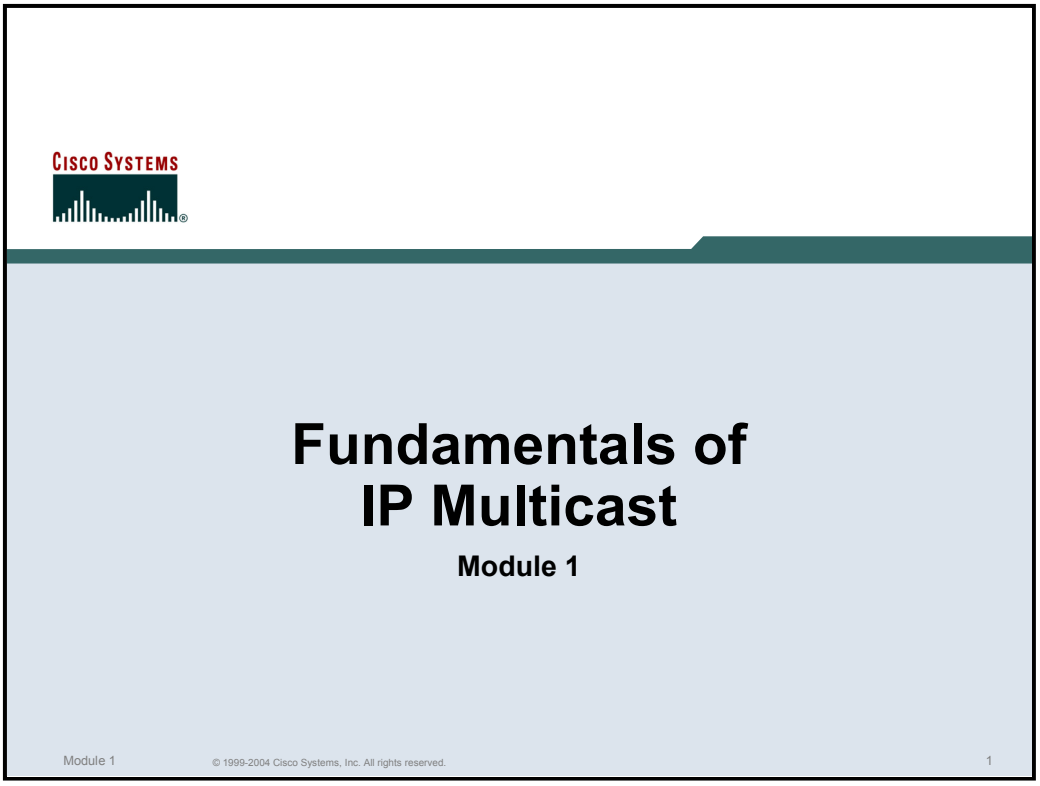
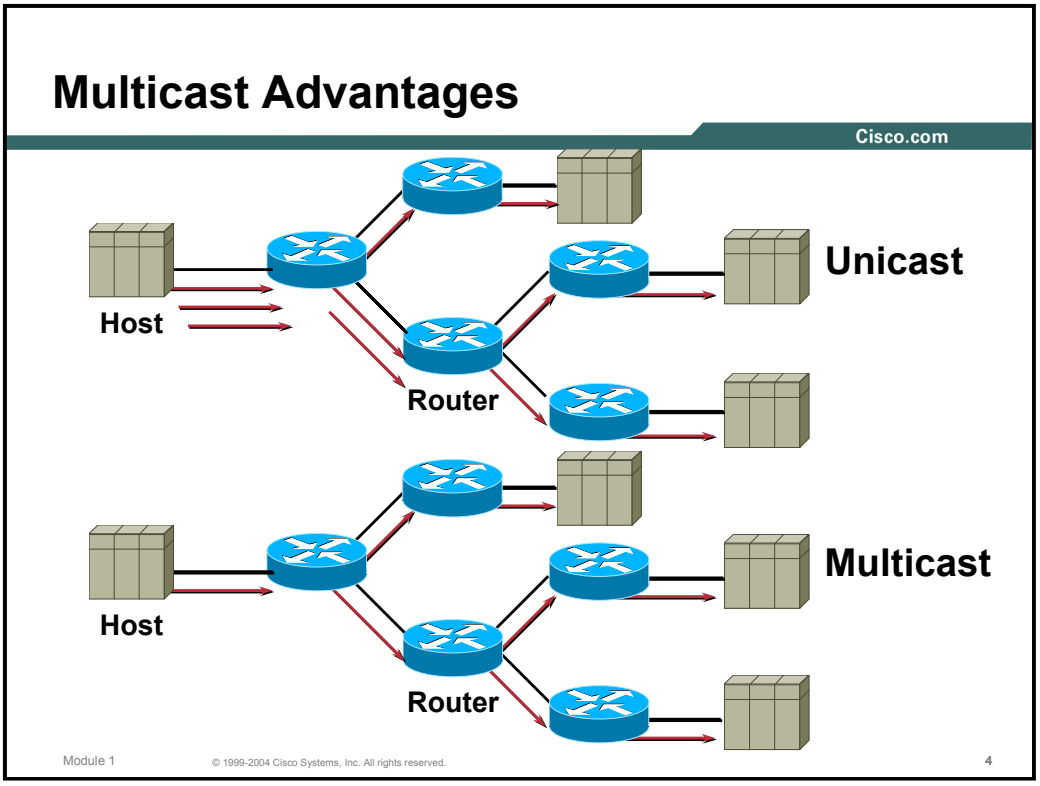


# Agenda

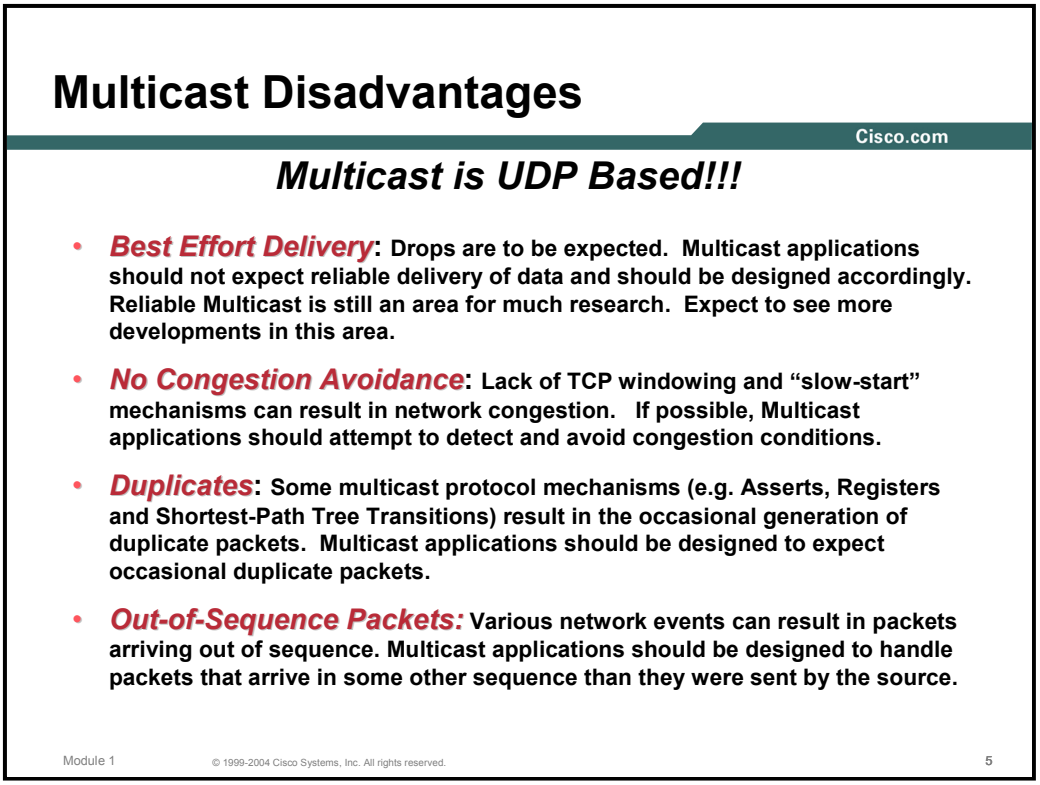

Cisco.com

- Why Multicast
- Multicast Applications
- Multicast Service Model
- Multicast Distribution Trees
- Multicast Forwarding
- Multicast Protocol Basics

# Why Multicast

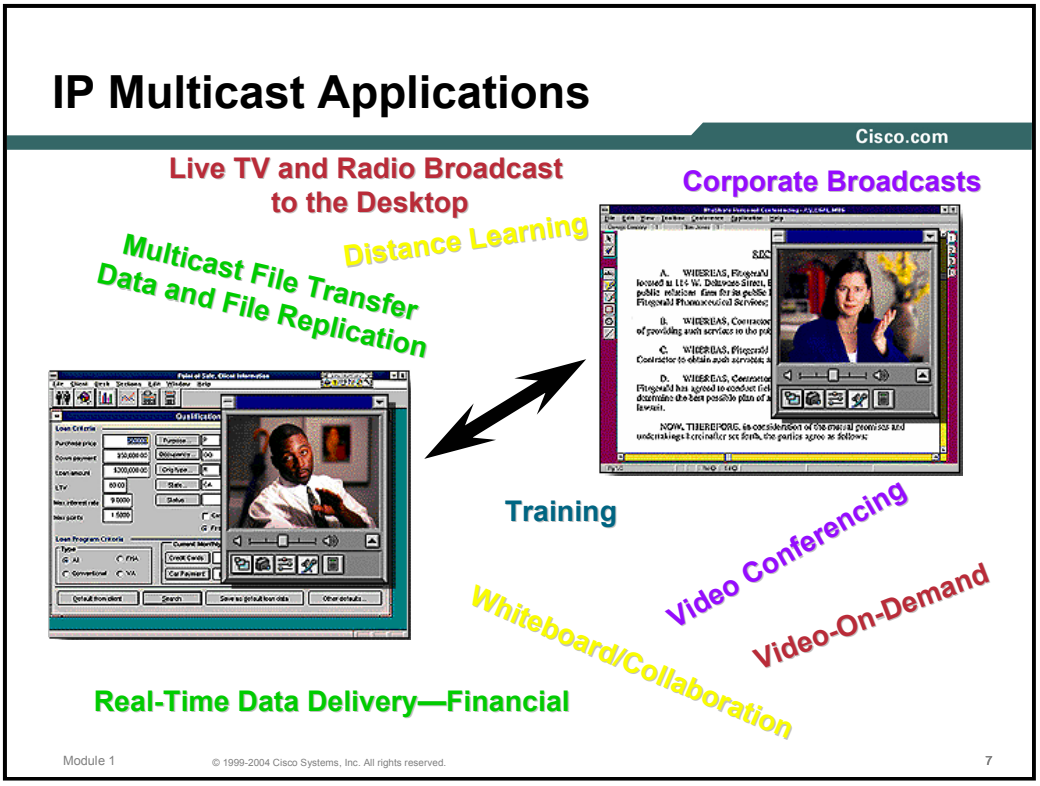



- **Unicast transmission sends multiple copies of data, one copy for each receiver**
  - Ex: host transmits 3 copies of data and network forwards each to 3 separate receivers
  - Ex: host can only send to one receiver at a time
- **Multicast transmission sends a single copy of data to multiple receivers**
  - Ex: host transmits 1 copy of data and network replicates at last possible hop for each receiver, each packet exists only one time on any given network
  - Ex: host can send to multiple receivers simultaneously
- **Multicast transmission affords many advantages over unicast transmission in a one-to-many or many-to-many environment**
  - Enhanced Efficiency: available network bandwidth is utilized more efficiently since multiple streams of data are replaced with a single transmission
  - Optimized Performance: less copies of data require forwarding and processing
  - Distributed Applications: multipoint applications will not be possible as demand and usage grows because unicast transmission will not scale
    - Ex: traffic level and clients increase at a 1:1 rate with unicast transmission
    - Ex: traffic level and clients do not increase at a greatly reduced rate with multicast transmission

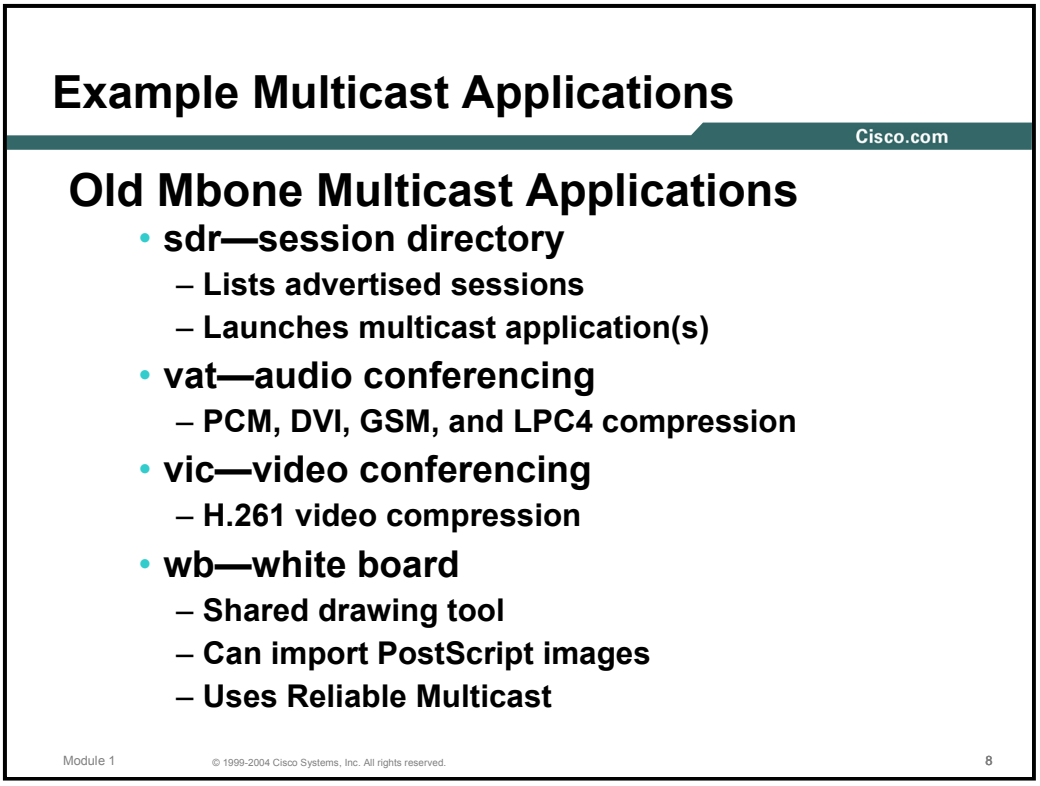



- **Multicast Disadvantages**
  - Most Multicast Applications are UDP based. This results in some undesirable side-effects when compared to similar unicast, TCP applications.
  - Best Effort Delivery results in occasional packet drops. Many multicast applications that operate in real-time (e.g. Video, Audio) can be impacted by these losses. Also, requesting retransmission of the lost data at the application layer in these sort of real-time applications is not feasible.
    - Heavy drops on Voice applications result in jerky, missed speech patterns that can make the content unintelligable when the drop rate gets high enough.
    - Moderate to Heavy drops in Video is sometimes better tolerated by the human eye and appear as unusual "artifacts" on the picture. However, some compression algorithms can be severely impacted by even low drop rates; causing the picture to become jerky or freeze for several seconds while the decompression algorithm recovers.
  - No Congestion Control can result in overall Network Degradation as the popularity of UDP based Multicast applications grow.
  - Duplicate packets can occasionally be generated as multicast network topologies change.
    - Applications should expect occasional duplicate packets to arrive and should be designed accordingly.

# Multicast Applications



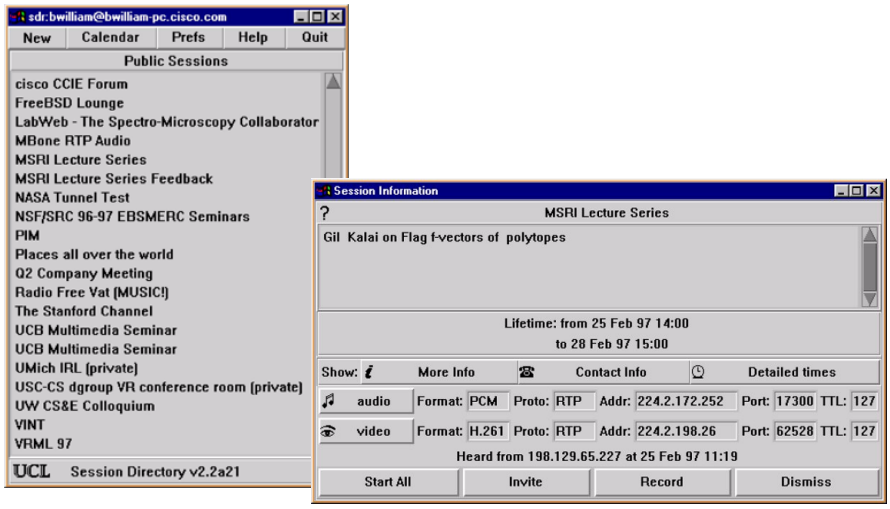

- **Many new multipoint applications are emerging as demand for them grows**
  - Ex: Real-time applications include live broadcasts, financial data delivery, whiteboard collaboration, and video conferencing
  - Ex: Non-real-time applications include file transfer, data and file replication, and video-on-demand
  - Note also that the latest version of Novell Netware uses Ipmc for file and print service announcements….see:
    - http//developer.novell.com/research/appnotes/1999/march/02/index.htm




- **Several MBONE multicast applications exist**
  - Ex: Session Directory is a tool that allows participants to view advertised multicast sessions and launch appropriate multicast applications to join an existing session
  - Ex: Audio Conferencing allows multiple participants to share audio interactively
  - Ex: Video Conferencing allows multiple participants to share video and audio interactively
  - Ex: White Boarding allows multiple participants to collaborate interactively in a text and graphical environment

## sdr—Session Directory

Cisco.com



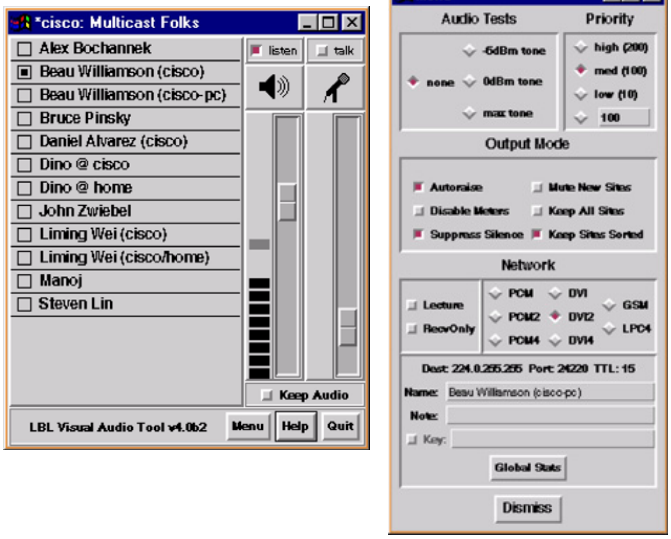





- **SDR - Session Directory (revised)**
  - The SDR tool allows Multimedia multicast sessions to be created by other users in the network. These multimedia sessions (video, audio, etc.) are announced by the SDR application via well-known multicast groups.
  - The window on the left shows an example of the SDR application in action. Each line is a multimedia session that has been created by some user in the network and is being announced (via multicast) by the creator's SDR application.
  - By clicking on one of these sessions, the window on the right is brought up. This window displays various information about the multimedia session including:
    - General session information
    - Session schedule
    - Media type (in this case audio and video)
    - Media format
    - Multicast group and port numbers
  - Using the window on the right, one can have SDR launch the appropriate multicast application(s) to join the session.

