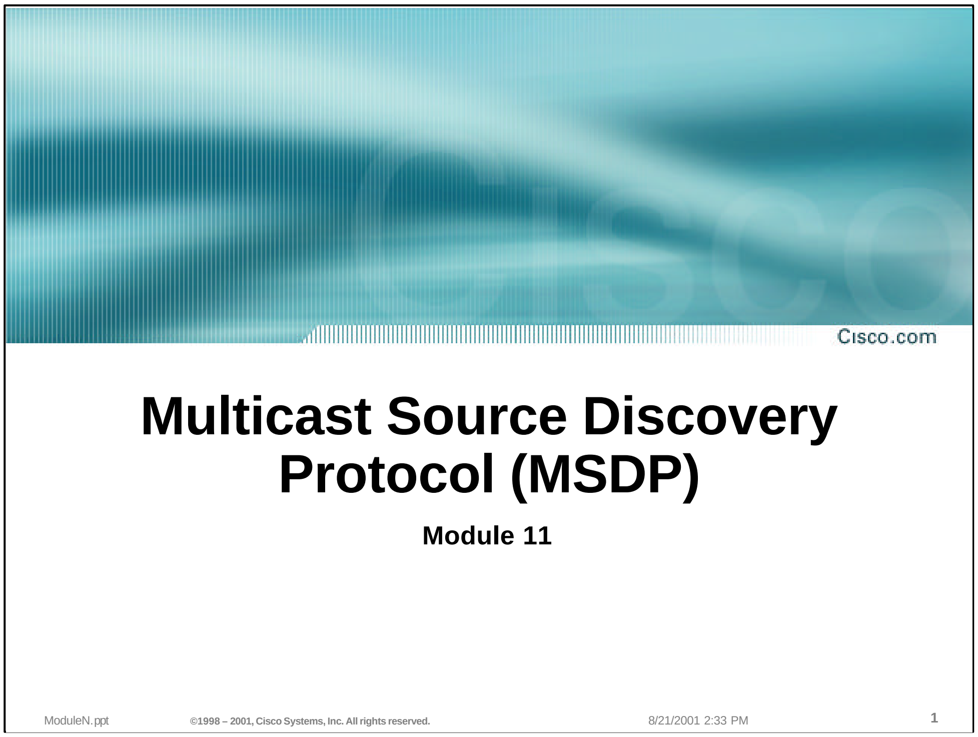
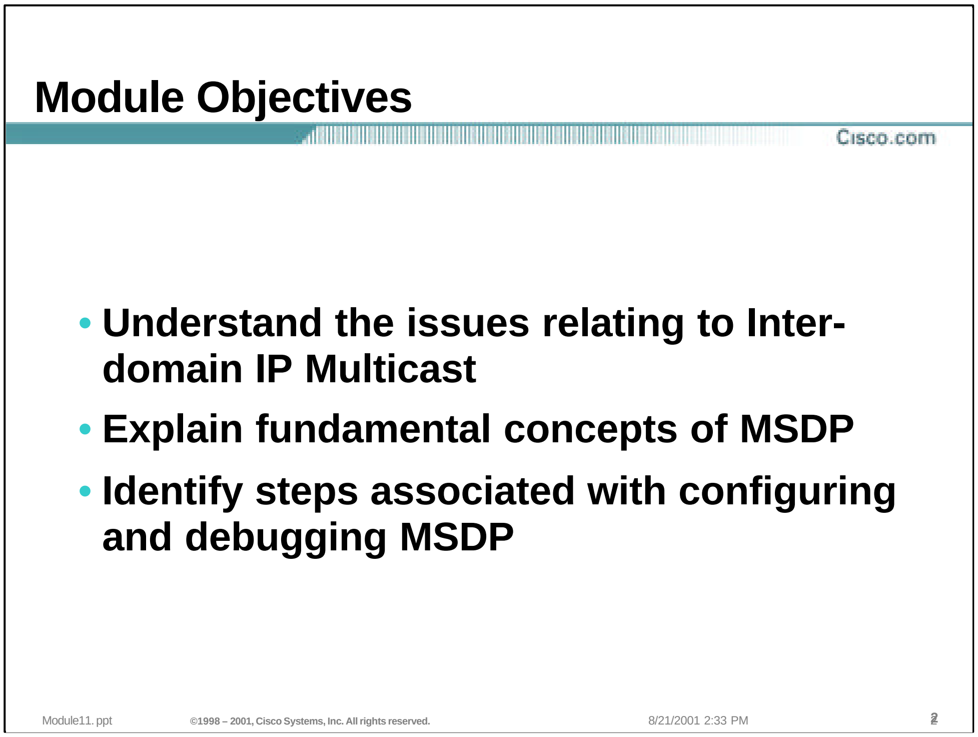


Cisco.com

Multicast Source Discovery Protocol (MSDP)

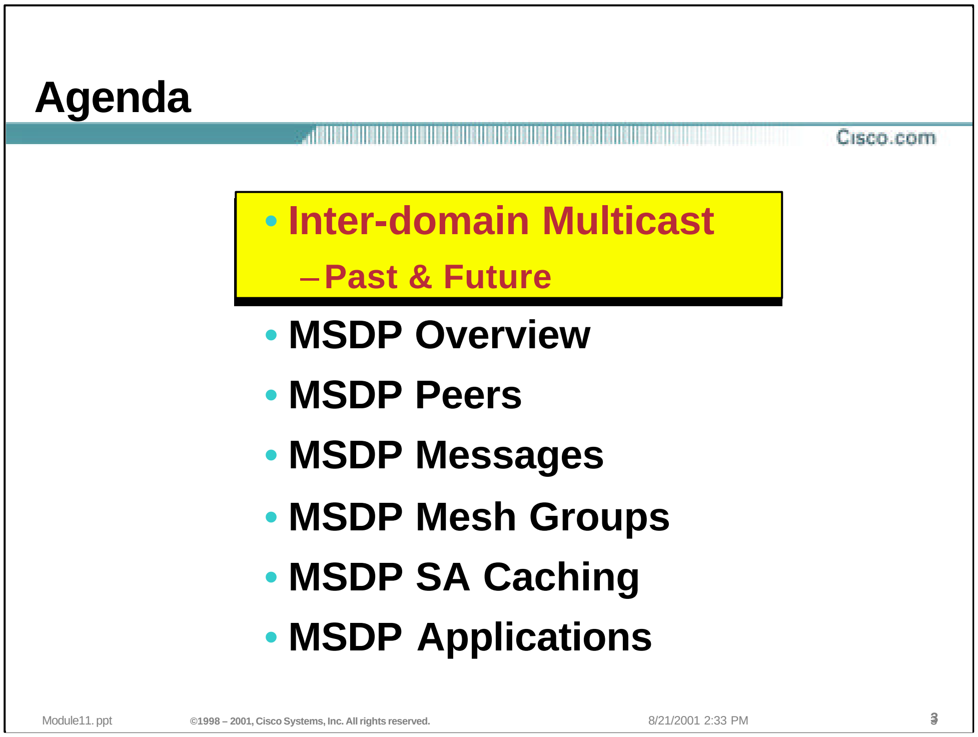
Module 11



Module Objectives

Cisco.com

- Understand the issues relating to Inter-domain IP Multicast
- Explain fundamental concepts of MSDP
- Identify steps associated with configuring and debugging MSDP

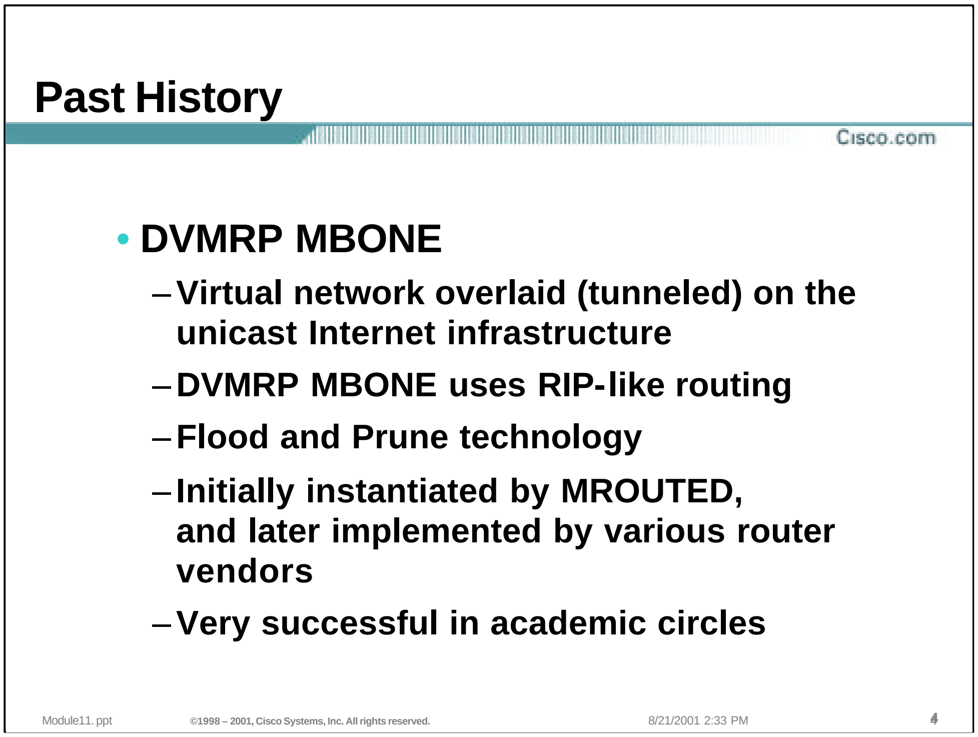


Agenda

Cisco.com

- **Inter-domain Multicast**
  - **Past & Future**
- **MSDP Overview**
- **MSDP Peers**
- **MSDP Messages**
- **MSDP Mesh Groups**
- **MSDP SA Caching**
- **MSDP Applications**

Module11. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 3



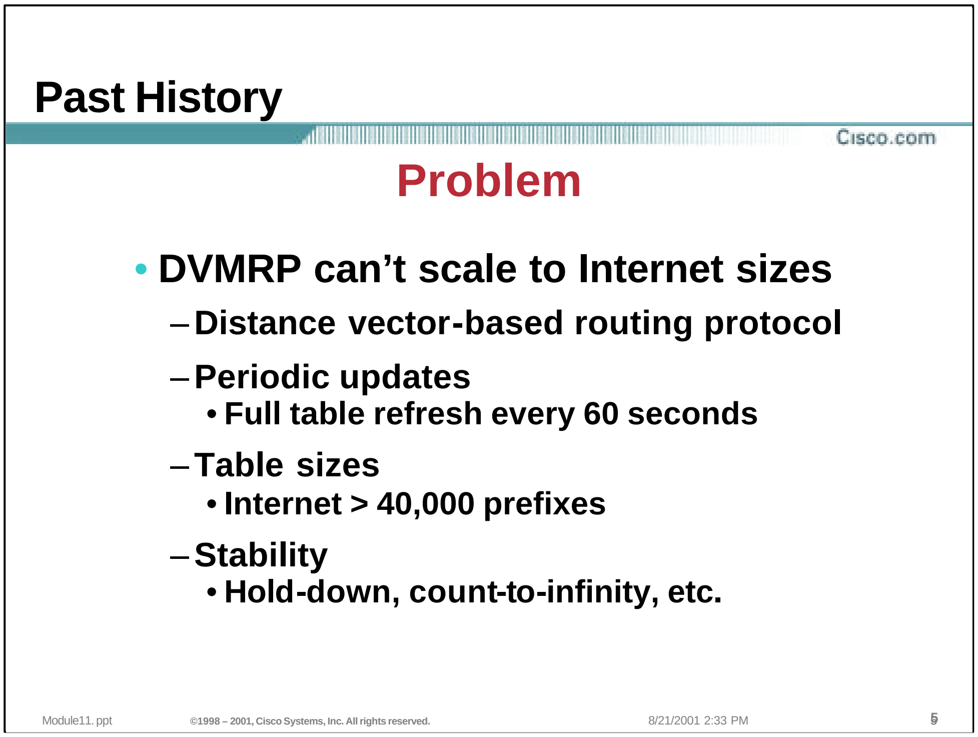
Past History

Cisco.com

- DVMRP MBONE
  - Virtual network overlaid (tunneled) on the unicast Internet infrastructure
  - DVMRP MBONE uses RIP-like routing
  - Flood and Prune technology
  - Initially instantiated by MROUTED,
    and later implemented by various router vendors
  - Very successful in academic circles

Module11. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 4

- DVMRP MBone
  - Historically, the very limited amount of multicast traffic that flowed across the Internet used DVMRP Tunnels to interconnect multicast enabled portions of the Internet together.
  - Unfortunately, DVMRP basically is an extension to the RIP unicast routing protocol and has all of the problems associated with RIP as a routing protocol.
  - DVMRP uses a Flood and Prune methodology where traffic is periodically flooded to every part of the network and pruned back where it is unwanted.
  - The first versions of DVMRP was the “mrouted” program that runs on Unix platforms. Later implementations of DVMRP were developed by commercial router vendors.
  - Initially, the DVMRP MBone was limited to academic sites and was managed by a handful of dedicated academic types that kept it running smoothly. The occasional outages were not considered to be a problem since the MBone was largely seen as an academic experiment.



