwarded to Router Port**
Sender side:**Multicast Traffic is forwarded to Router Port**

◆ **MAC address duplication**:

If the switch identify the multicast group based on MAC address only, the switch may not be able to differentiate multiple groups that has same MAC address.(source address identification of MLDv2 snooping has same issue.)

MLD Snooping: immediately Stopping traffic

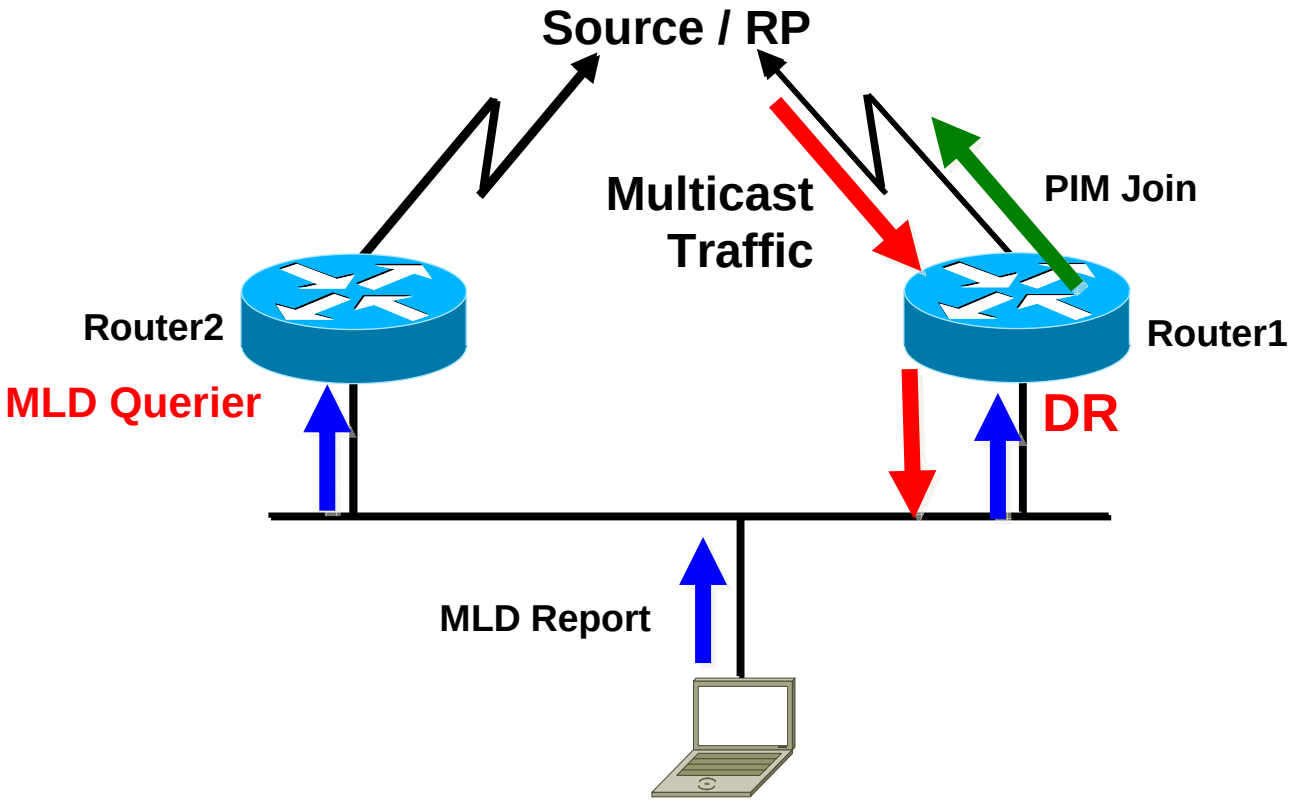
In IP/TV or video-surveillance environment, listener frequently Join/Leave to the group, router should shorten the time to stop forwarding traffic after receiving leave.