## vat—Audio Conferencing

Cisco.com



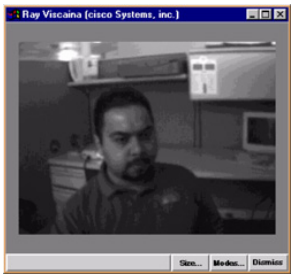





- **Vat - Audio Conferencing Tool**
  - This is an example of the vat audio conferencing tool. The window on the left is the main window for the session. It contains a speaker gain slider widget and an output VU bar-graph meter along with a microphone gain slider widget and VU meter. When one wishes to address the conference, one usually presses the right mouse button on the workstation.
  - The window on the right is a menu that can be brought up by pressing the “Menu” button on the main window. This menu allows various parameters about the session to be adjusted including encoding format.
  - Notice that there are several members of this session listed in the main window even though only the second person is talking. (Indicated by the blackened square next to the name.) This points out that all members of the session are multicast sources even though they may never speak and only listed to the session. This is because vat uses the RTP/RTCP model to transport Real-Time audio data. In this model, all members of the session multicast member information and reception statistics to the entire group in an RTCP “back-channel” .
  - Most all multimedia multicast applications use the RTP/RTCP model including:
    - vat (and its cousin application rat)
    - vic
    - wb - (Whiteboard)
    - IP/TV

## vic—Video Conferencing

Cisco.com





Module 1

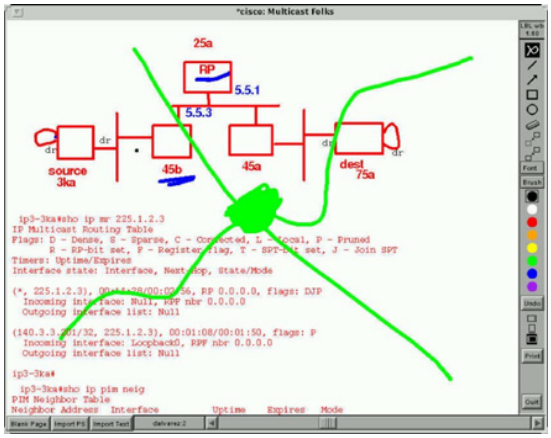


11

- **vic - Video Conferencing Tool**
  - This is an example of the “vic” video conferencing tool. The window on the right is the main window for the video conferencing session. Notice that multiple video streams are being received, each with its own “thumbnail” image.
  - The window on the left is a larger version of the thumbnail image from the main window.

## wb—White Board

Cisco.com

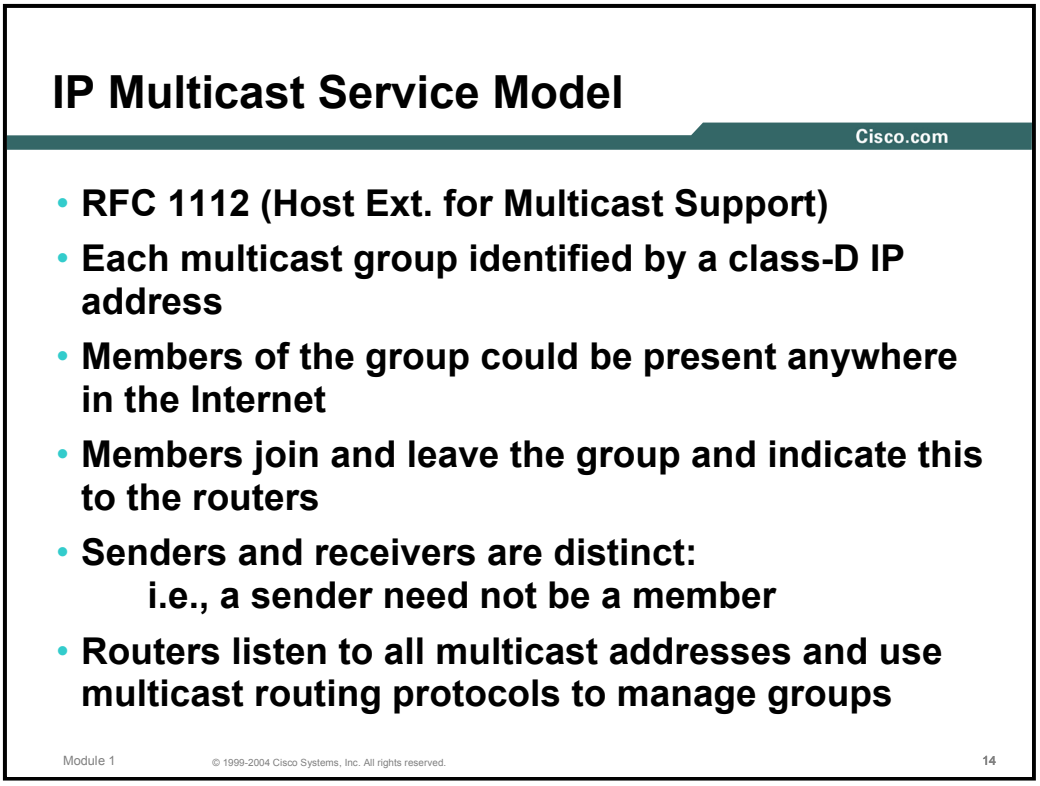




Module 1  12

- **wb - Whiteboard**
  - Just as its name implies, this is a form of electronic Whiteboard that can be shared by members of the multicast group.
- **“White Board” uses a form of Reliable Multicast**
  - Reliable Multicasting is necessary to insure no loss of critical graphic information occurs.
  - Most multimedia multicast applications simply use UDP, “Best Effort” datagram delivery mechanisms because of the time critical nature of the media. However, “wb” needs a reliable method to distribute the graphic images drawn on the electronic “Whiteboard”.

# Multicast Service Model




- **RFC 1112 is the Internet Group Management Protocol (IGMP)**
  - Allows hosts to join a group that receives multicast packets
  - Allows users to dynamically register (join/leave multicast groups) based on applications they execute
  - Uses IP datagrams to transmit data
- **Addressing**
  - Class D IP addresses (224-239) are dynamically allocated
  - Multicast IP addresses represent receiver groups, not individual receivers
- **Group Membership**
  - Receivers can be densely or sparsely distributed throughout the Internet
  - Receivers can dynamically join/leave a multicast session at any time using IGMP to manage group membership within the routers
  - Senders are not necessarily included in the multicast group they are sending to
  - Many applications have the characteristic of receivers also becoming senders eg RTCP streams from IP/TV clients and Tibco RV
- **Multicast Routing**
  - Group distribution requires packet distribution trees to efficiently forward data to multiple receivers
  - Multicast routing protocols effectively direct multicast traffic along network paths
    - Multicast Extension to Open Shortest Path First (MOSPF - 1584)
    - Core Based Tree (CBT)

## IP Multicast Addressing

Cisco.com

- **Group Addresses (224.0.0.0 – 239.255.255.255)**
  - **Class D address**
  - **High-order 3 bits are set (224.0.0.0)**
- **Link-Local addresses designated by IANA**
  - **Reserved use: 224.0.0.0 through 224.0.0.255**
    - **224.0.0.1—all multicast systems on subnet**
    - **224.0.0.2—all routers on subnet**
    - **See "http://www.iana.org/assignments/multicast-addresses"**
- **Transient addresses, assigned and reclaimed dynamically**
  - **Global scope: 224.0.1.0-238.255.255.255**
  - **Limited Scope: 239.0.0.0-239.255.255.255**
  - **Site-local scope: 239.253.0.0/16**
  - **Organization-local scope: 239.192.0.0/14**



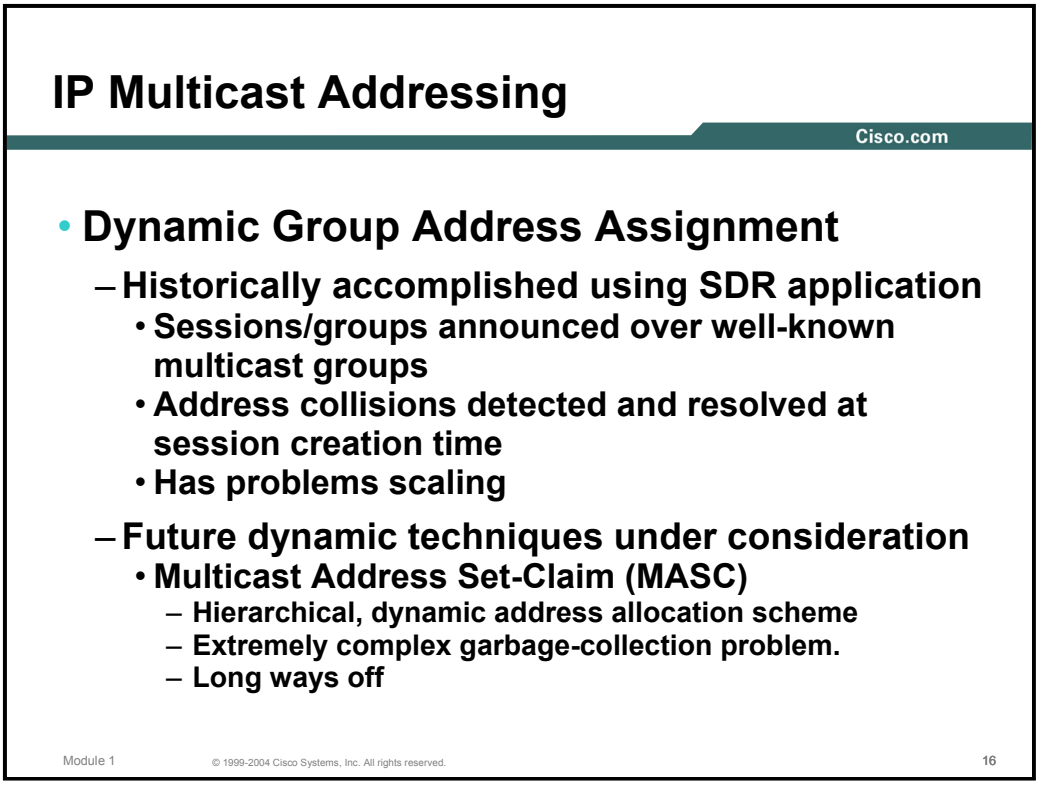


- **IP Addresses use the Class D address space**
  - Class D addresses are denoted by the high 4 bits set to 1110.
- **Local Scope Addresses**
  - Addresses 224.0.0.0 through 224.0.0.255
  - Reserved by IANA for network protocol use

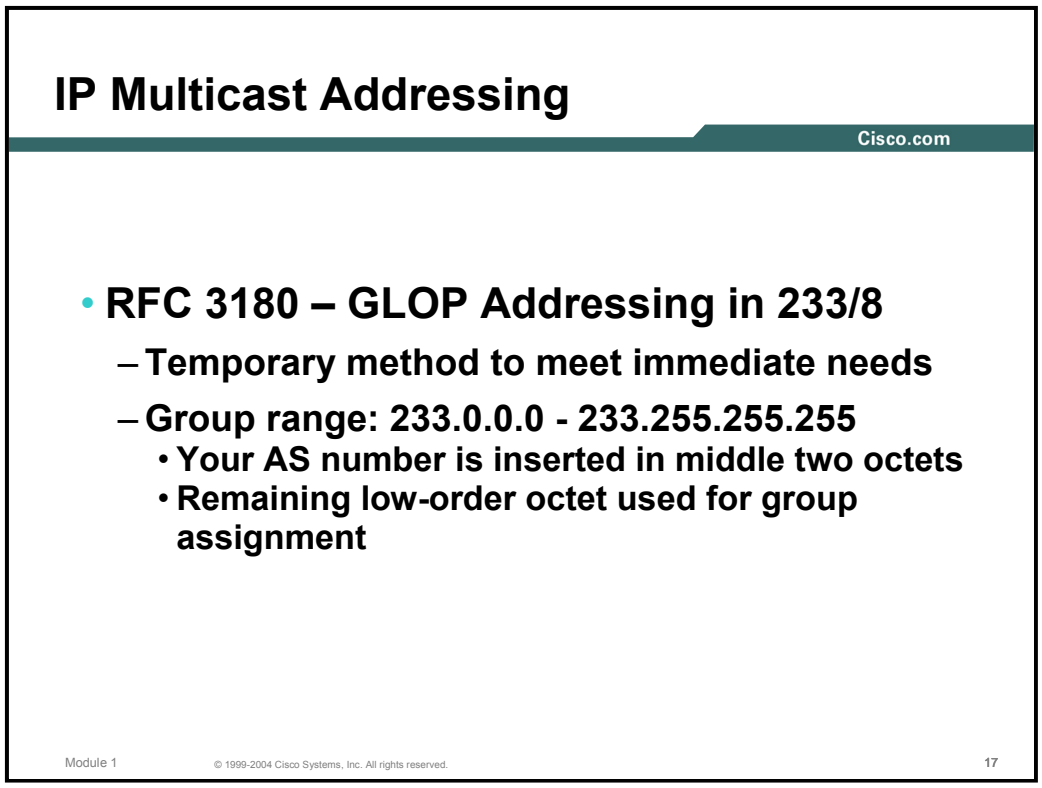
    Examples:

    | | |
    |---|---|
    | 224.0.0.1 | All Hosts |
    | 224.0.0.2 | All Multicast Routers |
    | 224.0.0.3 | All DVMRP Routers |
    | 224.0.0.5 | All OSPF Routers |
    | 224.0.0.6 | All OSPF DR |

  - Multicasts in this range are never forwarded off the local network regardless of TTL
  - Multicasts in this range are usually sent 'link local' with TLL = 1.
- **Global Scope Addresses**
  - Addresses 224.0.1.0 through 238.255.255.255
  - Allocated dynamically throughout the Internet
- **Administratively Scoped Addresses**
  - Addresses 239.0.0.0 through 239.255.255.255
  - Reserved for use inside of private Domains.

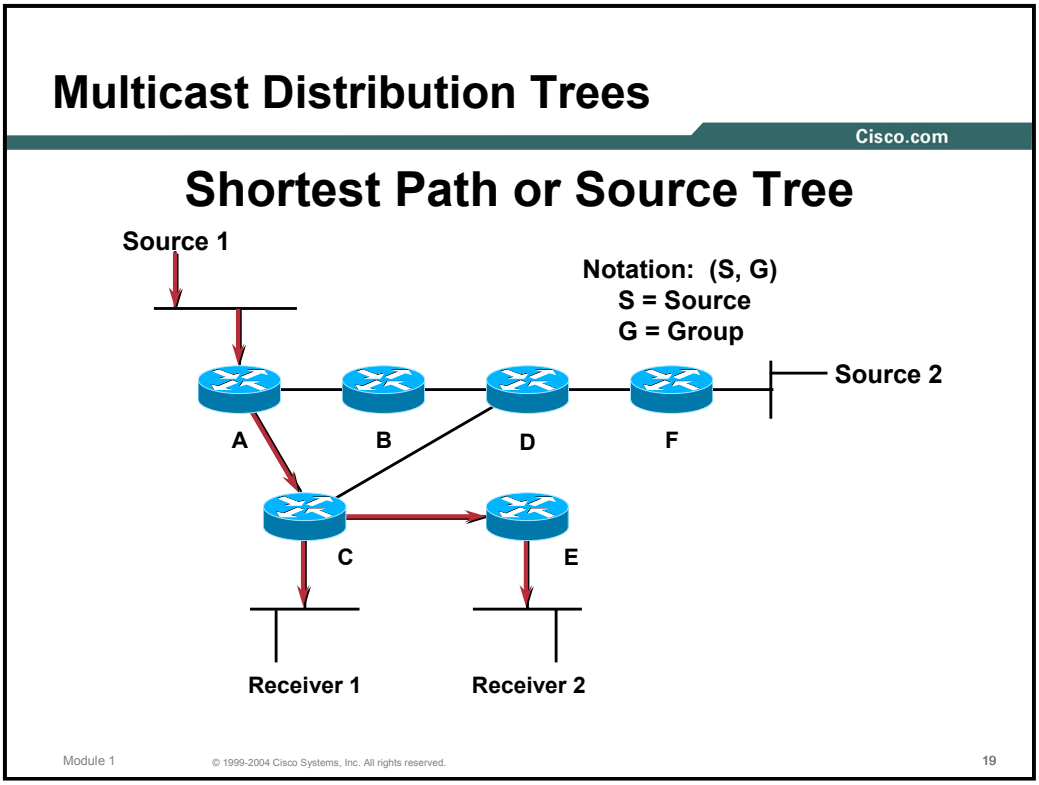



- **Dynamic Group Address Assignment**
  - SDR
    - This was typically accomplished using the SDR application which would detect collisions in IP multicast group address assignment at the time new sessions were being created and pick an unused address.
    - While it was sufficient for use on the old MBone when the total number of multicast sessions in the Internet was quite low, SDR has severe scaling problems that preclude it from continuing to be used as the number of sessions increase.
  - Multicast Address Set-Claim (MASC)
    - MASC is new proposal for a dynamic multicast address allocation that is being developed by the “malloc” Working Group of the IETF.
    - This new proposal will provide for dynamic allocation of the global IP Multicast address space in a hierarchical manner.
    - In this proposal, domains “lease” IP multicast group address space from their parent domain. These leases are good for only a set period. It is possible that the parent domain may grant a completely different range at lease renewal time due to the need to reclaim address space for use elsewhere in the Internet.
    - As one can imagine, this is a very non-trivial mechanism and is a long ways from actual implementation.

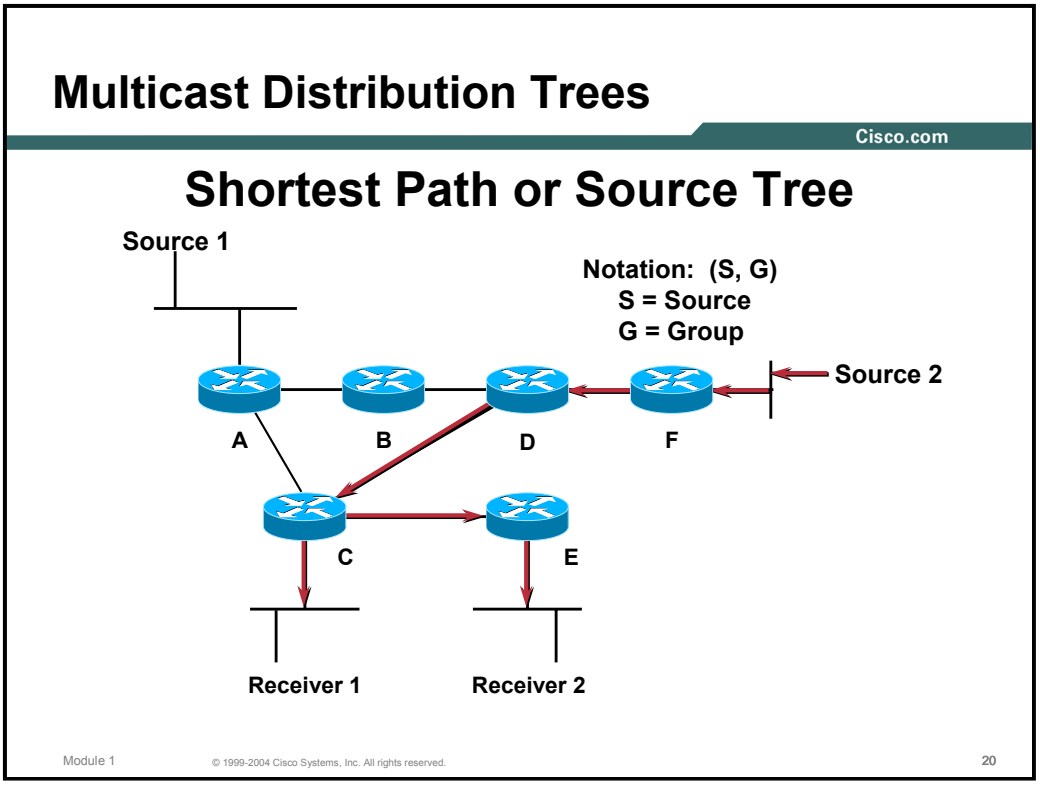



- **Static Group Address Assignment**
  - Until MASC has been fully specified and deployed, many content providers in the Internet require “something” to get going in terms of address allocation. This is being addressed with a temporary method of static multicast address allocation.
  - This special allocation method is defined in:
    - RFC 3180 “ GLOP Addressing in 233/8”
  - The basic concept behind this methodology is as follows:
    - Use the 233/8 address space for static address allocation
    - The middle two octets of the group address would contain your AS number
    - The final octet is available for group assignment.