Past History

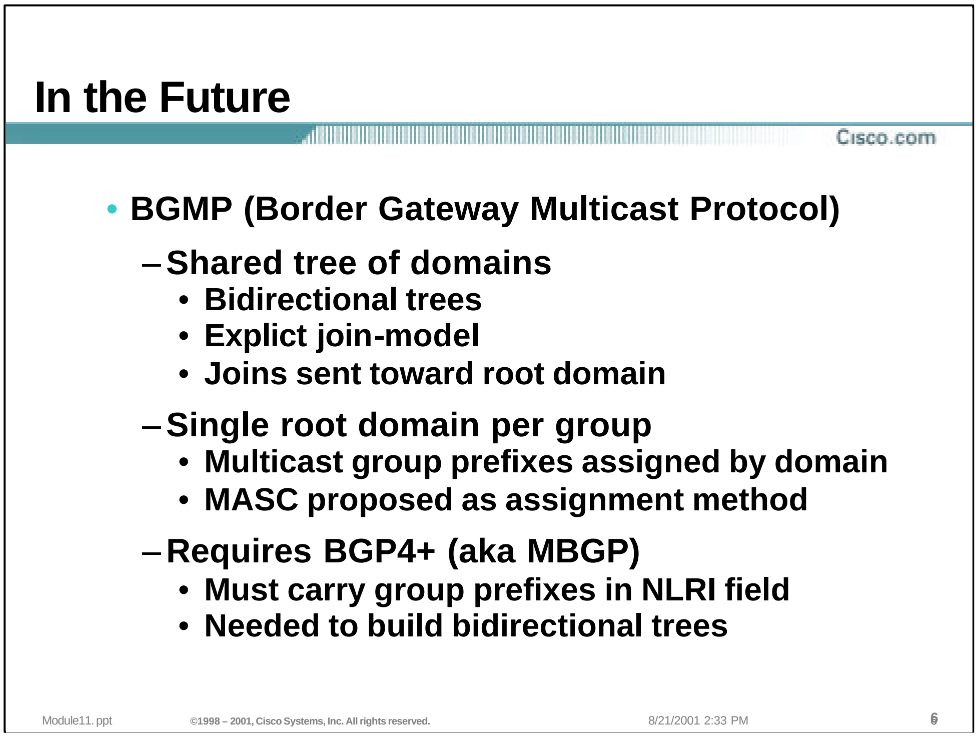
Cisco.com

Problem

- DVMRP can’t scale to Internet sizes
  - Distance vector-based routing protocol
  - Periodic updates
    - Full table refresh every 60 seconds
  - Table sizes
    - Internet > 40,000 prefixes
  - Stability
    - Hold-down, count-to-infinity, etc.

Module11. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 5

- **DVMRP Problems**
  - **DVMRP has problems scaling to any significant size, particularly to the size of the Internet. These problems include:**
    - DVMRP is based on a Distance Vector routing protocol (RIP).
    - Periodic updates of the entire routing table are sent every 60 seconds. This is fine for networks where the routing table is relatively small but is unthinkable for really large networks such as the Internet were the number of prefixes (routes) exceed 40,000.
    - Distance Vector based protocols suffer from some well know stability issues including route Holddown, Count-to-Infinity and other problems.



In the Future

- BGMP (Border Gateway Multicast Protocol)
  - –Shared tree of domains
    - Bidirectional trees
    - Explict join-model
    - Joins sent toward root domain
  - –Single root domain per group
    - Multicast group prefixes assigned by domain
    - MASC proposed as assignment method
  - –Requires BGP4+ (aka MBGP)
    - Must carry group prefixes in NLRI field
    - Needed to build bidirectional trees

- **Future — Border Gateway Multicast Protocol (BGMP)**

  Work is underway in the IETF to define a new protocol (other than PIM) that will provide for scalable Inter-domain IP Multicast. This protocol is the Border Gateway Multicast Protocol (BGMP) which has the following characteristics:

  – **Shared Tree of Domains**

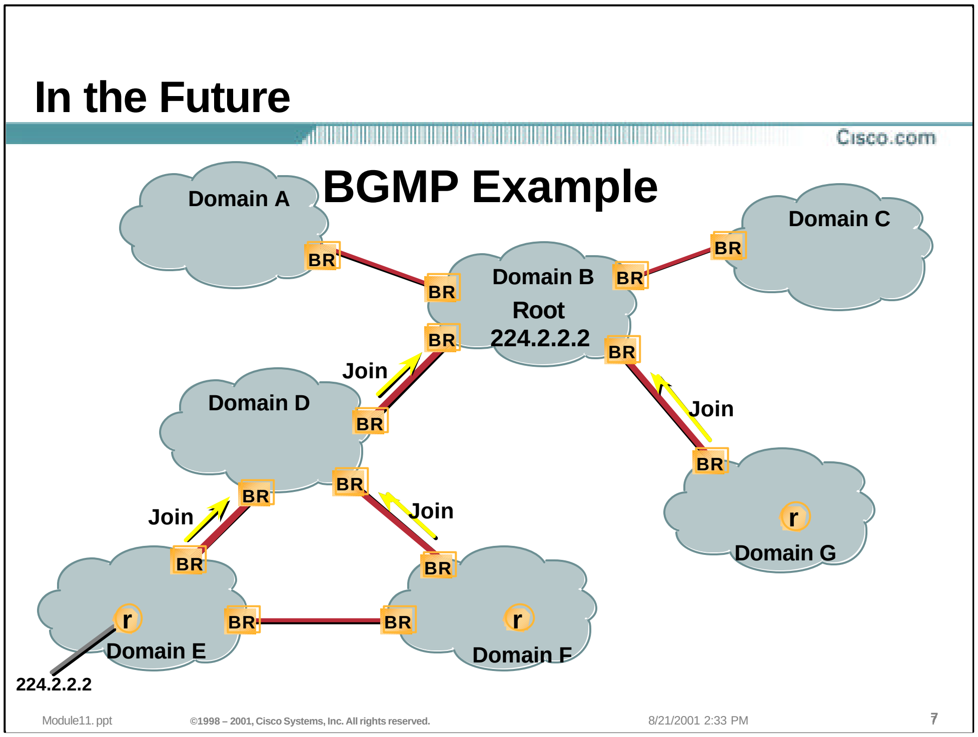
  - BGMP uses Bidirectional Shared Trees to interconnect multicast domains.
  - These trees are built using an Explicit join-model where the Join messages are sent toward the root domain.

  – **Single Root Domain per Group**

  - Each multicast domain serves as the “root” domain for a contiguous range of multicast addresses. This has the potential for Group state aggregation.
  - The address allocation method proposed for dynamically allocation these contiguous ranges is called Multicast Address Set-Claim (MASC).

  – **BGMP requires BGP4+ (MBGP)**

  - In order to build the Bidirectional Shared Trees towards a root domain, Multicast Group addresses must be distributed using MBGP in special Multicast Group NLRI. (This is not to be confused with unicast prefixes distributed in Multicast NLRI for the purpose of RPF calculation.)