- **MLD Fast-Leave**:
  When the router/switch receives the leave, immediately stop the traffic without sending multicast-address-specific query. In this case, it must be 1host/1port.

- **MLD Host-tracking**:
  Router/switch is tracking all listener's address that joining to the group, and last listener leave from the group on that port, immediately stop the traffic.

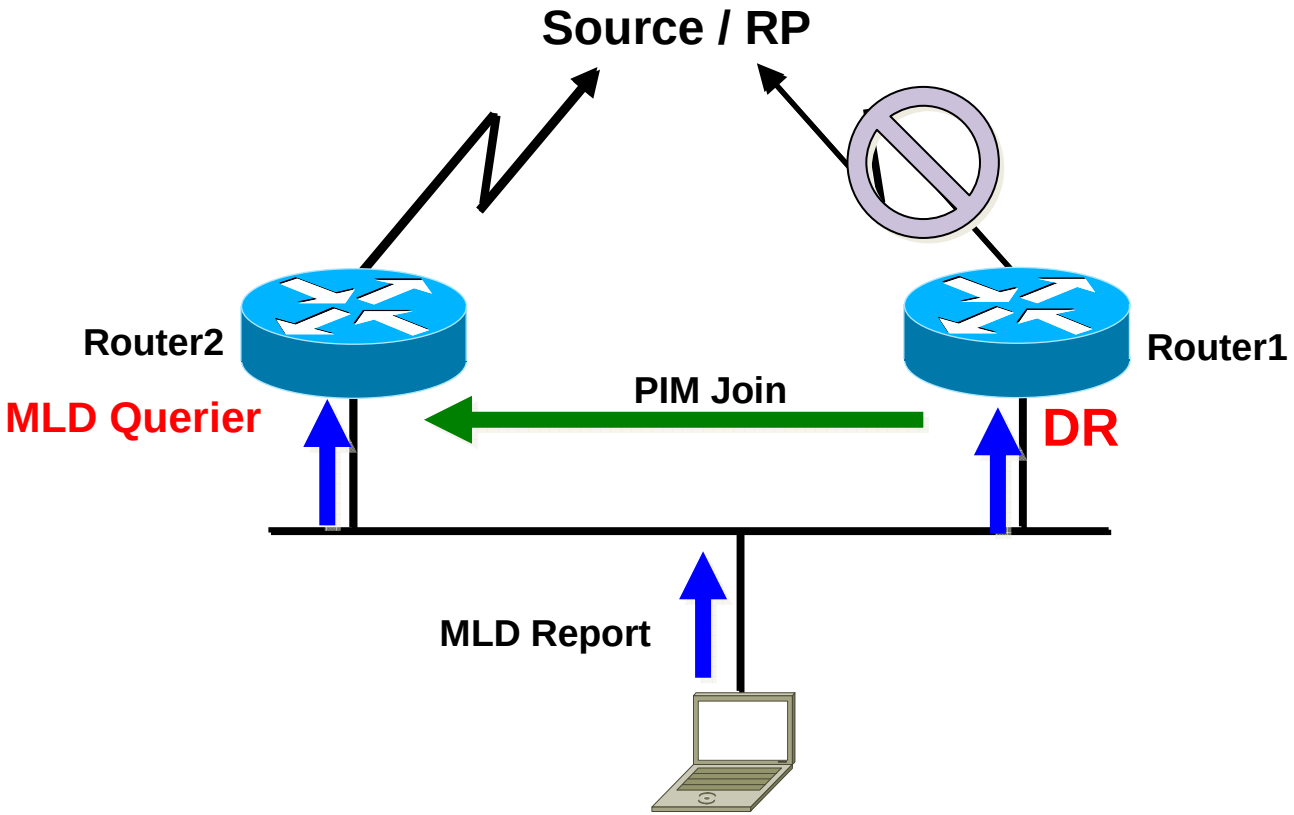
Failure Case Study

DR on Receiver Side



User expects that Router2 as Primary Multicast Forwarder on that segment, but Router1 has become DR unexpectedly and Router1 become forwarder.
=> Need to configure DR-Priority or Link-Local Address