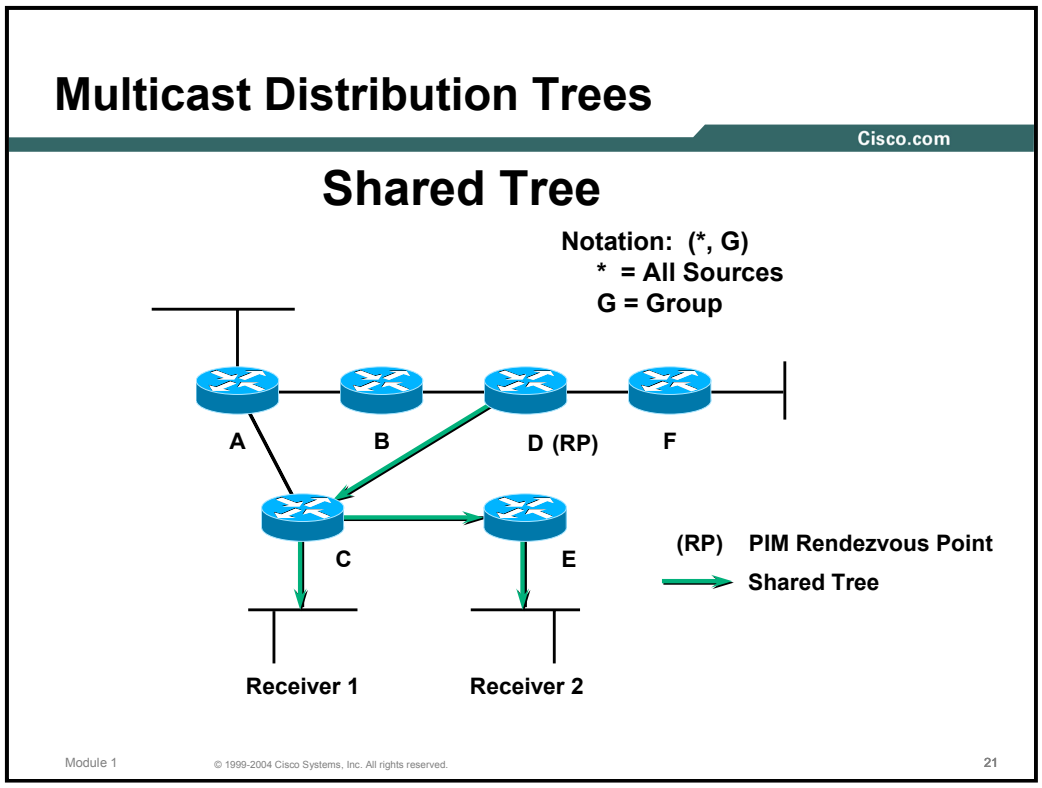
# Multicast Distribution Trees



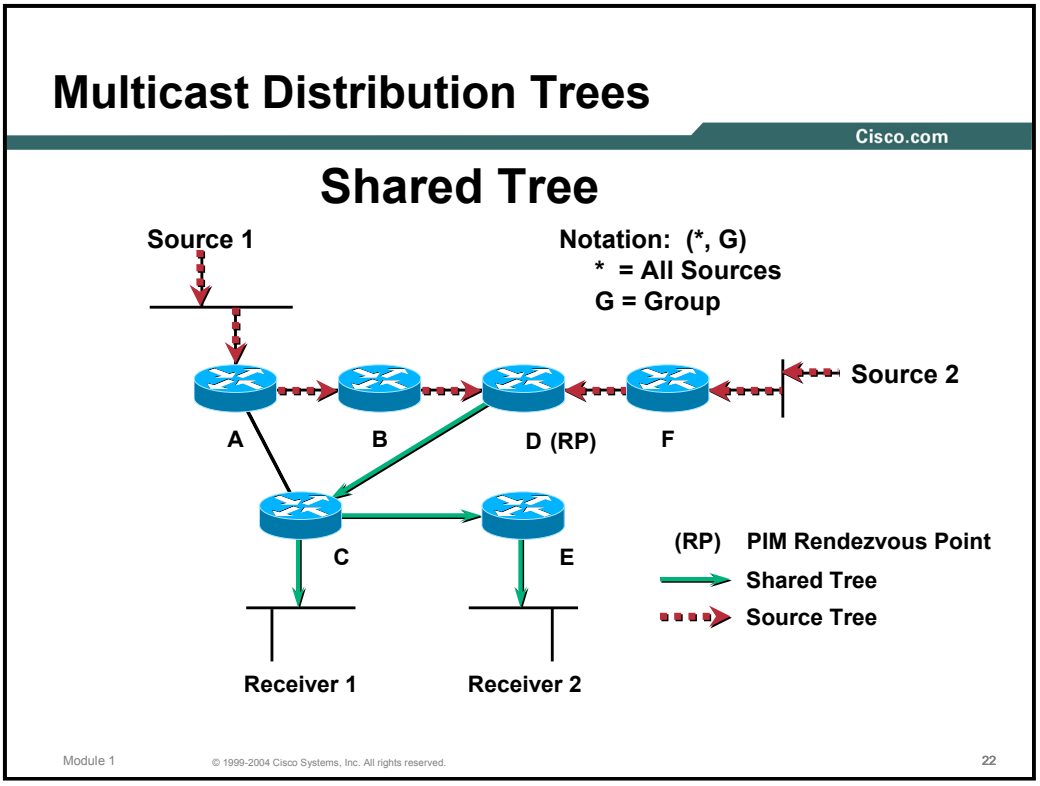

- **Shortest Path Trees — aka Source Trees**
  - A Shortest path or source distribution tree is a minimal spanning tree with the lowest cost from the source to all leaves of the tree.
  - We forward packets on the Shortest Path Tree according to both the Source Address that the packets originated from and the Group address G that the packets are addressed to. For this reason we refer to the forwarding state on the SPT by the notation (S,G) (pronounced “S comma G”).
  - where:
    - “S” is the IP address of the source.
    - “G” is the multicast group address
  - Example 1:
    - The shortest path between Source 1 and Receiver 1 is via Routers A and C, and shortest path to Receiver 2 is one additional hop via Router E.



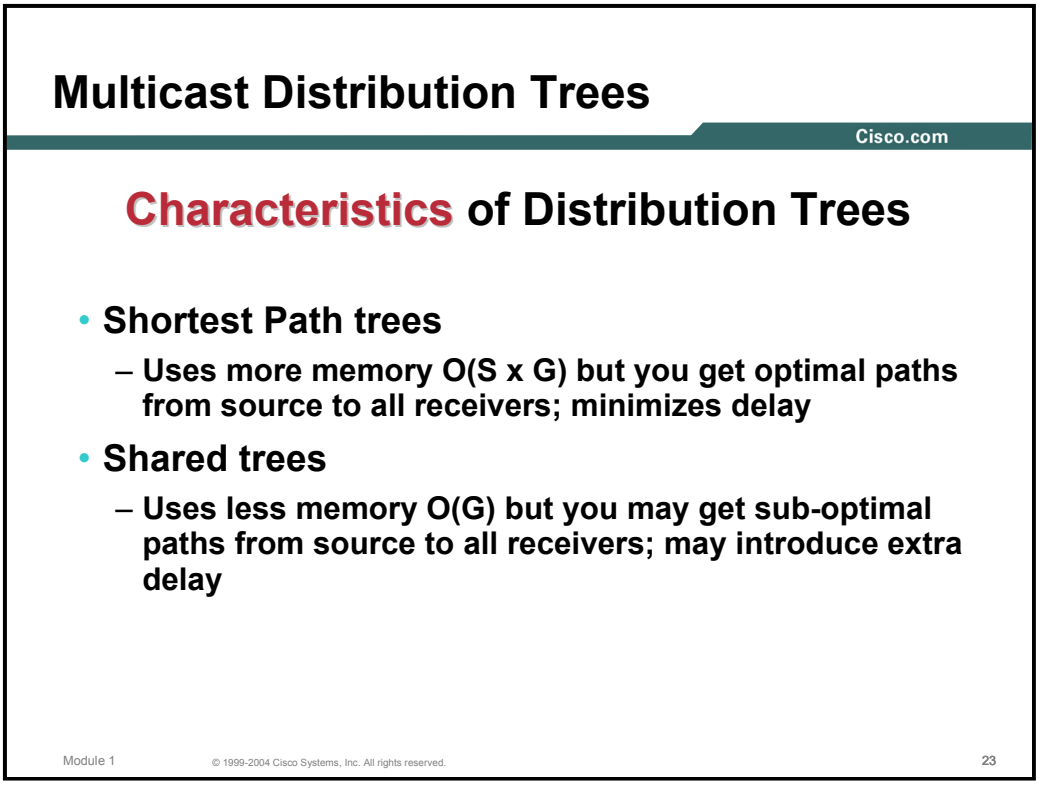

- **Shortest Path Trees — aka Source Trees (cont.)**
  - Every SPT is routed at the source. This means that for every source sending to a group, there is a corresponding SPT.
  - Example 2:
    - The shortest path between Source 2 and Receiver 1 is via Routers D, F and C, and shortest path to Receiver 2 is one additional hop via Router E.




- **Shared Distribution Trees**
  - Shared distribution tree whose root is a shared point in the network down which multicast data flows to reach the receivers in the network. In PIM-SM, this shared point is called the Rendezvous Point (RP).
  - Multicast traffic is forwarded down the Shared Tree according to just the Group address G that the packets are addressed to, regardless of source address. For this reason we refer to the forwarding state on the shared tree by the notation (*,G) (pronounced “star comma G”)
  - where:
    - “*” means any source
    - “G” is the group address

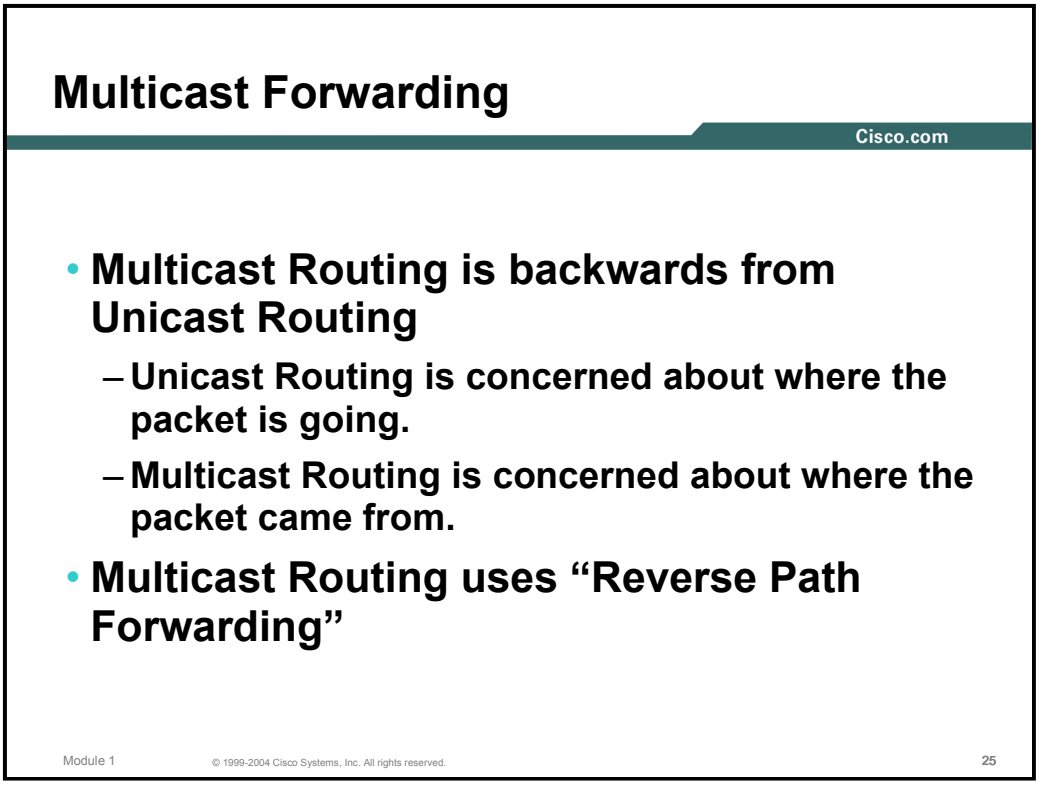



- **Shared Distribution Trees (cont.)**
  - Before traffic can be sent down the Shared Tree it must somehow be sent to the Root of the Tree.
  - In classic PIM-SM, this is accomplished by the RP joining the Shortest Path Tree back to each source so that the traffic can flow to the RP and from there down the shared tree. In order to trigger the RP to take this action, it must somehow be notified when a source goes active in the network.
    - In PIM-SM, this is accomplished by first-hop routers (i.e. the router directly connected to an active source) sending a special Register message to the RP to inform it of the active source.
  - In the example above, the RP has been informed of Sources 1 and 2 being active and has subsequently joined the SPT to these sources.

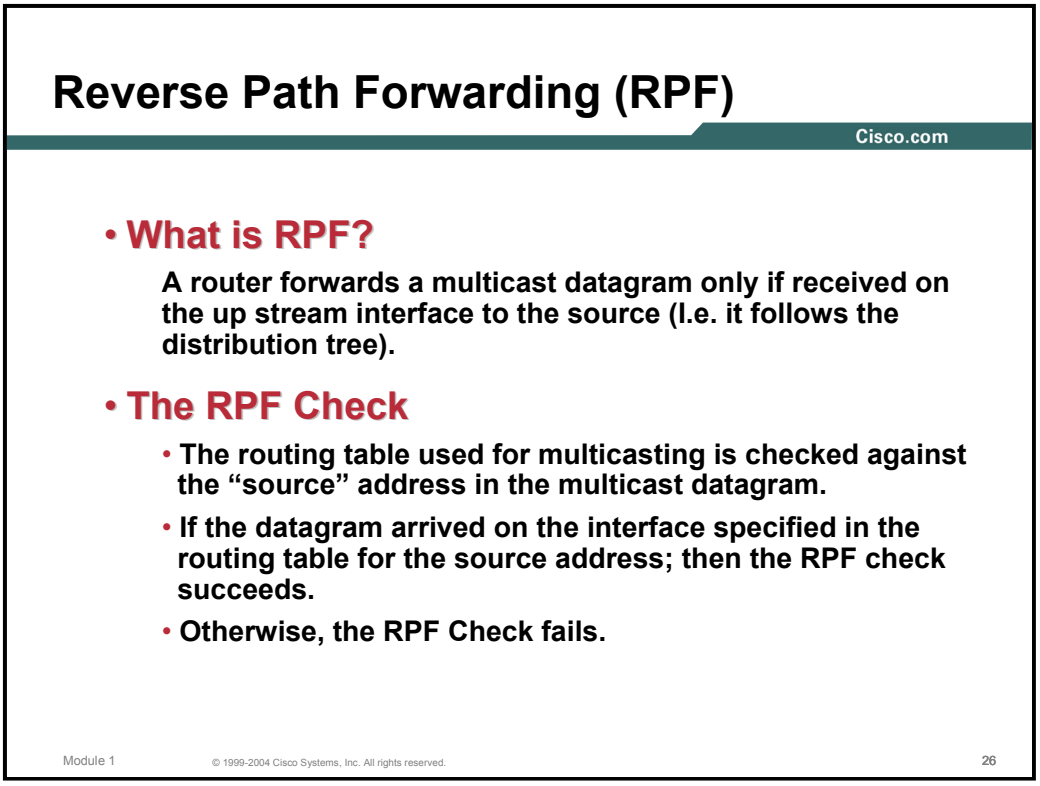



- **Source or Shortest Path Tree Characteristics**
  - Provides optimal path (shortest distance and minimized delay) from source to all receivers, but requires more memory to maintain
- **Shared Tree Characteristics**
  - Provides sub-optimal path (may not be shortest distance and may introduce extra delay) from source to all receivers, but requires less memory to maintain

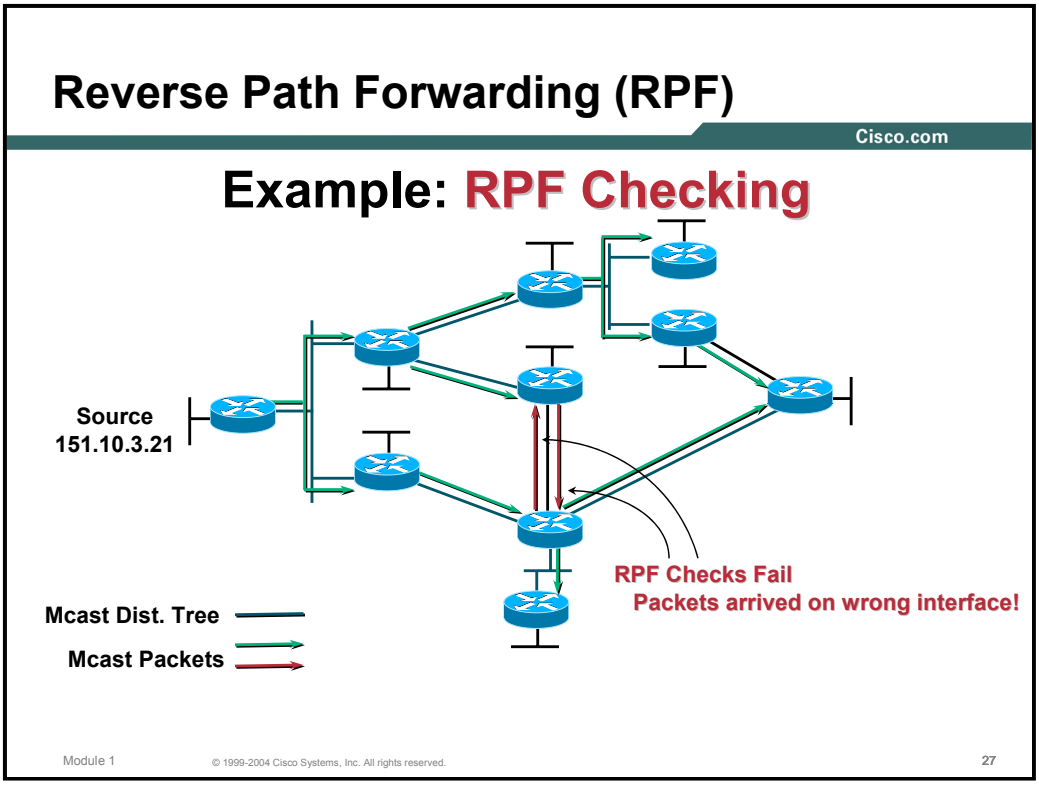
# Multicast Forwarding



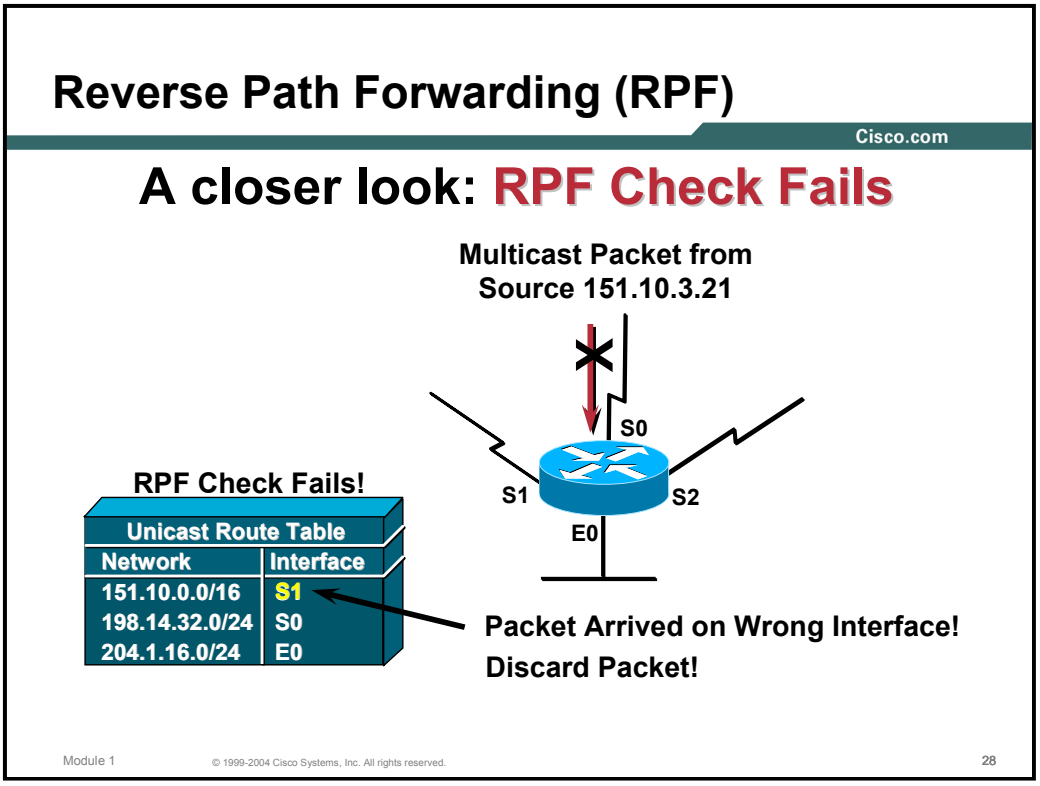

- **Multicast Forwarding**
  - Routers must know packet origin, rather than destination (opposite of unicast)
    - ... origination IP address denotes known source
    - ... destination IP address denotes unknown group of receivers
  - Multicast routing utilizes Reverse Path Forwarding (RPF)
    - ... Broadcast: floods packets out all interfaces except incoming from source; initially assuming every host on network is part of multicast group
    - ... Prune: eliminates tree branches without multicast group members; cuts off transmission to LANs without interested receivers
    - ... Selective Forwarding: requires its own integrated unicast routing protocol