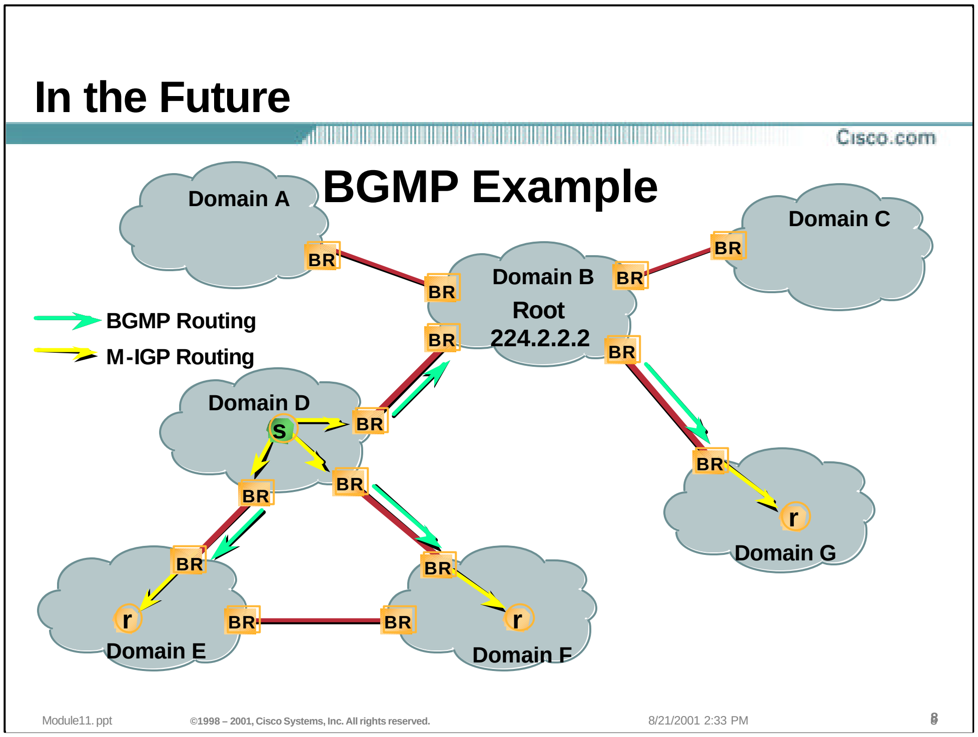


In the Future

Cisco.com

BGMP Example

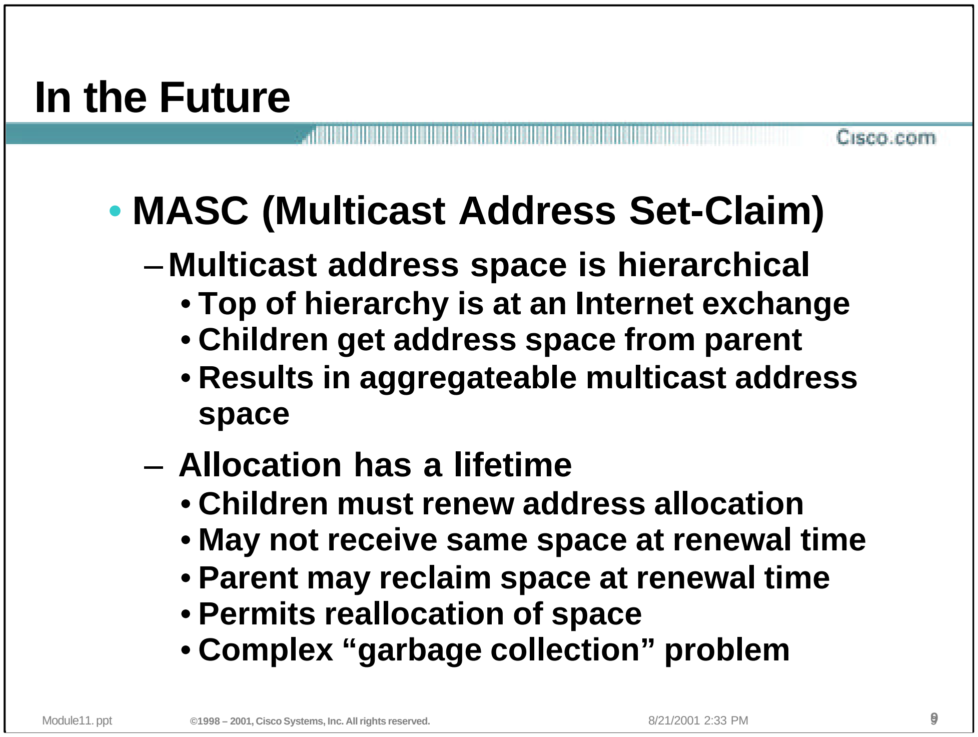
- **BGMP Example**
  - – **In the above example, multicast group 224.2.2.2 has been assigned to Domain B. Therefore, Domain B is the “root” domain for this multicast group and this fact is communicated to all other domains.**
  - – **Let’s now assume that are receiver in Domain E joins multicast group 224.2.2.2. The last-hop router “R” directly connected to the receiver would communicate this to the BGMP Border Routers (BR) by whatever method is appropriate for the multicast routing protocol (PIM, MOSPF, DVMRP) running in Domain E.**
  - – **When the BGMP BR learns that there is a receiver in its domain for group 224.2.2.2, it sends a BGMP Join for group 224.2.2.2 toward the root domain, Domain B. This Join travels domain by domain building a branch of the Bidirectional BGMP Shared Tree from the root domain to Domain E.**
  - – **If receivers in Domains F and G also join the multicast group, their last-hop routers also trigger their BGMP BR to join the Bidirectional Shared Tree by sending BGMP Joins toward the root domain.**
  - – **The end result is a Bidirectional Shared Tree that connects all domains that have active receivers for group 224.2.2.2.**



In the Future

Module11.ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 8

- **BGMP Example**
  - **Let’s now assume that a source in Domain D connected to first-hop router “S” goes active. This source traffic is routed by the Multicast IGP (PIM, MOSPF, DVMRP) to the BGMP BR’s that are on the Bidirectional Shared Tree for multicast group 224.2.2.2.**
  - **When the BGMP BR’s receive this traffic, they forward it up/down the Bidirectional Shared Tree.**
  - **The multicast traffic flows up and down the Bidirectional Shared Tree to the BGMP BR’s on all domains that are part of the Shared Tree.**
  - **The BGMP BR’s that receive the multicast traffic, forward it via the Multicast-IGP to the last-hop routers that have directly connected receivers.**
  - **Notice that in some cases, a domain is acting as a transient domain. This is the case for the “root” domain, Domain B. In this case, the multicast traffic must be forwarded by the Multicast IGP in Domain B from one BGMP BR to the other so that traffic will continue to flow down the Bidirectional Shared Tree to Domain G.**



In the Future

Cisco.com

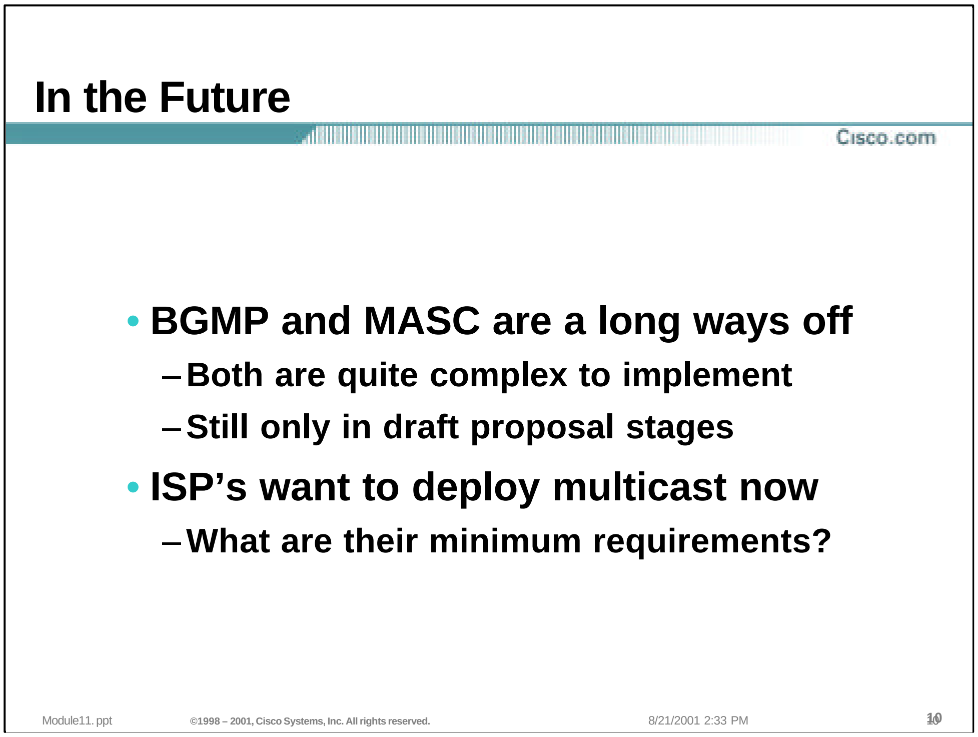
- MASC (Multicast Address Set-Claim)
  - Multicast address space is hierarchical
    - Top of hierarchy is at an Internet exchange
    - Children get address space from parent
    - Results in aggregateable multicast address space
  - Allocation has a lifetime
    - Children must renew address allocation
    - May not receive same space at renewal time
    - Parent may reclaim space at renewal time
    - Permits reallocation of space
    - Complex “garbage collection” problem

Module11. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 9

- **Multicast Address Set-Claim (MASC)**

  The IETF is also working on a new protocol to do dynamic allocation of blocks of multicast address space. This protocol is called the Multicast Address Set-Claim (MASC) protocol. This protocol has the following characteristics:

  - **Multicast Address space is hierarchical**
    - At the top of the hierarchy are one or more root MASC nodes that are responsible for the dynamic allocation of multicast group ranges. (The exact range is yet to be determined. It is very likely that MASC will, at least initially, only control a subset of the global IP multicast address space.)
    - Children request (set-claim) multicast group ranges from their parent MASC nodes. Parents allocate sub-ranges of the multicast group range(s) that they have allocated from their parent MASC nodes. This allocation scheme results in aggregateable group ranges.
  - **Allocation has a lifetime.**
    - Children must renew their allocation with their parent. However, there is no guarantee that they will get the same range at renewal time. It is possible that they may get a reduced range or a completely different range of addresses.
    - This permits parents to periodically “reclaim” address space from their children for the purpose of address reallocation.
    - This presents MASC with a complex “garbage collection” problem which is made even more complex by the fact that it must be performed in a distributed fashion across all MASC nodes in the Internet.



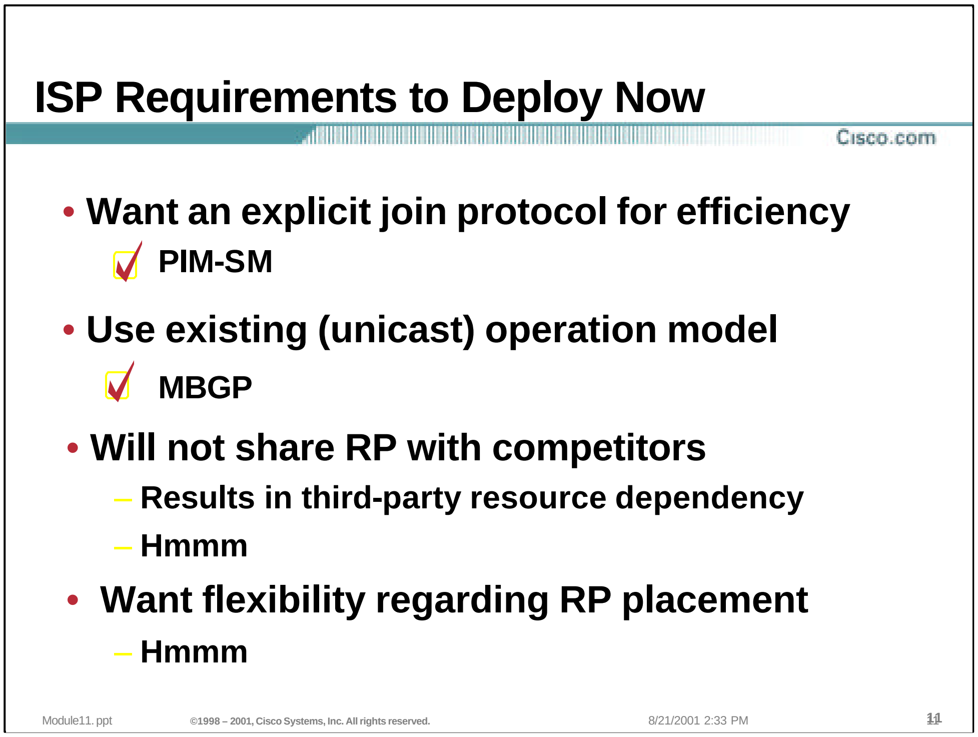
In the Future

Cisco.com

- BGMP and MASC are a long ways off
  - Both are quite complex to implement
  - Still only in draft proposal stages
- ISP’s want to deploy multicast now
  - What are their minimum requirements?

Module11. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 10

- **In the Future — A Summary**
  - **The complex nature of BGMP and MASC make them non-trivial protocols to implement. At the present time, both BGMP and MASC are only in the draft stages (and have been for quite some time).**
  - **Unfortunately, ISP’s want to deploy Inter-Domain multicast now.**



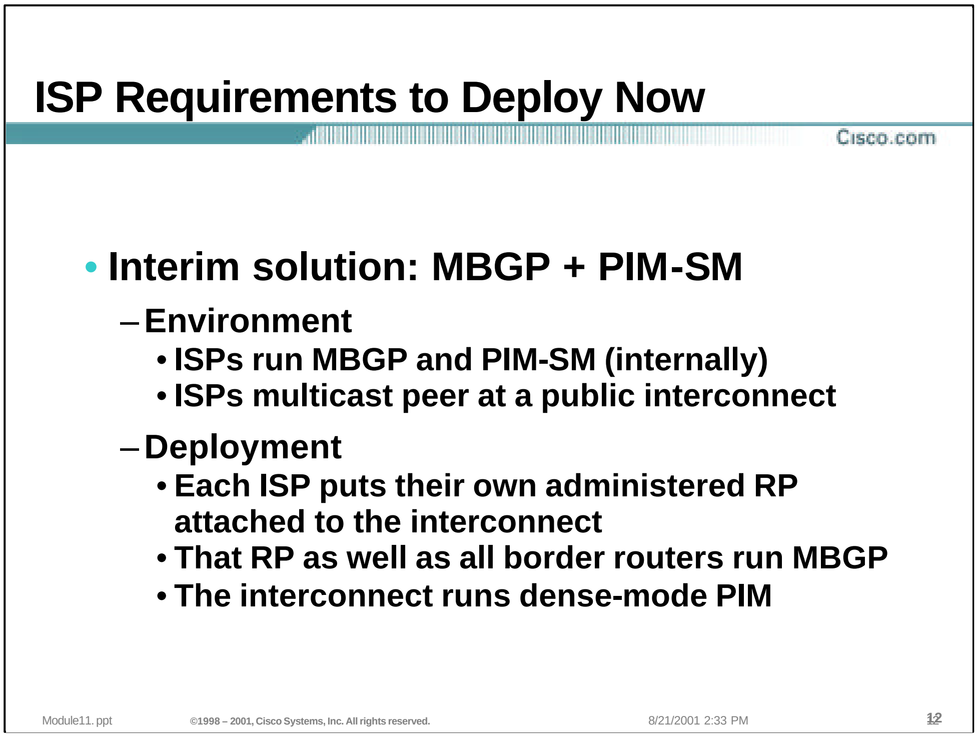
ISP Requirements to Deploy Now

Cisco.com

- Want an explicit join protocol for efficiency
  - ☑ PIM-SM
- Use existing (unicast) operation model
  - ☑ MBGP
- Will not share RP with competitors
  - – Results in third-party resource dependency
  - – Hmmm
- Want flexibility regarding RP placement
  - – Hmmm

