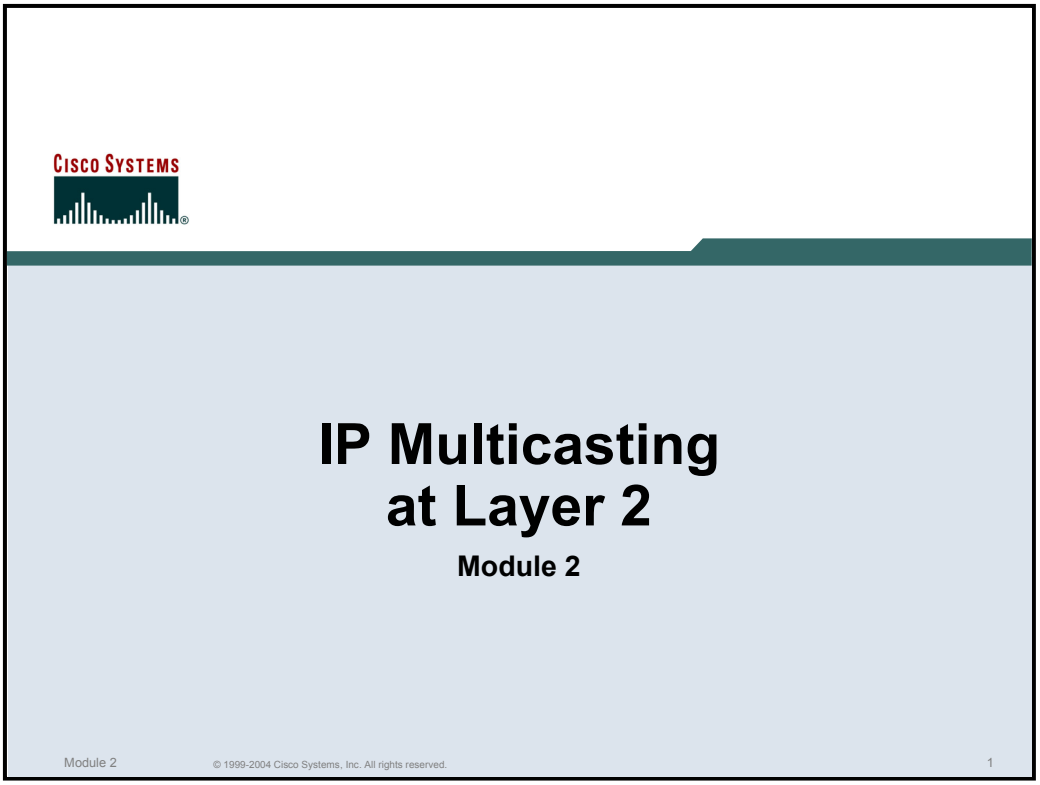

Cisco Systems
IP Multicasting
at Layer 2
Module 2
Module 2
© 1999-2004 Cisco Systems, Inc. All rights reserved.
1

Module Agenda

Cisco.com

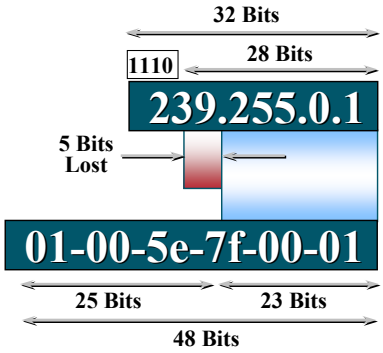
- MAC Layer Multicast Addresses
- IGMPv2
- IGMPv3
- L2 Multicast Frame Switching
  - IGMP Snooping
  - CGMP
  - PIM Snooping

MAC Layer Multicast Addresses

Layer 2 Multicast Addressing

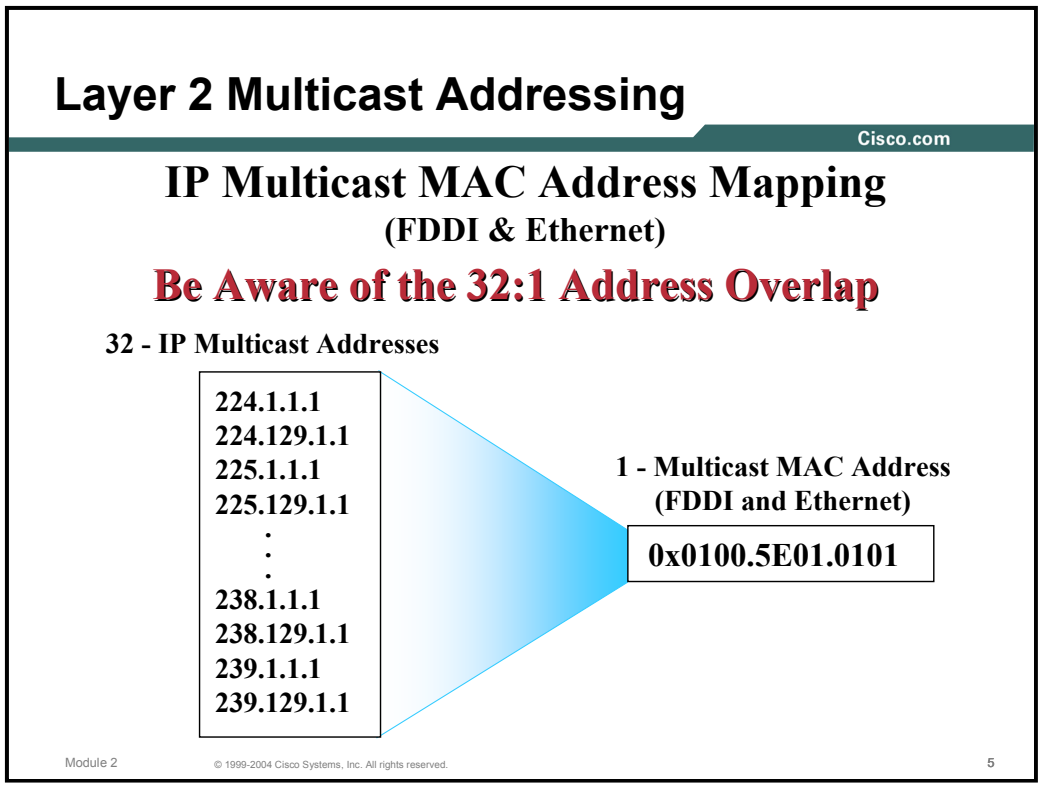
Cisco.com

IP Multicast MAC Address Mapping
(FDDI and Ethernet)

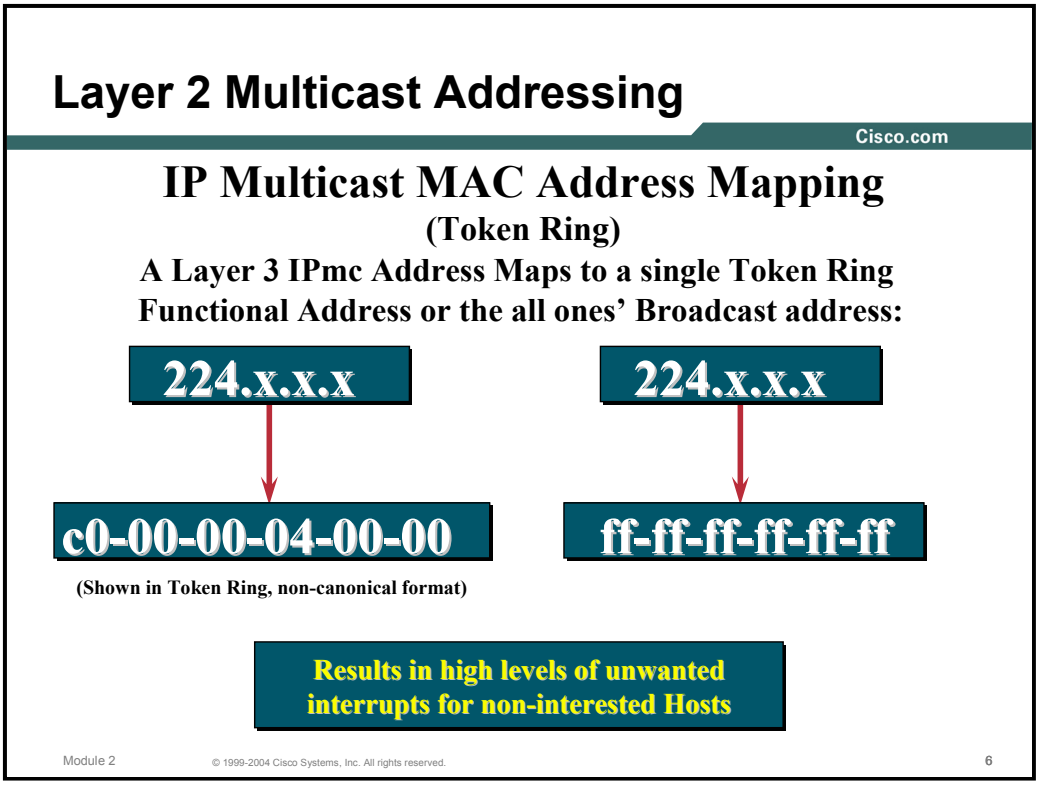


Module 2 4

- **Ethernet & FDDI Multicast Addresses**
  - The low order bit (0x01) in the first octet indicates that this packet is a Layer 2 multicast packet. Furthermore, the “0x01005e” prefix has been reserved for use in mapping L3 IP multicast addresses into L2 MAC addresses.
  - When mapping L3 to L2 addresses, the low order 23 bits of the L3 IP multicast address are mapped into the low order 23 bits of the IEEE MAC address. Notice that this results in 5 bits of information being lost.
- **A bit of History**
  - It turns out that this loss of 5 bits worth of information was not originally intended. When Dr. Steve Deering was doing his seminal research on IP Multicast, he approached his advisor with the need for 16 OUI’s to map all 28 bits worth of Layer 3 IP Multicast address into unique Layer 2 MAC addresses.
    - Note: An OUI (Organizationally Unique Identifier) is the high 24 bits of a MAC address that is assigned to “an organization” by the IEEE. A single OUI therefore provides 24 bits worth of unique MAC addresses to the organization.
  - Unfortunately, at that time the IEEE charged $1000 for each OUI assigned which meant that Dr. Deering was requesting that his advisor spend $16,000 so he could continue his research. Due to budget constraints, the advisor agreed to purchase a single OUI for Dr. Deering. However, the advisor also chose to reserve half of the MAC addresses in this OUI for other graduate research projects and granted Dr. Deering the other half.
  - This resulted in Dr. Deering having only 23 bits worth of MAC address space with which to map 28 bits of IP Multicast addresses. (It’s too bad that it wasn’t known back then how popular IP Multicast would become. If they had, Dr. Deering might have been able to “pass the hat” around to interested parties and collected enough money to purchase all 16 OUI’s. :-))

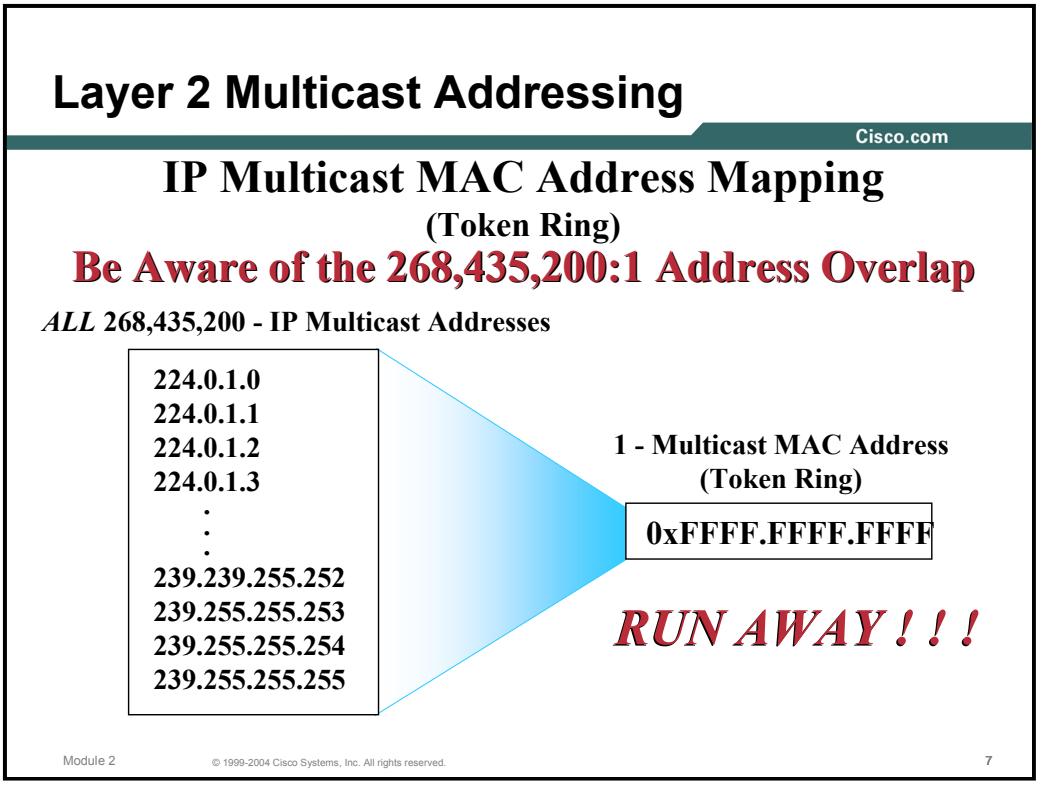


- **L2/L3 Multicast Address Overlap**
  - Since there are 28 bits of unique address space for an IP multicast address (32 minus the first 4 bits containing the 1110 Class D prefix) and there are only 23 bits plugged into the IEEE MAC address - there are 5 bits of overlap or 28-23 = 5. 2**5 = 32 therefore there is a 32:1 overlap of L3 addresses to L2 addresses - so beware several L3 addresses can map to the same L2 multicast address!
  - For example, all of the following IP multicast addresses map to the same L2 multicast of 01-00-5e-0a-00-01:
    - 224.10.0.1, 225.10.0.1, 226.10.0.1, 227.10.0.1
    - 228.10.0.1, 229.10.0.1, 230.10.0.1, 231.10.0.1
    - 232.10.0.1, 233.10.0.1, 234.10.0.1, 235.10.0.1
    - 236.10.0.1, 237.10.0.1, 238.10.0.1, 239.10.0.1
    - 224.138.0.1, 225.138.0.1, 226.138.0.1, 227.138.0.1
    - 228.138.0.1, 229.138.0.1, 230.138.0.1, 231.138.0.1
    - 232.138.0.1, 233.138.0.1, 234.138.0.1, 235.138.0.1
    - 236.138.0.1, 237.138.0.1, 238.138.0.1, 239.138.0.1



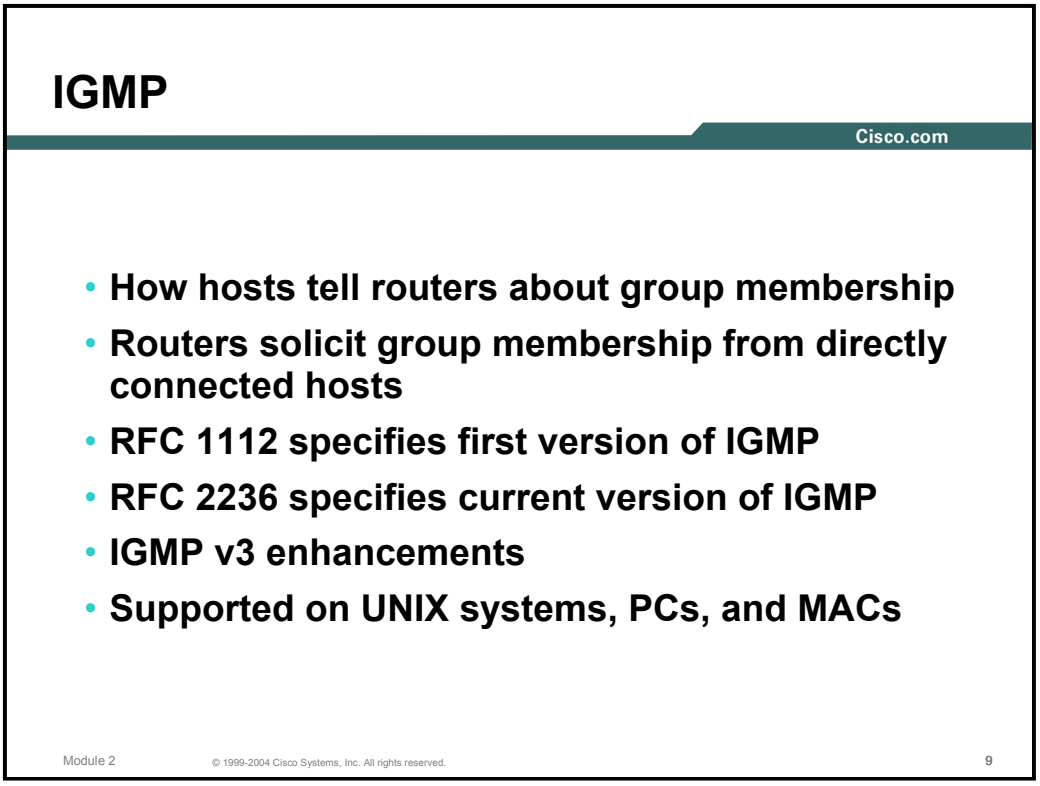
- **Token Ring MAC Addresses**
  - Because the bit order of bytes transmitted on Token-Ring are reversed, it is typical to see Token Ring MAC addresses written in their non-canonical form. For example, when transposed to canonical (Ethernet) form, the "0xc000.0004.0000" MAC address in the above slide would be "0x0300.0020.0000".
- **Token Ring Functional Addresses**
  - Token Ring Functional Addresses use a format of "0xc000.0004.xxxx" where the last 2 octets typically has at most, a single bit set. Many of the Functional Addresses are reserved for well-known Token-Ring MAC layer functions such as "Ring Error Monitor" and others. A bit in the 3rd Octet is used to signal that this is a Functional Address. In fact, the "0x5e" (canonical form) in the 3rd Octet of a normal Ethernet multicast address has a bit pattern that would confuse Token Ring end stations into thinking that the address was a Functional Address.

    Therefore, IP multicast address to L2 multicast address mapping cannot occur in Token Ring as it does in Ethernet.
- **Impact on Token-Ring End Stations**
  - Mapping all multicast addresses into a single L2 address forces the the main CPU in end systems to perform filtering of wanted vs. unwanted multicast packets instead of being handled in hardware by the Token Ring NIC card. This creates significant performance issues on Token-Ring end systems when multicasting traffic is present on the ring.
  - This is a very good reason, among many others, for users considering the Ethernet versus Token Ring debate to strongly consider Ethernet if MultiMedia Applications and IPmc is being deployed or planned.

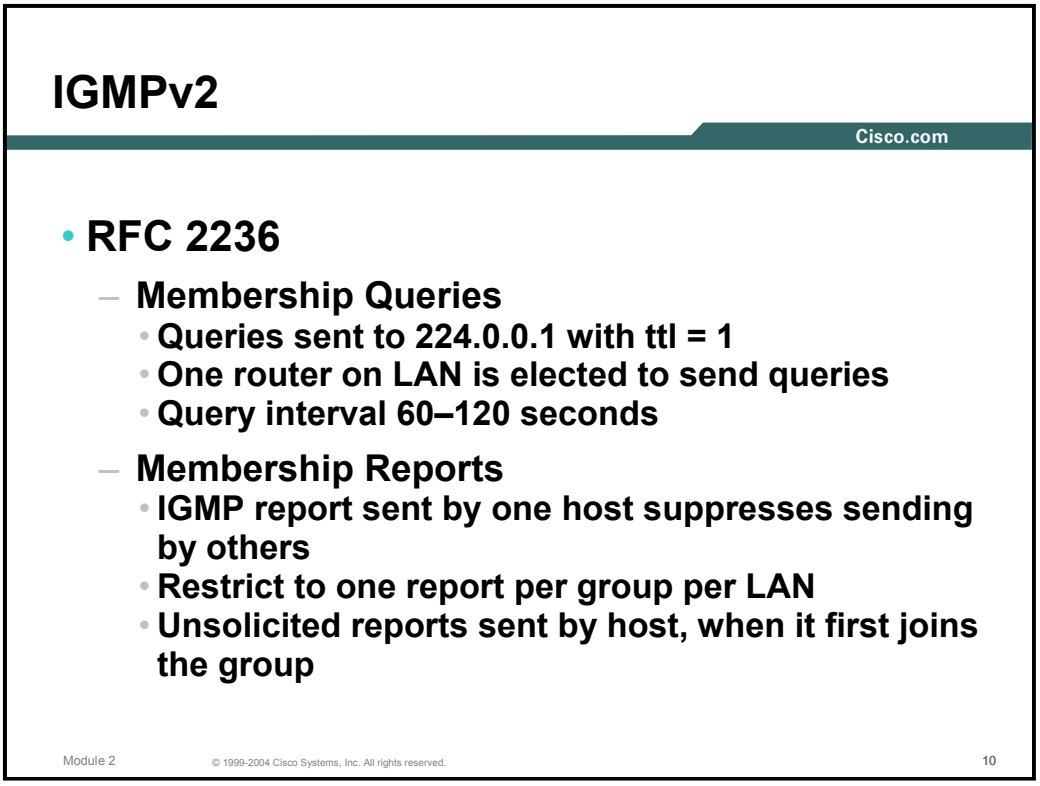


- **L2/L3 Multicast Address Overlap**
  - Unfortunately, all 28 significant bits of an IP multicast address (32 minus the first 4 bits) map into a single Token Ring MAC address. This has the disasterous result of a 2**28 = 268,435,200 ambiguity!
  - Because al L3 addresses map into the same L2 multicast address, constraint of multicast traffic at L2 is impossible on Token Ring networks.
  - A migration from Token-Ring to Ethernet should be considered by network administrators contemplating any extensive use of IP multicast.