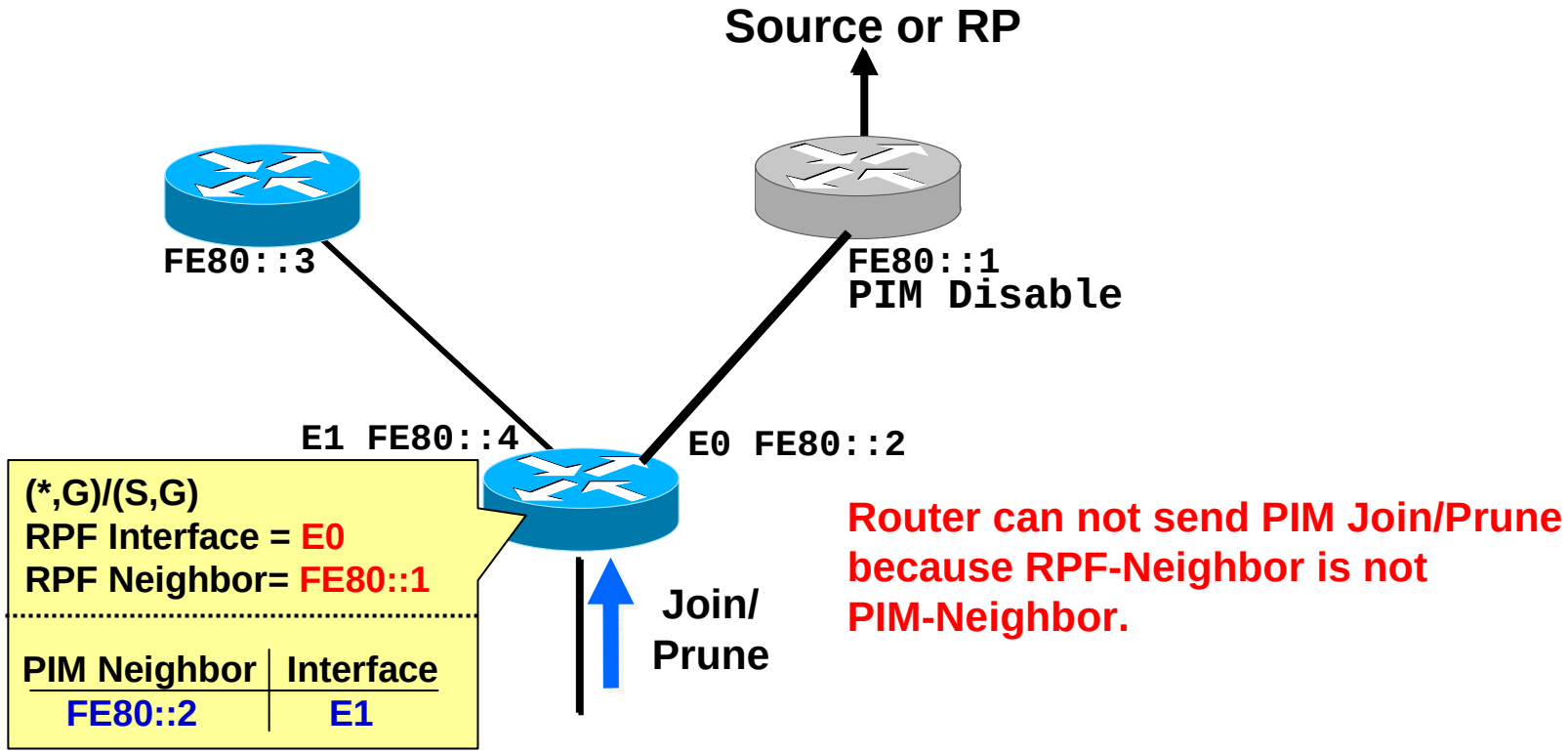
DR on Receiver Side



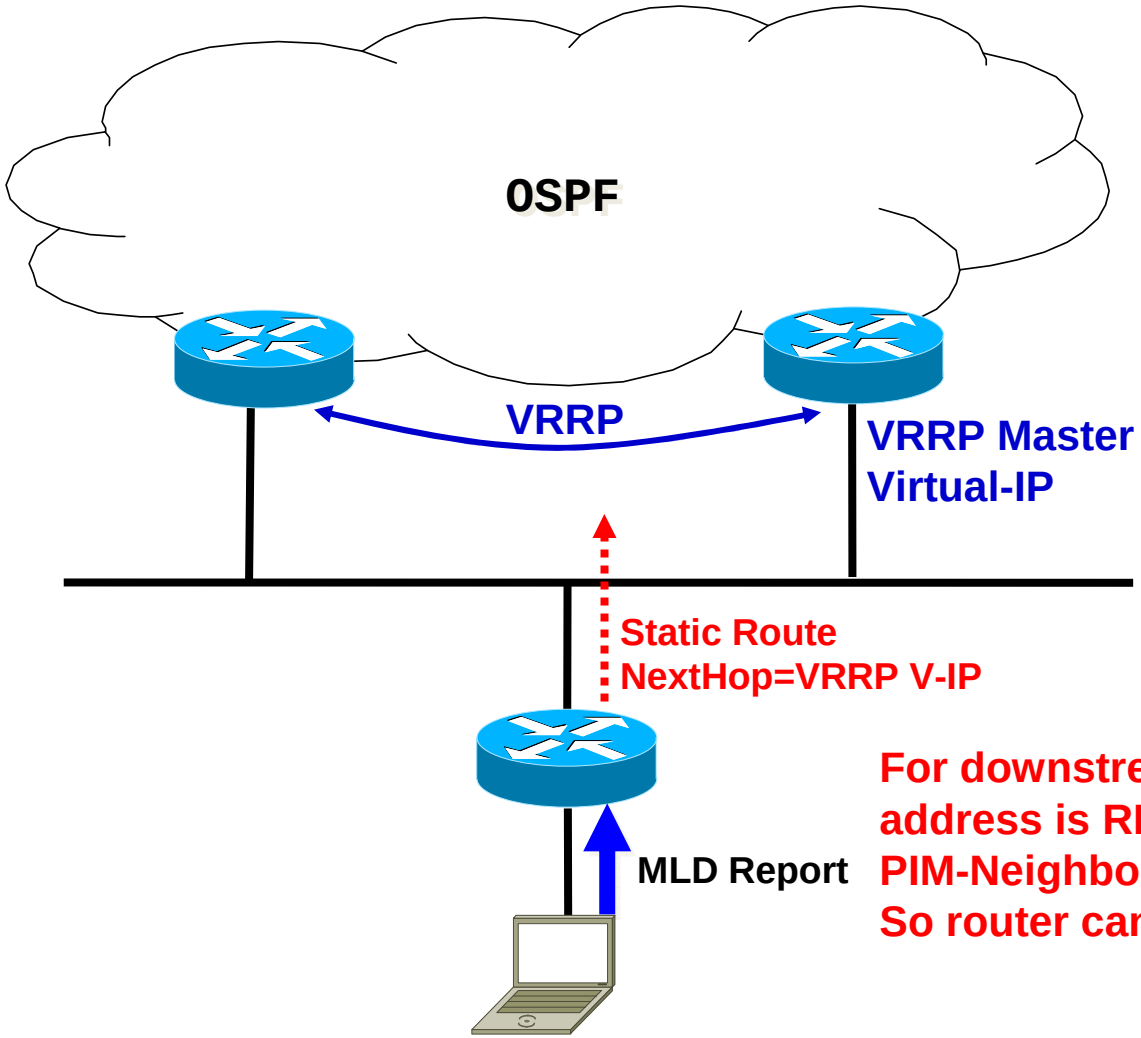
In above topology and uplink on Router1 become down, Router1 need to know alternative route via Router2. (generally LAN segment is configured as ” passive”) In some PIM implementation, above failover scenario does not work.