Module11.ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 11

- **ISP Requirements to deploy IP Multicast now**
  - **Explicit Join Model Protocol**
    - The Flood and Prune behavior of Dense mode protocols preclude them from being used by most ISPs. However, this requirement can be met using PIM-SM today.
  - **Existing unicast-like operations model.**
    - ISP's want to minimize the impact of multicast on their operations model. This can be achieved by using MBGP to carry both unicast and multicast routing information. Because MBGP uses the same configuration, maintenance and policy controls as BGP, virtually no new training is required of operations staff.
  - **Will not share an RP with their competitors.**
    - This can result in a third-party dependency which is unacceptable to all ISPs. Something new is require here.
  - **Want flexibility regarding the placement of their RPs.**
    - Interfacing multiple SM domains together has always been a problem. Historically, the workaround to this was to place all SM domain RP's at a common location. This is clearly unacceptable to all ISPs and requires something new to meet this requirement.



ISP Requirements to Deploy Now

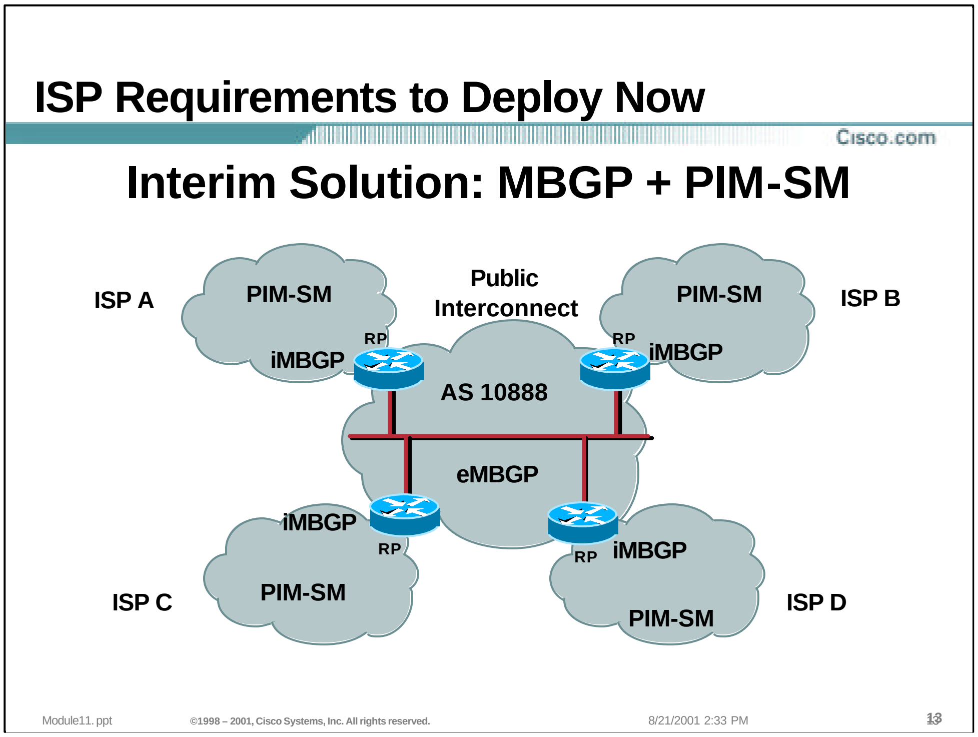
Cisco.com

- Interim solution: MBGP + PIM-SM
  - Environment
    - ISPs run MBGP and PIM-SM (internally)
    - ISPs multicast peer at a public interconnect
  - Deployment
    - Each ISP puts their own administered RP attached to the interconnect
    - That RP as well as all border routers run MBGP
    - The interconnect runs dense-mode PIM

- **Interim Solution: MBGP + PIM-SM**

  In order to move forward with Inter-domain multicast across the Internet, several ISPs agreed to an interim solution using only PIM-SM and MBGP.

  - **Environment:**
    - Each ISP ran MBGP and PIM-SM internally in their network.
    - Each ISP MBGP peered with the other multicast ISPs at a public multicast interconnect point.
  - **Deployment:**
    - Each ISP put their *single* RP at the interconnect point.
    - This router was MBGP with the other multicast routers at the interconnection point in order to exchange multicast NLRI for the purpose of RPF calculation.
    - To deal with the problems of connecting multiple PIM-SM domains together, the interconnect network was run in Dense mode. As a result, any multicast traffic sent by a source in one ISP domain would be flooded across the interconnect to all other ISP domains.



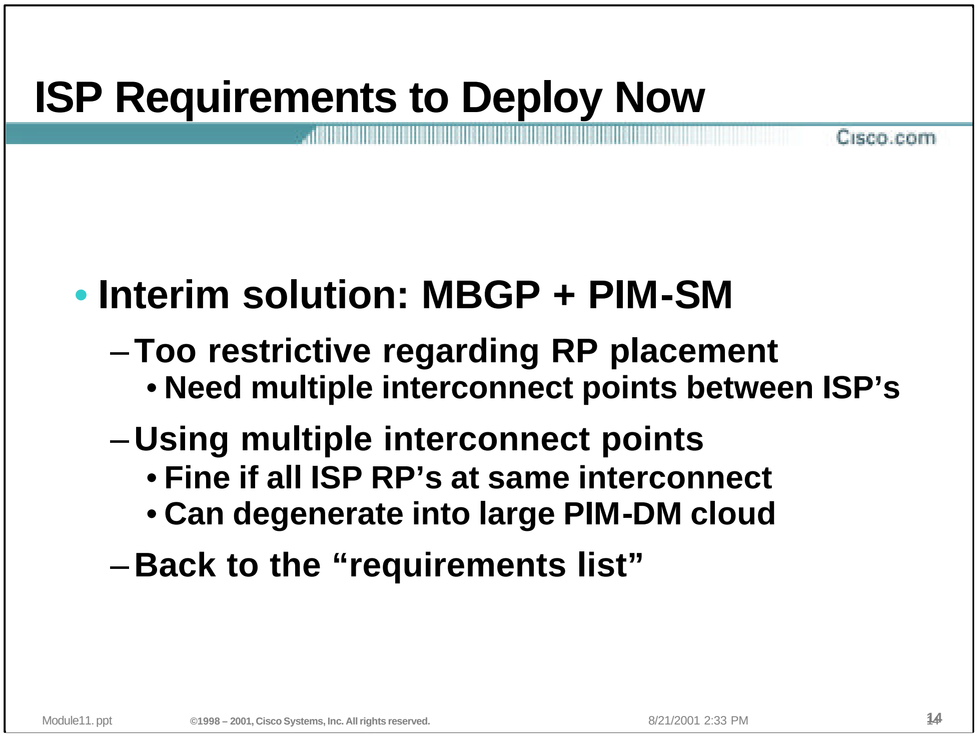
ISP Requirements to Deploy Now

Cisco.com

Interim Solution: MBGP + PIM-SM

- **Interim Solution: MBGP + PIM-SM**

  The drawing above illustrates the interim solution.



ISP Requirements to Deploy Now

Cisco.com

- Interim solution: MBGP + PIM-SM
  - –Too restrictive regarding RP placement
    - Need multiple interconnect points between ISP’s
  - –Using multiple interconnect points
    - Fine if all ISP RP’s at same interconnect
    - Can degenerate into large PIM-DM cloud
  - –Back to the “requirements list”

Module11. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 14

- **Interim Solution: MBGP + PIM-SM**

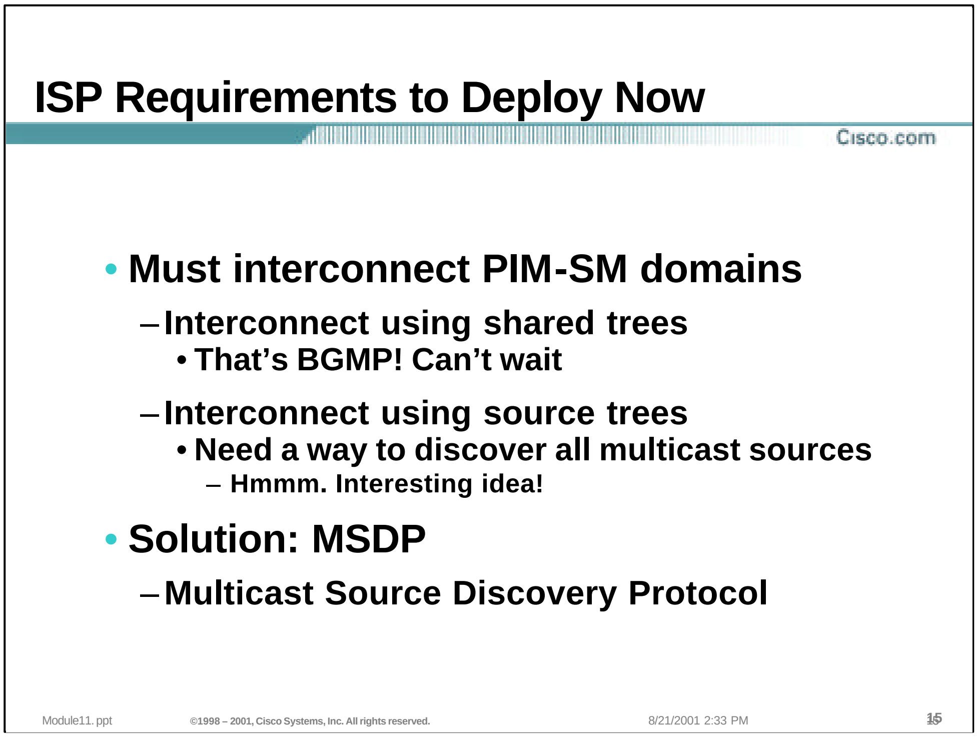
  While the interim solution allow the ISP’s to move forward and do some initial testing of native multicast across portions of the Internet, it was clearly not a solution that would scale.

  – **Too restrictive regarding RP placement.**

  - For redundancy purposes, multiple RP’s per ISP was needed.

  – **Multiple Interconnection Points**

  - Not all ISP’s share the same interconnection point. Therefore, in order to make the model work, it was necessary to link (generally via tunnels) multiple interconnection points into a single interconnection AS.
  - As the size of this interconnection AS grew, native multicast across the Internet would degenerate into a large PIM -DM cloud.