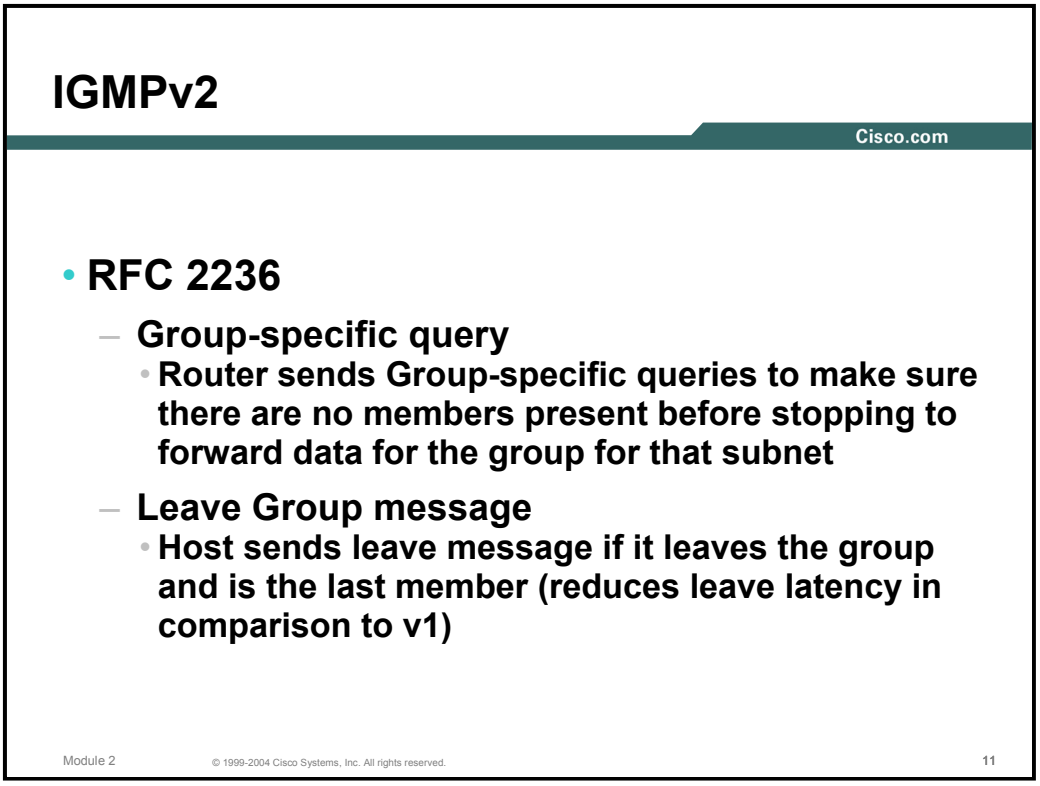
IGMPv2



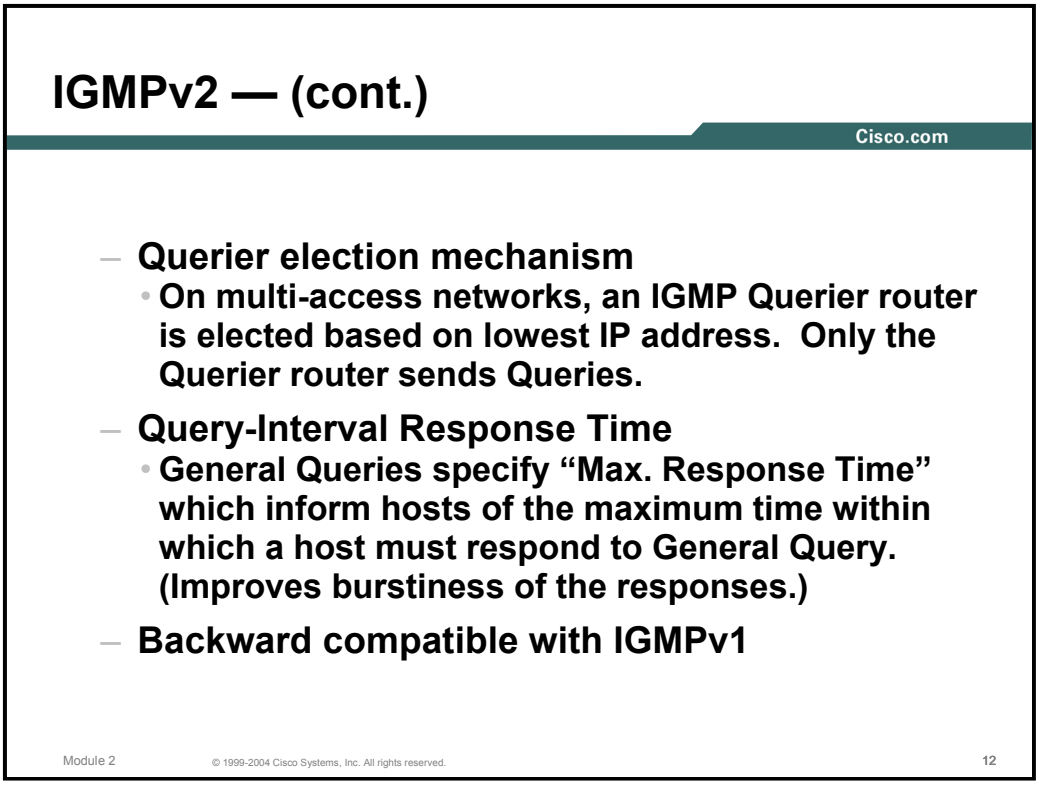
- **IGMP**
  - The primary purpose of IGMP is to permit hosts to communicate their desire to receive multicast traffic to the IP Multicast router(s) on the local network. This, in turn, permits the IP Multicast router(s) to “Join” the specified multicast group and to begin forwarding the multicast traffic onto the network segment.
  - The initial specification for IGMP (v1) was documented in RFC 1112, “Host Extensions for IP Multicasting”. Since that time, many problems and limitations with IGMPv1 have been discovered. This has lead to the development of the IGMPv2 specification which was ratified in November, 1997 as RFC 2236.
  - Even before IGMPv2 had been ratified, work on the next generation of the IGMP protocol, IGMPv3, had already begun. However, the IGMPv3 specification is still in the working stage and has only been implemented by a few vendors.



- **IGMP Membership Queries**
  - IGMPv1 Membership Queries are sent by the router to the “All-Hosts” (224.0.0.1) multicast address to solicit what multicast groups have active receivers on the local network.
- **IGMP Membership Reports**
  - IGMPv1 Membership Reports are sent by hosts wishing to receive traffic for a specific multicast group. Membership Reports are sent (with a TTL of 1) to the multicast address of the group for which the hosts wishes to receive traffic. Hosts either send reports asynchronously (when the wish to first join a group) or in response to Membership Queries. In the latter case, the response is used to maintain the group in an active state so that traffic for the group continues to be forwarded to the network segment.
- **Report Suppression**
  - Report suppression is used among group members so that all members do not have to respond to a query. This saves CPU and bandwidth on all systems.
  - The rule in multicast membership is that as long as one member is present, the group must be forwarded onto that segment. Therefore, only one member present is required to keep interest in a given group so report suppression is efficient.
- **TTL**
  - Since Membership Query and Report packets only have local significance, the TTL of these packets are always set to 1. This is also so they will not be accidentally forwarded off of the local subnet and cause confusion on other subnets.



- **IGMPv2**
  - As a result of some of the limitations discovered in IGMPv1, work was begun on IGMPv2 in an attempt to remove these limitations. Most of the changes between IGMPv1 and IGMPv2 are primarily to address the issues of Leave and Join latencies as well as address ambiguities in the original protocol specification. (IGMPv2 is almost to standard status.)
  - The following sections define some of the more significant changes.
- **Group Specific Queries**
  - A Group Specific query was added in v2 to allow the router to only query membership in a single group instead of all groups. This is an optimized way to quickly find out if any members are left in a group without asking all groups for a report.
  - The difference between the Group Specific query and the General Query is that a General Query is multicast to the “All-Hosts” (224.0.0.1) address while a Group Specific query for Group “G”, is multicast to the Group “G” multicast address.
- **Leave Group message**
  - A Leave Group message was also added in IGMPv2. This allows end systems to tell the router they are leaving the group which reduces the leave latency for the group on the segment when the member leaving is the last member of the group.
  - The standard is loosely written on when leave group messages should and must be sent. This is an important consideration when discussing CGMP.

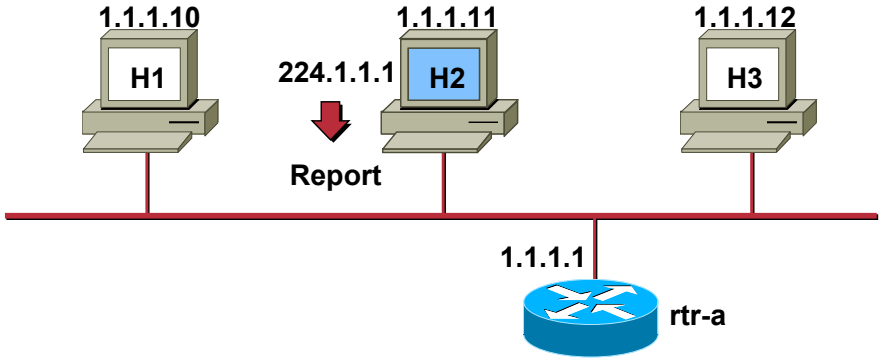


- **Querier Election**
  - IGMP itself now has a Querier election mechanism unlike v1. The lowest unicast IP address of the IGMP-speaking routers will be elected as the Querier. All IGMP speaker come up thinking they will be the querier but must immediately relinquish that role if a lower IP address query is heard on the same segment.

- **Query-Interval Response Time**
  - The Query-Interval Response time has also been added to control the burstiness of reports. This value is indicated in queries to convey to the membership how much time they have to respond to a query with a report.

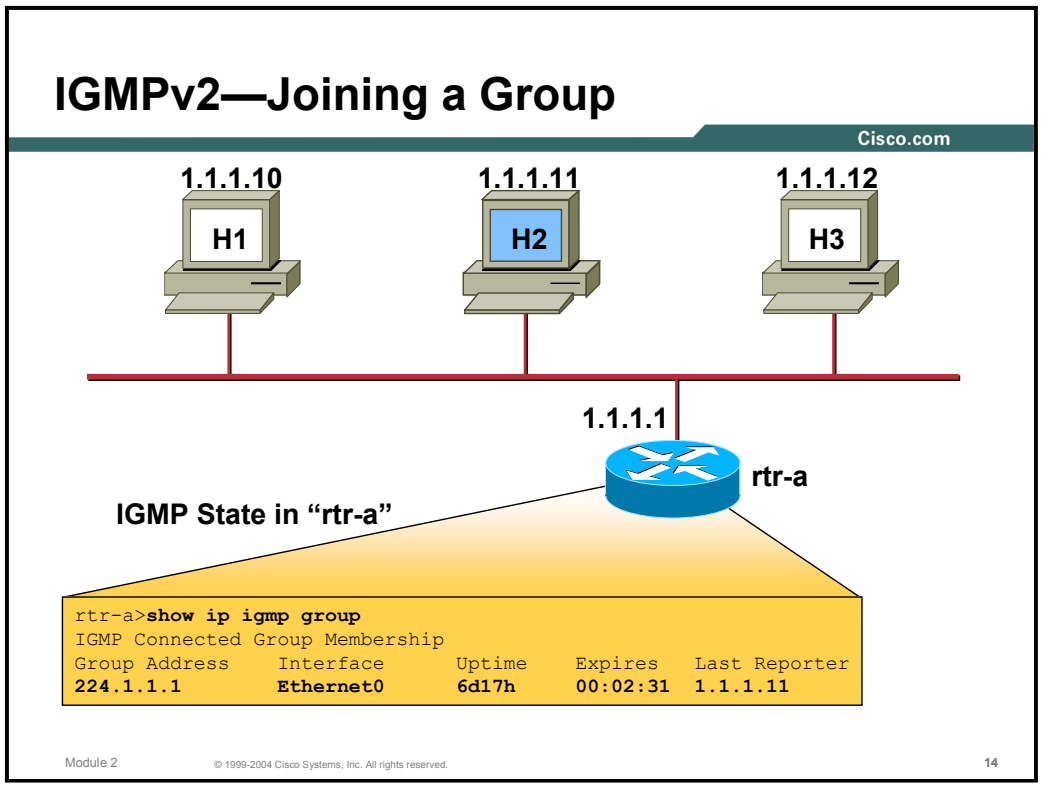
IGMPv2—Joining a Group

Cisco.com

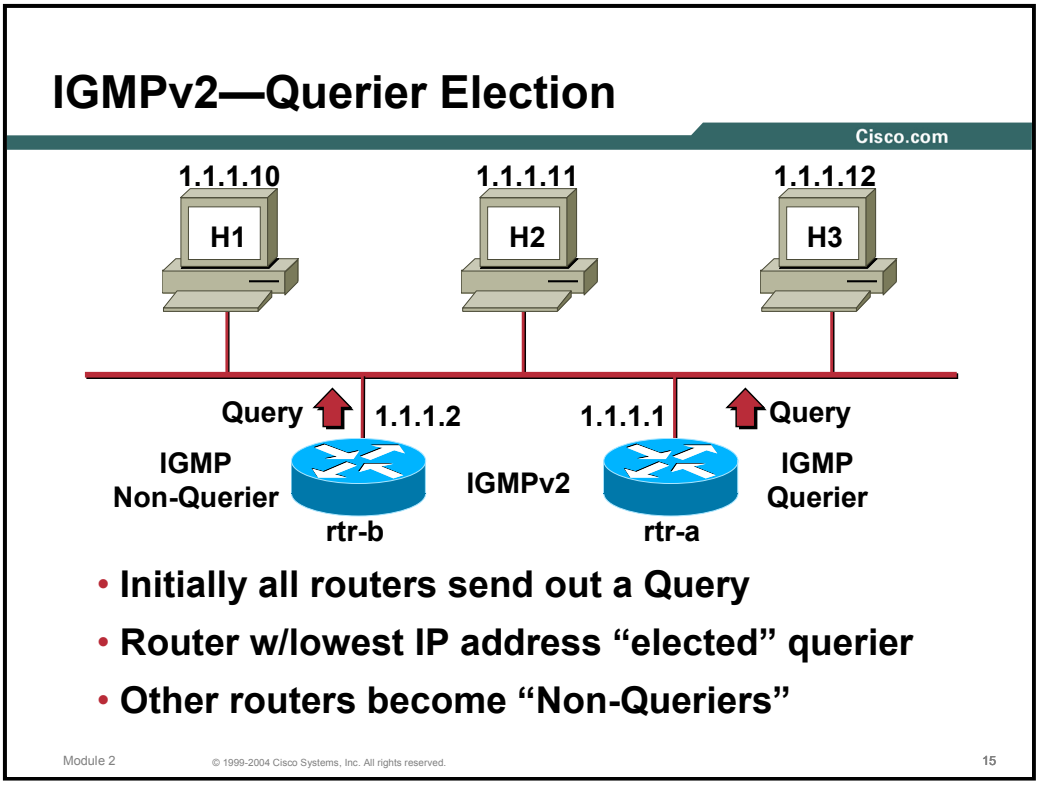


- **Joining member sends report to 224.1.1.1 immediately upon joining (same as IGMPv1)**

- **Asynchronous Joins**
  - Members joining a group do not have to waited for a query to join; they send in an unsolicited report indicating their interest. This reduces join latency for the end system joining if no other members are present.



- **IGMP State in “rtr-a”**
  - Group 224.1.1.1 is active on Ethernet 0 and
    - Has been active on this interface for 6 days and 17 hours.
    - It expires (and will be deleted) in 2 minutes and 31 seconds if an IGMP Host Membership report for this group is not heard in that time.
    - The last Host to report membership was 1.1.1.11 (H2).



- **Querier Election**
  - In IGMPv1 there was no formal IGMP querying router election process in within IGMPv1 itself - it was left up to the multicast routing protocol and different protocols used different mechanisms. This would often result in multiple queriers on a single multiaccess network. With the definition of IGMPv2 a formal querying router election process was specified within the IGMPv2 protocol itself.
  - In IGMPv2 each router on a multiaccess network will initially assume it is the querier and begin sending queries. Each router will see the queries from the other IGMPv2 routers and will examine the IP address of these queries. All IGMPv2 routers will then defer to the router with the lowest IP address.
  - In other words, the IGMPv2 router with the lowest IP address will become the querying router.
  - Finally, if the currently elected Query Router fails to issue a query within a specified time limit, a timer in the other IGMPv2 routers will "time-out" and cause them to re-initiate the Query Election process.
- **Group Specific Queries**
  - IGMPv2 also added the concept of Group Specific Queries. This is accomplished by sending the IGMPv2 Membership Query to the Group's multicast address as opposed to sending to the "All Hosts" (224.0.0.1) multicast address as is done for IGMPv2 General Queries.
- **Query Interval**
  - Membership queries are sent every 60 seconds (default).

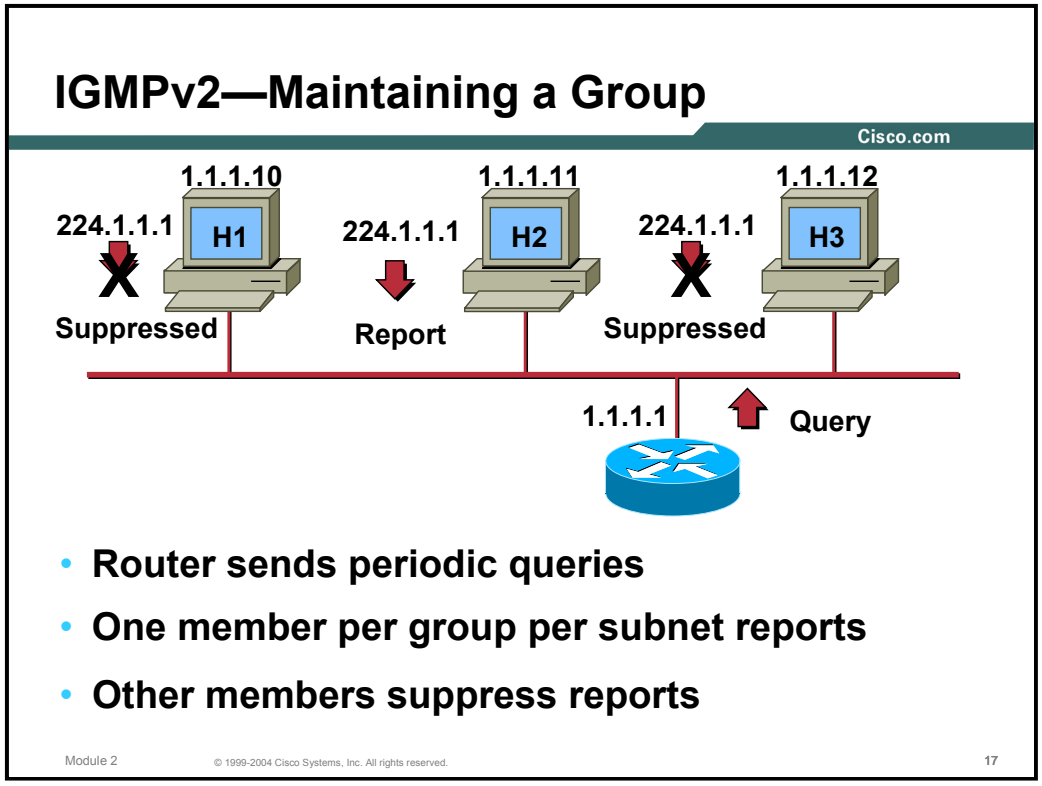
IGMPv2—Querier Election

Cisco.com

Determining which router is the IGMP Querier

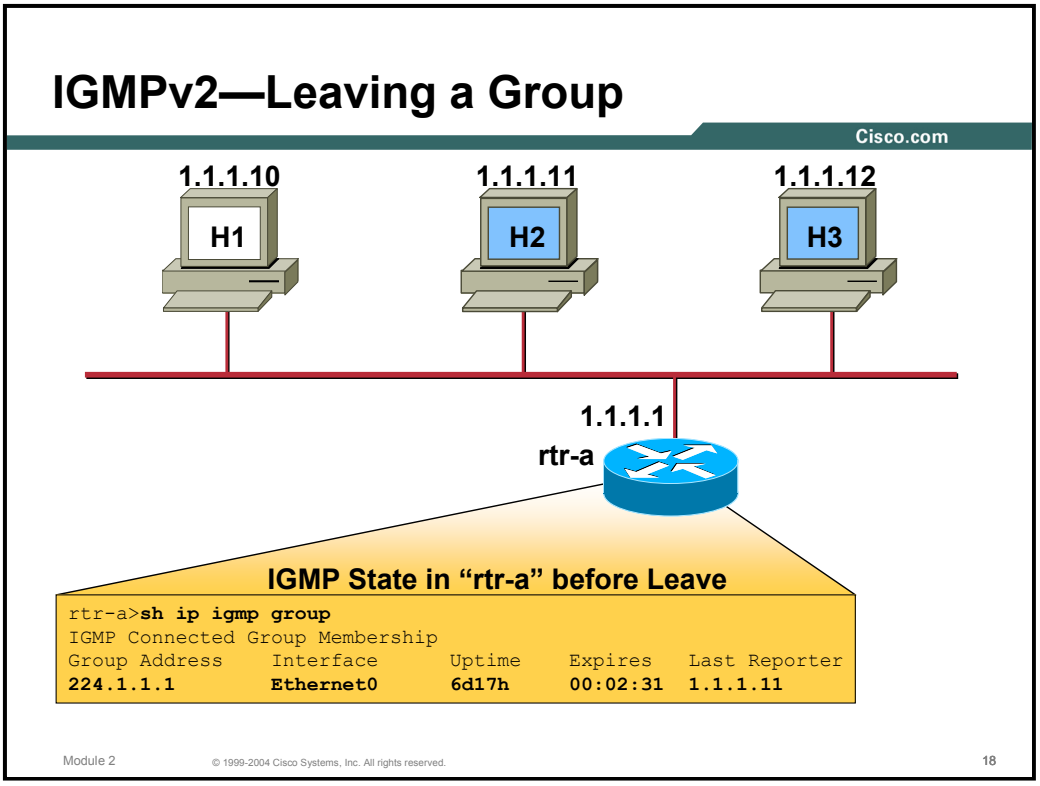
```
rtr-a>show ip igmp interface e0
Ethernet0 is up, line protocol is up
  Internet address is 1.1.1.1, subnet mask is 255.255.255.0
  IGMP is enabled on interface
  Current IGMP version is 2
  CGMP is disabled on interface
  IGMP query interval is 60 seconds
  IGMP querier timeout is 120 seconds
  IGMP max query response time is 10 seconds
  Inbound IGMP access group is not set
  Multicast routing is enabled on interface
  Multicast TTL threshold is 0
  Multicast designated router (DR) is 1.1.1.1 (this system)
  IGMP querying router is 1.1.1.1 (this system)
  Multicast groups joined: 224.0.1.40 224.2.127.254
```


- **Verifying the IGMPv2 Querier**
  - Use the “show ip igmp interface” command to determine which router is the IGMPv2 Querier on the multiaccess network.
  - Note that the “Designated Router” is a different function and is listed separately in the display above.