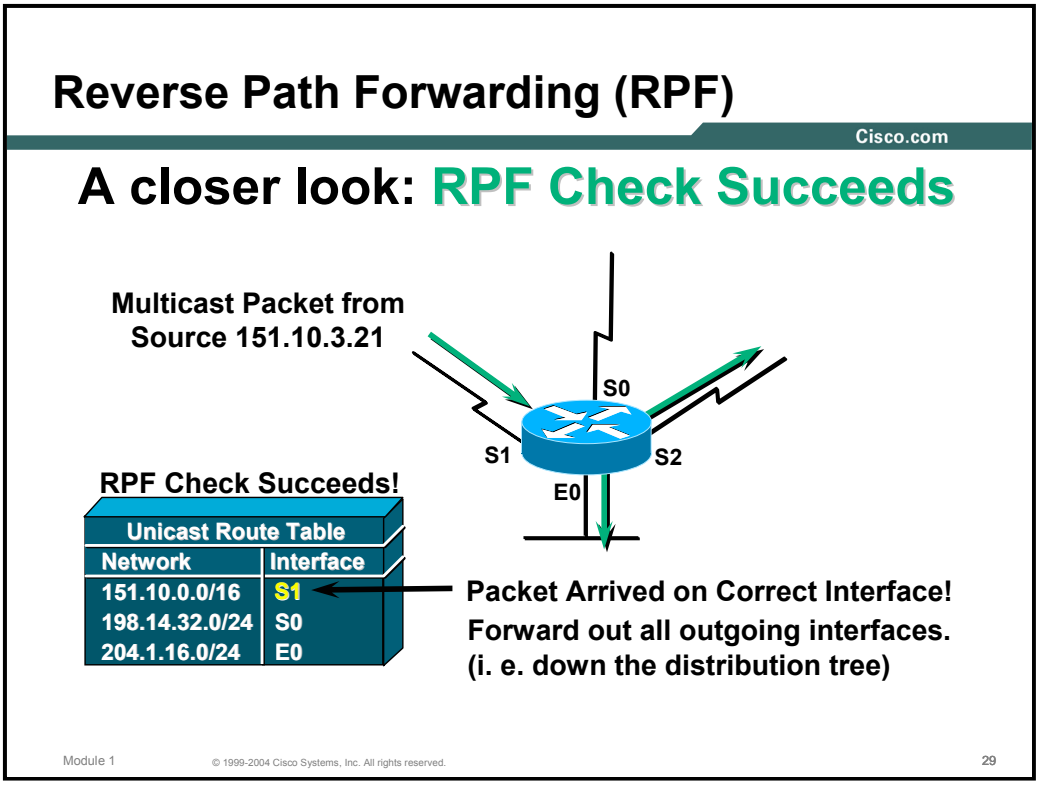

- **Reverse Path Forwarding**
  - Routers forward multicast datagrams received from incoming interface on distribution tree leading to source
  - Routers check the source IP address against their multicast routing tables (RPF check); ensure that the multicast datagram was received on the specified incoming interface
  - Note that changes in the unicast topology will not necessarily immediately reflect a change in RPF…this depends on how frequently the RPF check is performed on an Ipmc stream - every 5 seconds is current Cisco default.



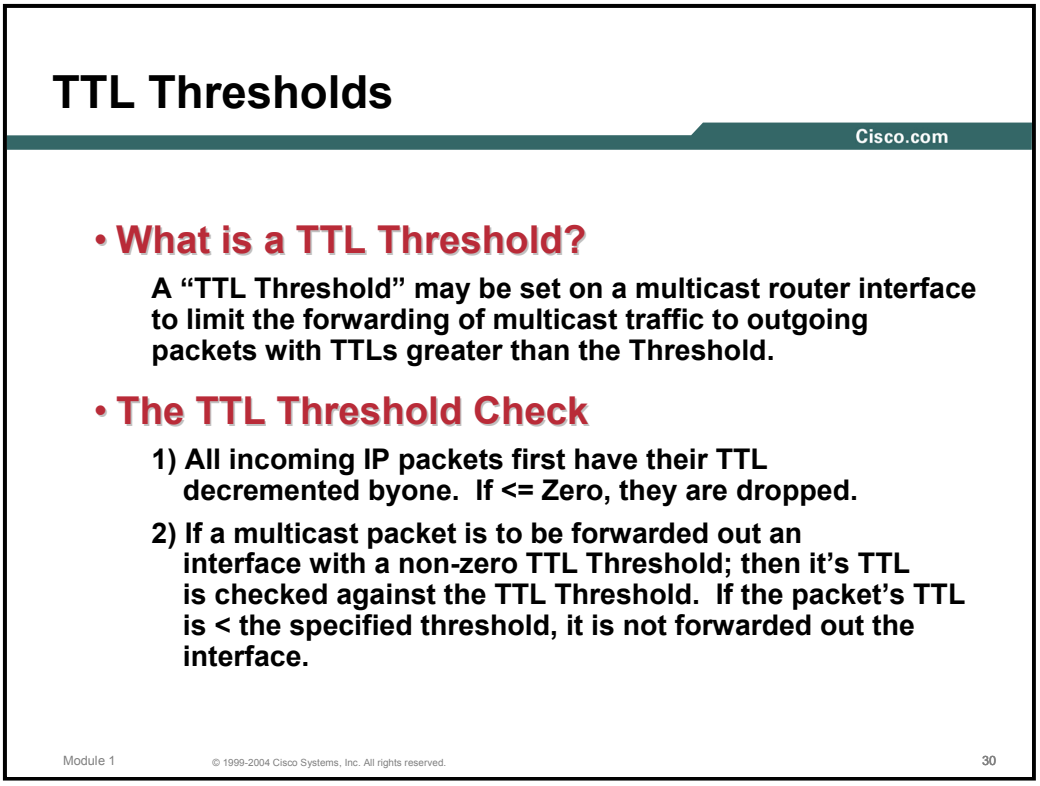

- **Multicast Forwarding: RPF Checking**
  - Source floods network with multicast data
  - Each router has a designated incoming interface (RPF interface) on which multicast data can be received from a given source
  - Each router receives multicast data on one or more interfaces, but performs RPF check to prevent duplicate forwarding
- **Example: Router receives multicast data on two interfaces**
  - 1) performs RPF Check on multicast data received on interface E0; RPF Check succeeds because data was received on specified incoming interface from source 151.10.3.21; data forwarded through all outgoing interfaces on the multicast distribution tree
  - 2) performs RPF Check on multicast data received on interface E1; RPF Check fails because data was not received on specified incoming interface from source 151.10.3.21; data silently dropped



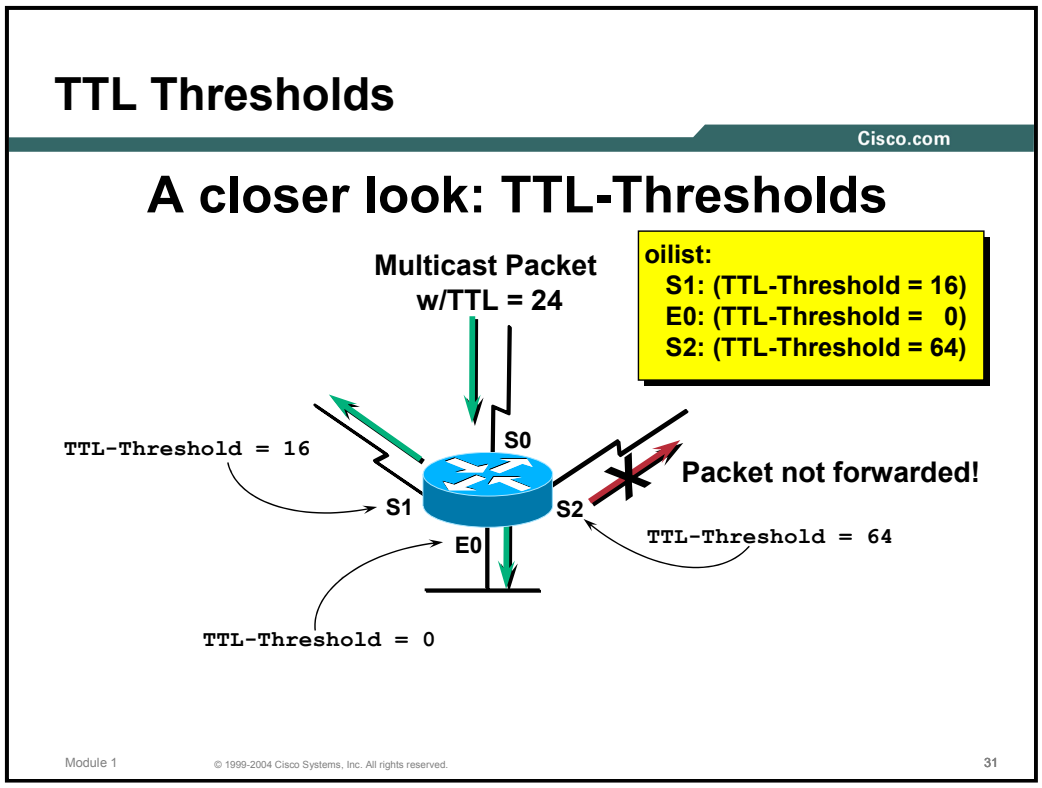

- **Multicast Forwarding: RPF Check Fails**
  - Ex: Router can only accept multicast data from Source 151.10.3.21 on interface S1
    - ... multicast data is silently dropped because it arrived on an interface not specified in the RPF check (S0)




- **Multicast Forwarding: RPF Check Succeeds**
  - Ex: Router can only accept multicast data from Source 151.10.3.21 on interface S1
    - ... multicast data is forwarded out all outgoing on the distribution tree because it arrive on the incoming interface specified in the RPF check (S1)




- **TTL-Thresholds**
  - Non-Zero, Multicast, TTL-Thresholds may be set on any multicast capable interface.
  - IP multicast packets whose TTLs (after being decremented by one by normal router packet processing) are less than the TTL-Threshold on an outgoing interface, will be not be forwarded out that interface.
  - Zero Multicast TTL implies NO threshold has been set.
- **TTL-Threshold Application**
  - Frequently used to set up multicast boundaries to prevent unwanted multicast traffic from entering/exiting the network.




- **TTL-Threshold Example**
  - In the above example, the interfaces have been configured with the following TTL-Thresholds:

    S1: TTL-Threshold = 16
    E0: TTL-Threshold = 0 (none)
    S2: TTL-Threshold = 64
  - An incoming Multicast packet is received on interface S0 with a TTL of 24.
  - The TTL is decremented to 23 by the normal router IP packet processing.
  - The outgoing interface list for this Group contains interfaces S1, E0 & S2.
  - The TTL-Threshold check is performed on each outgoing interface as follows:

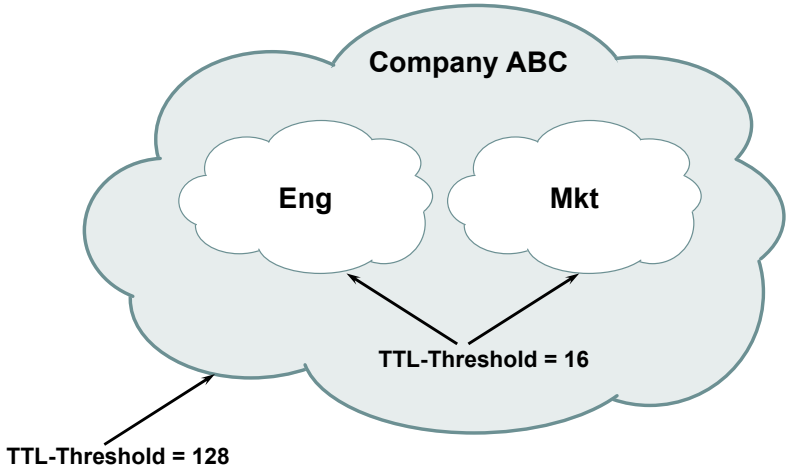
    S1: TTL (23) > TTL-Threshold (16). FORWARD

    E0: TTL (23) > TTL-Threshold (0). FORWARD

    S2: TTL (23) < TTL-Threshold (64). DROP

## TTL Threshold Boundaries

Cisco.com



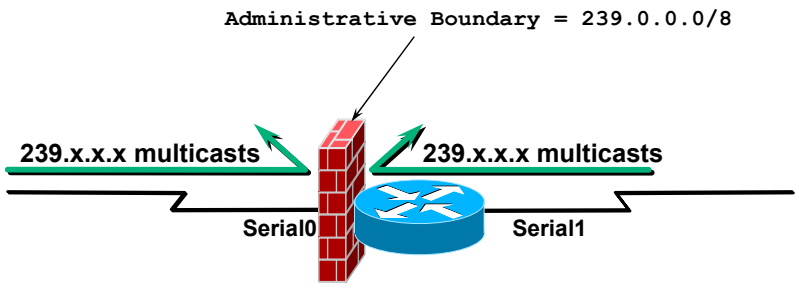

Module 1  32

- **TTL-Threshold Boundaries**
  - TTL-Thresholds may be used as boundaries around portions of a network to prevent the entry/exit of unwanted multicast traffic. This requires multicast applications to transmit their multicast traffic with an initial TTL value set so as to not cross the TTL-Threshold boundaries.

    In the example above, the Engineering or Marketing departments can prevent department related multicast traffic from leaving their network by using a TTL of 15 for their multicast sessions. Similarlly, Company ABC can prevent private multicast traffic from leaving their network by using a TTL of 127 for their multicast sessions.

## Administrative Boundaries

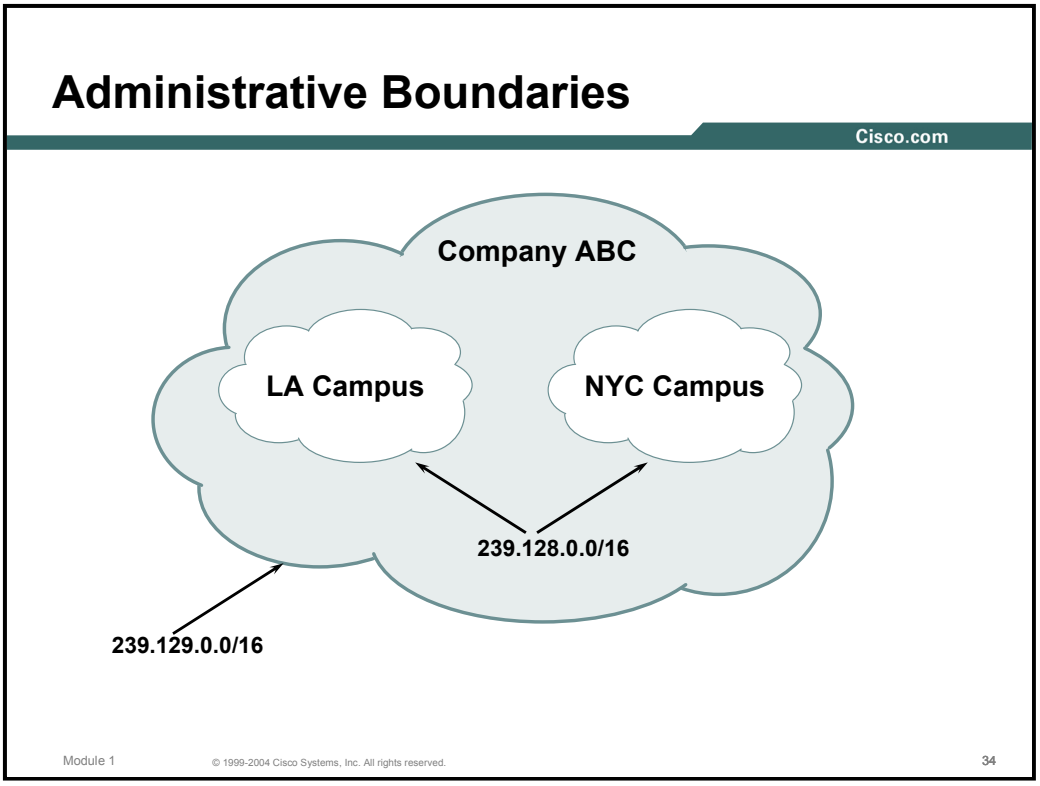
Cisco.com




• **Configured using the '`ip multicast boundary <acl>`' interface command**

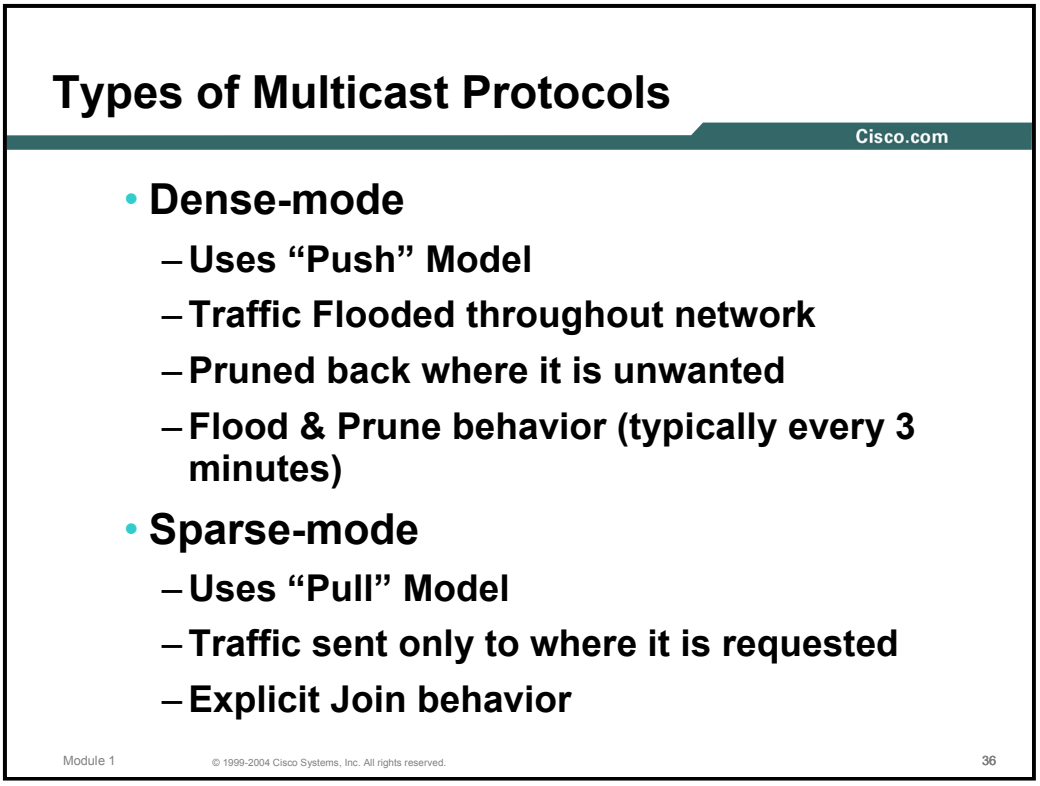
Module 1 © 1999-2004 Cisco Systems, Inc. All rights reserved. 33

- **Administrative Boundaries**
  - Administratively-scoped multicast address ranges may also be used as boundaries around portions of a network to prevent the entry/exit of unwanted multicast traffic. This requires multicast applications to transmit their multicast traffic with a group address that falls within the Administrative address range so that it will not cross the Administrative boundaries.
  - In the example above, the entire Administratively-Scoped address range, (239.0.0.0/8) is being blocked from entering or leaving the router via interface Serial0. This is often done at the border of a network where it connects to the Internet so that potentially sensitive company Administratively-Scoped multicast traffic can leave the network. (Nor can it enter the network from the outside.)
  - Administrative multicast boundaries can be configured in Cisco IOS by the use of the 'ip multicast boundary' interface command.