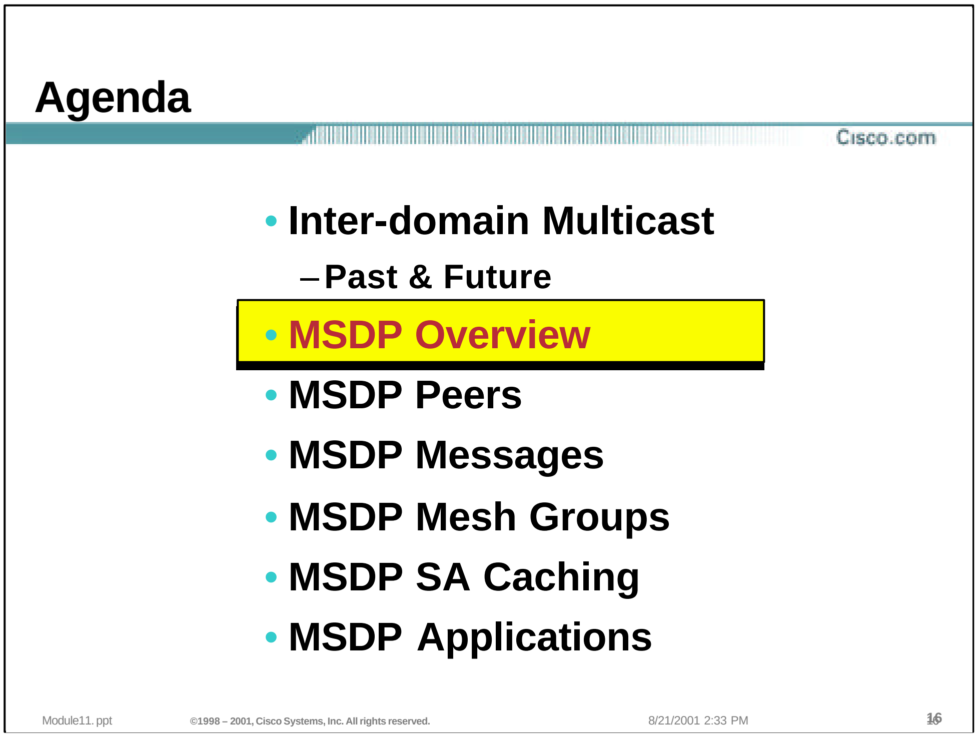
ISP Requirements to Deploy Now

Cisco.com

- Must interconnect PIM-SM domains
  - Interconnect using shared trees
    - That's BGMP! Can't wait
  - Interconnect using source trees
    - Need a way to discover all multicast sources
      - Hmmm. Interesting idea!
- Solution: MSDP
  - Multicast Source Discovery Protocol

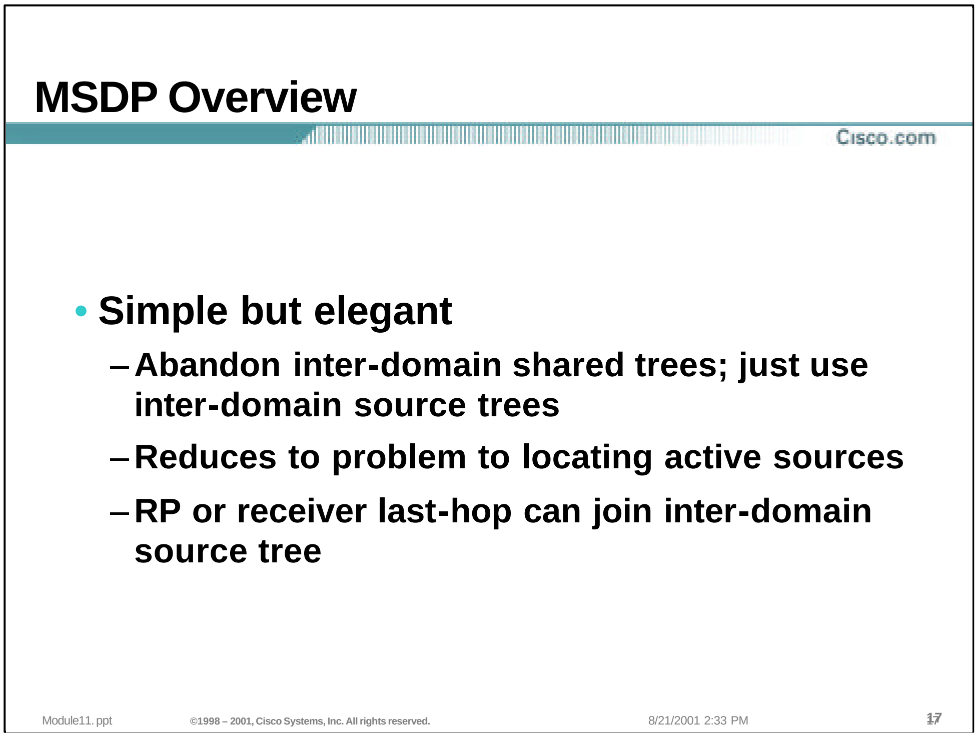
Module11. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 15

- **Must somehow interconnect multiple PIM-SM domains**
  - **Solution 1: Interconnect using Shared Trees**
    - When all is said and done, this solution results in something as complex as BGMP. In either case, ISP's didn't want to wait for something that complex to be developed.
  - **Solution 2: Interconnect using Source Trees**
    - PIM-SM has the ability to send explicit (S,G) Joins toward the source and to join the source tree. This is, of course, the basic way that an RP receives (S,G) traffic from sources within their domains. (That is to say, by explicitly joining the Source Tree.)
    - If RP's could somehow *learn* of the existence of active sources in other PIM-SM domains, it could simply send (S,G) Joins toward those sources and join their Source Trees. This resulted in the concepts of the Multicast Source Discovery Protocol (MSDP).



Agenda

- Inter-domain Multicast
  - Past & Future
- MSDP Overview
- MSDP Peers
- MSDP Messages
- MSDP Mesh Groups
- MSDP SA Caching
- MSDP Applications



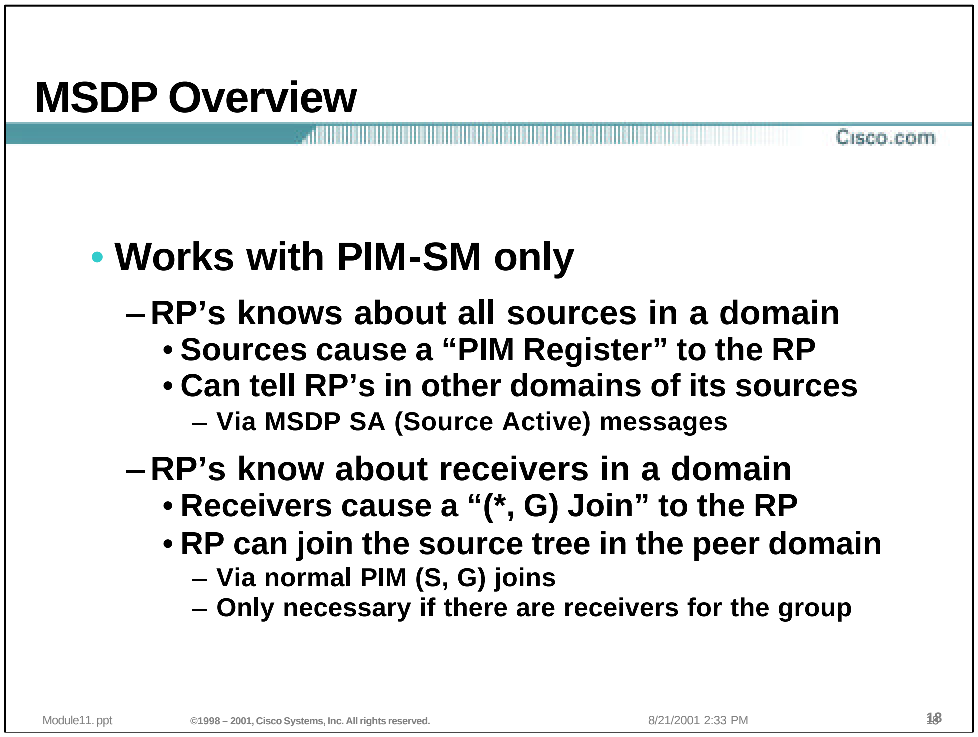
MSDP Overview

Cisco.com

- Simple but elegant
  - Abandon inter-domain shared trees; just use inter-domain source trees
  - Reduces to problem to locating active sources
  - RP or receiver last-hop can join inter-domain source tree

Module11. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 17

- **MSDP Overview**
  - **By abandoning the notion of inter-domain Shared-Trees and using only inter-domain Source-Trees, the complexity of interconnecting PIM-SM domains is reduced considerably.**
  - **The remaining problem becomes one of communicating the existence of active sources between the RP's in the PIM-SM domains.**
  - **The RP can join the inter-domain source-tree for sources that are sending to groups for which the RP has receivers. This is possible because the RP is the root of the Shared-Tree which has branches to all points in the domain where there are active receivers.**
    - Note: If the RP either has no Shared-Tree for a particular group or a Shared-Tree whose outgoing interface list is Null, it does not send a Join to the source in another domain.
  - **Once a last-hop router learns of a new source outside the PIM-SM domain (via the arrival of a multicast packet from the source down the Shared-Tree), it too can send a Join toward the source and join the source tree.**



MSDP Overview

- Works with PIM-SM only
  - –RP’s knows about all sources in a domain
    - Sources cause a “PIM Register” to the RP
    - Can tell RP’s in other domains of its sources
      - – Via MSDP SA (Source Active) messages
  - –RP’s know about receivers in a domain
    - Receivers cause a “(*, G) Join” to the RP
    - RP can join the source tree in the peer domain
      - – Via normal PIM (S, G) joins
      - – Only necessary if there are receivers for the group

- **MSDP Overview**

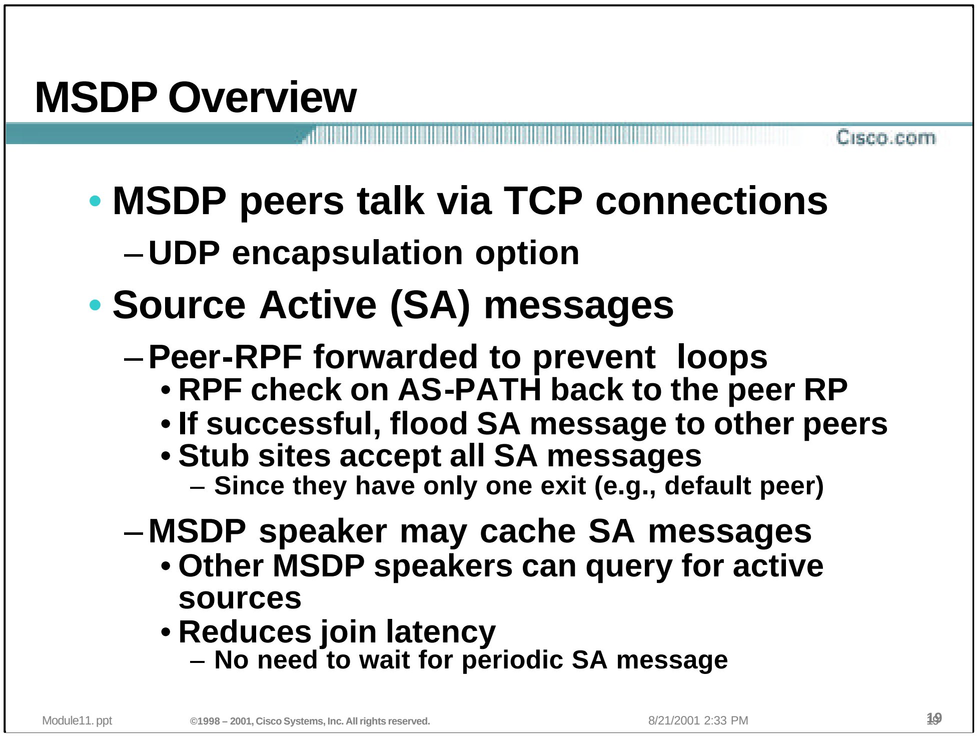
  The entire concept of MSDP depends on the RP’s in the inter-connected domains being the be-all, know-all oracle that is aware of all sources and receivers in its local domain. As a result, MSDP can only work with PIM-SM.