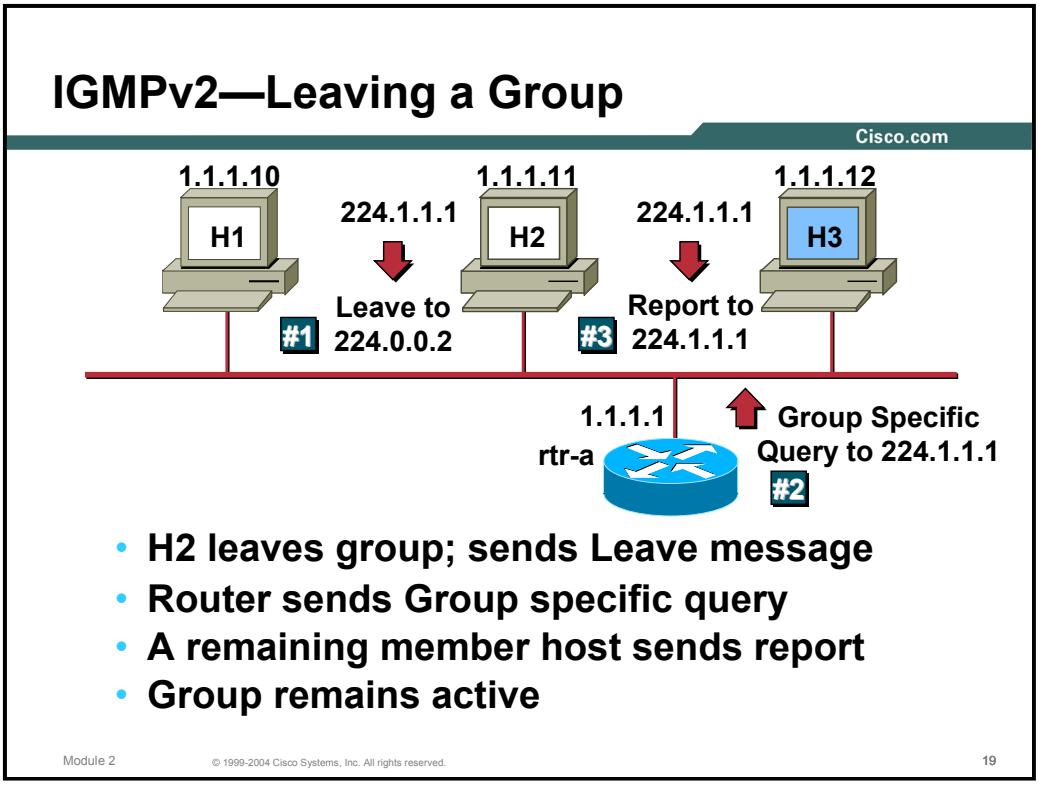


- **Query-Response Process**
  - The router multicasts periodic IGMPv1 Membership Queries to the “All-Hosts” (224.0.0.1) group address.
  - Only one member per group responds with a report to a query. This is to save bandwidth on the subnet network and processing by the hosts. This is process is called “Response Suppression”. (See section below.)

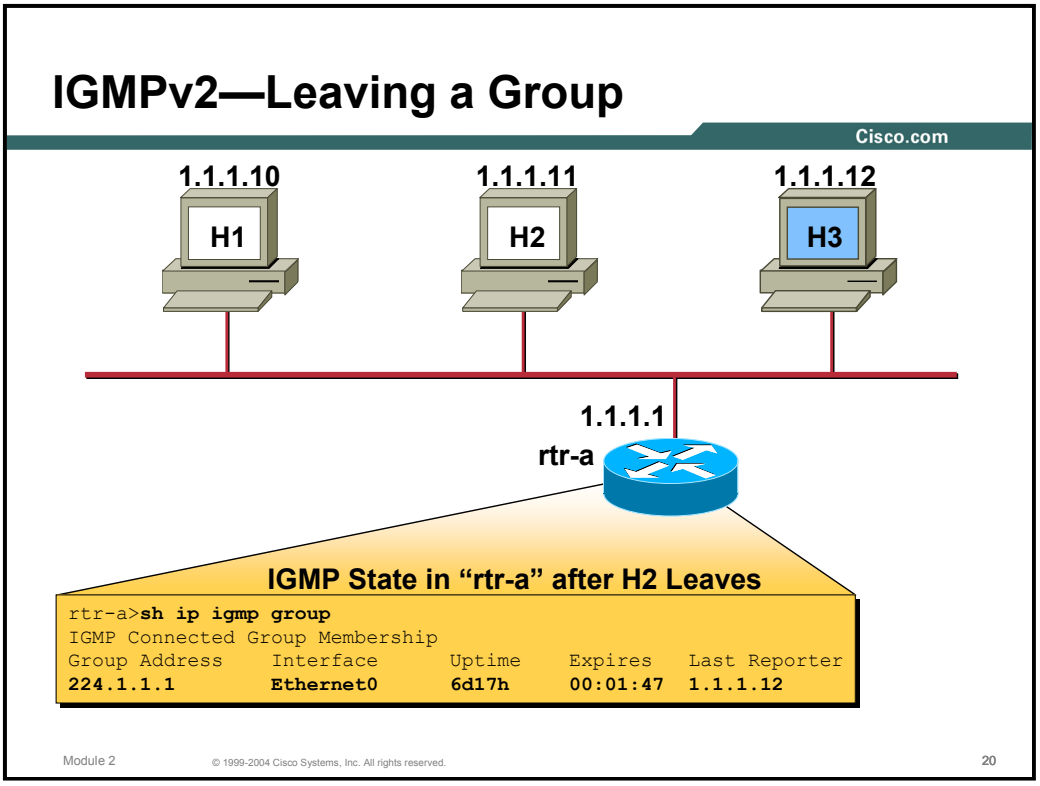
- **Response Suppression Mechanism**
  - The “Report Suppression” mechanism is accomplished as follows:
    - When a host receives the Query, it starts a count-down timer for each multicast group of which it is a member. The count-down timers are each initialized to a random count within a given time range. (In IGMPv1 this was a fixed range of 10 seconds. Therefore the count-down timers were randomly set to some value between 0 and 10 seconds.)
    - When a count-down timer reaches zero, the host sends a Membership Report for the group associated with the count-down timer to notify the router that the group is still active.
    - However, if a host receives a Membership Report before its associated count-down timer reaches zero, it cancels the count-down timer associated with the multicast group, thereby suppressing its own report.
    - In the example shown in the slide, H2’s time expired first so it responded with its Membership Report. H1 and H3 cancelled their timers associated with the group; thereby suppressing their reports.



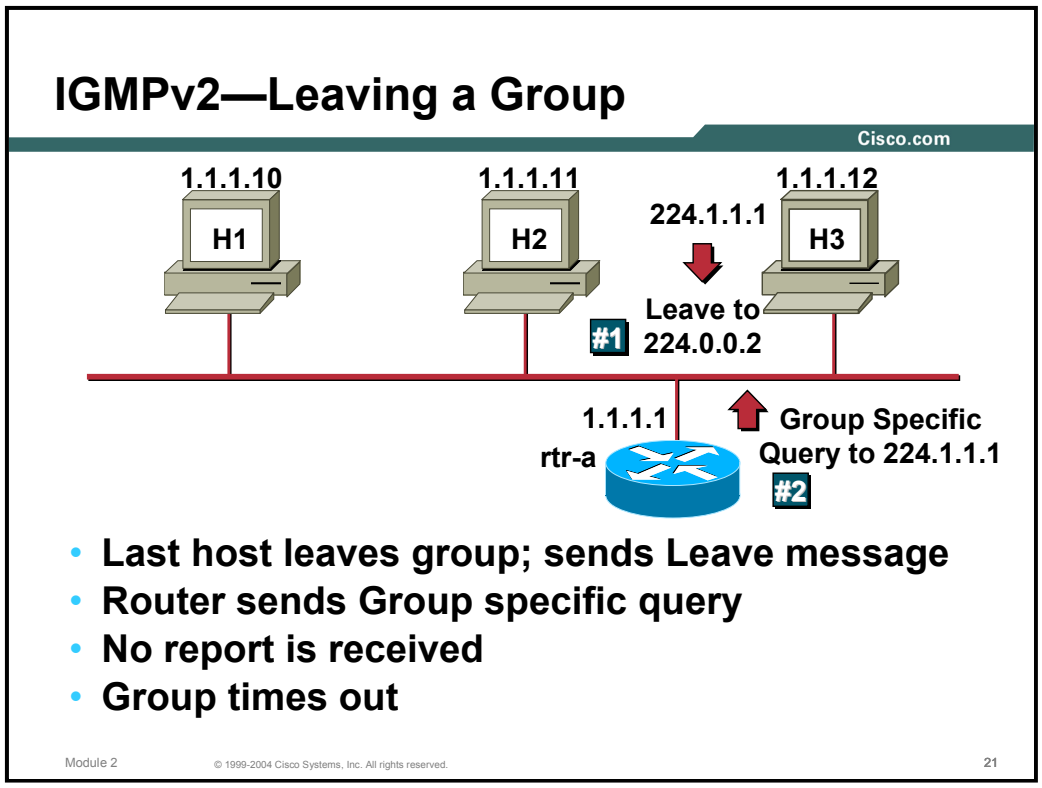
- **IGMPv2 Leaves**
  - In the above example, notice that the router is aware that there one or more members of group 224.1.1.1 active on Ethernet0 and that Host 2 responded with a Group Membership Report for this group during the last General Query interval. (Indicated by the IP address of Host 2 in the Last Reporter field.)



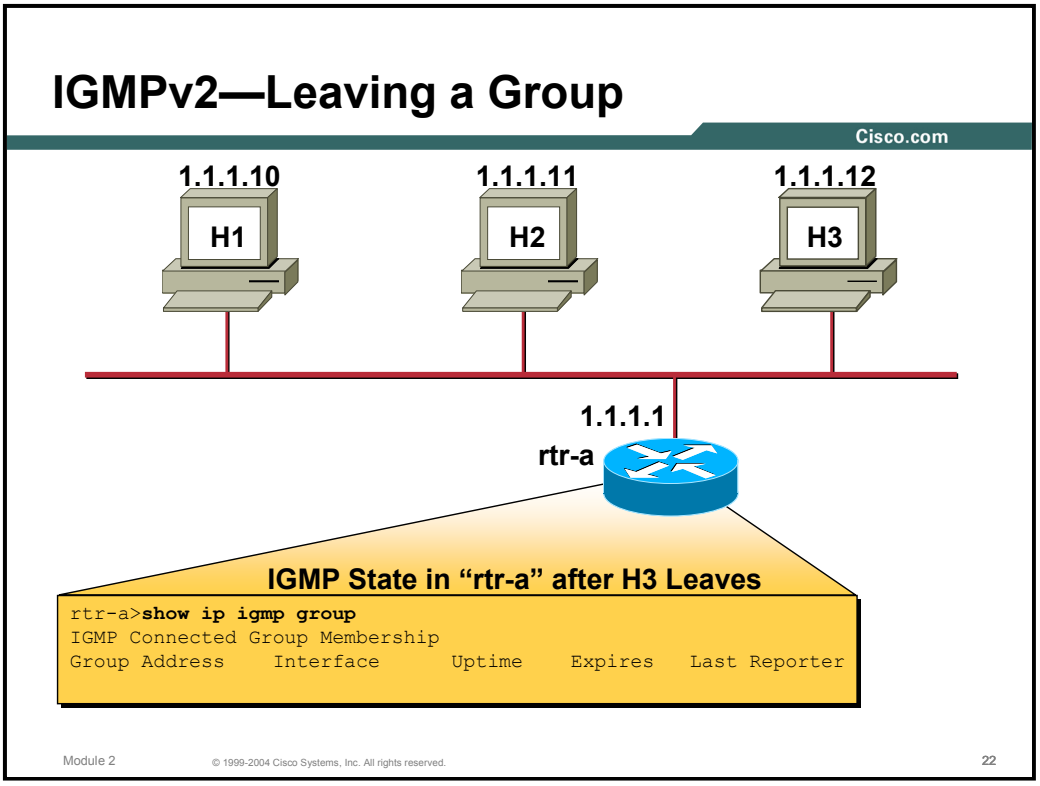
- **IGMPv2 Leaves**
  - In IGMPv1, hosts would leave passively - i.e.. they do not explicitly say they are leaving - they just stop reporting. However, IGMPv2 has explicit “Leave Group” messages.
  - When the IGMPv2 Query router receives a “Leave Message”, it responds by sending a “Group Specific Query” for the associated group to see if there are still other hosts wishing to receive traffic for the group. This process helps to reduce overall “Leave Latency”.
  - When CGMP is in use, the IGMPv2 “Leave Message” mechanism also helps the router to better manage the CGMP state in the switch. This also improves the leave latency for the specific host at layer 2.
  - (Note: Due to the wording of the current IGMPv2 draft specification, hosts may chose to NOT send Leave messages if they are not the last host to leave the group. This can adversely affect CGMP performance.)
- **Example :**
  - H2 and H3 are members of group 224.1.1.1
  - #1 - H2 leaves
  - #2 - Router sends group specific query to see if any other group members are present.
  - #3 - H3 hasn’t left yet so it responds with a Report message.
  - Router keeps sending multicast for 224.1.1.1 since there is >= 1 member present



- **IGMPv2 Leaves**
  - At this point, the group is still active. However, the router shows that Host 3 is the last host to send an IGMP Group Membership Report.

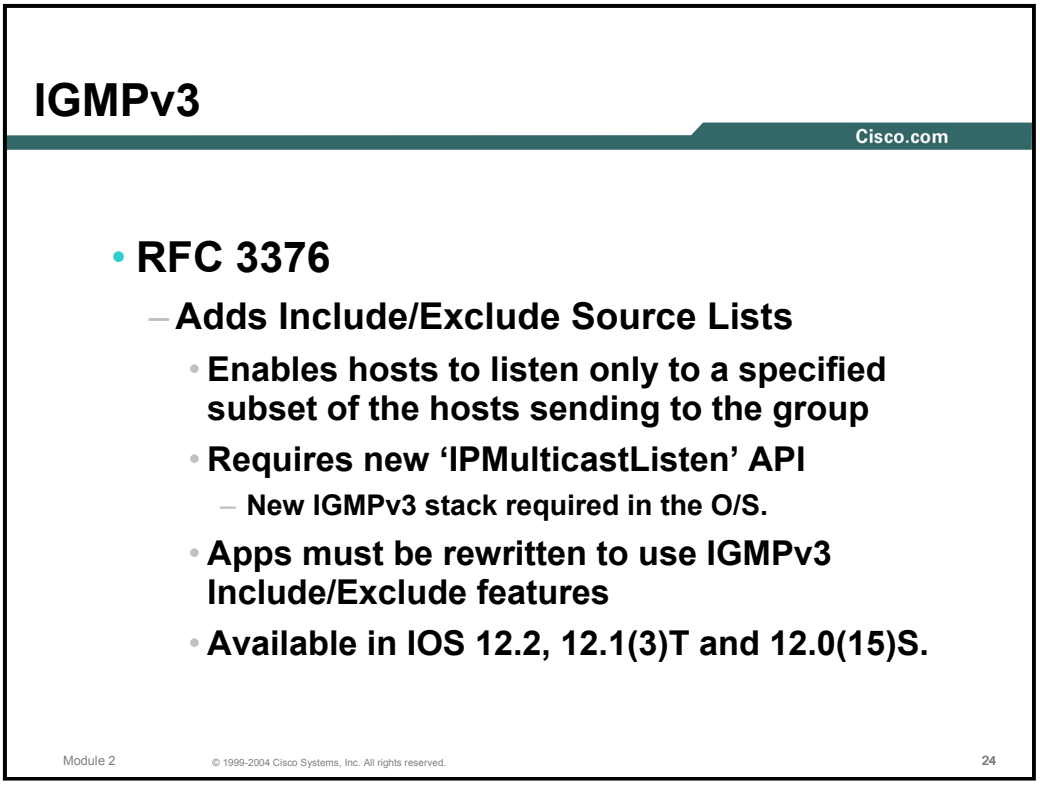


- **IGMPv2 Leaves**
- **Example (continued):**
  - H3 is the only remaining member of group 224.1.1.1
  - #1 - H3 leaves
  - #2 - Router sends group specific query to see if any other group members are present.
  - H3 was the last remaining member of the group so no IGMP Membership Report for group 224.1.1.1 is received and the group times out. (This typically takes from 1-3 seconds from the time that the Leave message is sent until the Group Specific Query times out and traffic stops flowing.)

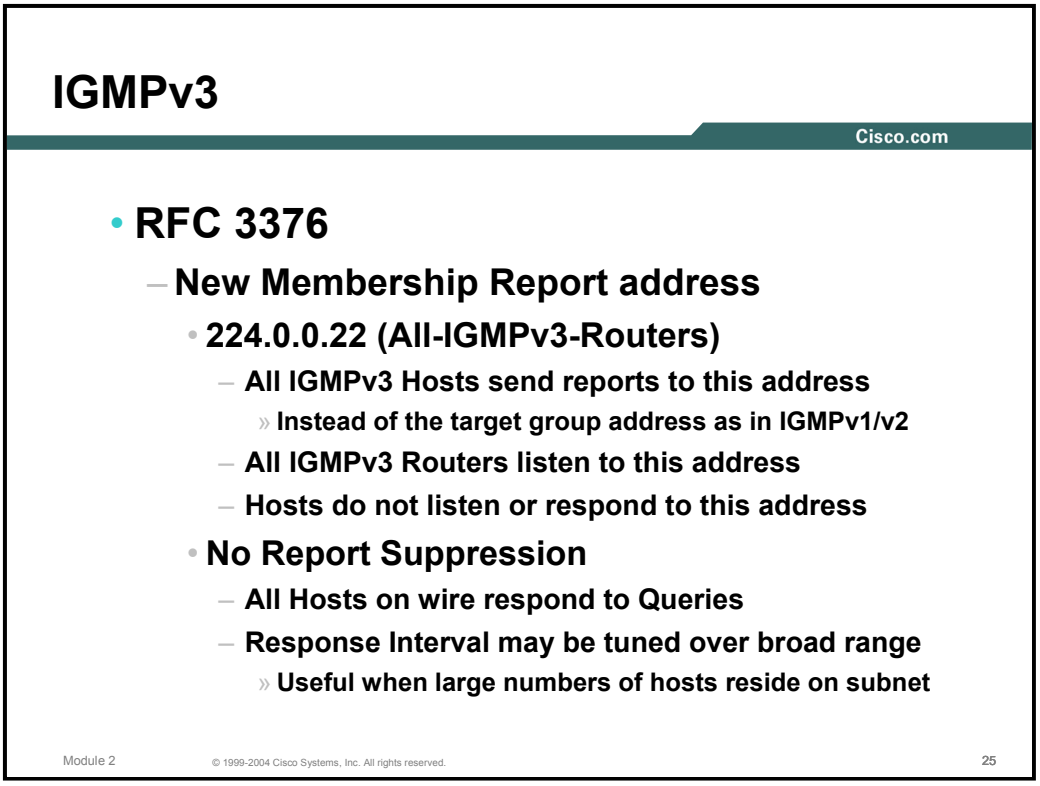


- **IGMPv2 Leaves**
  - At this point, all hosts have left the 224.1.1.1 group on Ethernet0. This is indicated by “rtr-a” above.in the output of the show ip igmp group command.

IGMPv3



- **IGMPv3**
  - The IDMR is completing work on IGMPv3.
  - The key change in IGMPv3 is the addition of Group records each containing a list of sources to Include or Exclude. This permits a host to signal which set of hosts that they wish to receive group traffic.
  - IGMPv3 requires that the 'IPMulticastListen' API be changed to accommodate the Include/Exclude filter list. This means that the IGMP stack in the OS will have to be updated to support IGMPv3.
  - In order to take advantage of the benefits of IGMPv3, applications must be (re)written to support the new API.



- **IGMPv3**
  - IGMPv3 is assigned its own “All IGMPv3 Routers” link-local multicast group address, 224.0.0.22.
    - IGMPv3 hosts no longer send their reports to the target multicast group address. Instead, they send their IGMPv3 Membership Reports to the “All IGMPv3 Routers” multicast address.
    - Routers listen to the 224.0.0.22 address in order to receive and maintain IGMP membership state for every member on the subnet! This is a radical change over the behavior in IGMPv1/v2 where the routers only maintained group state on a subnet basis.
    - Hosts do not listen to 224.0.0.22 and therefore do not hear other hosts’ IGMPv3 membership reports.
  - IGMPv3 drops the Report Suppression mechanism that was used in IGMPv1/v2.
    - All IGMPv3 hosts on the wire respond to Queries by sending and IGMPv3 membership reports containing their total IGMP state for all groups in the report.
    - In order to prevent huge bursts of IGMPv3 Reports, the Response Interval may now be tuned over a much greater range than before. This permits the network engineer to adjust the burstiness of IGMPv3 Reports when there is a large number of hosts on the subnet.