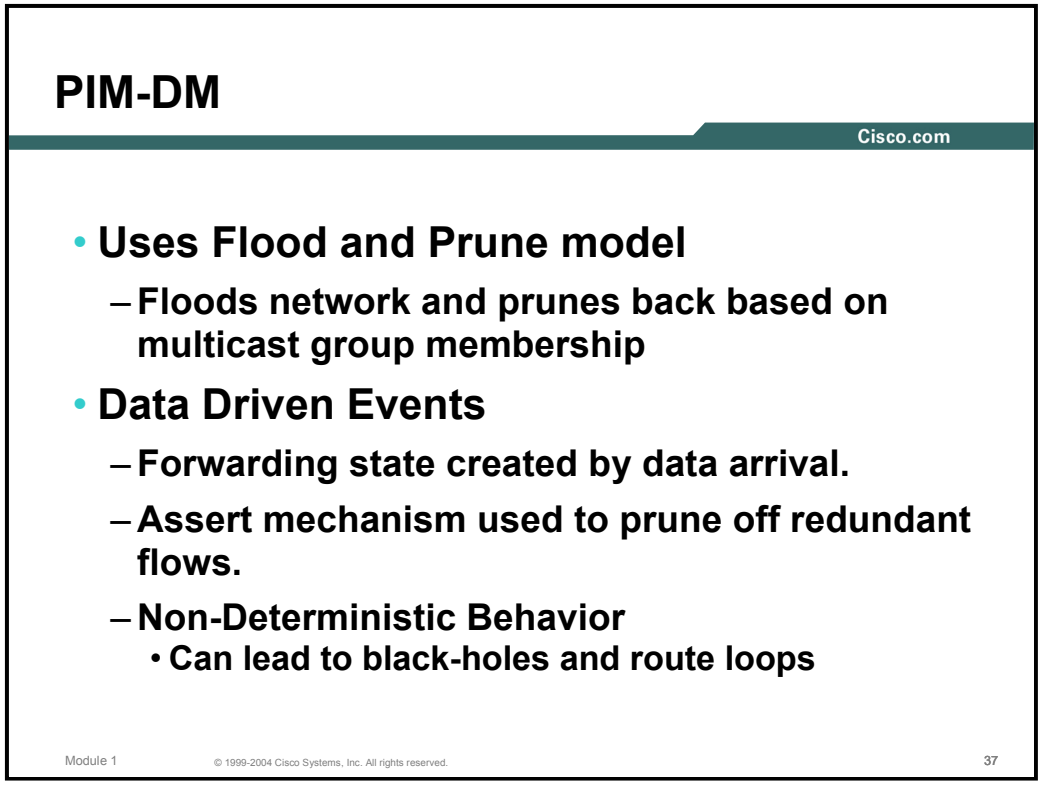



- **Administrative Boundaries**
  - Administratively-scoped multicast address ranges generally used in more than one location.
  - In the example above, the Administratively-Scoped address range, (239.128.0.0/16) is being used by both the LA campus and the NYC campus. Multicast traffic originated in these address ranges will remain within each respective campus and not onto the WAN that exists between the two campuses.
  - This is often sort of configuration is often used so that each campus can source high-rate multicasts on the local campus and not worry about it being accidentally “leaked” into the WAN and causing congestion on the slower WAN links.
  - In addition to the 239.128.0.0/16 range, the entire company network has a Administrative boundary for the 239.129.0.0/16 multicast range. This is so that multicasts in these ranges do not leak into the Internet.
    - Note: The Admin.-Scoped address range (239..0.0/8) is similar to the 10.0.0.0 unicast address range in that it is reserved and is not assigned for use in the Internet.

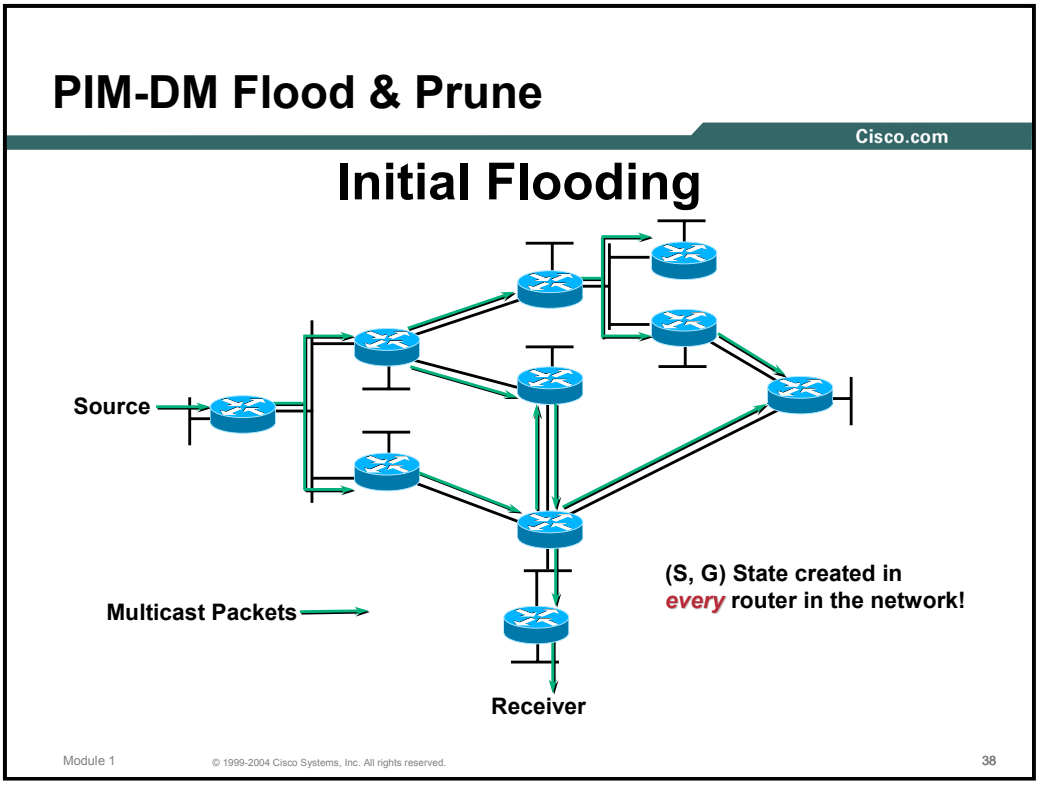
# Multicast Protocol Basics



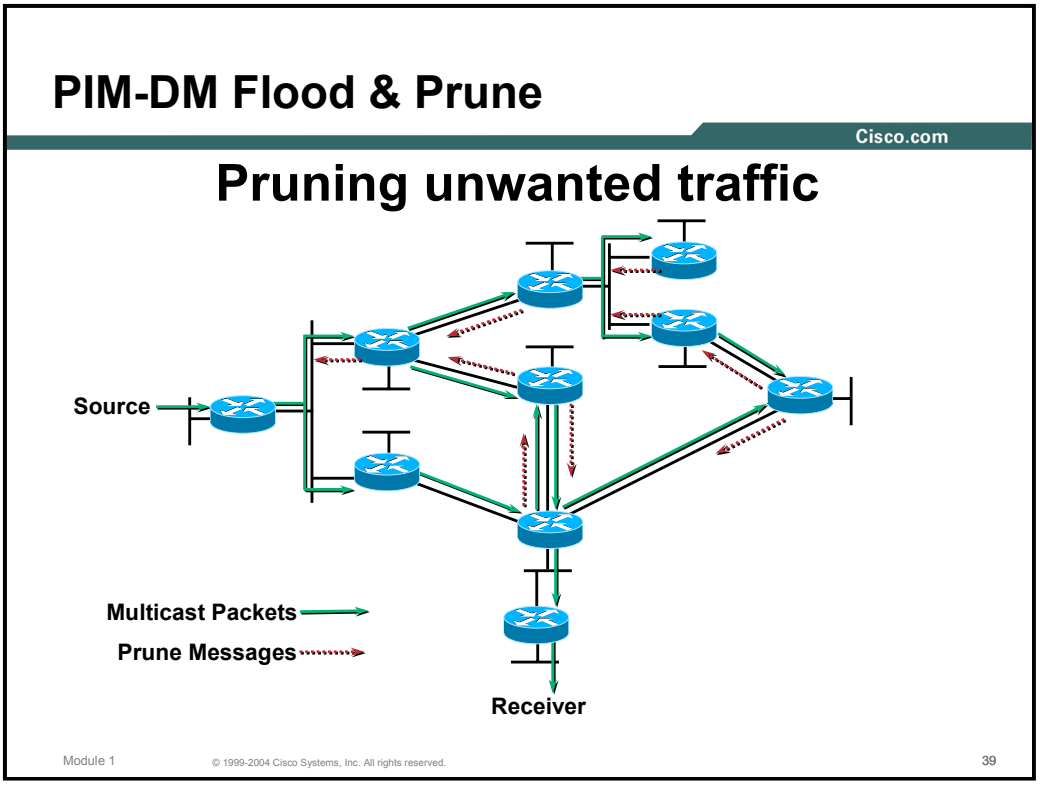

- **Dense-mode multicast protocols**
  - Initially flood/broadcast multicast data to entire network, then prune back paths that don't have interested receivers
- **Sparse-mode multicast protocols**
  - Assumes no receivers are interested unless they explicitly ask for it



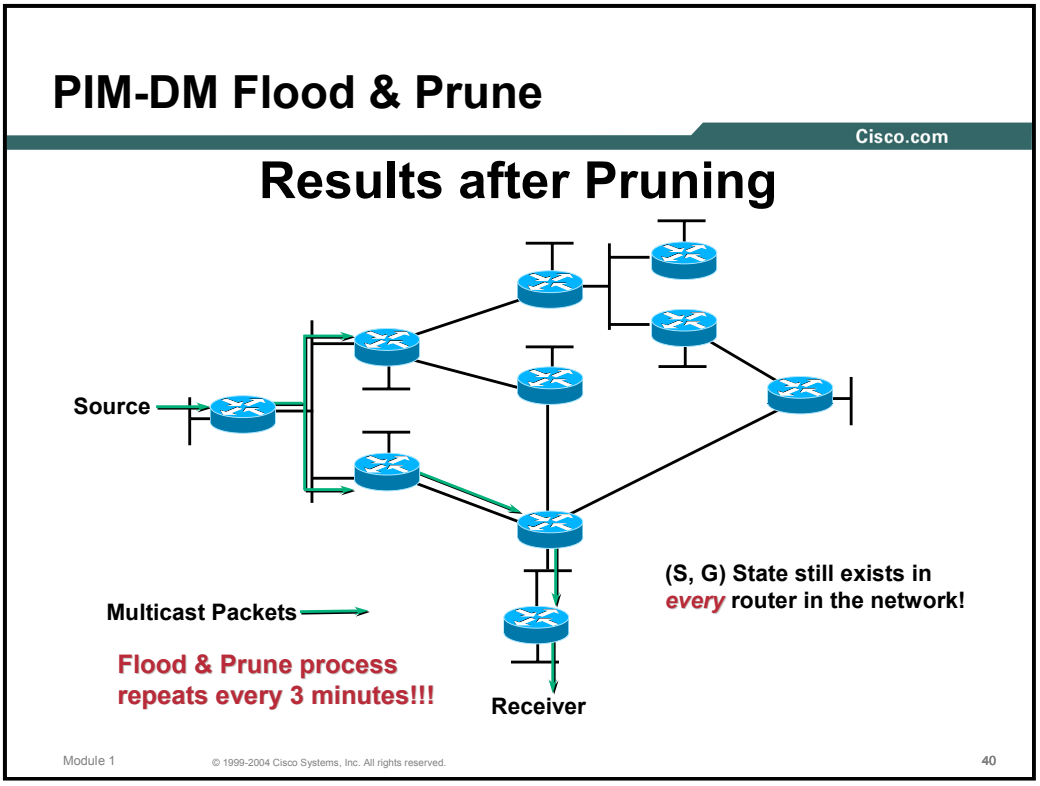

- **Protocol Independent Multicast (PIM) Dense-mode (Internet-draft)**
  - Uses Reverse Path Forwarding (RPF) to flood the network with multicast data, then prune back paths based on uninterested receivers
    - Interoperates with DVMRP
  - Data Driven Events
    - The Flood-and-Prune mechanisms of PIM-DM result in multicast routing state being created by the arrival of multicast data. When the first multicast packet from source “S” to group “G” arrives, an (S,G) multicast forwarding entry is created in the mroute table.
    - PIM-DM is heavily dependent on the PIM Assert mechanism to prune off redundant paths. This coupled with the normal Prune mechanism results in the desired multicast tree being established.
    - The dependency on Data Driven Events result in non-deterministic behavior. This is especially true during network topology changes and can sometimes lead to black holes and multicast route loops.
- **Appropriate for**
  - Testing router performance in labs.



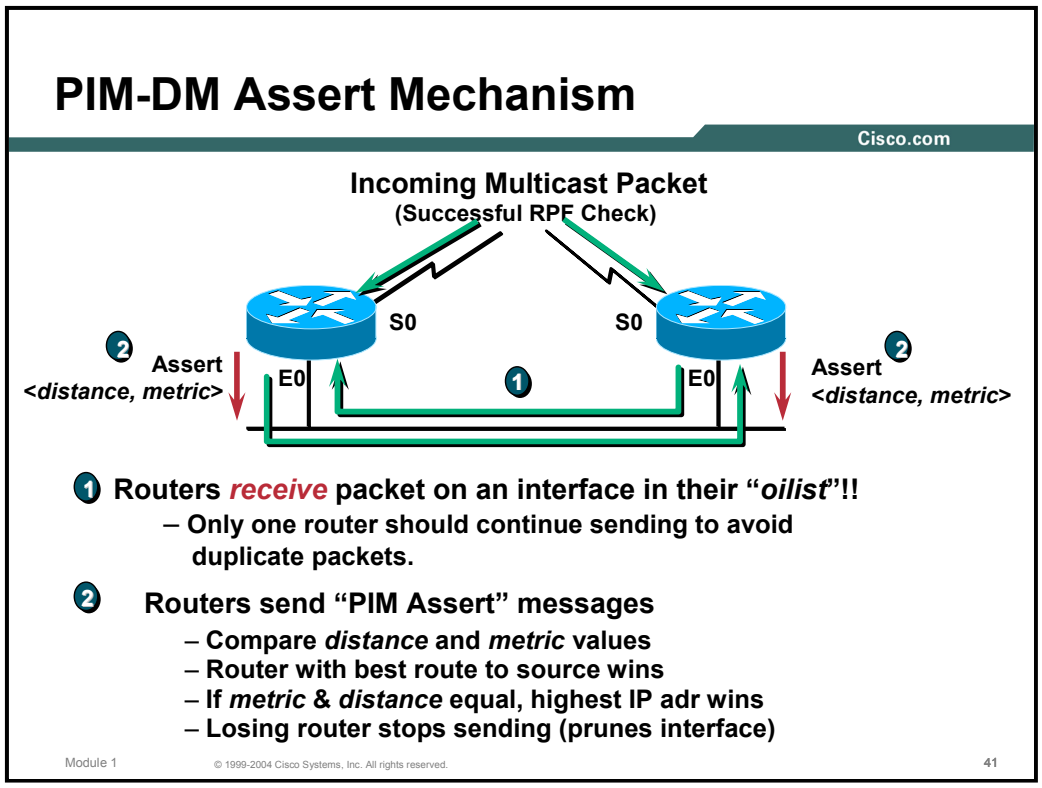

- **PIM-DM Initial Flooding**
  - PIM-DM is similar to DVMRP in that it initially floods multicast traffic to all parts of the network.
  - However unlike DVMRP, which pre-builds a “Truncated Broadcast Tree” that is used for initial flooding, PIM-DM initially floods traffic out ALL non RPF interfaces where there is:
    - Another PIM-DM neighbor or
    - A directly connected member of the group
  - The reason that PIM-DM does not use “Truncated Broadcast Trees” to pre-build a spanning tree for each source network is that this would require running a separate routing protocol as does DVMRP. (At the very least, some sort of Poison-Reverse messages would have to be sent to build the TBT.) Instead, PIM-DM uses other mechanisms to prune back the traffic flows and build Source Trees.
- **Initial Flooding Example**
  - In this example, multicast traffic being sent by the source is flooded throughout the entire network.
  - As each router receives the multicast traffic via its RPF interface (the interface in the direction of the source), it forwards the multicast traffic to all of its PIM-DM neighbors.
  - Note that this results in some traffic arriving via a non-RPF interface such as the case of the two routers in the center of the drawing. (Packets arriving via the non-RPF interface are discarded.) These non-RPF flows are normal for the initial flooding of data and will be corrected by the normal PIM-DM pruning mechanism.



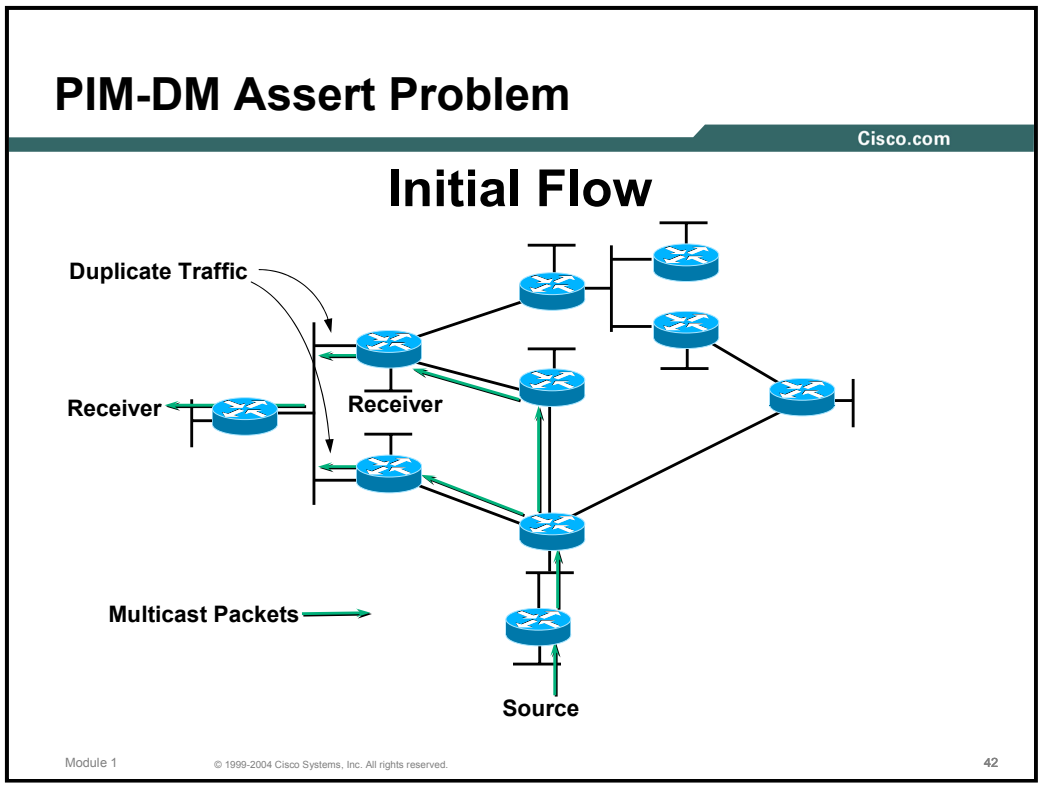

- **Pruning unwanted traffic**
  - In the example above, PIM Prunes (denoted by the dashed arrows) are sent to stop the flow of unwanted traffic.
  - Prunes are sent on the RPF interface when the router has no downstream members that need the multicast traffic.
  - Prunes are also sent on non-RPF interfaces to shutoff the flow of multicast traffic that is arriving via the wrong interface (i.e. traffic arriving via an interface that is not in the shortest path to the source.)
    - An example of this can be seen at the second router from the receiver near the center of the drawing. Multicast traffic is arriving via a non-RPF interface from the router above (in the center of the network) which results in a Prune message.