  – **RP’s know about all sources in a domain.**

  - Whenever a source goes active in a PIM-SM domain, the first-hop router immediately informs the RP of this via the use of PIM Register messages.
  - (S,G) state for the source is kept alive in the RP by normal PIM-SM mechanisms as long the source is actively sending. As a result, the RP can inform other RP’s in other PIM-SM domains of the existence of active sources in its local domain. This is accomplished via MSDP Source Active (SA) messages.

  – **RP’s know about receivers in a domain.**

  - Receivers cause last-hop routers to send (*, G) Joins to the RP to build branches of a Shared-Tree for a group.
  - If an RP has (*, G) state for a group and the outgoing interface list of the (*, G) entry is not Null, it knows it has active receivers for the group. Therefore, when it receives an SA message announcing an active source for group G in another domain, it can send (S, G) Joins toward the source in the other domain.



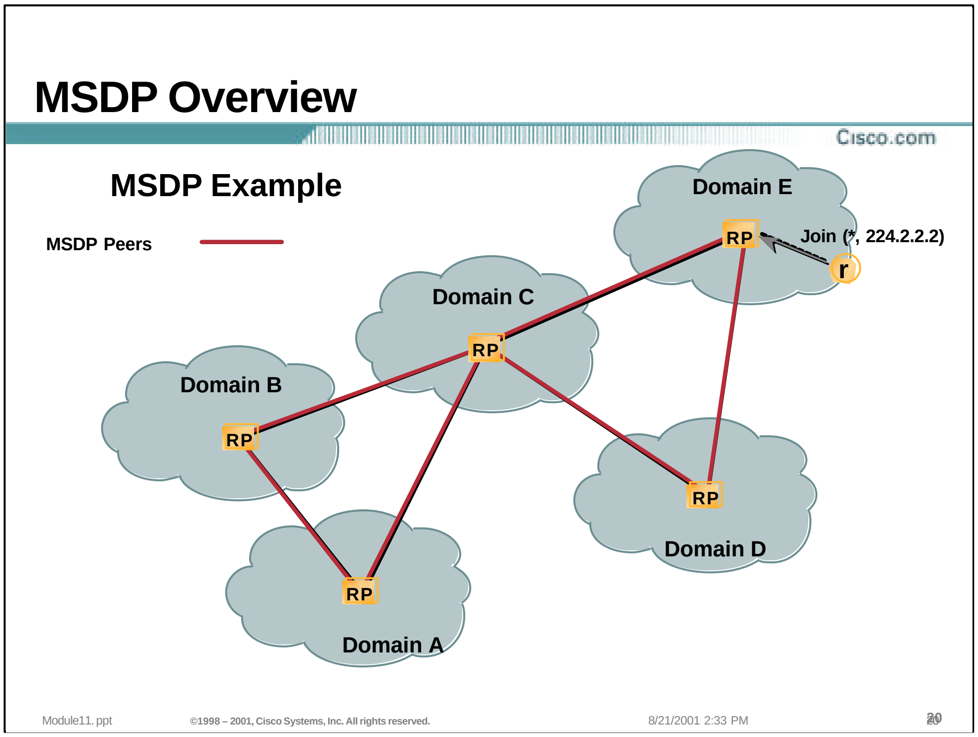
MSDP Overview

Cisco.com

- MSDP peers talk via TCP connections
  - UDP encapsulation option
- Source Active (SA) messages
  - Peer-RPF forwarded to prevent loops
    - RPF check on AS-PATH back to the peer RP
    - If successful, flood SA message to other peers
    - Stub sites accept all SA messages
      - Since they have only one exit (e.g., default peer)
  - MSDP speaker may cache SA messages
    - Other MSDP speakers can query for active sources
    - Reduces join latency
      - No need to wait for periodic SA message

Module11.ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 19

- **MSDP Overview**
  - **MSDP Peers (typically RP’s) are connected via TCP sessions.**
    - Note: The MSDP specification describes a UDP encapsulation option but this is not currently available in the IOS implementation.
  - **Source Active (SA) Messages**
    - RP’s periodically originate SA messages for sources that are active in their local domain. These SA messages are sent to all active MSDP peers.
    - When a MSDP speaker receives an SA messages from one of its peers, it is RPF forwarded to all of its other peers. An RPF check is performed on the arriving SA message (using the originating RP address in the SA message) to insure that it was received via the correct AS-PATH. Only if this RPF check succeeds is the SA message flooded downstream to its peers. This prevents SA messages from looping through the Internet.
    - Stub domains (i.e. domains with only a single MSDP connection) do not have to perform this RPF check since there is only a single entrance/exit.
    - MSDP speakers may cache SA messages. Normally, these messages are not stored to minimize memory usage. However, by storing SA messages, join latency can be reduced as RP’s do not have to wait for the arrival of periodic SA messages when the first receiver joins the group. Instead, the RP can scan its SA cache to immediately determine what sources are active and send (S, G) Joins.
    - Non-caching MSDP speakers can query caching MSDP speakers in the same domain for information on active sources for a group.



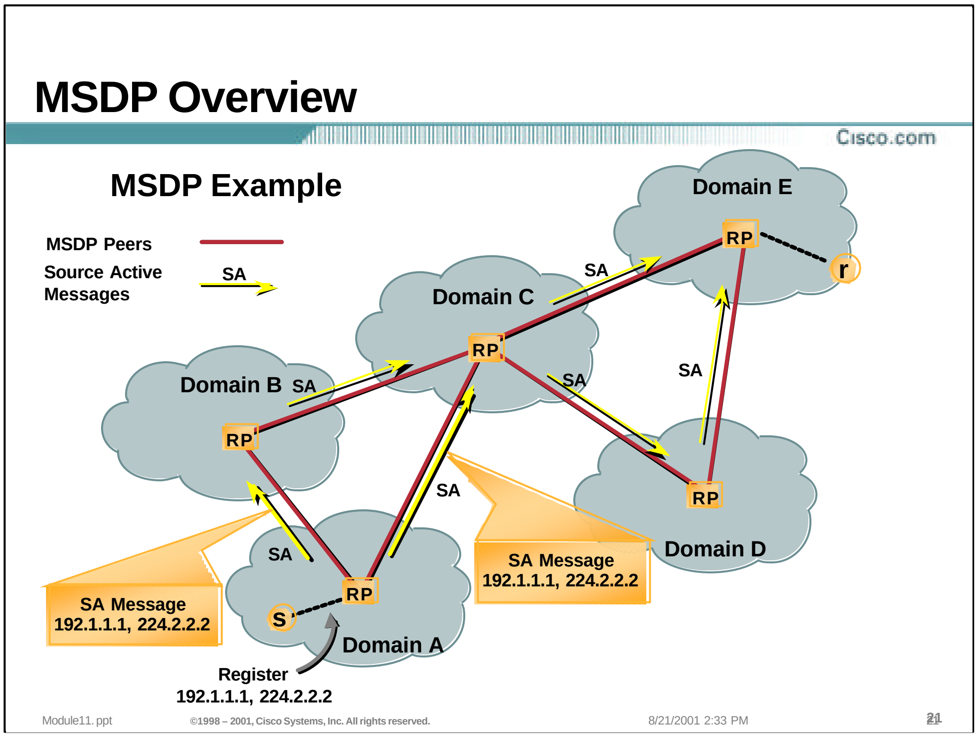
MSDP Overview

Cisco.com

MSDP Example

Module11.ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 20

- **MSDP Example**
  - **In the example above, PIM-SM domains A through E each have an RP which is an MSDP speaker. The solid lines between these RP's represents the MSDP peer sessions via TCP and not actual physical connectivity between the domains.**
    - *Note: The physical connectivity between the domains is not shown in the drawing above.*
  - **Assume that a receiver in Domain E joins multicast group 224.2.2.2 which in turn, causes its DR to send (*, G) Join for this group to the RP.**
  - **This builds a branch of the Shared-Tree from the RP in Domain E to the DR as shown.**
  - **When a source goes active in Domain A, the first-hop router (S) sends a PIM Register message to the RP. This informs the RP in Domain A that a source is active in the local domain. The RP responds by originating an (S, G) SA message for this source and send them to its MSDP peers in domains B and C. (The RP will continue to send these SA messages periodically as long as the source remains active.)**
  - **When the RP's in domains B and C receive the SA messages, they are RPF checked and forwarded downstream to their MSDP peers. These SA messages continue to travel downstream and eventually reach the MSDP peers (the RP's) in domains D and E.**
    - Note: The SA message traveling from domain B to domain C, will fail the RPF check at the domain C RP (MSDP speaker) and will be dropped. However, the SA message arriving at domain C from domain A will RPF correctly and will be processed and forwarded on to domains D and E.



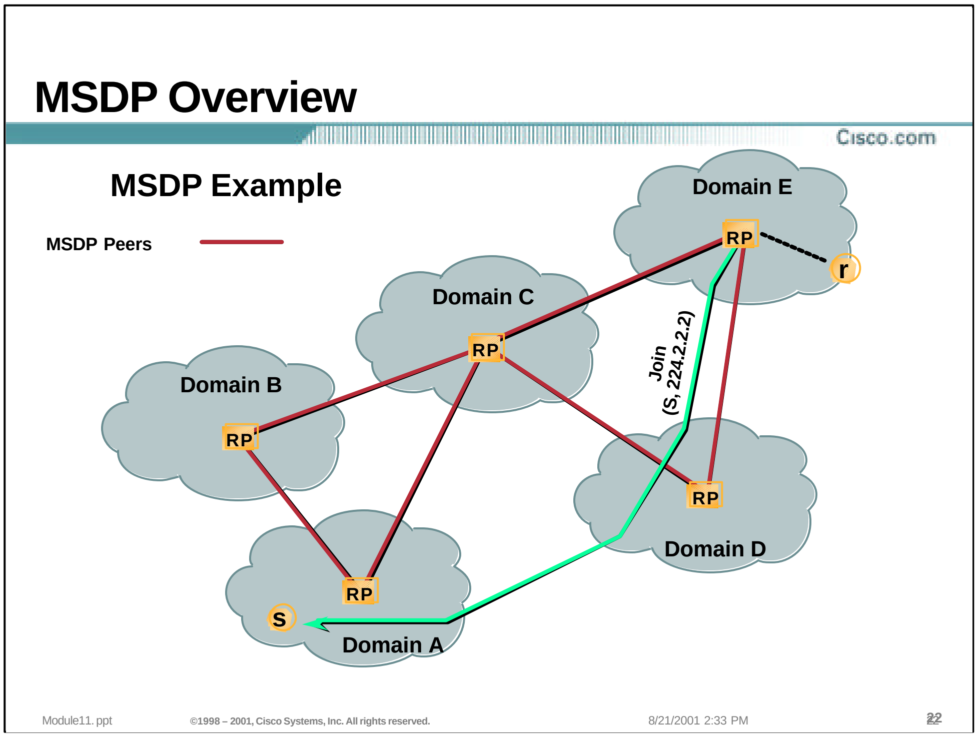
MSDP Overview

Cisco.com

MSDP Example

Module11.ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 21

- **MSDP Example**
  - **In the example above, PIM-SM domains A through E each have an RP which is an MSDP speaker. The solid lines between these RP's represents the MSDP peer sessions via TCP and not actual physical connectivity between the domains.**
    - *Note: The physical connectivity between the domains is not shown in the drawing above.*
  - **Assume that a receiver in Domain E joins multicast group 224.2.2.2 which in turn, causes its DR to send (*, G) Join for this group to the RP.**
  - **This builds a branch of the Shared-Tree from the RP in Domain E to the DR as shown.**
  - **When a source goes active in Domain A, the first-hop router (S) sends a PIM Register message to the RP. This informs the RP in Domain A that a source is active in the local domain. The RP responds by originating an (S, G) SA message for this source and send them to its MSDP peers in domains B and C. (The RP will continue to send these SA messages periodically as long as the source remains active.)**
  - **When the RP's in domains B and C receive the SA messages, they are RPF checked and forwarded downstream to their MSDP peers. These SA messages continue to travel downstream and eventually reach the MSDP peers (the RP's) in domains D and E.**
    - Note: The SA message traveling from domain B to domain C, will fail the RPF check at the domain C RP (MSDP speaker) and will be dropped. However, the SA message arriving at domain C from domain A will RPF correctly and will be processed and forwarded on to domains D and E.