IGMPv3 — Query Packet Format

Cisco.com

Type = 0x11
IGMP Query

Max. Resp. Time
Max. time to send a response
if < 128, Time in 1/10 secs
if > 128, FP value (12.8 - 3174.4 secs)

Group Address:
Multicast Group Address
(0.0.0.0 for General Queries)

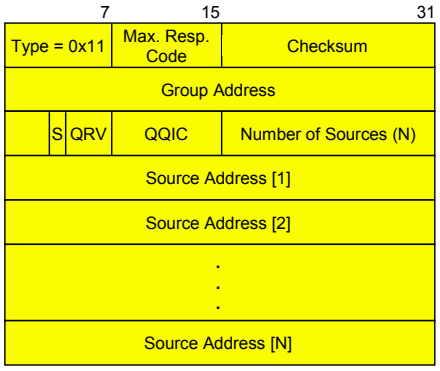
S Flag
Suppresses processing by routers

QRV (Querier Robustness Value)
Affects timers and # of retries

QQIC (Querier's Query Interval)
Same format as Max. Resp. Time

Number of Sources (N)
(Non-zero for Group-and-Source Query)

Source Address
Address of Source



Module 2 26

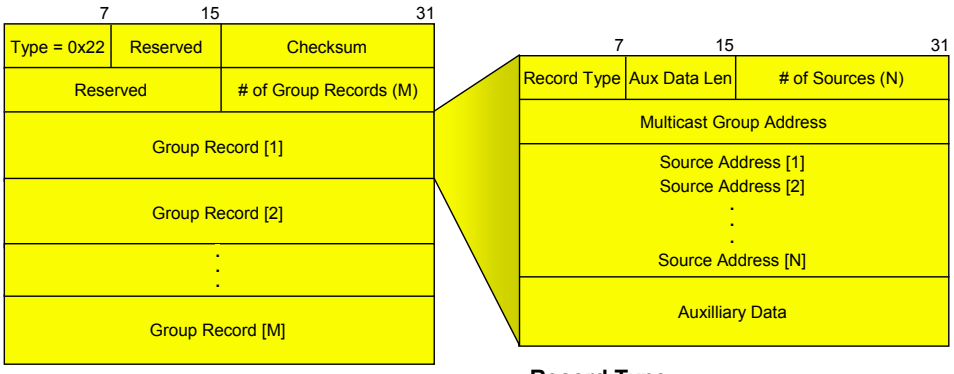
- **Type**
  - The same IGMPv2 type code 0x11 is used as the IGMPv3 Membership Query Type code.
- **Max. Response Time (1/10 seconds)**
  - This field has been reformatted to permit longer times to be expressed. If the value is < 128, the the time is absolute (.1- 12.8 seconds). If the value is > 128, it is interpreted as a floating-point number as follows:

```
+-+-+-+-+-+-+-+-+
|1| exp | mant  | value = (mant|0x10)<<(exp+3)
+-+-+-+-+-+-+-+-+
```

- **Group Address**
  - This field is identical to the IGMPv2 version of this field. It is set to 0.0.0.0 for General Queries.
- **S Flag**
  - Indicates that the routers that receive this message should not process it.
- **QRV (Querier Robustness Value)**
  - This value causes all hosts to adjust their Robustness Values which in turn affect various timers and retry counts. Increasing this value provides more protocol robustness at the expense of latency.
- **QQIC (Querier' Query Interval)**
  - This field indicates the Query Interval in use by the Querying router. It's format is the same as the Max. Response Time field.
- **Number of Sources**
  - The number of Source Addresses in the Group-and-Source-Specific Query.

IGMPv3 — Report Packet Format

Cisco.com



# of Group Records (M)
Number of Group Records in Report

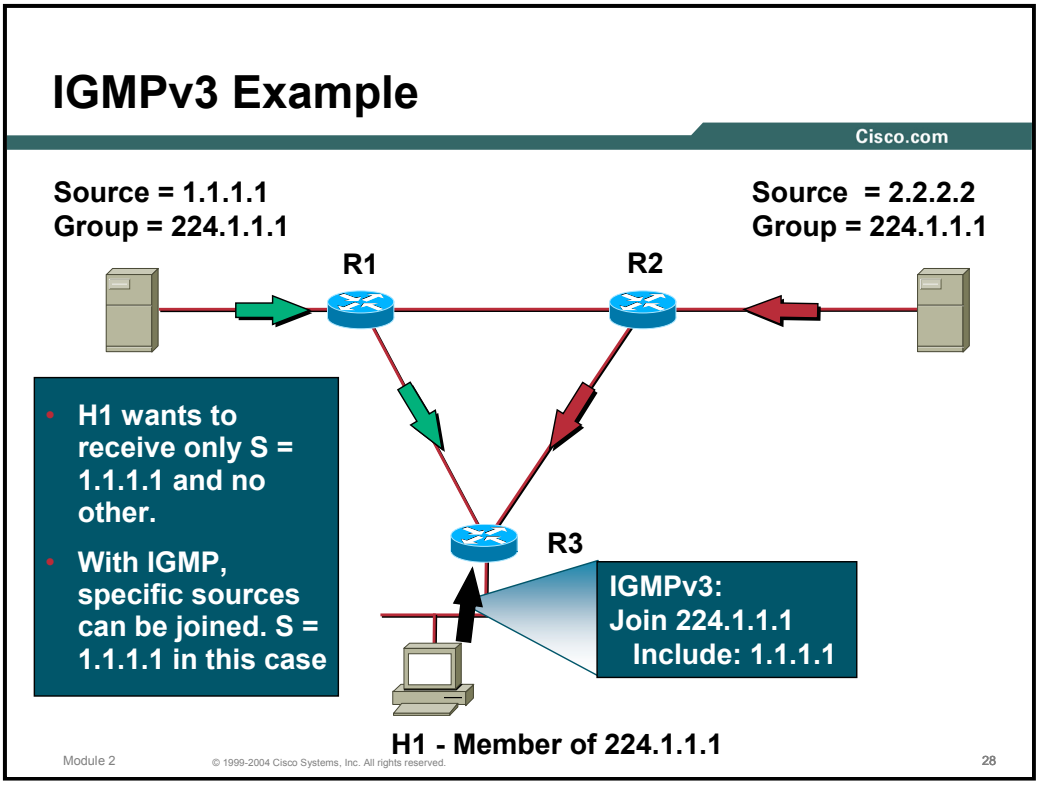
Group Records 1 - M
Group address plus list of zero or more sources to Include/Exclude (See Group Record format)

Record Type
Include, Exclude, Chg-to-Include, Chg-to-Exclude, Add, Remove

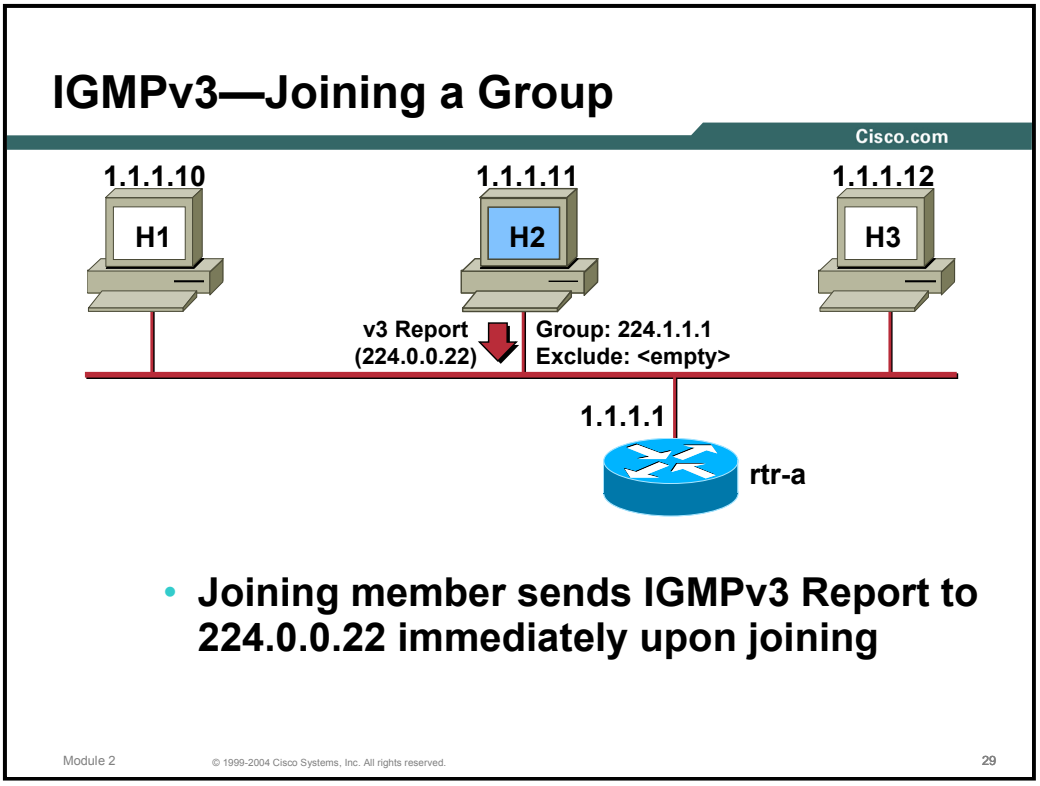
# of Sources (N)
Number of Sources in Record

Source Address 1- N
Address of Source

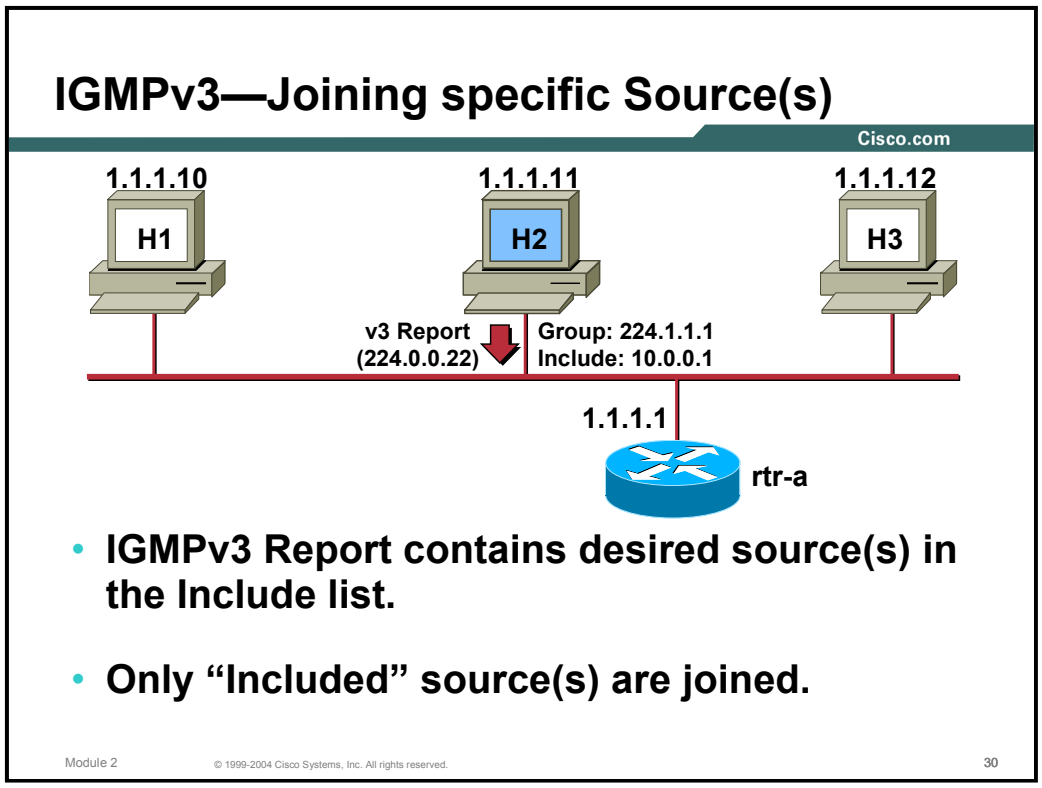
- **# of Group Records**
  - Indicates the number of Group records that are contained in the Membership Report. IGMPv3 Membership Reports can contain IGMP state on a number of Groups and Sources within the group. The source information specifies which Sources to "Include" or "Exclude".
- **Aux. Data Length (Group Records)**
  - Indicates the size of the Auxilliary Data area.
- **Multicast Address (Group Records)**
  - The multicast group address of the joined Group.
- **# of Sources (Group Records)**
  - Indicates the number of Sources in the list.



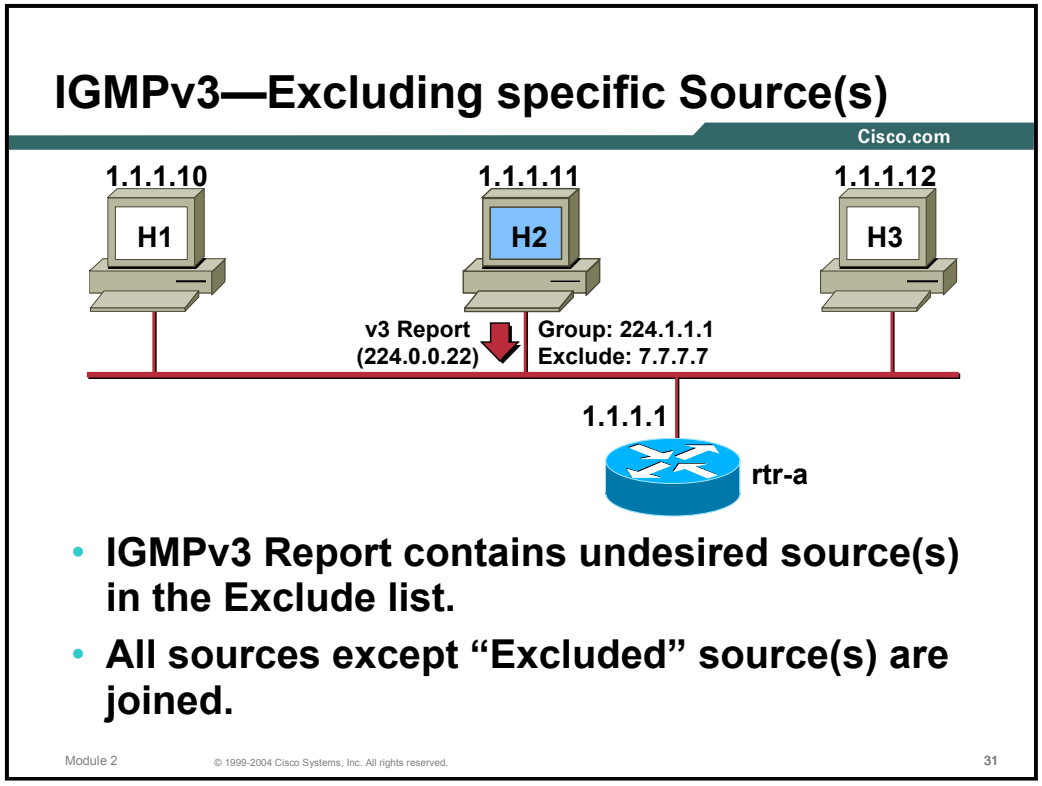
- **IGMPv3 Example**
  - In this example, host “H1” wishes to join group 224.1.1.1 but only wishes to receive traffic from Source 1.1.1.1. The IGMPv3 host can signal the designated router, “R3”, that it is only interested in multicast traffic from Source 1.1.1.1 for Group 224.1.1.1. Router “R3” could then potentially “prune” the unwanted source, 2.2.2.2,.



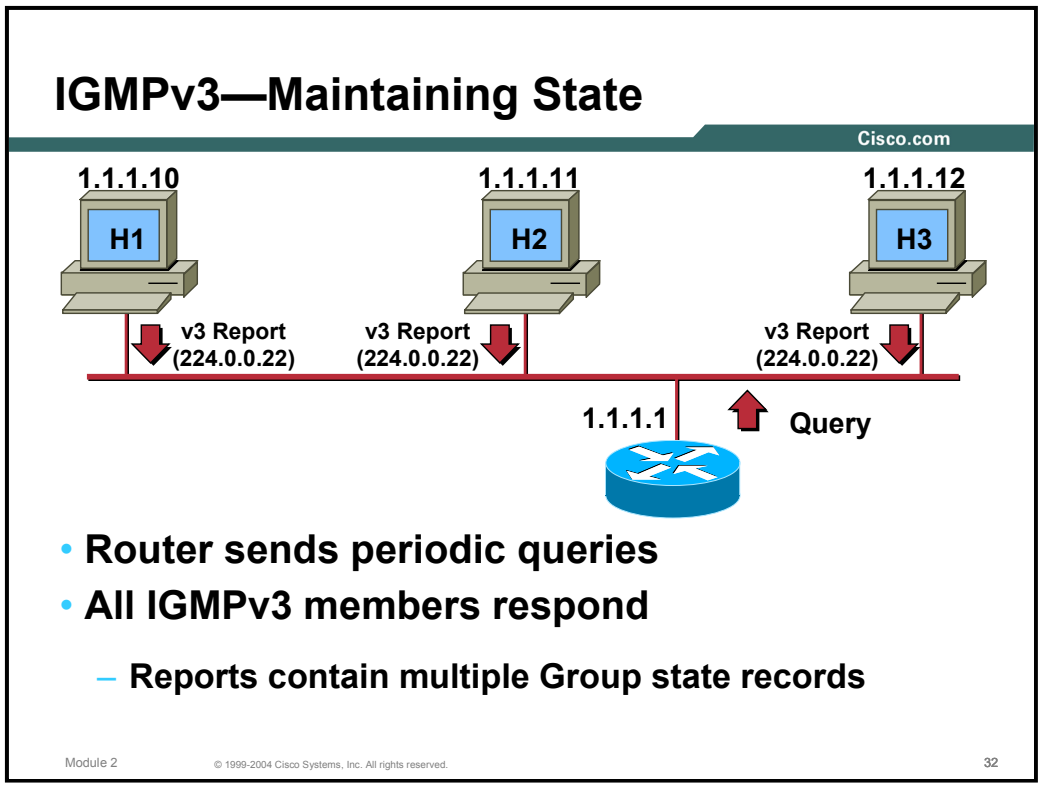
- **Asynchronous Joins**
  - Members joining a group do not have to waited for a query to join; they send in an unsolicited IGMPv3 Membership Report indicating their interest. This reduces join latency for the end system joining if no other members are present.
    - In the example above, Host 2 is joining multicast group 224.1.1.1 and is willing to receive any and all sources in this group.



- **Joining only specific Source(s)**
  - Hosts may signal the router that it wishes to receive only a specific set of sources sending to the group. This is done by using an “Include” list in the Group record of the Report. When an Include list is in use, only the specific sources listed in the Include list are joined.
    - In the example above, Host 2 is joining multicast group 224.1.1.1 and only wants to receive source 10.0.0.1 sending to the group.



- **Joining only specific Source(s)**
  - Hosts may signal the router that it wishes to receive all sources sending to the group except a specific set of undesired sources. This is done by using an "Exclude" list in the Group record of the Report. When an Exclude list is in use, all sources in the group are joined except the sources listed in the Exclude list.
    - In the example above, Host 2 is joining multicast group 224.1.1.1 and wish to receive multicast traffic from any source in the group except source 7.7.7.7.



- **Query-Response Process**
  - The router multicasts periodic Membership Queries to the “All-Hosts” (224.0.0.1) group address.
  - All hosts on the wire respond by sending back an IGMPv3 Membership Report that contains their complete IGMP Group state for the interface.

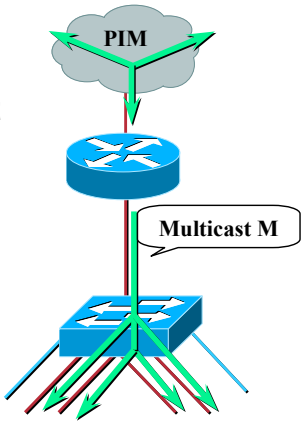
L2 Multicast Frame Switching

L2 Multicast Frame Switching

Cisco.com

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



Module 2 34

- **L2 Multicast Switching**
  - For most L2 Switches, Multicast traffic is normally treated like an unknown MAC address or Broadcast frame which causes the frame to be flooded out every port within a VLAN at rates of over 1 Mbps. This is fine for unknowns and broadcasts but as we have seen earlier, IP Multicast hosts may join and be interested in only specific multicast groups. Again, on most L2 Switches, all this traffic is forwarded out all ports resulting in wasted bandwidth on both the segments and on the end stations.

    One way around this on Catalyst Switches is using the Command Line Interface to program the switch manually to associate a multicast MAC address with say ports 5,6,7 so only ports 5,6,and 7 receive the multicast traffic destined for the multicast group. This works fine but again we know IP Multicast hosts dynamically join and leave groups using IGMP to signal to the Multicast Router. This static way of entering the multicast information is not very scaleable. Dynamic configuration of the Switches' forwarding tables would be a better idea, and cut down on user administration.