PIM-Neighbor

When the router find RPF Neighbor based on unicast routing info, Router can not send PIM-Join/Prune to the RPF neighbor if that RPF Neighbor is not recognized as PIM-Neighbor.
=> PIM should be enabled on all links in the network

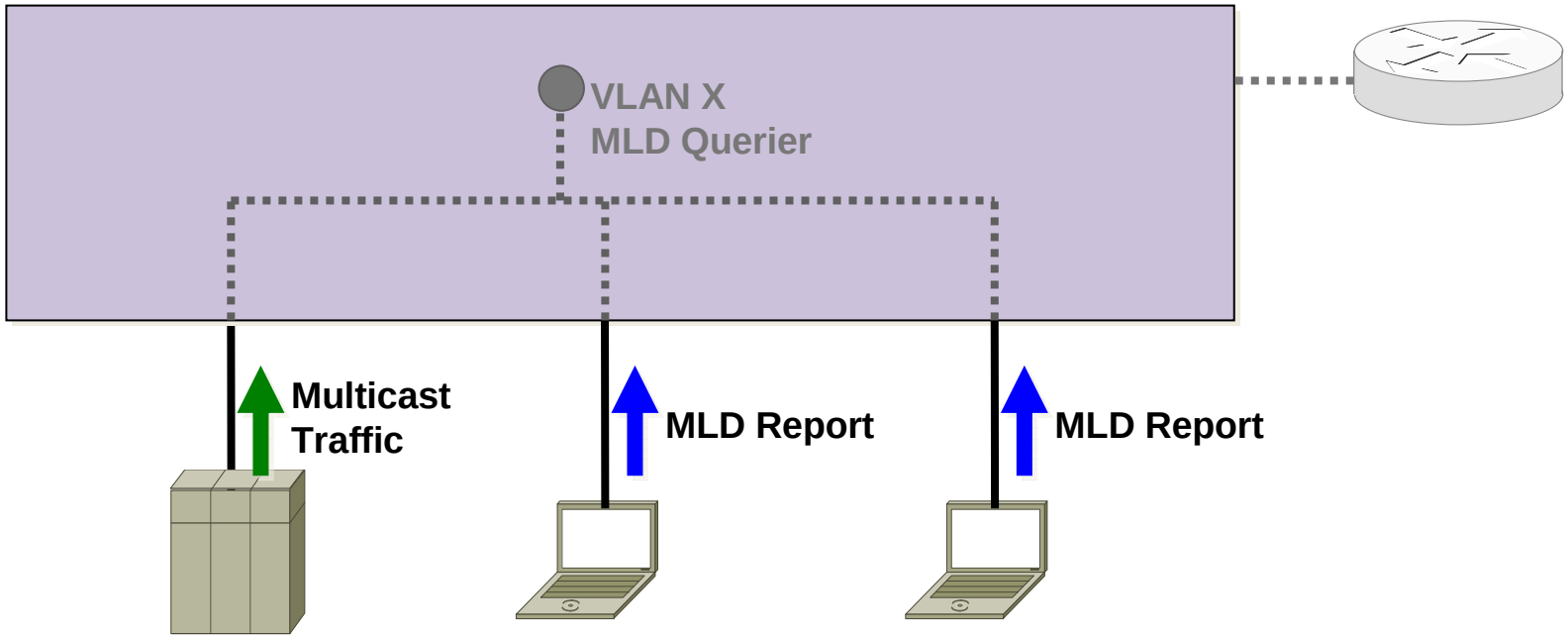


VRRP and PIM



For downstream router, VRRP virtual address is RPF Neighbor, but it is not PIM-Neighbor.
So router can not send any PIM messages.

L2 Switch only



There is no MLD querier in the vlan, MLD-snooping does not work without router port.
=> Need to connect MLD Router, or need to configure MLD querier

www.alcatel-lucent.com