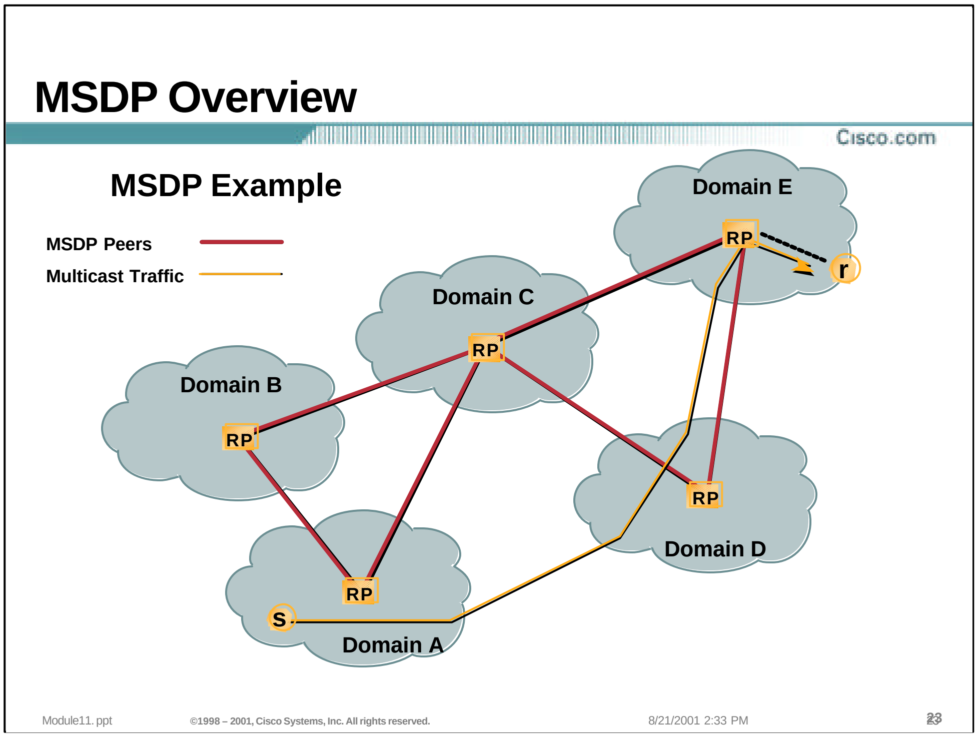


MSDP Overview

Cisco.com

MSDP Example

- **MSDP Example**
  - **Once the SA message arrives at the RP (MSDP speaker) in domain E, it sees that it has an active branch of the Shared-Tree for group 224.2.2.2. It responds to the SA message by sending an (S, G) Join toward the source.**
    - IMPORTANT: The (S, G) Join will follow the normal inter-domain routing path from the RP to the source. This inter-domain routing path is not necessarily the same path as that used by the MSDP connections. *In order to emphasis this point, the (S, G) Join is shown following a different path between domains.*

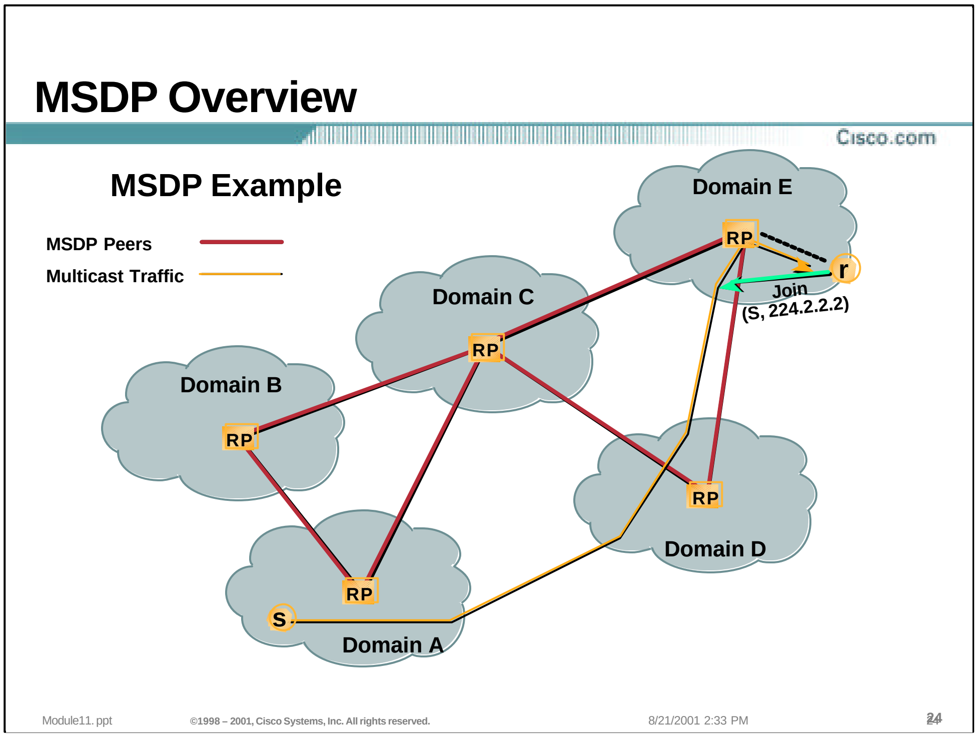


MSDP Overview

Cisco.com

MSDP Example

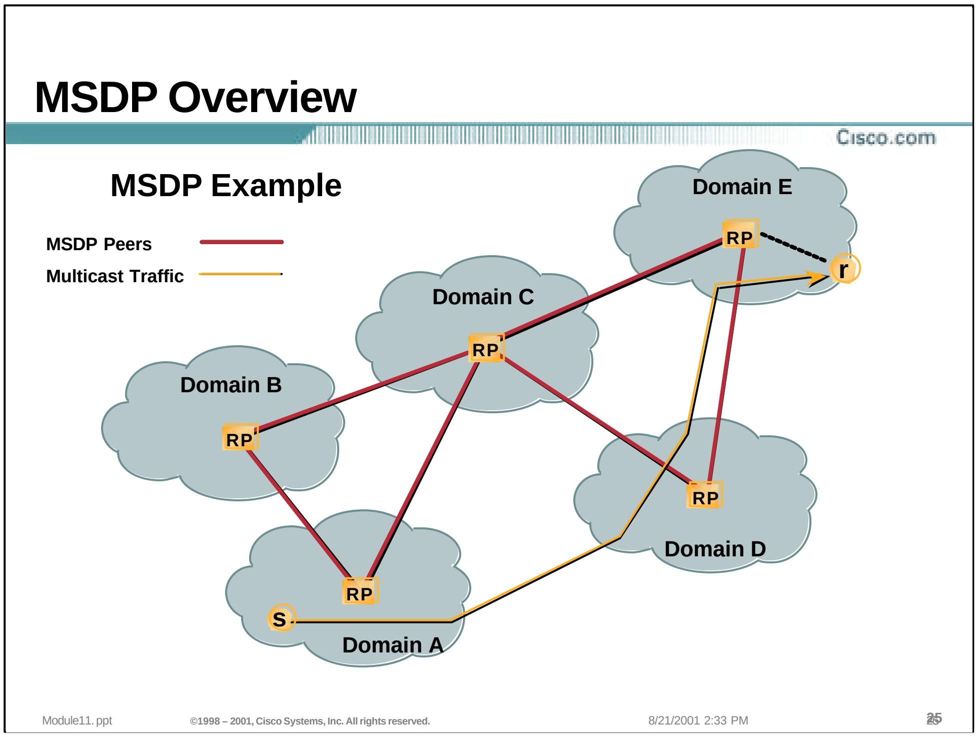
- **MSDP Example**
  - **Once the (S, G) Join message reaches the first-hop router (S) in domain A, (S, G) traffic begins to flow to the RP in domain E via the Source Tree shown.**
    - IMPORTANT: The (S, G) traffic will not flow over the TCP MSDP sessions. It will instead follow the path of the Source Tree that was built in the preceding step.



MSDP Overview

MSDP Example

- **MSDP Example**
  - **Once the (S, G) traffic reaches the last-hop router (R) in domain E, the last-hop router may optionally send an (S, G) Join toward the source in order to bypass the RP in domain E. This is shown in the above example.**

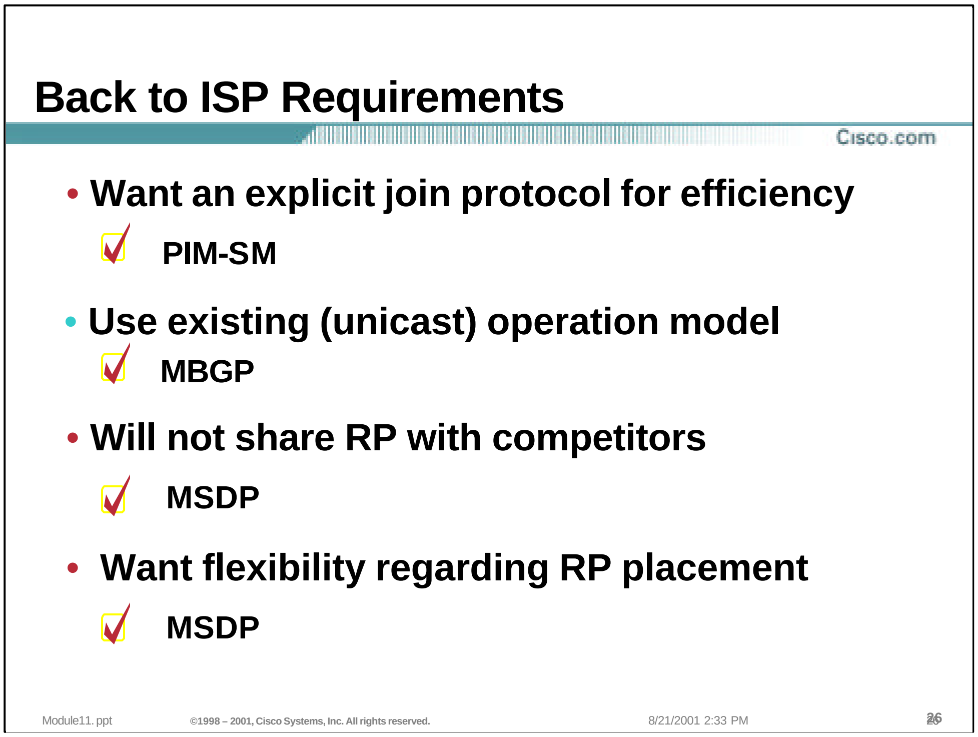


MSDP Overview

Cisco.com

MSDP Example

- **MSDP Example**
  - **At this point in the example, the (S, G) traffic is flowing to the last-hop router (R) in domain E via the Source-Tree as shown in the above example.**