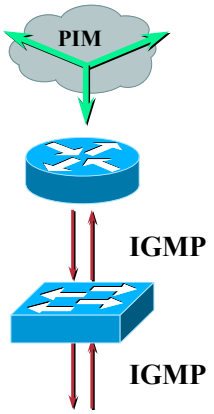
IGMP Snooping

L2 Multicast Frame Switching

Cisco.com

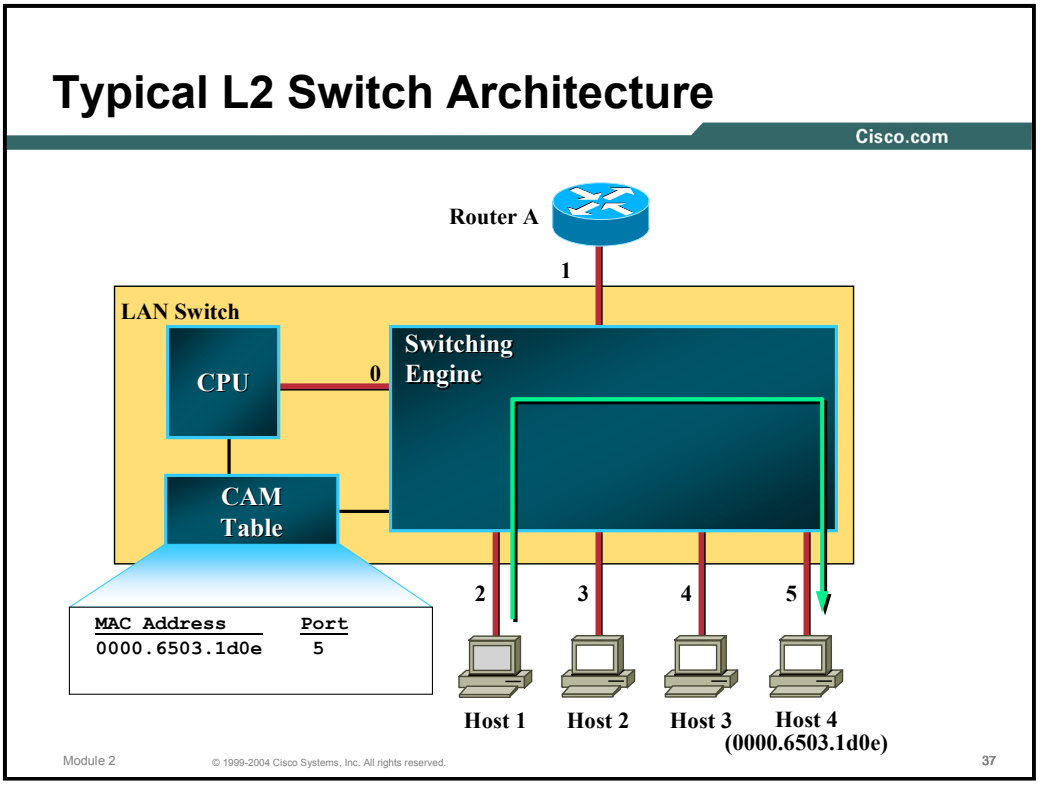
Solution 1: IGMP Snooping

- Switches become "IGMP" aware
- IGMP packets intercepted by the NMP or by special hardware ASICs
- Switch must examine contents of IGMP messages to determine which ports want what traffic
  - IGMP membership reports
  - IGMP leave messages
- Impact on switch:
  - Must process ALL Layer 2 multicast packets
  - Admin. load increases with multicast traffic load
  - Requires special hardware to maintain throughput

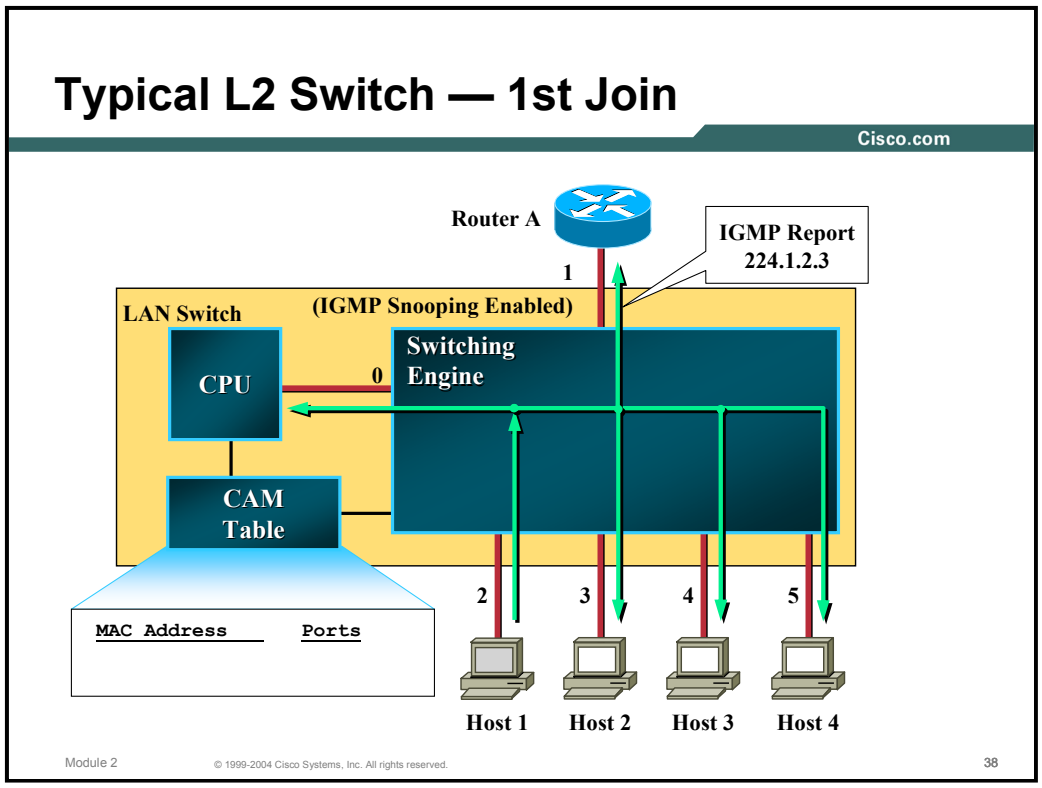


Module 2 36

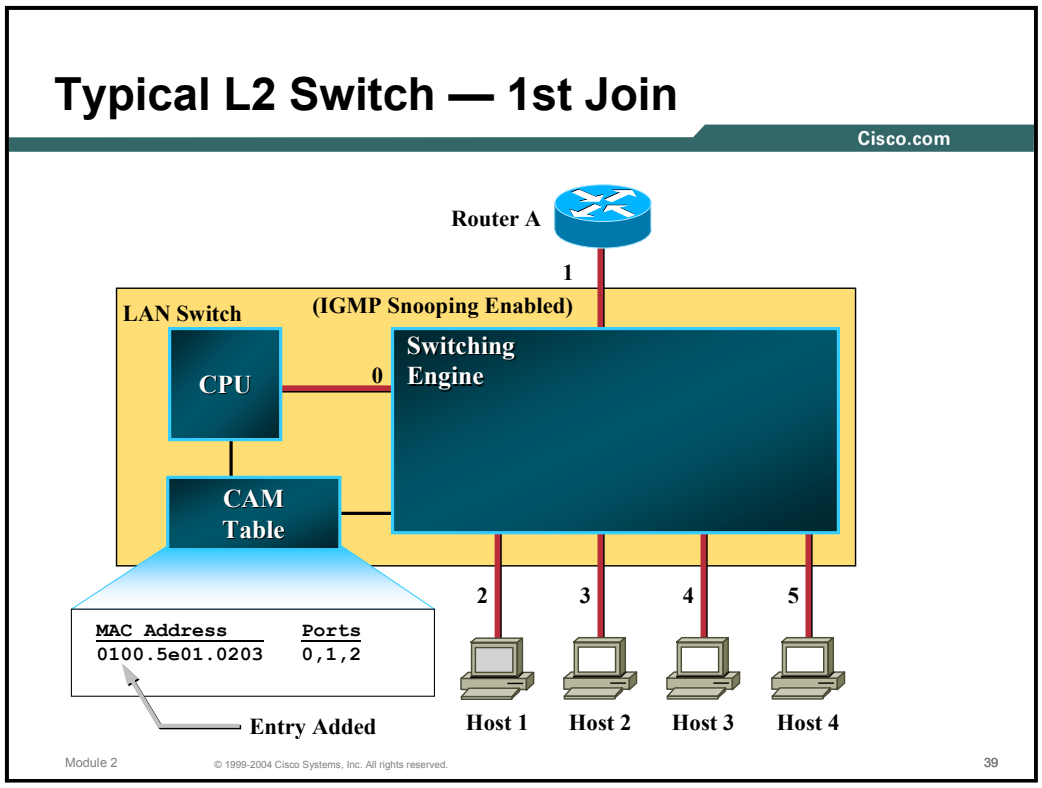
- **Solution 1: IGMP Snooping**
  - As its name implies, switch become IGMP "aware" and listen in on the IGMP conversations between hosts and routers.
  - This requires the processor in the switch to identify and intercept a copy of all IGMP packets flowing between router and hosts and vice versa. This includes:
    - IGMP Membership Reports
    - IGMP Leaves
  - If care is not taken as to how IGMP Snooping is implemented, a switch may have to intercept ALL layer 2 multicast packets in order to identify IGMP packets.
    - This can have a significant impact on the switch's performance.
    - Proper designs require special hardware to avoid this problem. This can directly affect the overall cost of the switch.



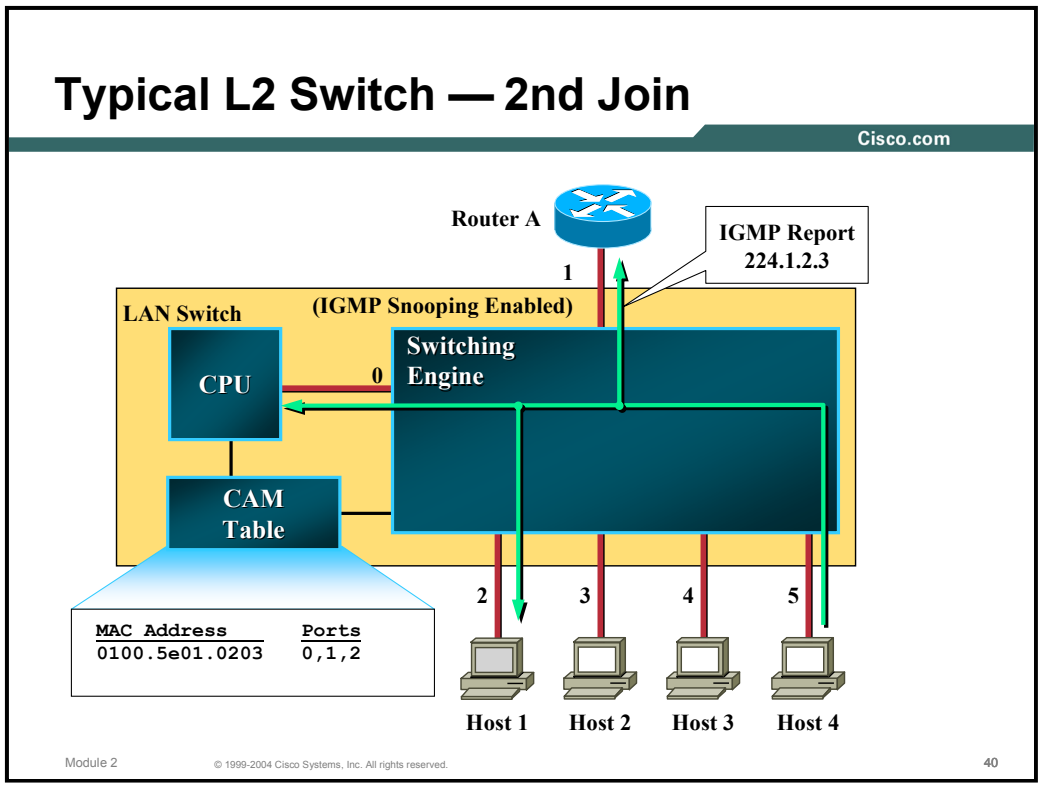
- **Typical Layer 2 Switch**
  - Most Layer 2 switches consist of the following components:
    - Switching Engine - Used to actually perform switching of packets from the input port to the output port(s) under the control of the Contents Addressable Memory (CAM) Table. If there is no entry in the CAM Table that matches the destination MAC address, the Switching Engine will flood the packet to all ports in an attempt to insure that the packet reaches the destination.
    - CAM Table - The information in this table is used to control the operation of the Switching Engine. Each entry in this table contains a Layer 2 destination MAC address and output port(s) where packets addressed to this destination should be switched.
    - CPU - The switch's main CPU populates the CAM Table with destination MAC addresses so that packets can be switched efficiently by the Switching Engine. The CPU "learns" the ports associated with a particular MAC address by watching arriving traffic sent by hosts. It then populates the CAM Table with this "learned" information. (Switches can typically also be instructed to populate the CAM Table with specific MAC address to port mapping information via configuration commands.)
  - In the example shown above, the switch has learned the port (port 5) associated with Host 4's MAC address (0000.6503.1d0e). This information has been stored by the CPU in the CAM Table.
  - Because of this CAM Table entry, packets arriving with Host 4's MAC address as the destination are being switched by the Switching Engine to port 5 as can be seen in the drawing above.
  - In the next few pages, we will see how this simply Layer 2 architecture might be used to implement IGMP Snooping and its potential impact on the switch.



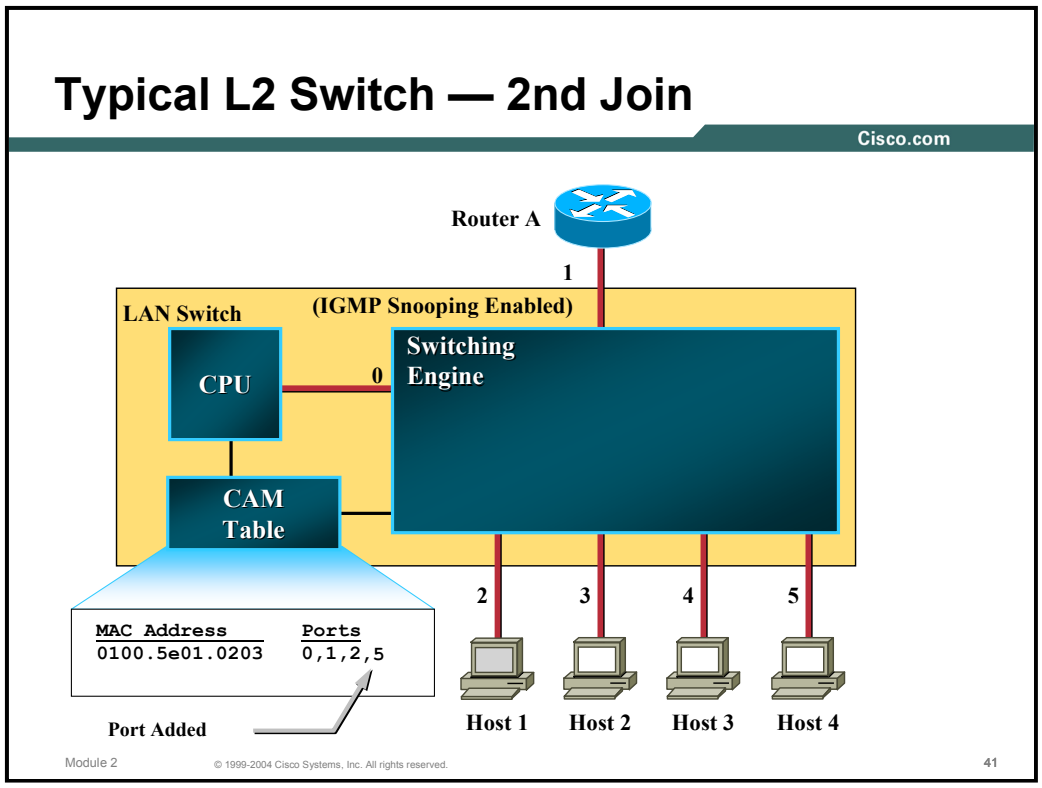
- **IGMP Snooping in L2 Switches**
  - In the above example, the CPU has been programmed to perform IGMP Snooping. This requires the CPU to listen to all IGMP traffic and then add an appropriate Layer 2 multicast MAC address to the CAM Table in order to constrain the IP Multicast traffic to only those ports that require the traffic.
  - Initially, when the first host (Host 1) joins group 224.1.2.3, there is no entry in the CAM table associated with the Layer 2 MAC address equivalent to this group address. Therefore, the initial IGMP Group Membership Report sent by Host 1 is flooded to all ports including the switch's CPU and the Router.
  - Overhearing this, the CPU populates the CAM table with an entry of 0x0100.5e01.0203 which is the L2 MAC address equivalent of IP multicast address 224.1.2.3. Additionally, this entry is populated with the port associated with Host 1 (port 2) as well as the Router and the CPU ports (ports 0 and 1). The CPU port must be included in order for the Switching Engine to continue to forward any further IGMP messages addressed to this group to the CPU for processing.



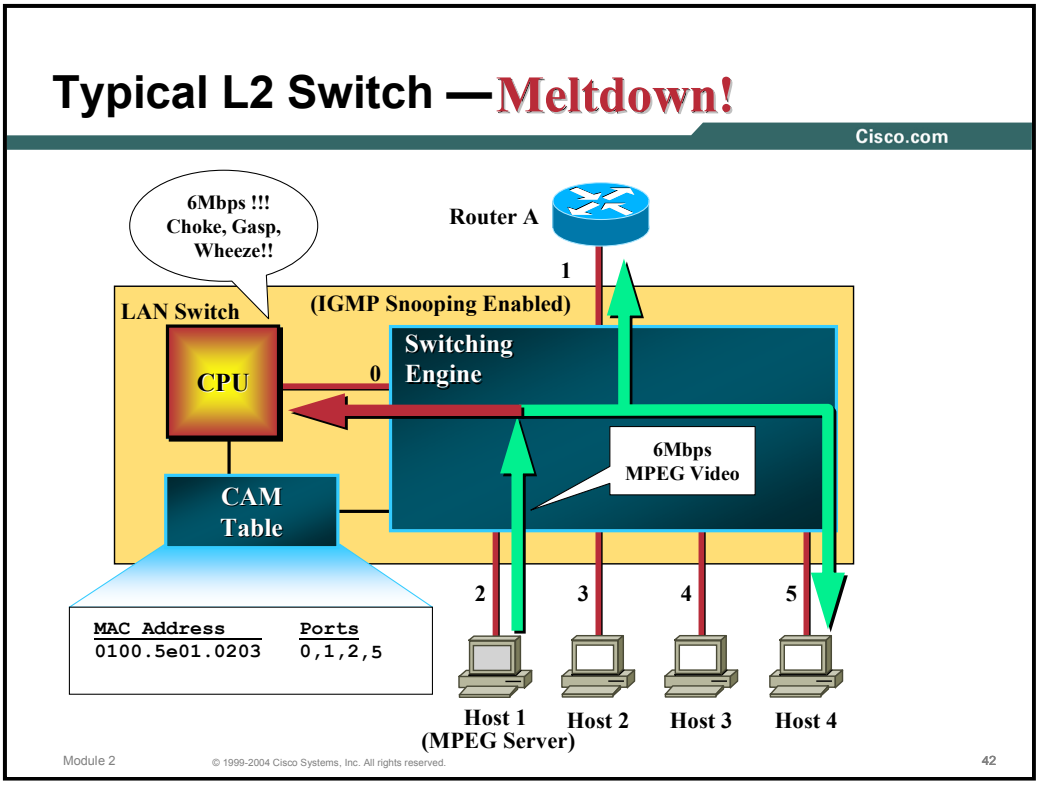
- **IGMP Snooping in L2 Switches**
  - In the above example, the CPU has been programmed to perform IGMP Snooping. This requires the CPU to listen to all IGMP traffic and then add an appropriate Layer 2 multicast MAC address to the CAM Table in order to constrain the IP Multicast traffic to only those ports that require the traffic.
  - Initially, when the first host (Host 1) joins group 224.1.2.3, there is no entry in the CAM table associated with the Layer 2 MAC address equivalent to this group address. Therefore, the initial IGMP Group Membership Report sent by Host 1 is flooded to all ports including the switch's CPU and the Router.
  - Overhearing this, the CPU populates the CAM table with an entry of 0x0100.5e01.0203 which is the L2 MAC address equivalent of IP multicast address 224.1.2.3. Additionally, this entry is populated with the port associated with Host 1 (port 2) as well as the Router and the CPU ports (ports 0 and 1). The CPU port must be included in order for the Switching Engine to continue to forward any further IGMP messages addressed to this group to the CPU for processing.



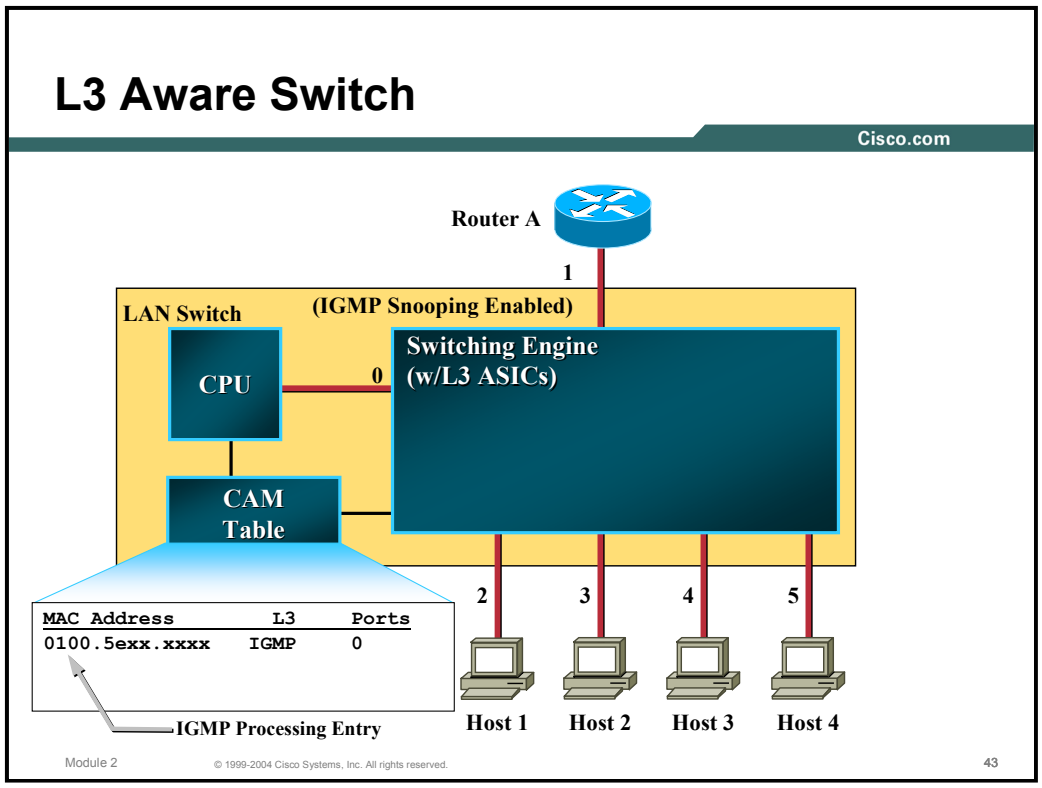
- **IGMP Snooping in L2 Switches**
  - Now let's assume that a second host (Host 4) also joins the group by sending an IGMP Report to group 224.1.2.3.
  - Because of the CAM Table entry for 0x0100.5e01.02.03, this IGMP Report is constrained to only Host 1, the router and the CPU.
  - When the CPU receives the IGMP Report, it simply adds the port (port 5) on which Host 4 is connected to the CAM Table entry. This results in ports 0, 1, 2 and 5 being associated with the multicast MAC address 0x0100.5e01.02.03.