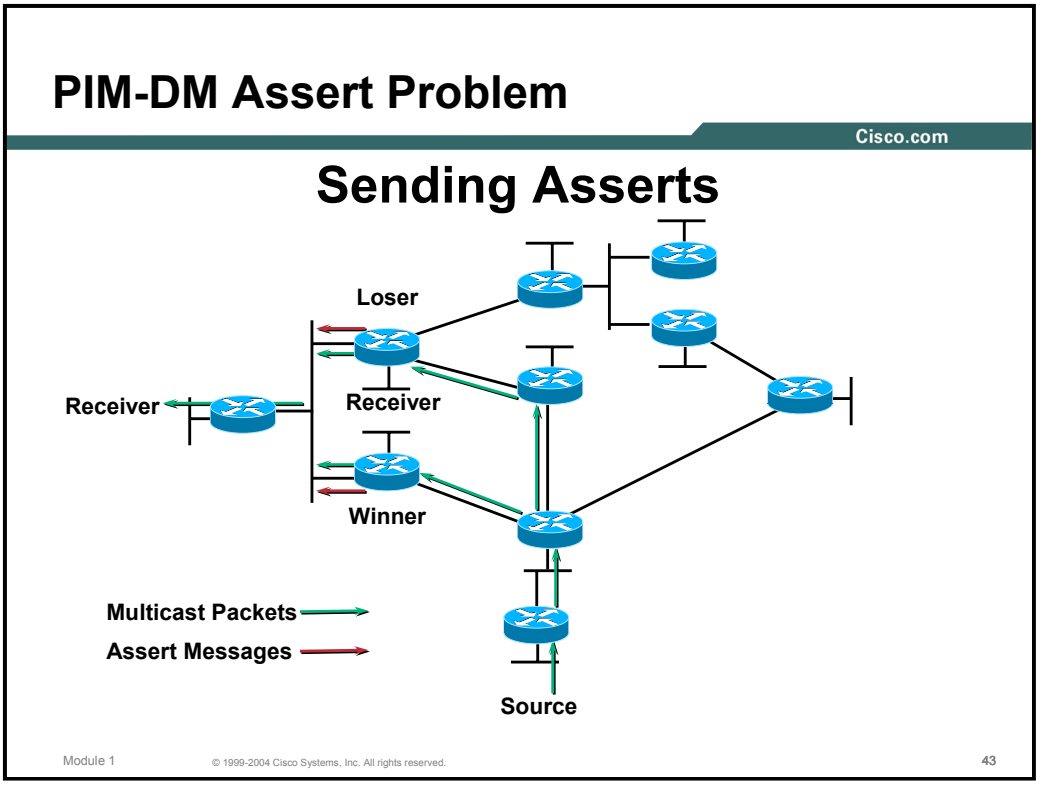

- **Results after Pruning**
  - In the final drawing in our example shown above, multicast traffic has been pruned off of all links except where it is necessary. This results in a Shortest Path Tree (SPT) being built from the Source to the Receiver.
  - Even though the flow of multicast traffic is no longer reaching most of the routers in the network, (S, G) state still remains in ALL routers in the network. This (S, G) state will remain until the source stops transmitting.
  - In PIM-DM, Prunes expire after three minutes. This causes the multicast traffic to be re-flooded to all routers just as was done in the “Initial Flooding” drawing. This periodic (every 3 minutes) “Flood and Prune” behavior is normal and must be taken into account when the network is designed to use PIM-DM.



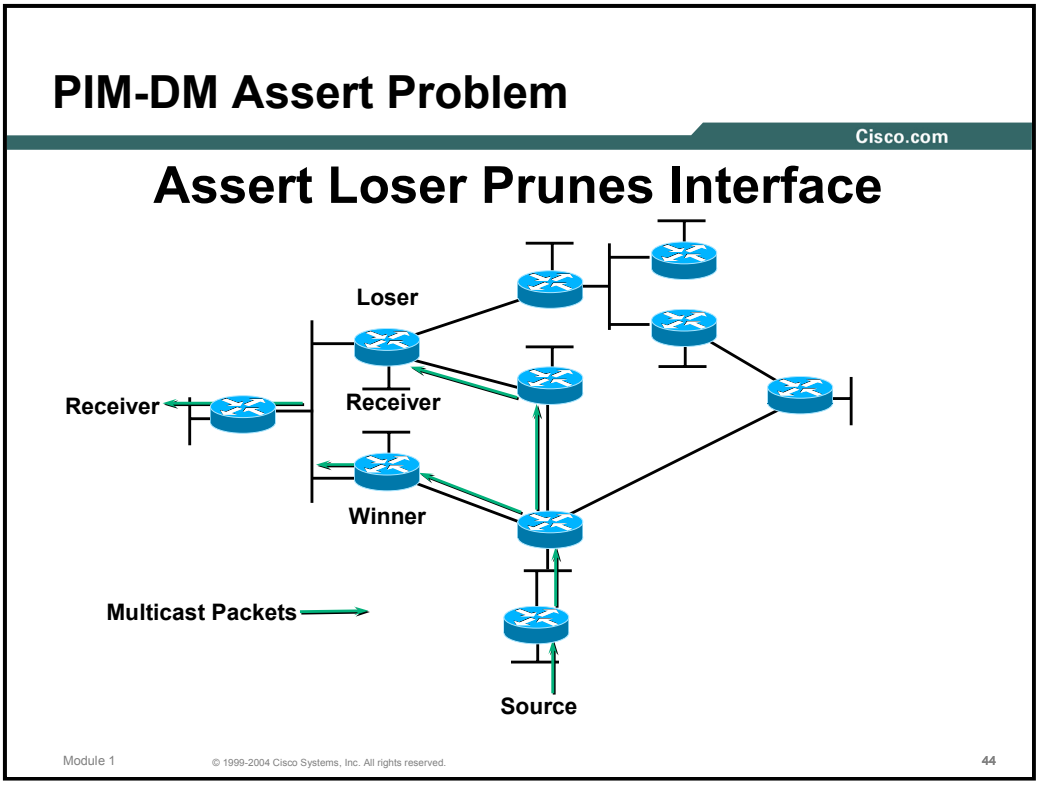

- **PIM Assert Mechanism**
  - The PIM Assert mechanism is used to shutoff duplicate flows onto the same multi-access network.
    - Routers detect this condition when they receive an (S, G) packet via a multi-access interface that it is in the (S, G) OIL.
    - This causes the routers to send Assert Messages.
  - Assert messages containing the Admin. Distance and metric to the source combined into a single assert value. (The Admin. Distance is the high-order part of this assert value.)
  - Routers compare these values to determine who has the best path (lowest value) to the source. (If both values are the same, the highest IP address is used as the tie breaker.)
  - The Losing routers (the ones with the higher value) Prunes its interface while the winning router continues to forward multicast traffic onto the LAN segment.



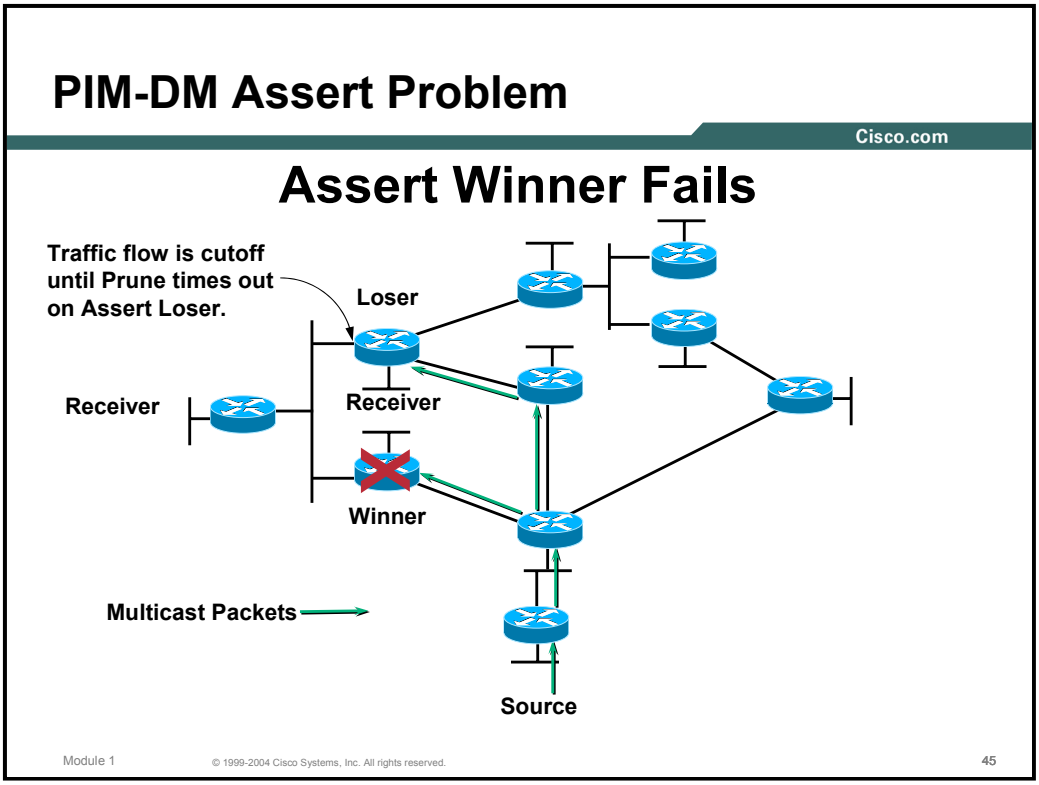

- **PIM-DM Assert Problem**
  - While the PIM Assert mechanism is effective in pruning off duplicate traffic, it is not without its weaknesses.
  - Consider the above example where duplicate traffic is flowing onto a LAN segment.



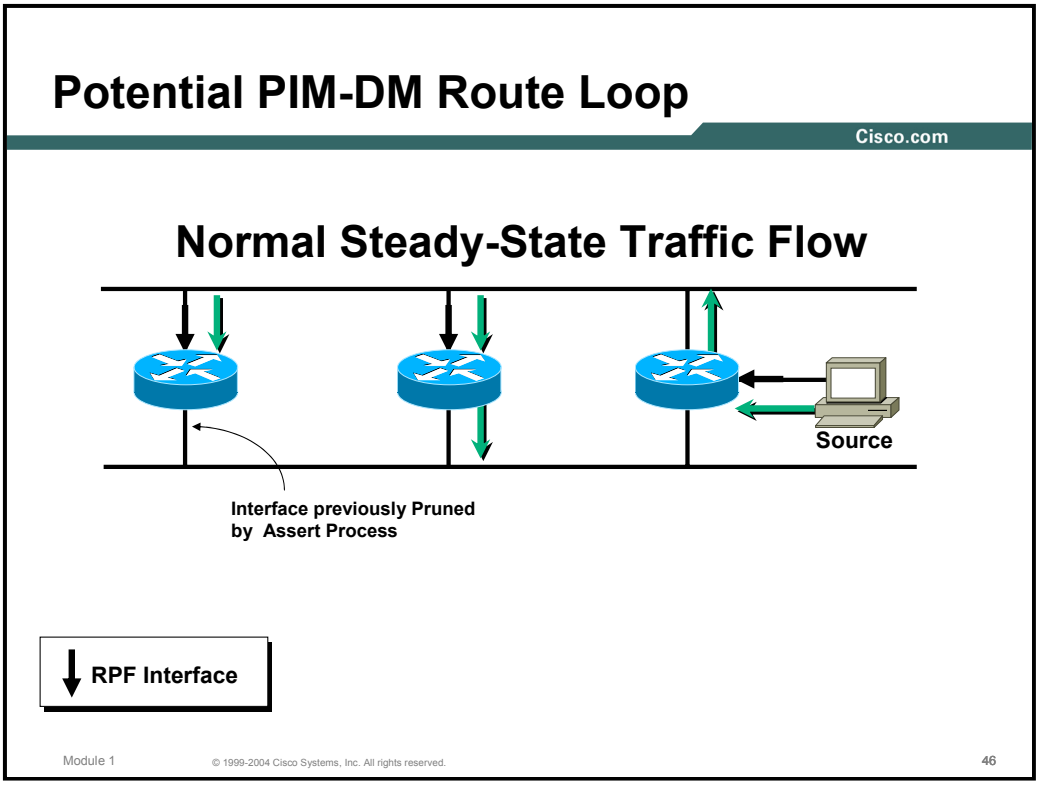

- **PIM-DM Assert Problem**
  - The normal PIM Assert mechanism takes place and the two routers exchange routing metrics to determine which one has the best route to the source.
  - In this case, the bottom router has the best metric and is the Assert Winner.



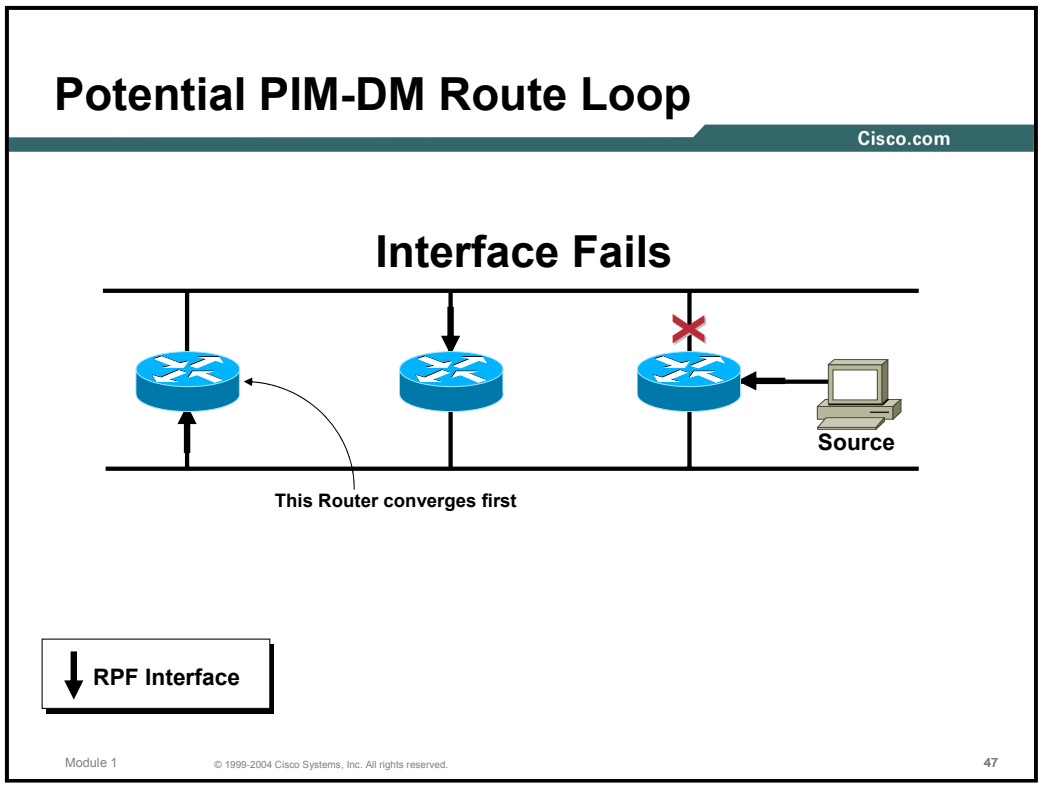

- **PIM-DM Assert Problem**
  - The normal PIM Assert mechanism takes place and the Assert Winner continues forwarding while the Assert Loser prunes it's interface and starts its prune timer.



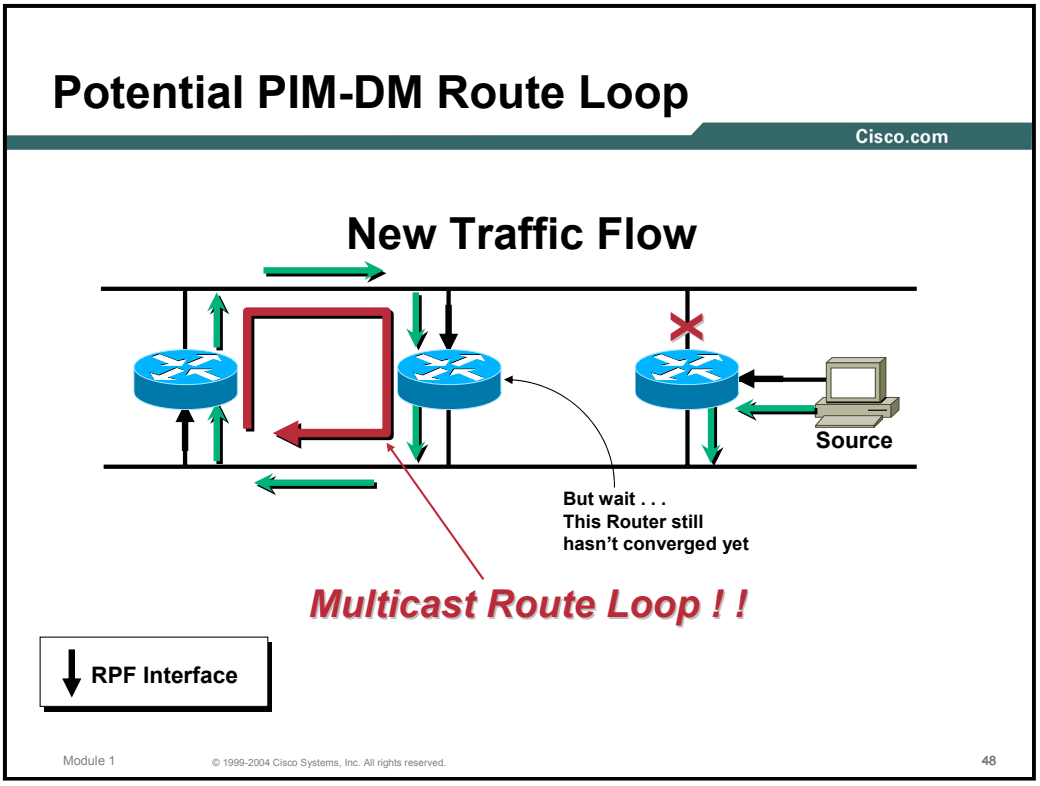

- **PIM-DM Assert Problem**
  - Let's now assume that the Assert Winner fails immediately after winning the Assert process.
  - Unfortunately, the Assert Loser has no way of knowing that the Assert Winner has failed and will wait 3 minutes before timing out its pruned interface. This results in a 3 minute (worst-case) loss of traffic.