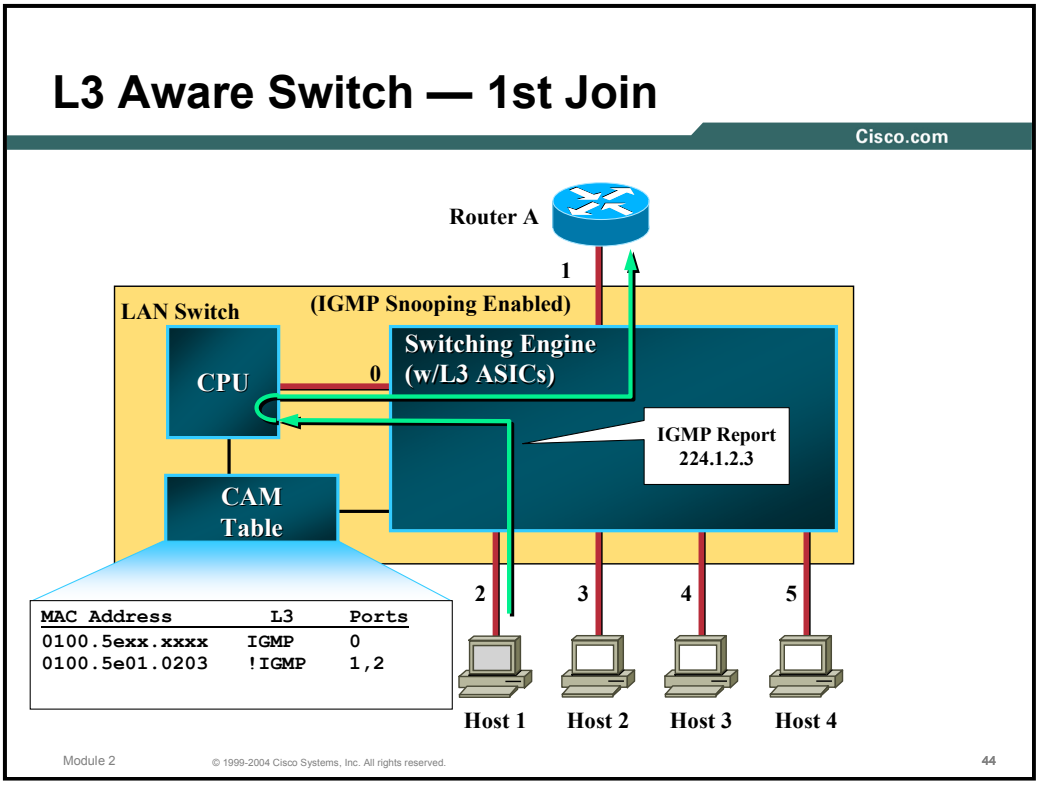
- **IGMP Snooping in L2 Switches**
  - Now let's assume that a second host (Host 4) also joins the group by sending an IGMP Report to group 224.1.2.3.
  - Because of the CAM Table entry for 0x0100.5e01.02.03, this IGMP Report is constrained to only Host 1, the router and the CPU.
  - When the CPU receives the IGMP Report, it simply adds the port (port 5) on which Host 4 is connected to the CAM Table entry. This results in ports 0, 1, 2 and 5 being associated with the multicast MAC address 0x0100.5e01.02.03.



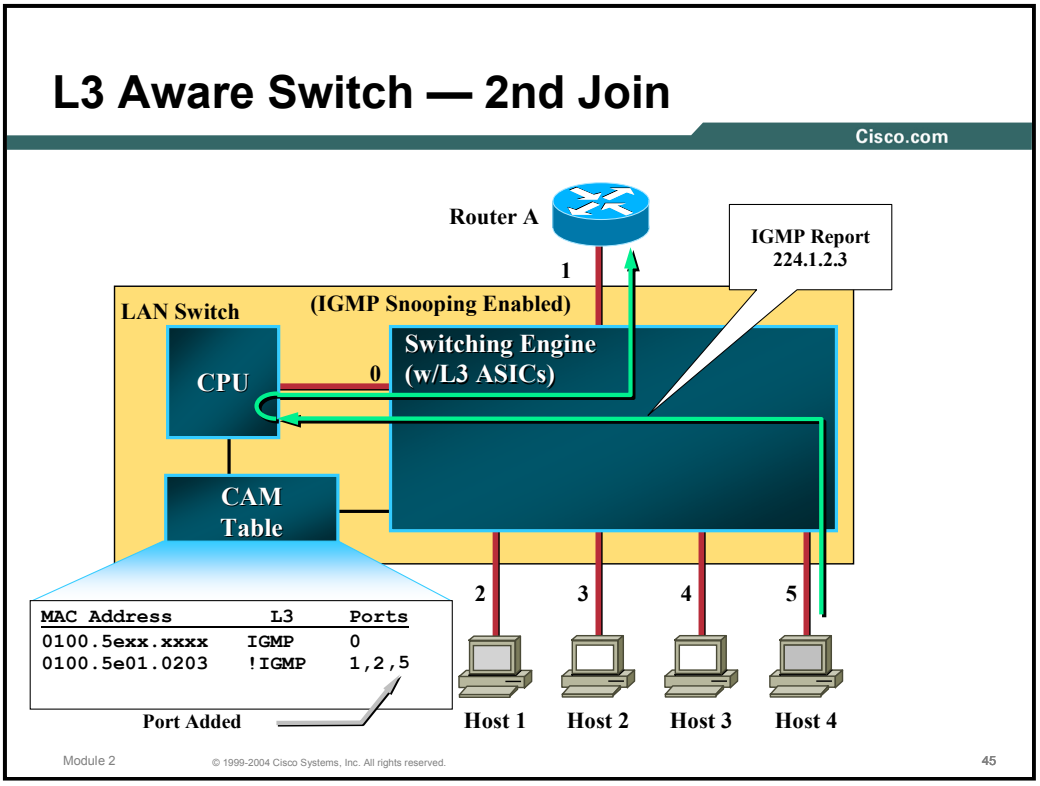
- **IGMP Snooping in L2 Switches**
  - Let us now assume that Host 1 begins transmitting a 1.5Mbps MPEG video stream to multicast group 224.1.2.3. Because the destination MAC address of this stream maps to 0x0100.5e01.02.03, the Switching Engine dutifully switches this traffic to Host 4, the Router and the CPU!
  - In most cases, the switch's CPU does not have sufficient horsepower to keep up with this high rate flow of multicast traffic and switch performance can suffer. In some cases, the switch can actually fail under such loads.
- **Summary**
  - IGMP Snooping can be (and often is) implemented in low-end, Layer-2 only switches using techniques similar to the above. While this is fine for extremely low data-rate multicast flows or carefully orchestrated vendor demonstrations of their switch's IGMP Snooping feature, it is generally inadequate for real-world use.



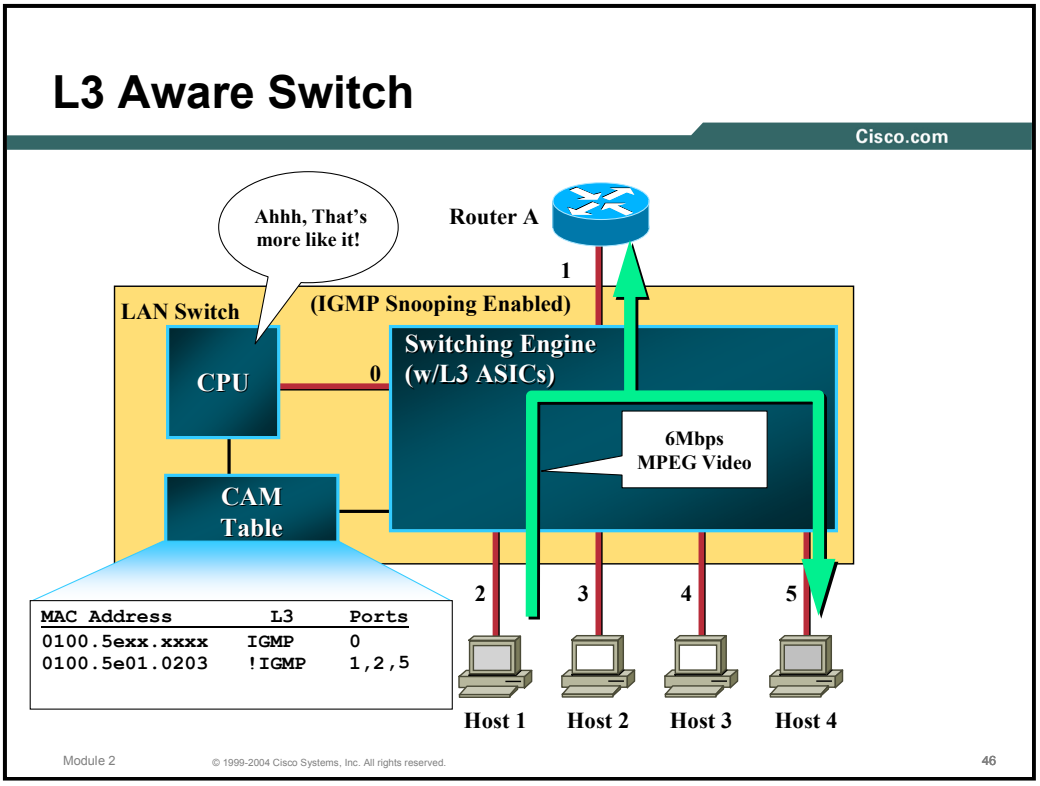
- **IGMP Snooping in L3-aware Switches**
  - In order to properly implement IGMP Snooping on a switch without suffering performance degradation, it is necessary to make the switch Layer 3 aware. This is typically accomplished adding Layer 3 ASIC's to the Switching Engine in addition to extending the CAM Table so that entries may contain additional Layer 3 information that can be used to make switching decisions. (In case it is not obvious, this means the switch will cost more money.)
  - In the above example, we have just such a Layer-3 aware switch that has been programmed to perform IGMP Snooping using some of the added Layer 3 capabilities in the switch's architecture. In order to accomplish this, the CPU populates the CAM Table with a special entry to capture any and all IGMP packets.
    - There can be many ways to do this but in the example above, the CAM Table entry contains a wildcard MAC address that will match on any IP multicast address. Furthermore, the Layer 3 part of the packet must contain an IGMP protocol packet in order for the entry to match and cause the packet to be switched to the CPU



- **IGMP Snooping in L3-aware Switches**
  - Let's assume that the first host (Host 1) now joins group 224.1.2.3 and signals this by sending an IGMP Report. This report matches on the first entry in the CAM Table and is switched to the CPU.
  - The CPU responds by forwarding the packet on to the Router (for normal IGMP processing) and then adds a second entry to the CAM table to switch 224.1.2.3 group traffic to Host 1 and the Router (ports 1 and 2).
  - This second entry will match IFF:
    - The packet is addressed to multicast MAC address 0x0100.5e01.0203 (the Layer 2 equivalent to group address 224.1.2.3) and
    - The packet is not and IGMP packet.



- **IGMP Snooping in L3-aware Switches**
  - Now let's assume that again Host 4 is the second host to join 224.1.2.3 and therefore sends an IGMP Report to 224.1.2.3. Once again, the IGMP Report matches on the first entry and is switched to the CPU.
  - The CPU responds by forwarding a copy of the IGMP Report to the Router and by adding the port associated with Host 4 (port 5) to the port list in the second CAM Table entry.



- **IGMP Snooping in L3-aware Switches**
  - In the final step of our example, Host 1 once again starts up the 1.5Mbps MPEG video stream to group 224.1.2.3.
  - Packets in this stream will not match on the first CAM Table entry but instead will match on the second entry. Therefore, the video stream is switched to only Host 4 and the Router and the CPU is not burdened with this unwanted data stream.
- **Summary**
  - In order to construct a switch that is capable of IGMP Snooping without suffering a performance hit, the switch must use special Layer 3 ASIC or some similar technique. This increases the overall cost of the switch.

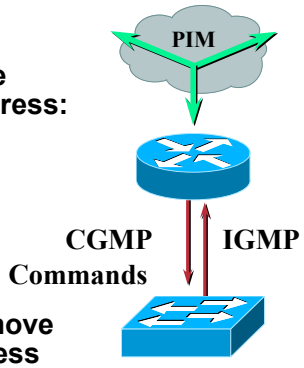
CGMP

L2 Multicast Frame Switching

Cisco.com

Solution 2: CGMP—Cisco Group Multicast Protocol

- Runs on both the switches and the router
- Router sends CGMP multicast packets to the switches at a well known multicast MAC address:
  - 0100.0cdd.dddd
- CGMP packet contains :
  - Type field—Join or Leave
  - MAC address of the IGMP client
  - Multicast address of the group
- Switch uses CGMP packet info to add or remove an entry for a particular multicast MAC address



Module 2 48

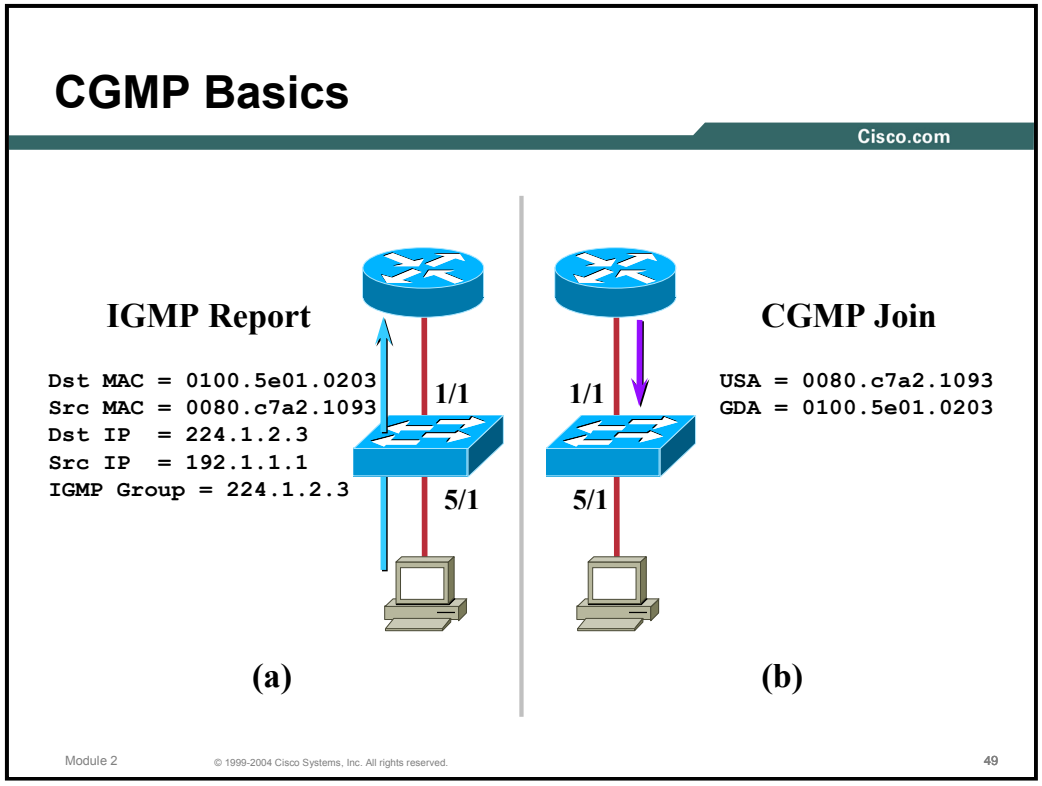
- **Solution 2: CGMP**
  - CGMP is based on a client server model where the router can be considered a CGMP server and the switch taking on the client role. There are software components running on both devices, with the router translating IGMP messages into CGMP commands which are then executed on the Catalyst 5000 NMP and used to program the EARL's forwarding tables with the correct Multicast entries.

    Since the hosts and routers use well-known IP Multicast Addresses, the EARL can be preprogrammed to direct IGMP Control packets both to the router and the NMP. We will see the NMPs use of these IGMP control packets in a later slide.

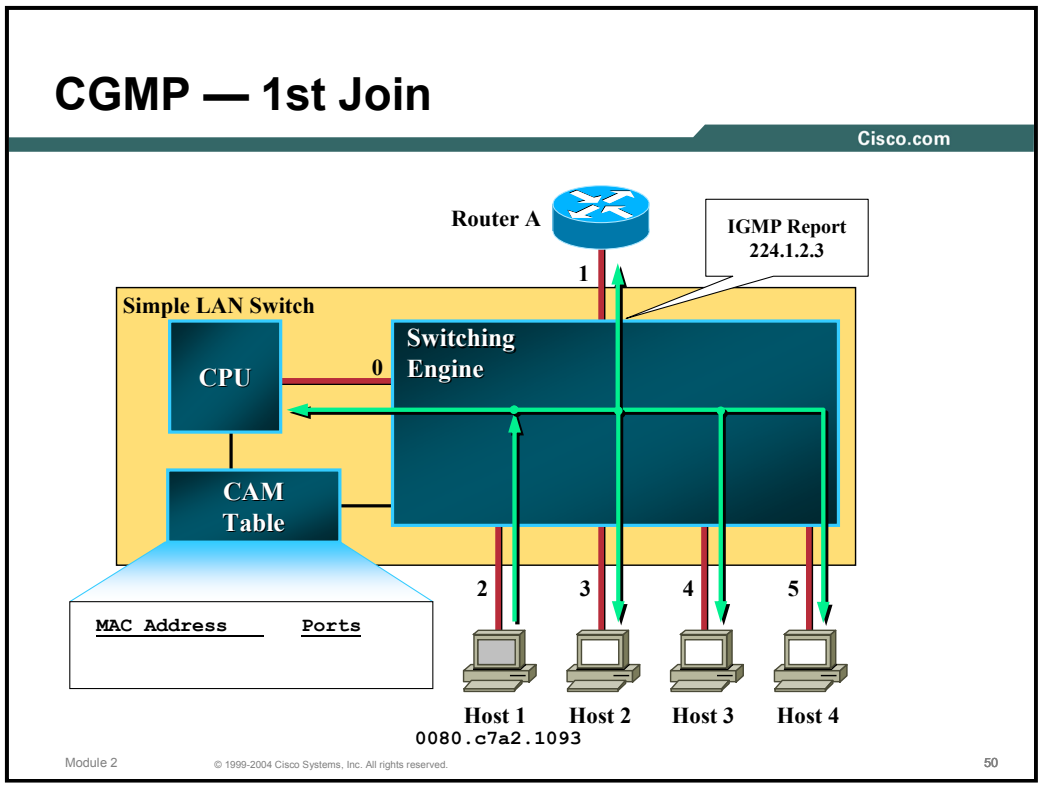
    The basis of CGMP is that the IP Multicast router sees all IGMP packets and therefore can inform the switch when specific hosts join or leave Multicast groups. The switch then uses this information to program it's forwarding table.

    When the router sees an IGMP control packet it creates a CGMP packet that contains the request type (Join or Leave), the Layer 2 Multicast MAC Address, and the actual MAC address of the client.

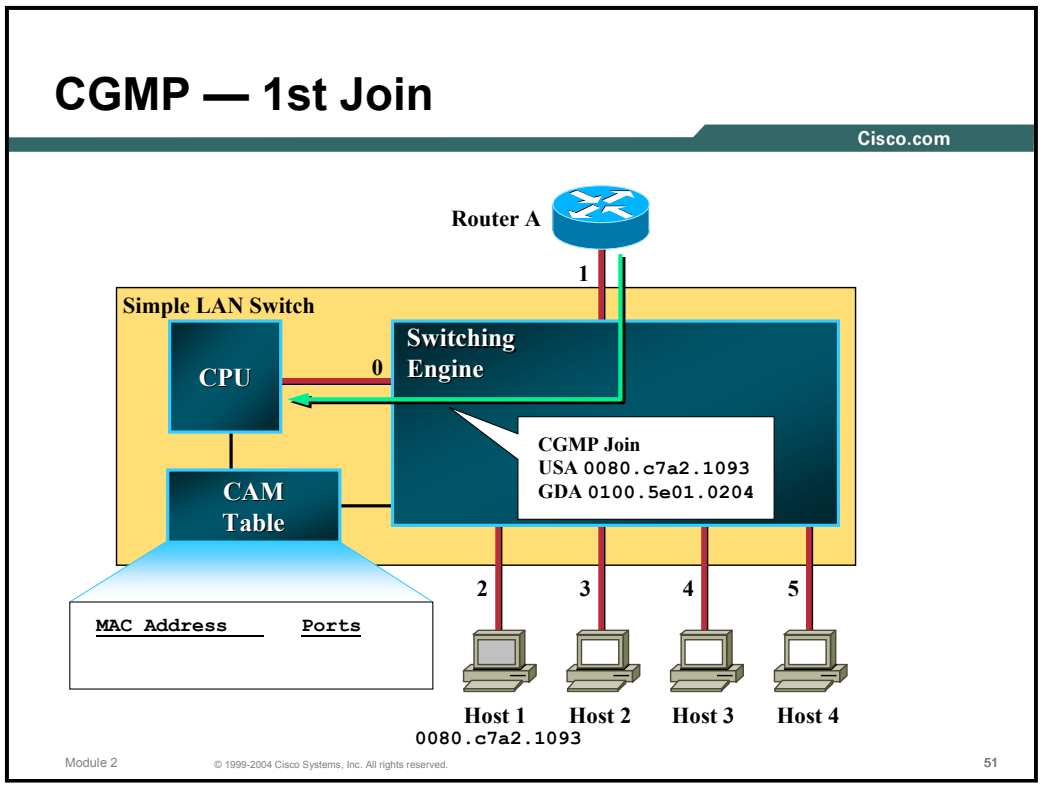
    This packet is sent to a well known address which all CGMP switches listen on. It is then interpreted and the proper entries created in the switch's CAM Table to constrain the forwarding of multicast traffic for this group.