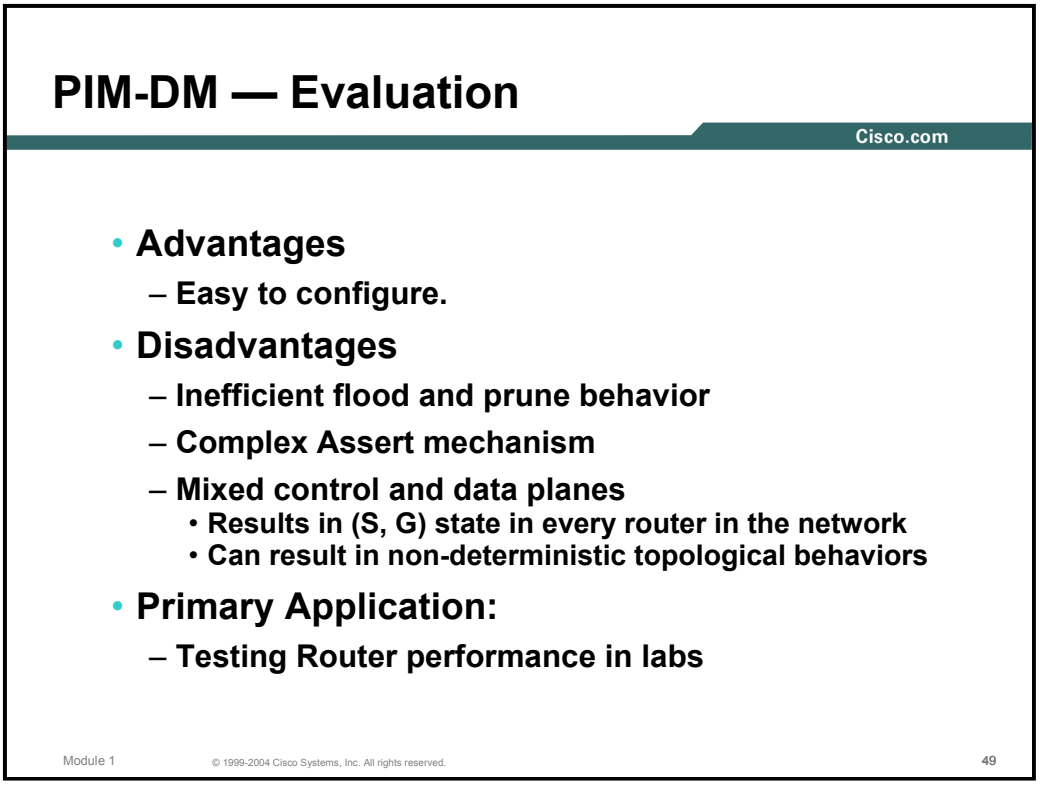

- **Potential PIM-DM Route Loops**
  - The non-deterministic behavior of PIM-DM along with its flood-and-prune mechanism can sometimes result in serious network outages including “blackholes” and multicast route loops.
  - The network in the above example is a simplified version of a frequently used network design whereby multiple routers are used to provide redundancy in the network.
  - Under normal steady-state conditions, traffic flows from the source via the RPF interfaces as shown.
    - Note that the routers have performed the Assert process and one interface on one router is in the pruned state.



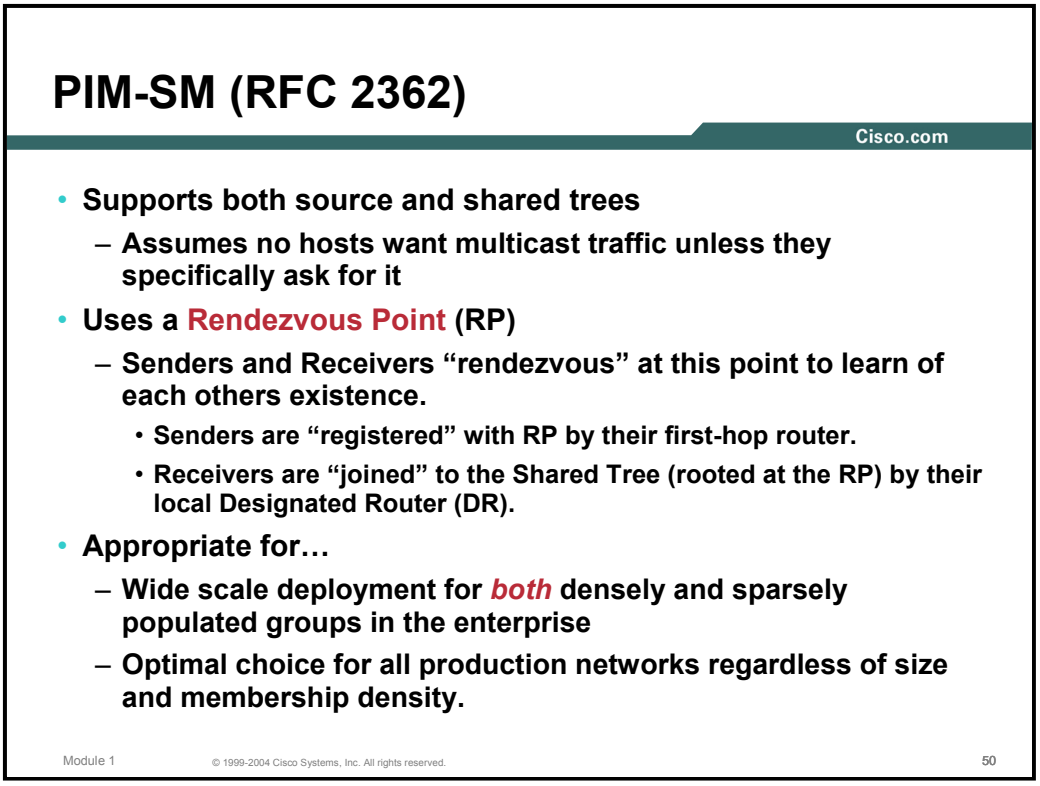

- **Potential PIM-DM Route Loops**
  - Now let's assume that the forwarding interface of the first-hop router fails as shown above.
  - Let's also assume that the unicast routing of router on the left converges first and PIM computes the new RPF interface as shown.



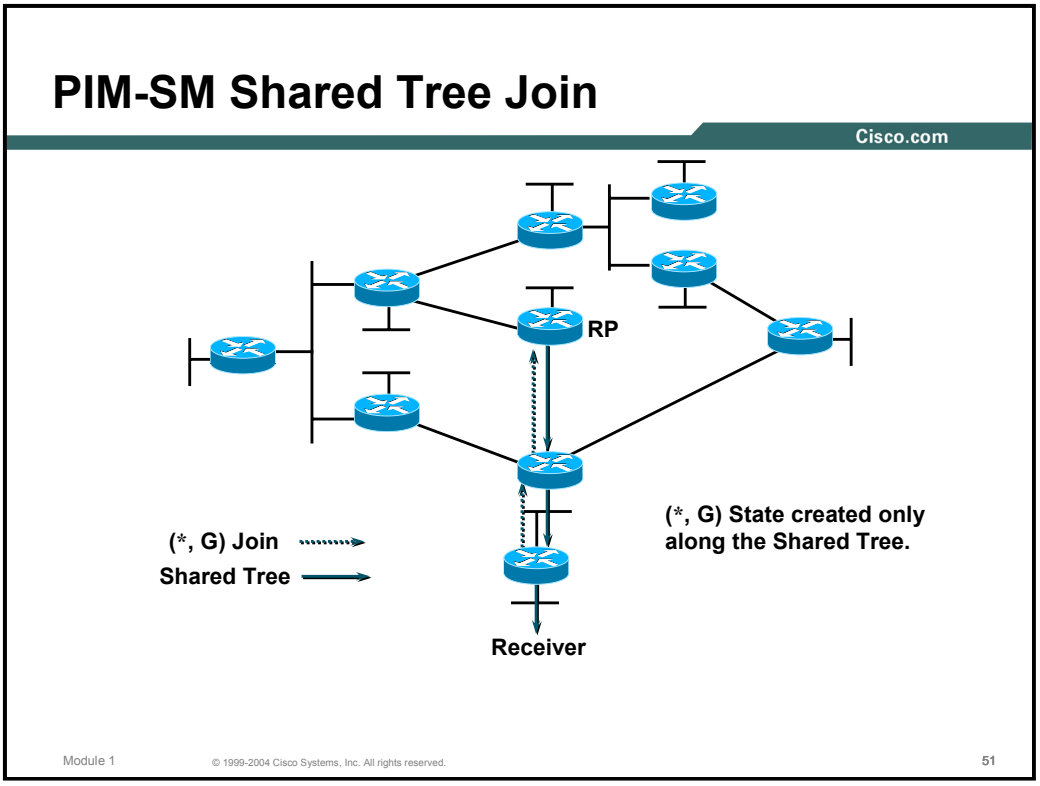

- **Potential PIM-DM Route Loops**
  - Unfortunately, the middle router has not yet converged and is still forwarding multicast traffic using the old RPF interface.
  - At this point, a multicast route loop exists in the network due to the transient condition of the two routers having opposite RPF interfaces.
  - During the time that this route loop exists, virtually all of the bandwidth on the network segments can be consumed. This situation will continue until the router in the middle of the picture finally converges and the new "correct" RPF interface is calculated.
  - Unfortunately, if the router needs some bandwidth to complete this convergence (as in the case when EIGRP goes active), then this condition will never be resolved!



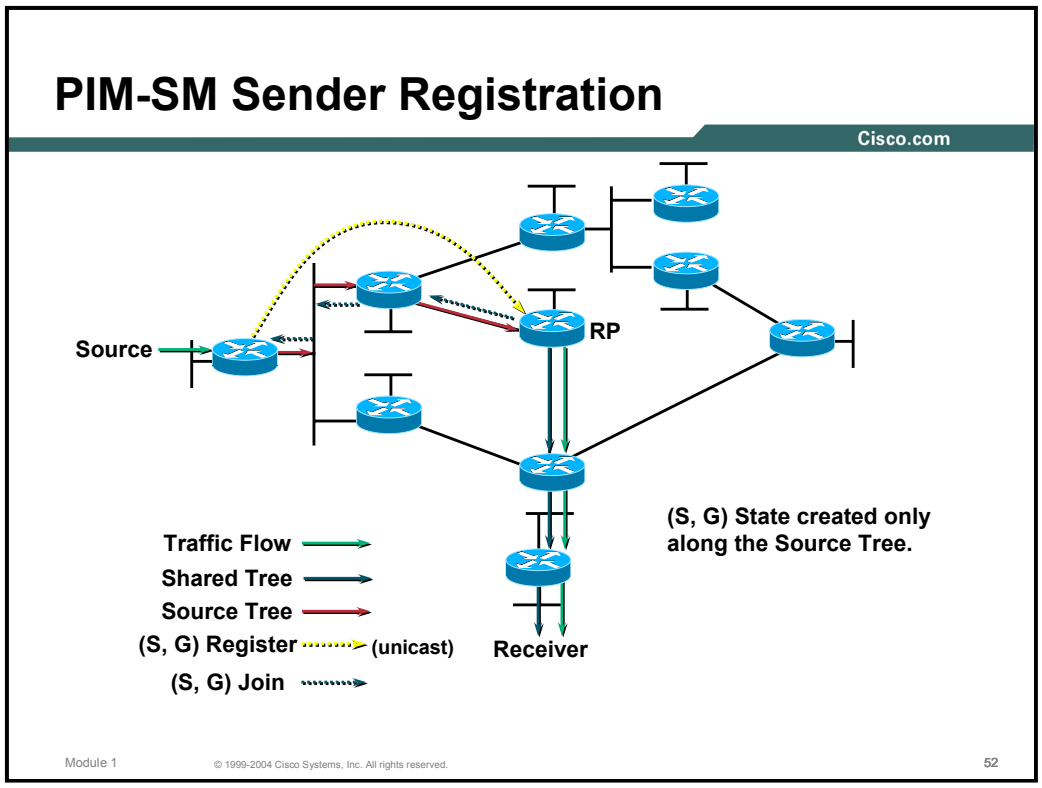

- **Evaluation: PIM Dense-mode**
  - Most effective for small pilot networks.
  - Advantages
    - Minimal number of commands required for configuration (two)
    - Simple mechanism for reaching all possible receivers and eliminating distribution to uninterested receivers
    - Interoperates with DVMRP
  - Disadvantages
    - Necessity to flood frequently because prunes expire after 3 minutes.
    - Assert mechanism can become quite complex in behavior under certain topological conditions making it difficult to debug.
    - Mixed control and data planes are a result of the lack of explicit Join mechanisms. This in turn causes PIM-DM to have a non-deterministic behavior because it is dependent on data driven events to create/maintain multicast forwarding state.
  - Primary Application:
    - Testing router forwarding performance in labs.



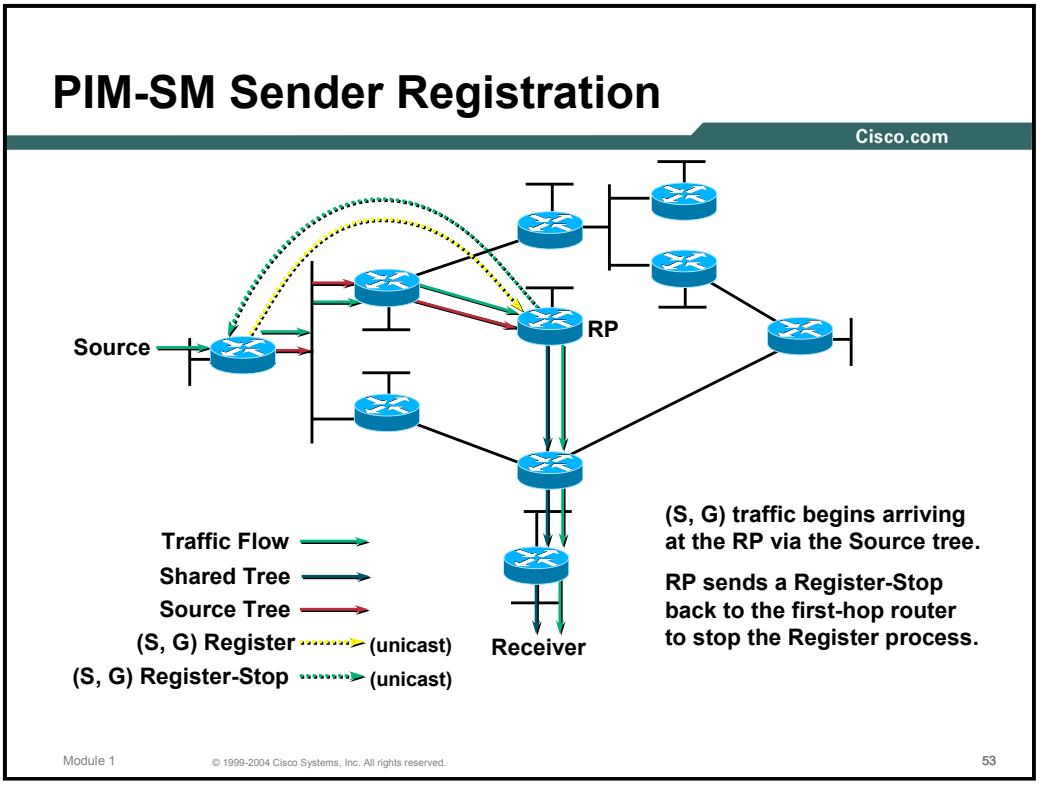

- **Protocol Independent Multicast (PIM) Sparse-mode (RFC 2362)**
  - Utilizes a rendezvous point (RP) to coordinate forwarding from source to receivers
    - Regardless of location/number of receivers, senders register with RP and send a single copy of multicast data through it to registered receivers
    - Regardless of location/number of sources, group members register to receive data and always receive it through the RP
  - Appropriate for
    - Wide scale deployment for both densely and sparsely populated groups in the Enterprise
    - Optimal choice for all production networks regardless of size and membership density.



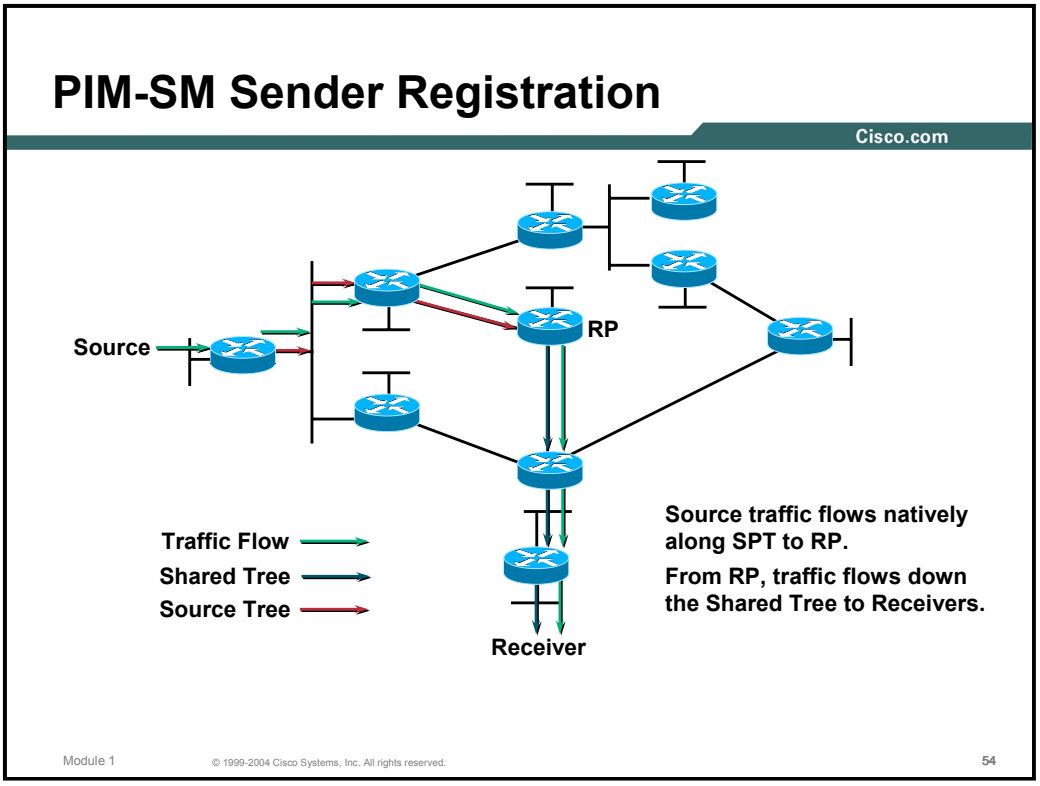

- **PIM-SM Shared Tree Joins**
  - In this example, there is an active receiver (attached to leaf router at the bottom of the drawing) has joined multicast group “G”.
  - The leaf router knows the IP address of the Rendezvous Point (RP) for group G and when it sends a (*,G) Join for this group towards the RP.
  - This (*, G) Join travels hop-by-hop to the RP building a branch of the Shared Tree that extends from the RP to the last-hop router directly connected to the receiver.
  - At this point, group “G” traffic can flow down the Shared Tree to the receiver.