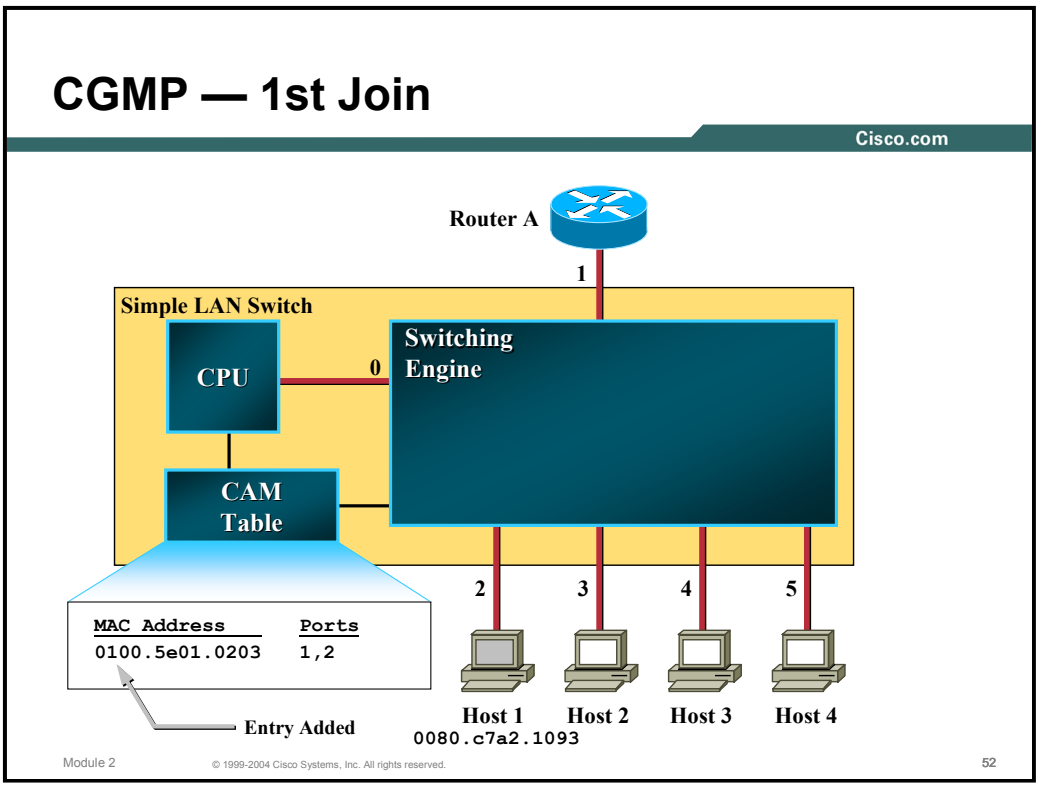
- **CGMP Example**
  - In this example - the client will asynchronously send an IGMP Membership Report when it wants to join the group.
  - The Router converts this IGMP Membership Report into a CGMP Join containing:
    - USA - Unicast Source Address
    - GDA - Group Destination Address
  - The CGMP Join is multicast to a well-known (non-IP) multicast MAC address which the switch listens on.



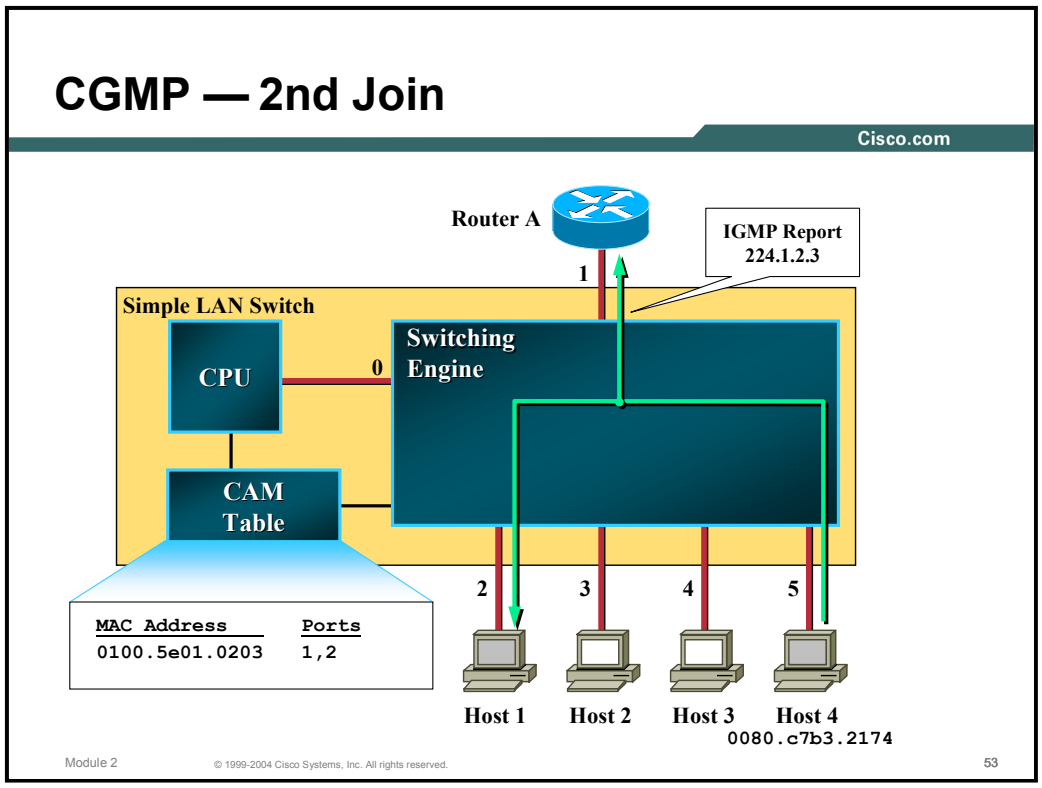
- **CGMP Implementation in L2 switches**
  - Because the switch relies on the Router to assist in the process of constraining IP multicast traffic at Layer 2, it can be implemented very easily in low-end, Layer2 only switches.
  - In the above CGMP example, the first host (Host 1) joins multicast group 224.1.2.3 by sending an IGMP Membership Report.
  - Because there is no matching entry in the CAM Table, the IGMP Membership Report is flooded to all ports including the Router who processes the IGMP Report.



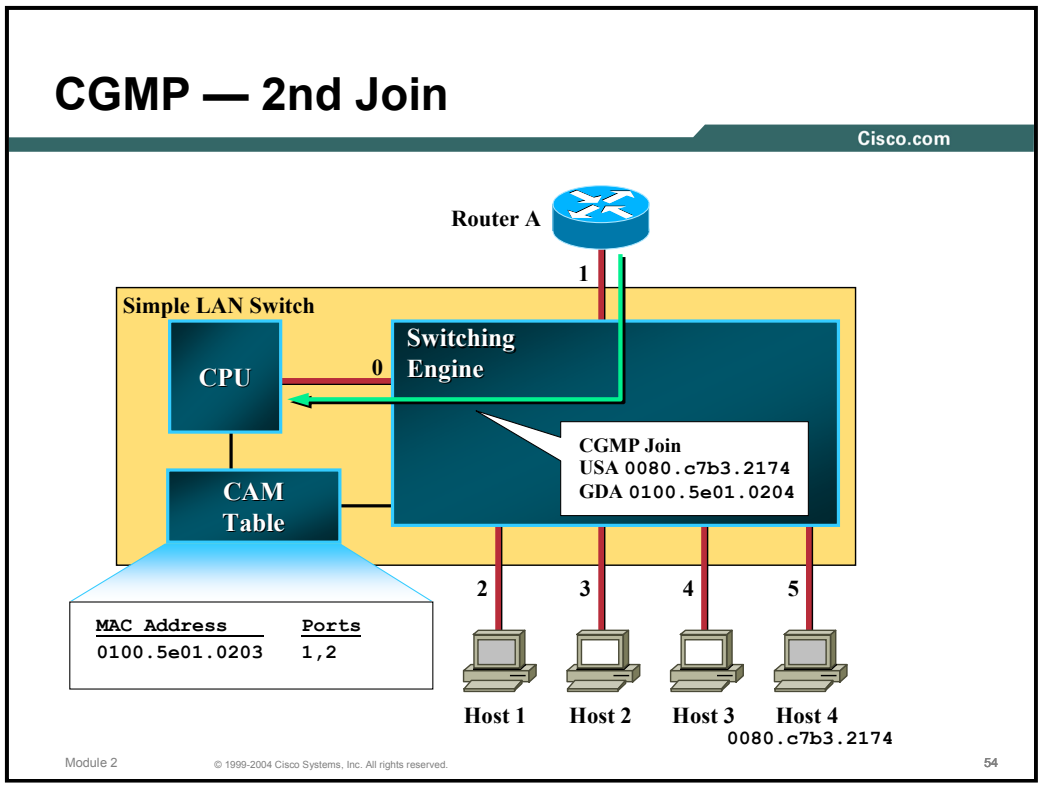
- **CGMP Implementation in L2 switches**
  - In addition to performing normal IGMP processing of the IGMP Membership Report, the Router also converts it into a CGMP Join message containing the MAC address of the host that sent the IGMP Report (Host 1) in the USA field and the Layer 2 MAC address equivalent of group 224.1.2.3 in the GDA field. This CGMP Join message is then multicast back to the switch.
  - When the switch receives the CGMP Join, it uses the host address in the USA field to determine the port where the Host resides. This is done by scanning the CAM table for the hosts MAC address to obtain the associated port number. (This step is not shown in the example above.)
  - The CPU then populates its CAM Table with an entry containing the multicast MAC address from the GDA field and the port number of the host that joined along with the port numbers of any routers connected to the switch.
    - Note: The CPU has many ways to determine which ports have routers attached. These include listening for DVMRP Probes, PIM Hellos, and IGMP Queries.



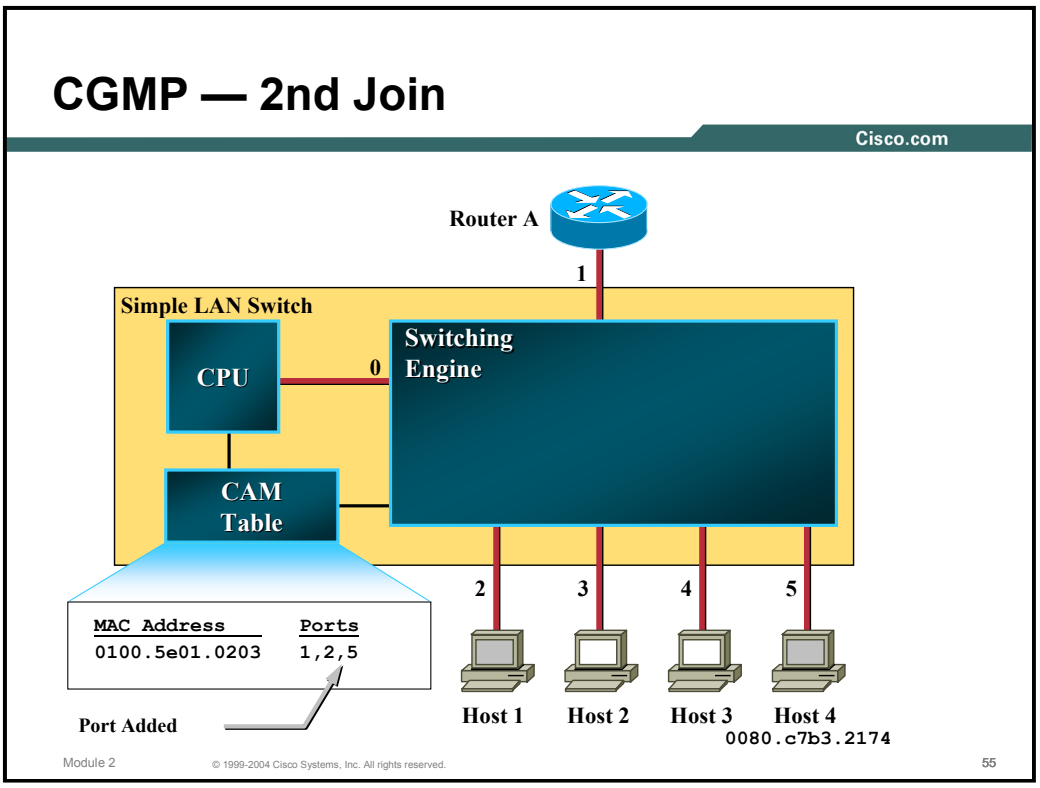
- **CGMP Implementation in L2 switches**
  - In addition to performing normal IGMP processing of the IGMP Membership Report, the Router also converts it into a CGMP Join message containing the MAC address of the host that sent the IGMP Report (Host 1) in the USA field and the Layer 2 MAC address equivalent of group 224.1.2.3 in the GDA field. This CGMP Join message is then multicast back to the switch.
  - When the switch receives the CGMP Join, it uses the host address in the USA field to determine the port where the Host resides. This is done by scanning the CAM table for the hosts MAC address to obtain the associated port number. (This step is not shown in the example above.)
  - The CPU then populates its CAM Table with an entry containing the multicast MAC address from the GDA field and the port number of the host that joined along with the port numbers of any routers connected to the switch.
    - Note: The CPU has many ways to determine which ports have routers attached. These include listening for DVMRP Probes, PIM Hellos, and IGMP Queries.



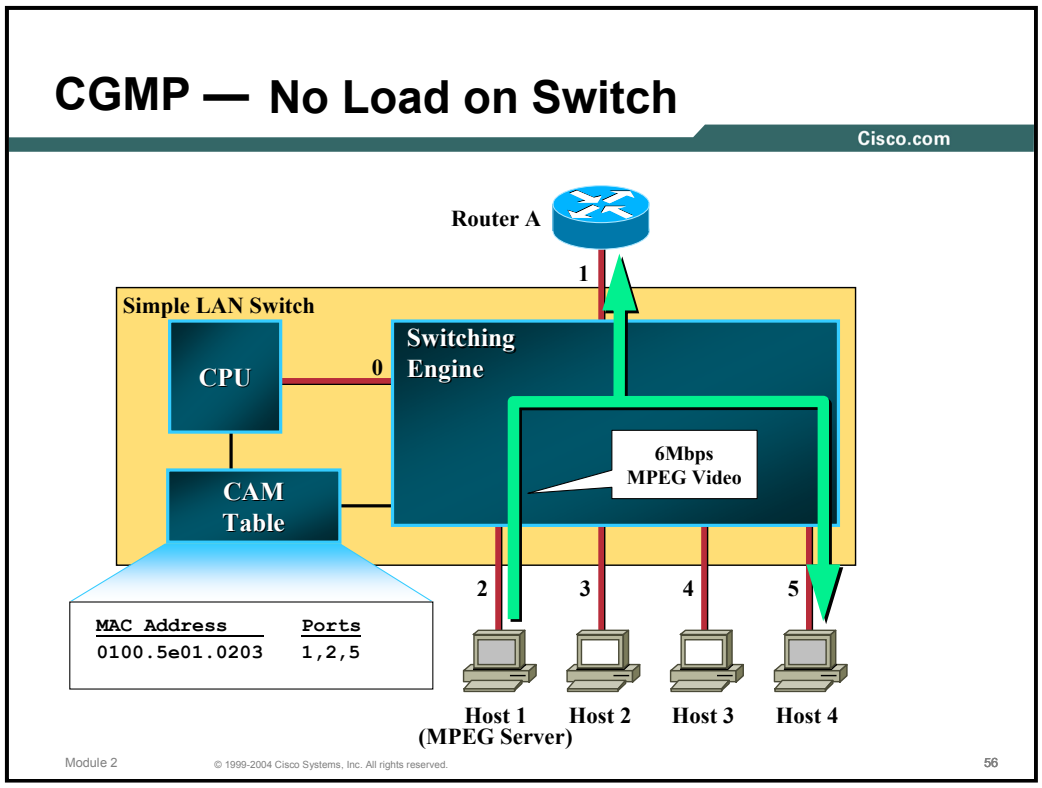
- **CGMP Implementation in L2 switches**
  - Next, let's assume that (once again) Host 4 is the second host to join group 224.1.2.3 and signals this by sending an IGMP Report to 224.1.2.3.
  - Because the IGMP Report is sent to group 224.1.2.3, the MAC destination address is 0x0100.5e01.0203 which matches on the first entry in the CAM Table shown above. This results in the IGMP Report being sent to Host 1 and the Router.



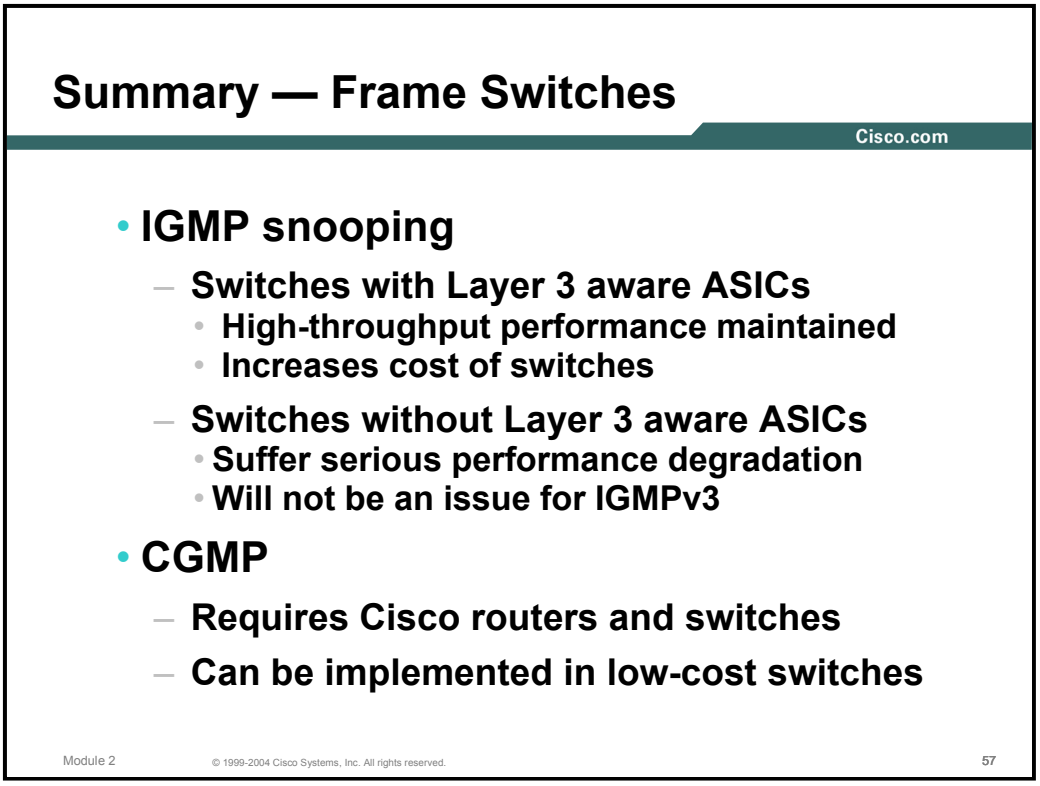
- **CGMP Implementation in L2 switches**
  - In addition to performing normal IGMP processing of the IGMP Membership Report, the Router again converts it to a CGMP Join message containing the MAC address of Host 4 in the USA field and the Layer 2 MAC address equivalent of group 224.1.2.3 in the GDA field. The resulting CGMP Join message is then multicast back to the switch.
  - When the switch receives this CGMP Join, it again uses the host address in the USA field to determine the port where the Host resides. (In this case, port 5.)
  - The CPU then adds port 5 to the port list in the existing CAM Table entry associated with the multicast MAC address from the GDA field.



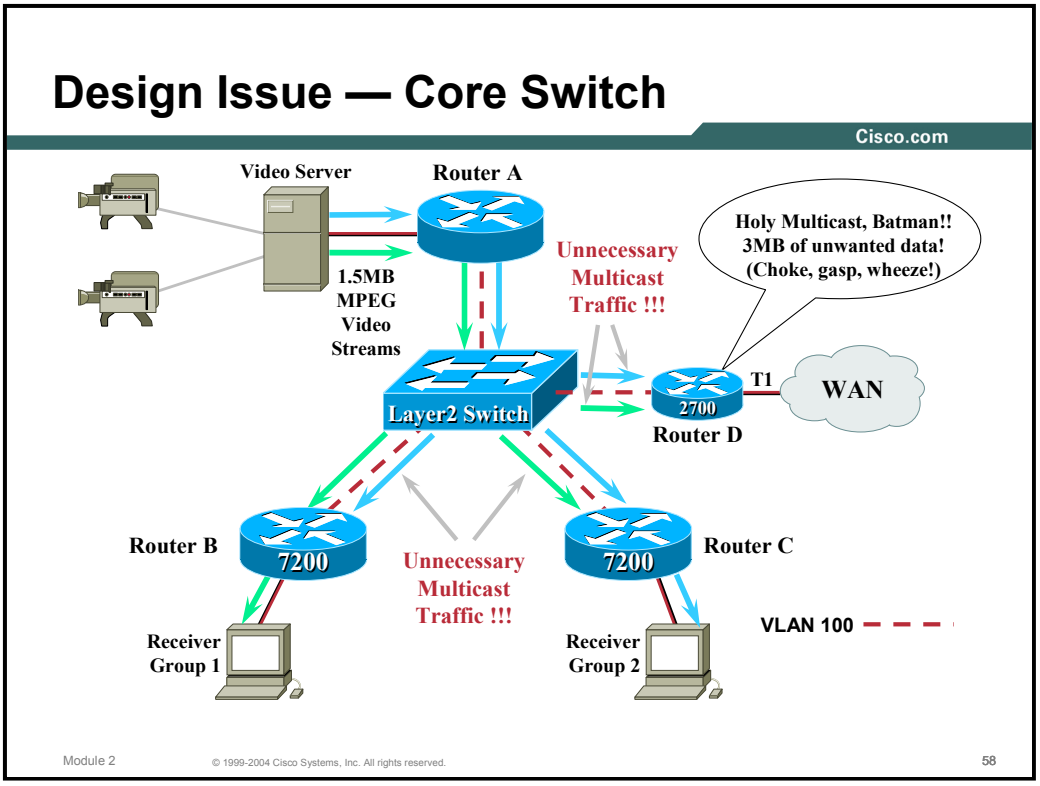
- **CGMP Implementation in L2 switches**
  - In addition to performing normal IGMP processing of the IGMP Membership Report, the Router again converts it to a CGMP Join message containing the MAC address of Host 4 in the USA field and the Layer 2 MAC address equivalent of group 224.1.2.3 in the GDA field. The resulting CGMP Join message is then multicast back to the switch.
  - When the switch receives this CGMP Join, it again uses the host address in the USA field to determine the port where the Host resides. (In this case, port 5.)
  - The CPU then adds port 5 to the port list in the existing CAM Table entry associated with the multicast MAC address from the GDA field.



- **CGMP Implementation in L2 switches**
  - In our final drawing of the example, Host 1 again begins sourcing its 1.5Mbps MPEG video stream to group 224.1.2.3.
  - When this stream hits the switch, it matches on the first entry in the CAM Table and is switched to Host 4 and the Router.
    - Note that because the CPU's port is not included in this entry, the high-rate video stream is not being sent to the CPU and hence does not impact the performance of the switch.