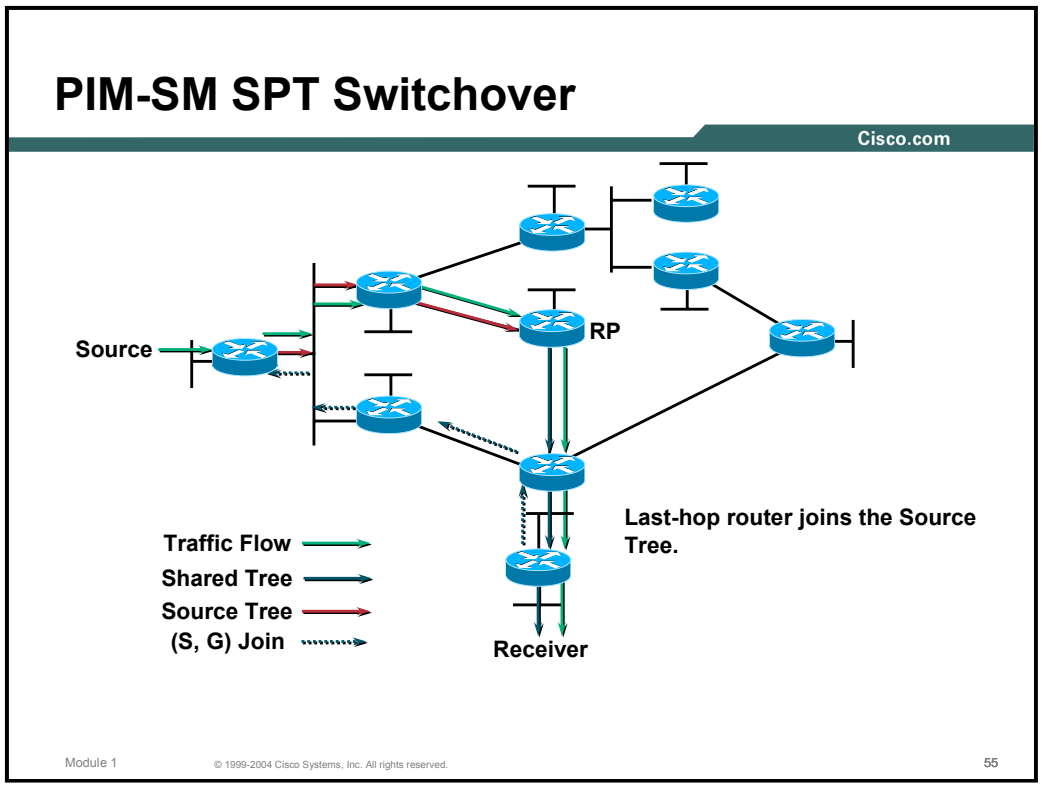

- **PIM-SM Sender Registration**
  - As soon as an active source for group G sends a packet the leaf router that is attached to this source is responsible for “Registering” this source with the RP and requesting the RP to build a tree back to that router.
  - The source router encapsulates the multicast data from the source in a special PIM SM message called the Register message and unicasts that data to the RP.
  - When the RP receives the Register message it does two things
    - It de-encapsulates the multicast data packet inside of the Register message and forwards it down the Shared Tree.
    - The RP also sends an (S,G) Join back towards the source network S to create a branch of an (S, G) Shortest-Path Tree. This results in (S, G) state being created in all the router along the SPT, including the RP.



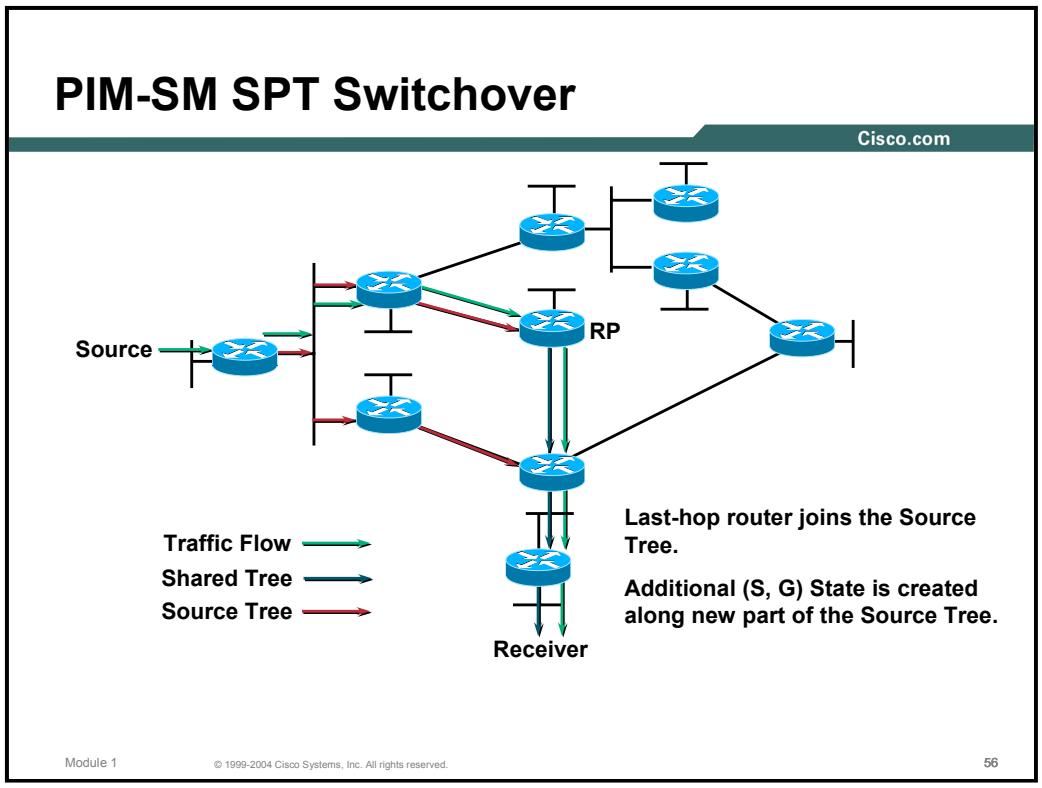

- **PIM-SM Sender Registration (cont.)**
  - As soon as the SPT is built from the Source router to the RP, multicast traffic begins to flow natively from source S to the RP.
  - Once the RP begins receiving data natively (i.e. down the SPT) from source S it sends a 'Register Stop' to the source's first hop router to inform it that it can stop sending the unicast Register messages.



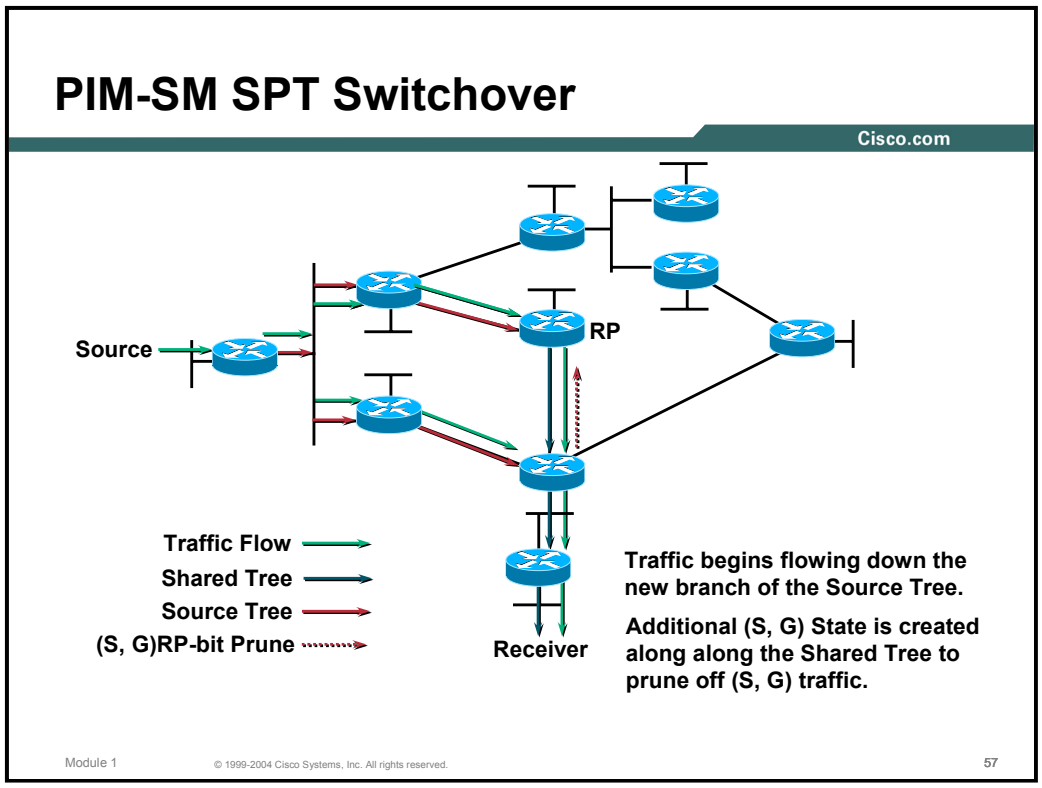

- **PIM-SM Sender Registration (cont.)**
  - At this point, multicast traffic from the source is flowing down the SPT to the RP and from there, down the Shared Tree to the receiver.



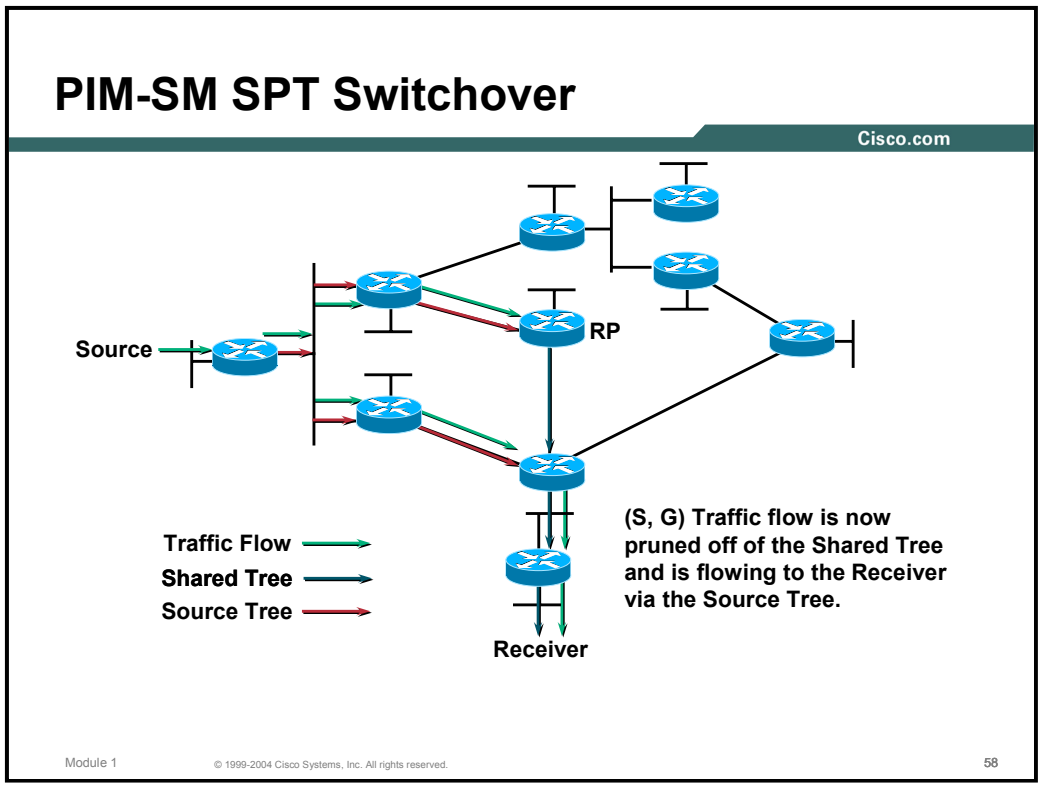

- **PIM-SM Shortest-Path Tree Switchover**
  - PIM-SM has the capability for last-hop routers (i.e. routers with directly connected members) to switch to the Shortest-Path Tree and bypass the RP if the traffic rate is above a set threshold called the “SPT-Threshold”.
    - The default value of the SPT-Threshold” in Cisco routers is zero. This means that the default behaviour for PIM-SM leaf routers attached to active receivers is to immediately join the SPT to the source as soon as the first packet arrives via the (*,G) shared tree.
  - In the above example, the last-hop router (at the bottom of the drawing) sends an (S, G) Join message toward the source to join the SPT and bypass the RP.
  - The (S, G) Join message travels hop-by-hop to the first-hop router (i.e. the router connected directly to the source) thereby creating another branch of the SPT.



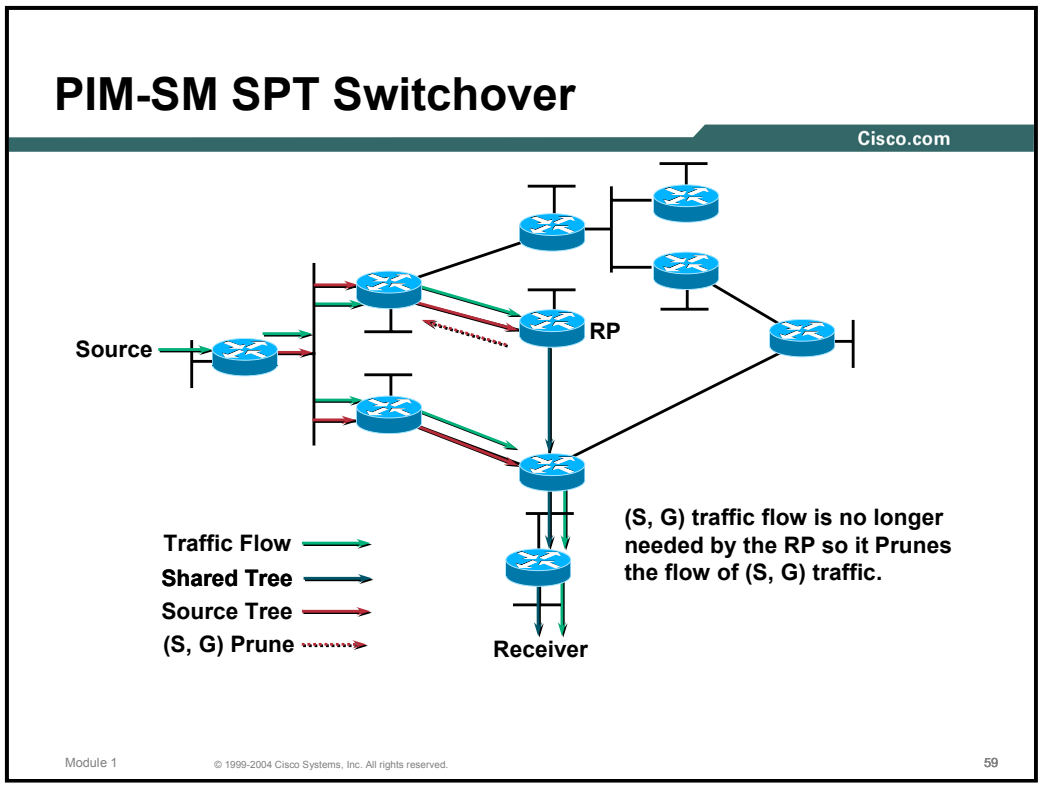

- **PIM-SM Shortest-Path Tree Switchover**
  - The (S, G) Join messages creates (S, G) state in all the routers along this branch of the SPT.



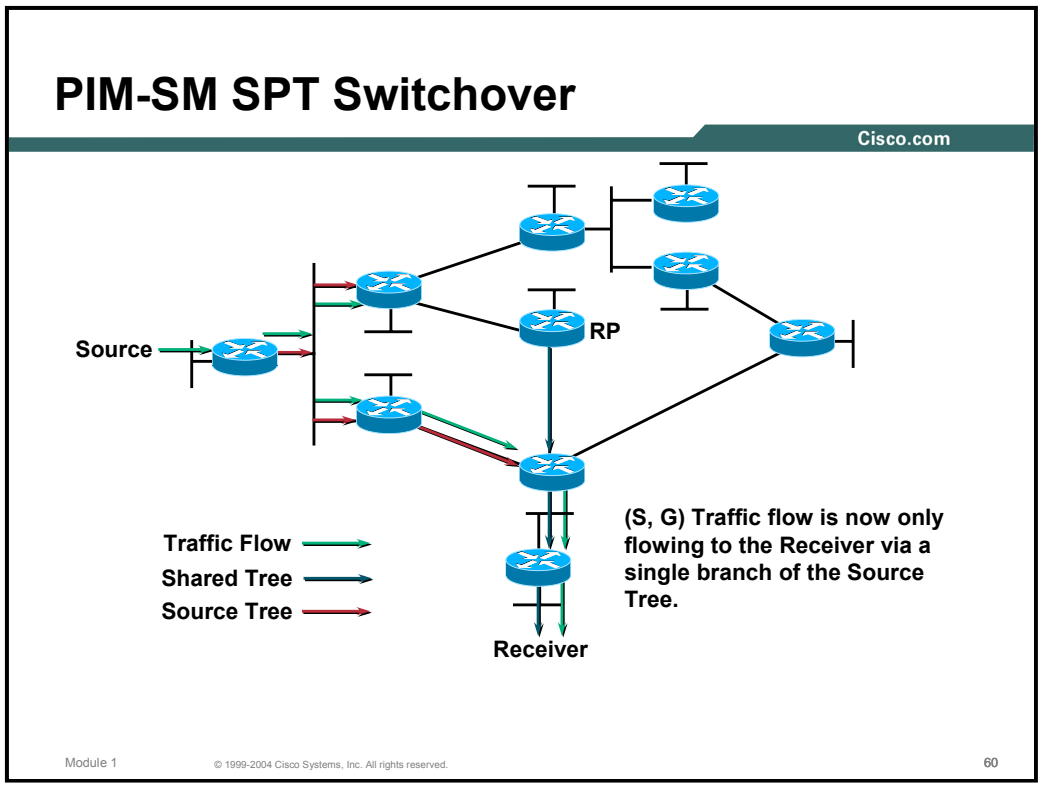

- **PIM-SM Shortest-Path Tree Switchover**
  - Once the branch of the Shortest-Path Tree has been built, (S, G) traffic begins flowing to the receiver via this new branch.
  - Next, special (S, G)RP-bit Prune messages are sent up the Shared Tree to prune off the redundant (S,G) traffic that is still flowing down the Shared Tree.
    - If this were not done, (S, G) traffic would continue flowing down the Shared Tree resulting in duplicate (S, G) packets arriving at the receiver.




- **PIM-SM Shortest-Path Tree Switchover**
  - As the (S, G)RP-bit Prune message travels up the Shared Tree, special (S, G)RP-bit Prune state is created along the Shared Tree that selectively prevents this traffic from flowing down the Shared Tree.
  - At this point, (S, G) traffic is now flowing directly from the first-hop router to the last-hop router and from there to the receiver.




- **PIM-SM Shortest-Path Tree Switchover**
  - At this point, the RP no longer needs the flow of (S, G) traffic since all branches of the Shared Tree (in this case there is only one) have pruned off the flow of (S, G) traffic.
  - As a result, the RP will send (S, G) Prunes back toward the source to shutoff the flow of the now unnecessary (S, G) traffic to the RP
    - Note: This will occur IFF the RP has received an (S, G)RP-bit Prune on all interfaces on the Shared Tree.




- **PIM-SM Shortest-Path Tree Switchover**
  - As a result of the SPT-Switchover, (S, G) traffic is now only flowing from the first-hop router to the last-hop router and from there to the receiver. Notice that traffic is no longer flowing to the RP.
  - As a result of this SPT-Switchover mechanism, it is clear that PIM SM also supports the construction and use of SPT (S,G) trees but in a much more economical fashion than PIM DM in terms of forwarding state.

## PIM-SM Frequently Forgotten Fact

**"The default behavior of PIM-SM is that routers with directly connected members will join the Shortest Path Tree as soon as they detect a new multicast source."**

## PIM-SM — Evaluation

Cisco.com

- **Advantages:**
  - **Traffic only sent down "joined" branches**
  - **Can switch to optimal source-trees for high traffic sources dynamically**
  - **Unicast routing protocol-independent**
  - **Basis for inter-domain multicast routing**
    - **When used with MBGP and MSDP**
- **Disadvantages**
  - **Few if any**
- **Primary Application**
  - **All Production Multicast Networks with sparse or dense distribution of receivers**

Module 1  62

- **Evaluation: PIM Sparse-mode**
  - Can be used for sparse of dense distribution of multicast receivers (no necessity to flood)
  - Advantages
    - Traffic sent only to registered receivers that have explicity joined the multicast group
    - RP can be switched to optimal shortest-path-tree when high-traffic sources are forwarding to a sparsely distributed receiver group
    - Interoperates with DVMRP
  - Potential issues
    - Requires RP during initial setup of distribution tree (can switch to shortest-path-tree once RP is established and determined suboptimal)
  - Primary Application
    - All production networks.

## Protocol Summary

Cisco.com

# CONCLUSION

**"Sparse Mode Good, Dense Mode Bad!"**

Source: ***"The Caveman's Guide to IP Multicast", ©2000, R. Davis***