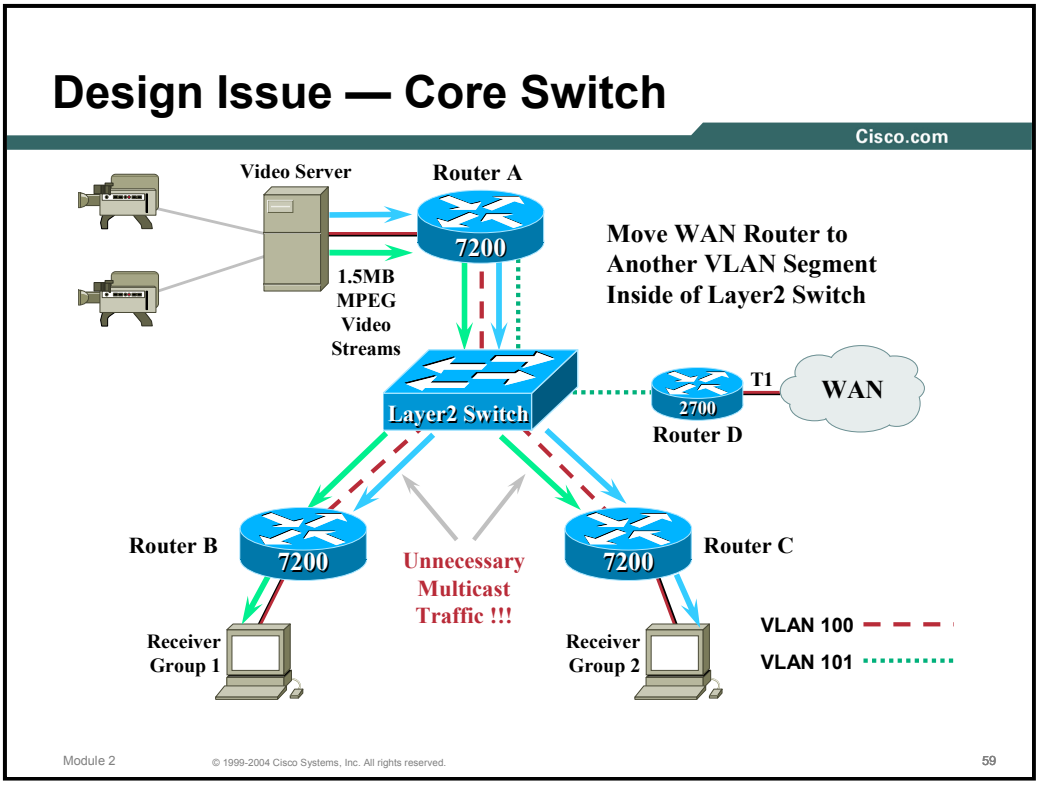


- **Summary**
  - IGMP Snooping can actually provide some performance optimizations over CGMP. However, it requires switches that are implemented with more costly Layer 3 aware ASIC's in order to avoid performance impacts.
  - CGMP is a proprietary protocol that is only implemented on Cisco routers and switches and does not have quite as many performance optimizations that IGMP Snooping can offer. However, it is the ONLY choice if one desires to provide Layer 2 multicast traffic constraint on low-end switches such as the Cisco Catalyst 1900 or other equivalent switches.



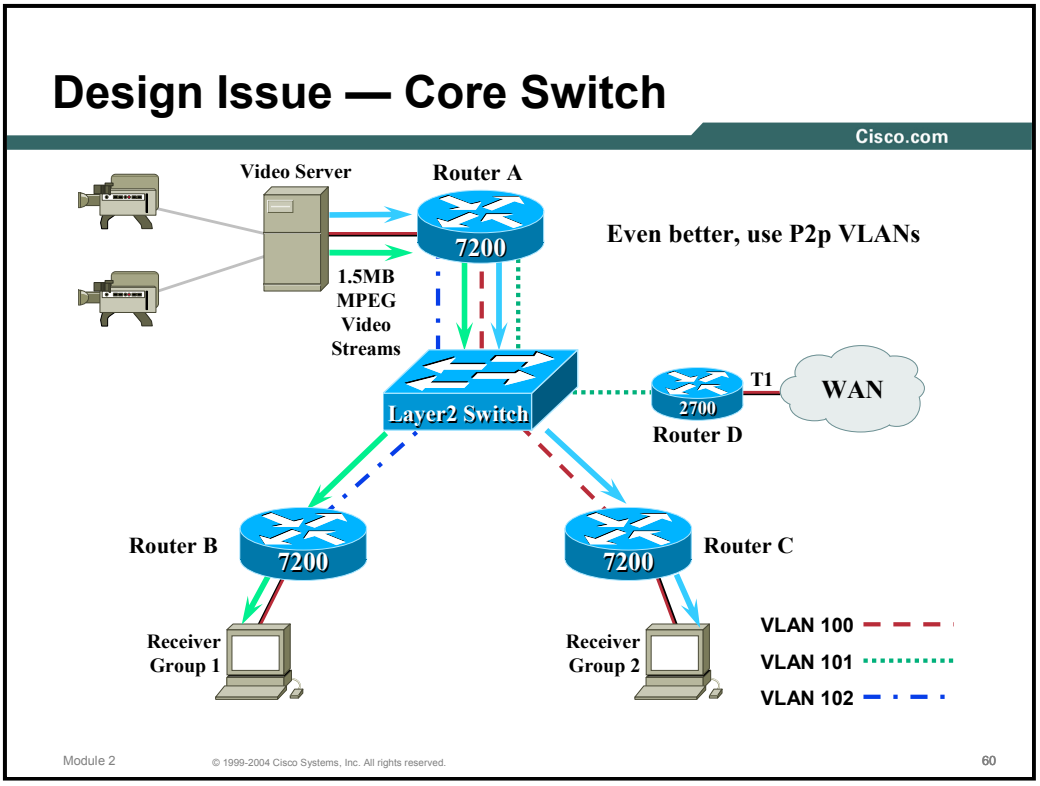
- **Layer 2 Design Issues — Core Switch Issues**
  - In the case of a core network composed of several routers on an Ethernet segment, IGMP Snooping and CGMP provide absolutely no help in the constraint of multicast traffic flows. This is because routers do not send IGMP Membership Reports for desired multicast flows. (They use PIM control messages or some other routing protocol control messages instead.)
- **Example:**
  - Consider the network shown in the drawing above. Three campus routers are connected via 100Mbps Ethernet to a core switch. A video server connected to Router A is sourcing two 1.5Mbps MPEG video multicast streams, one to Group 1 and another to Group 2.
  - Router B has a directly connected member of Group 1 and therefore needs the 1.5Mbps Group 1 video stream.
  - Router C has a directly connected member of Group 2 and therefore needs the 1.5Mbps Group 1 video stream.
  - Because both Routers B & C are on the same VLAN (albeit on different ports on the switch), they each receive both Group 1 & 2 video streams even though they only need one.
  - Even worse, Router D has been connected to this core backbone VLAN for the purpose of supplying remote sites with unicast connectivity and low rate multicast. (i.e. there is no intention of sending MPEG video to the remote sites.) Unfortunately, the little 2700 will also receive both of the high-rate video streams for a total of 3Mbps of unwanted traffic!

    While the 2700 is capable of “fast-dropping” the unwanted traffic in the fast-switching path, it still has a significant impact on the performance of the router.

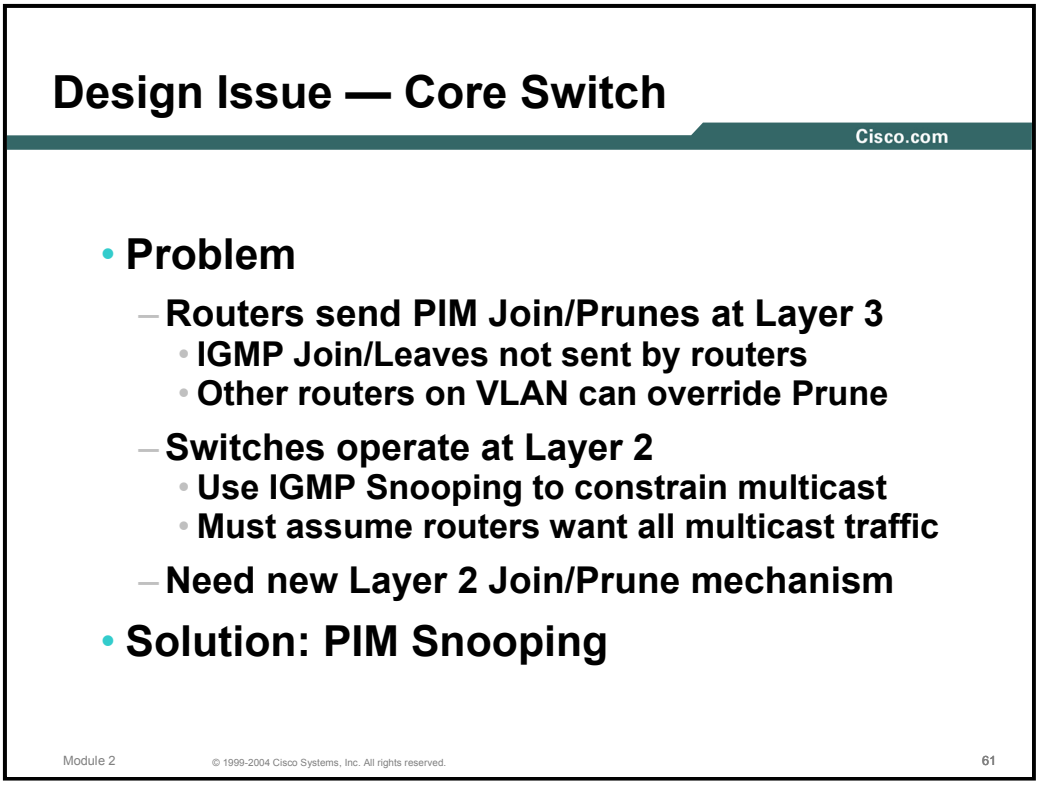


- **Layer 2 Design Issues — Core Switch Issues**
  - While today's technology can not solve this problem (it would basically require the switch to run PIM which means it must become a router and not a switch), this problem can be address by proper network design.
- **Solution**
  - By connecting Router D to a separate VLAN off of Router A (this could be accomplished using another port on Router A and a separate VLAN in the Layer2 Switch), Router D is able to prune off any unwanted traffic.
    - Exercise 1: Why is this now possible? (See answer below.)
  - Unfortunately, we still have unwanted traffic flowing to Routers B & C.

Answer to Exercise 1: Because there is no other router on the LAN segment, Router A is able to Prune off the traffic flow without the Prune being overridden by another router on the LAN segment.



- **Solution (cont.)**
  - The better solution would be to use point-to-point VLANs to interconnect the routers.



- **Design Issue — Core Switch**
  - Routers do not send IGMP Membership reports or IGMP Leaves to communicate their desire to receive multicast traffic. Instead, they communicate at Layer 3 using PIM Join/Prune messages. However, when sending a PIM (*,G) Prune message to indicate they do not want a particular multicast group, another router on the VLAN can override the Prune.
  - Switches only operate at Layer 2 (otherwise they would be routers). They listen to IGMP messages to constrain the flow of multicast traffic to hosts that wish to receive a particular multicast group. Because routers do not send IGMP membership reports, the switches must assume that the routers want all multicast traffic. (This is an assumption of the basic multicast model defined in RFC 1112.)
  - In order to constrain multicast traffic between routers on a core LAN segment, routers and switches need some form of Layer 2 Join/Prune communication that permits the routers to inform the switch of which groups it has interest.

PIM Snooping

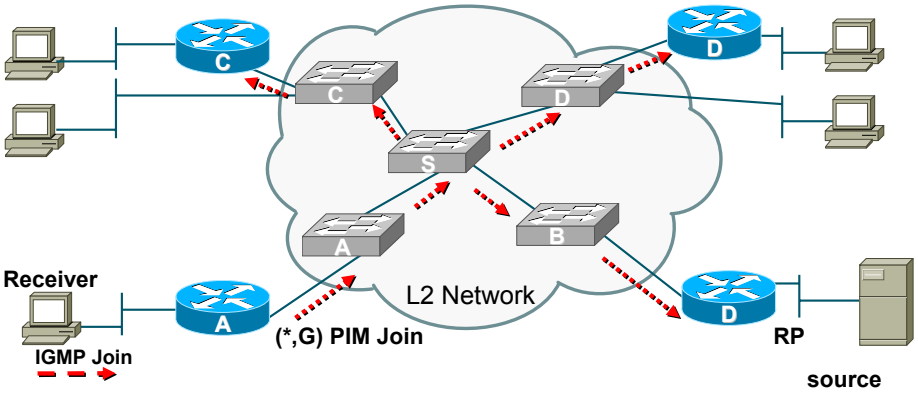
Cisco.com

- **Constrain multicast traffic among multicast router ports of a VLAN**
  - **IGMP Snooping constrains on host ports**
  - **IGMP Snooping floods on multicast router ports**
  - **Effective in core, IXP, metro-ethernet.**

Without PIM Snooping

Cisco.com

PIM Join Flow



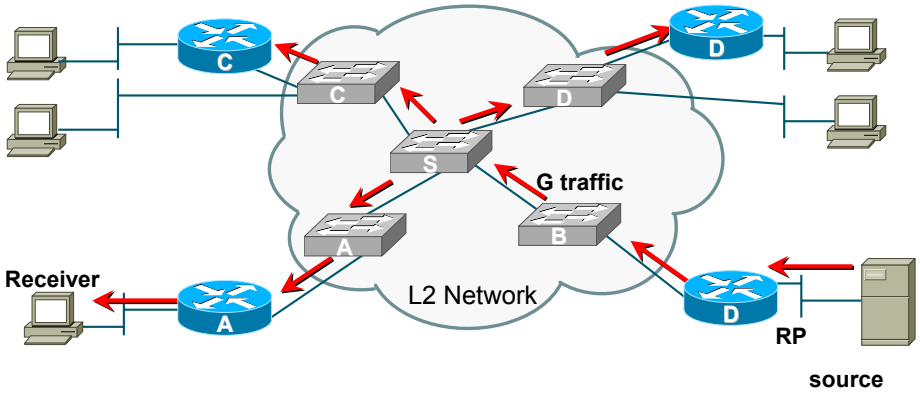
PIM Joins are flooded in the vlan

- **Without PIM snooping**
  - PIM join/prune messages are flooded in the VLAN. This can lead to router's doing Join suppression. That is if router C also need's traffic, it will not initiate a Join message on seeing that Router A had already requested for traffic

Without PIM Snooping

Cisco.com

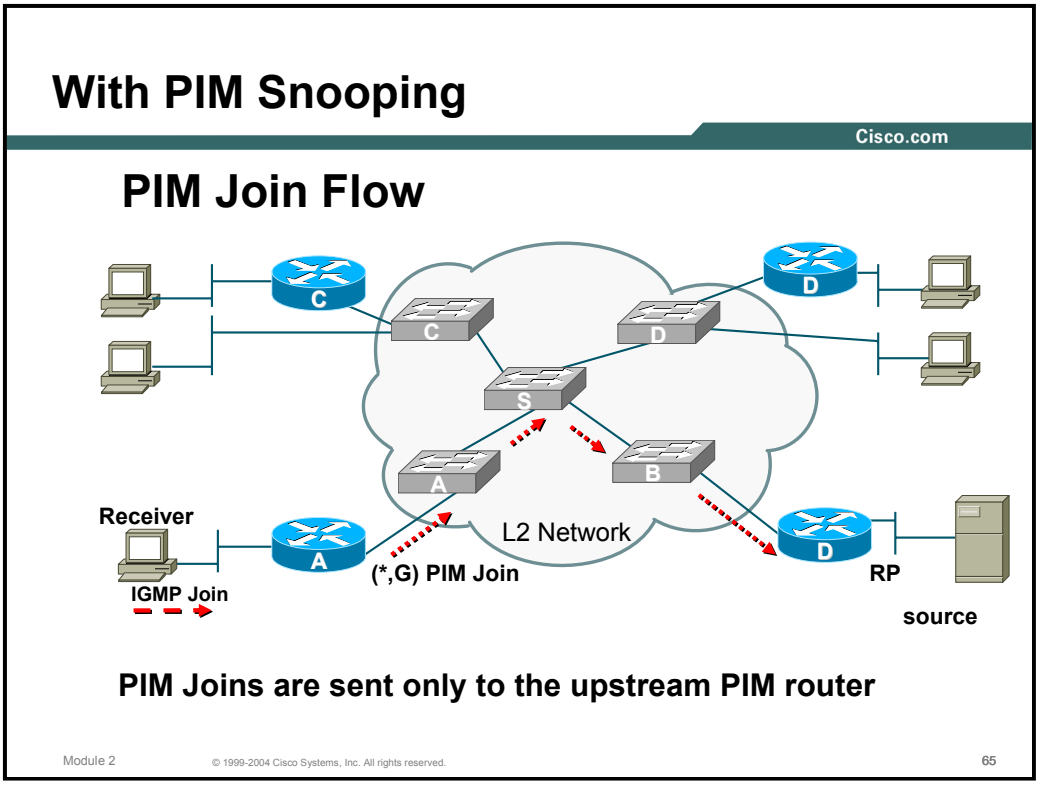
Traffic Flow



Traffic is unnecessarily flooded to Routers C and D also

Module 2 © 1999-2004 Cisco Systems, Inc. All rights reserved. 64

- **Without PIM Snooping**
  - The data traffic, show in solid lines, get flooded on all multicast router ports by switch-S. Routers C and D are hit with unnecessary traffic leading to wastage of bandwidth and cpu cycles on routers C and D.

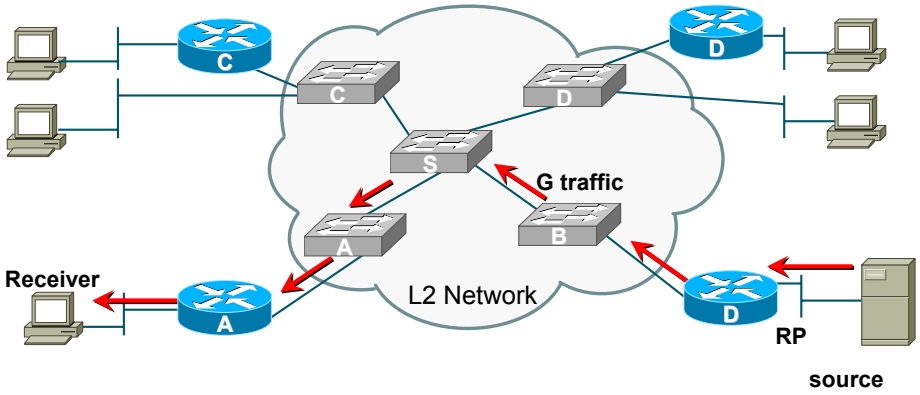


- **With PIM Snooping**
  - Switch C knows from the payload of the join-prune message who the upstream router is. Here it is Router B. From the hello message (of B) it knows the port behind which router B is located. Switch-S forwards the join-prune only out of that port and hence we avoid join-suppression by routers C and D.

With PIM Snooping

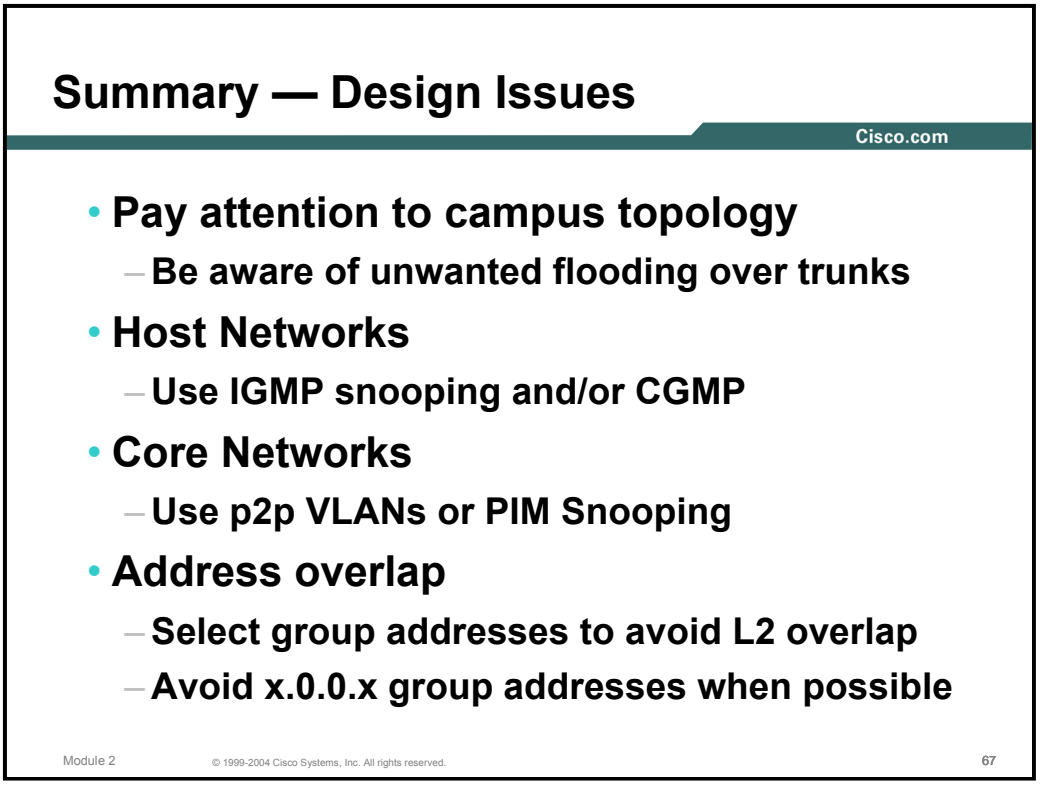
Cisco.com

Traffic Flow



Traffic from router B is sent only to Router A

- **With PIM Snooping**
  - Traffic is not sent to Routers C and D with PIM snooping running saving precious bandwidth of the links. Routers C and D are saved from the unnecessary processing of the data packets.



- **Design Issues — Summary**
  - Topology
    - Watch your campus topology when designing you network for multicast and beware of the possibility of unwanted traffic over inter-switch trunks.
  - Use IGMP Snooping and/or CGMP
    - This will help constrain multicast traffic to hosts that have requested it.
    - Keep in mind that not all situations are covered by IGMP Snooping or CGMP and that traffic is not always constrained under certain conditions.
  - 224.0.0.x Flooding
    - Watch out for vendor switches that do not flood multicast traffic in these ranges. Misbehaved or misconfigured hosts can cause this critical traffic to be shutoff in switches that do not flood this traffic.
  - Address Overlap
    - Try to select multicast addresses so that different applications don't map their multicast streams into the same L2 MAC address due to the 32:1 overlap of IP group addresses at Layer 2.
    - Avoid *.0.0.* and *.128.0.* multicast addresses when possible as these ranges are flooded by Cisco switches.