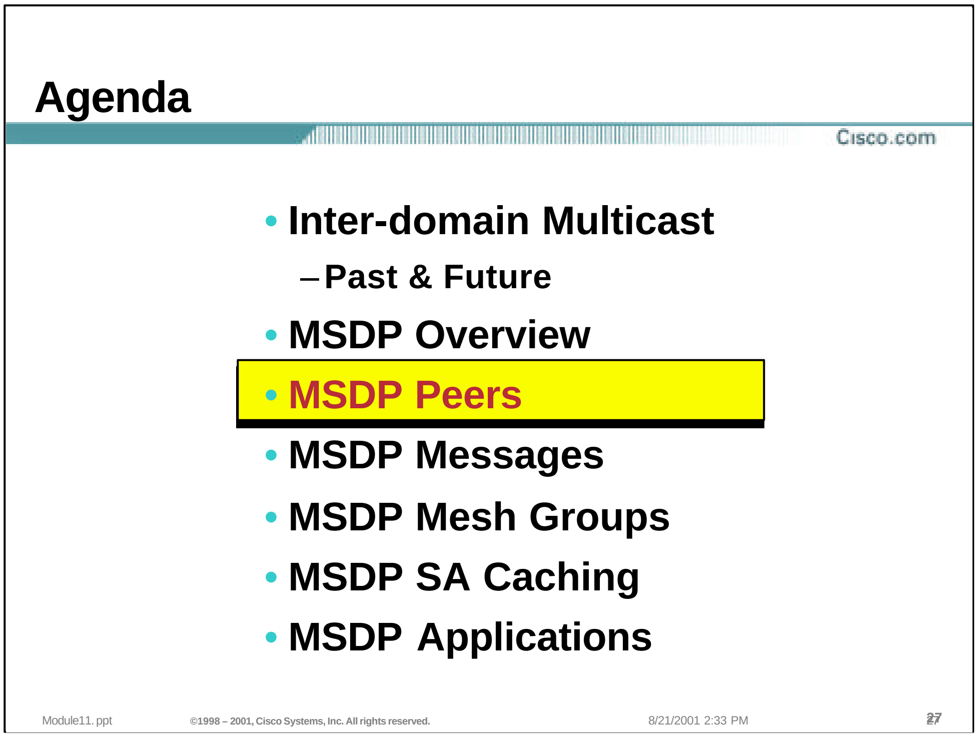
Back to ISP Requirements

Cisco.com

- Want an explicit join protocol for efficiency
  - ☑ PIM-SM
- Use existing (unicast) operation model
  - ☑ MBGP
- Will not share RP with competitors
  - ☑ MSDP
- Want flexibility regarding RP placement
  - ☑ MSDP

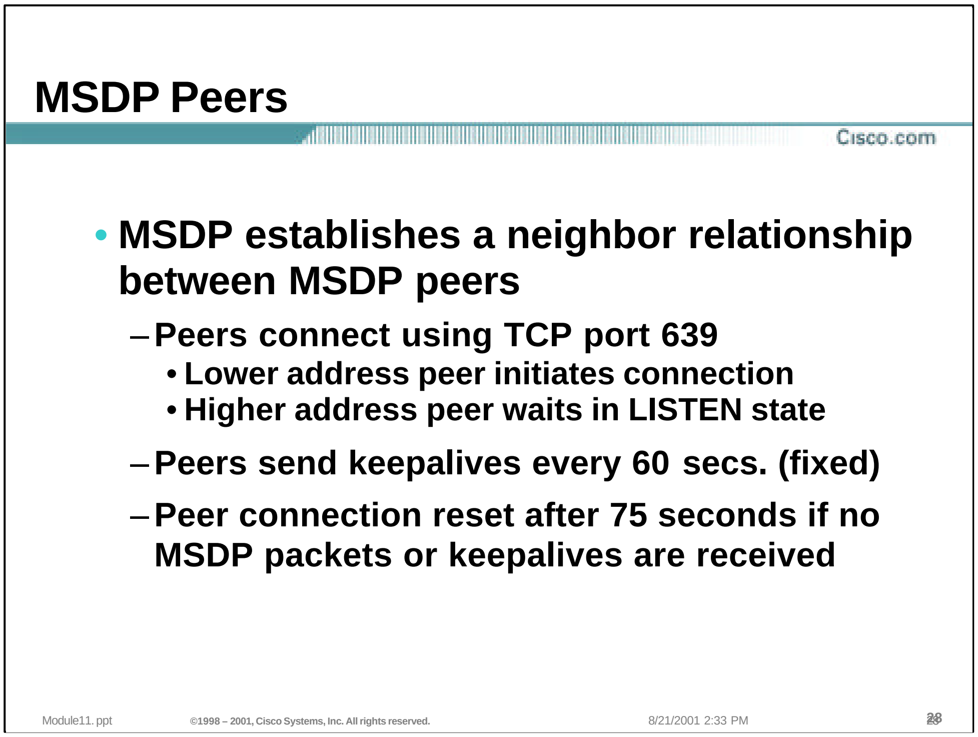
Module11. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 26

- **ISP Requirements Revisited**
  - **The requirements have all now been met using a combination of PIM-SM, MBGP and MSDP.**
  - **Want an Explicit Join Protocol**
    - Again, PIM-SM meets this requirement.
  - **Want to use an existing unicast operations model.**
    - The extension of BGP to MBGP (BGP4+) permits both unicast and multicast traffic flows to be configured and managed using the same exist set of tools.
  - **Will not share RP's with competitors**
    - MSDP permits each PIM-SM domain to have its own RP for each group.
  - **Want flexibility regarding placement of the RPs**
    - RP's can be placed anywhere in the PIM-SM domain as long as they are linked via MSDP to other RP's in other domains. (Note: It is not necessary to have a full mesh of MSDP connections. It is sufficient to have a single MSDP connection that leads to the rest of the MSDP speakers (RP's) in the Internet.)



Agenda

- Inter-domain Multicast
  - Past & Future
- MSDP Overview
- MSDP Peers
- MSDP Messages
- MSDP Mesh Groups
- MSDP SA Caching
- MSDP Applications

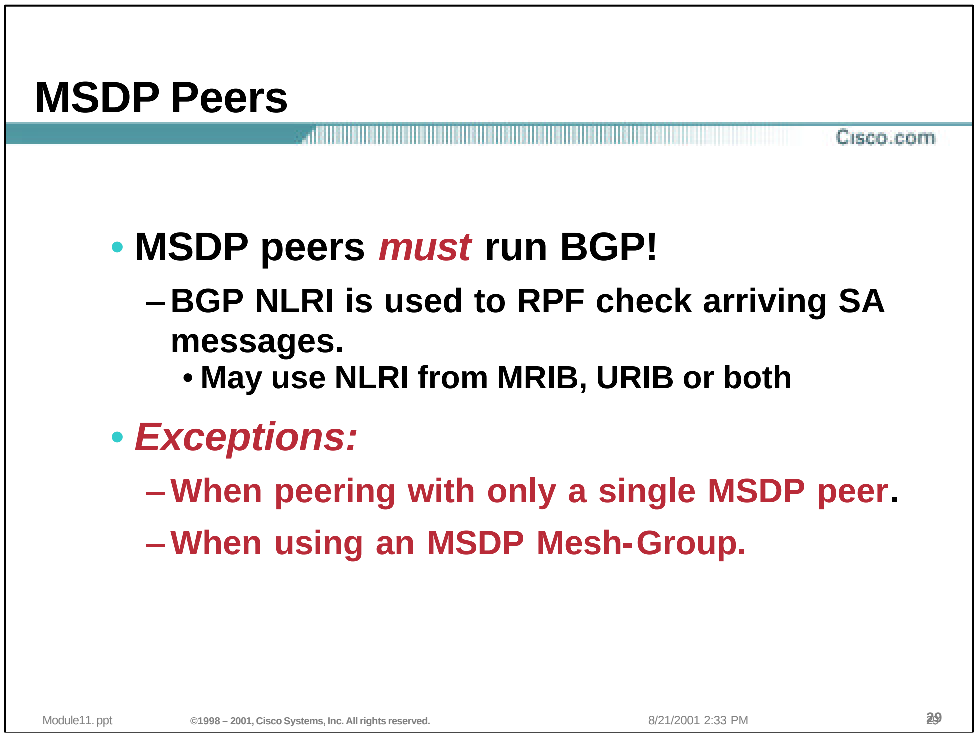


MSDP Peers

- MSDP establishes a neighbor relationship between MSDP peers
  - –Peers connect using TCP port 639
    - Lower address peer initiates connection
    - Higher address peer waits in LISTEN state
  - –Peers send keepalives every 60 secs. (fixed)
  - –Peer connection reset after 75 seconds if no MSDP packets or keepalives are received

Module11. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 28

- **MSDP Peers**
  - **– Like BGP, MSDP establishes neighbor relationships with other MSDP peers.**
    - MSDP peers connect using TCP port 639. The lower IP address peer takes the active role of opening the TCP connection. The higher IP address peer waits in LISTEN state for the other to make the connection.
    - MSDP peers send Keepalives every 60. The arrival of data performs the same function as the Keepalive and keeps the session from timing out.
    - If no Keepalive or data is received for 75 seconds, the TCP connection is reset and reopened.

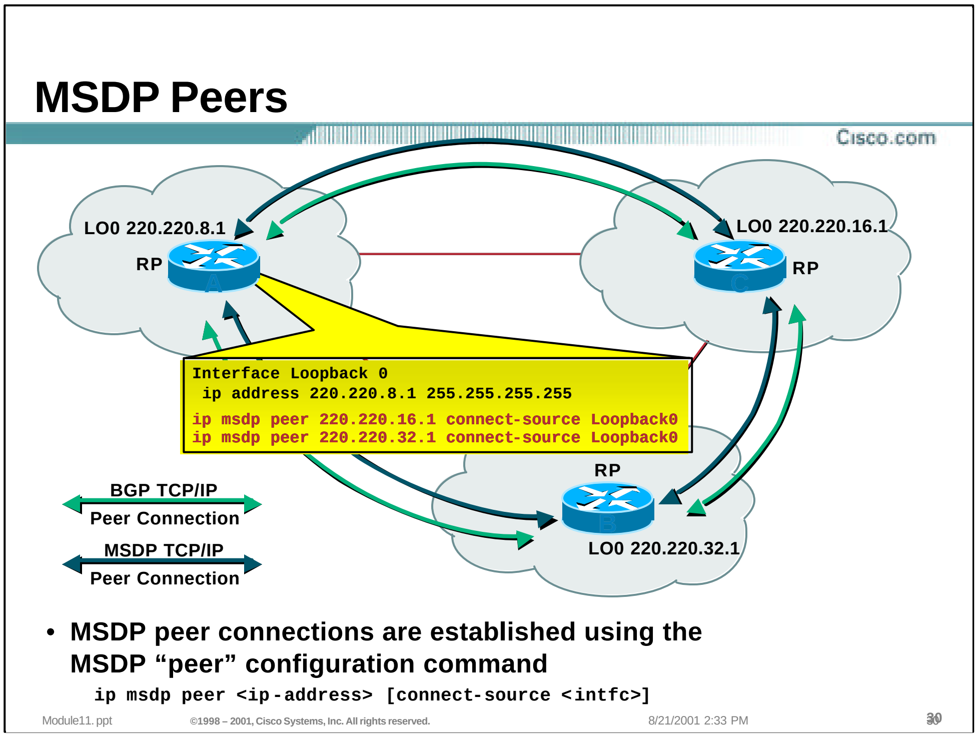


MSDP Peers

Cisco.com

- MSDP peers *must* run BGP!
  - BGP NLRI is used to RPF check arriving SA messages.
    - May use NLRI from MRIB, URIB or both
- *Exceptions:*
  - When peering with only a single MSDP peer.
  - When using an MSDP Mesh-Group.

- **MSDP Peers**
  - **MSDP speakers *must* run BGP.**
    - This requirement is due to the fact that the SA message RPF check mechanism uses AS-PATH information contained in the MBGP M-RIB or U-RIB.
    - There are some special cases where the requirement to perform an RPF check on the arriving SA message is suspended. This is the case when there is only a single MSDP peer connection or if the MSDP mesh groups are in use. In these cases, (M)BGP is not necessary.



MSDP Peers

```
Interface Loopback 0
 ip address 220.220.8.1 255.255.255.255
ip msdp peer 220.220.16.1 connect-source Loopback0
ip msdp peer 220.220.32.1 connect-source Loopback0
```

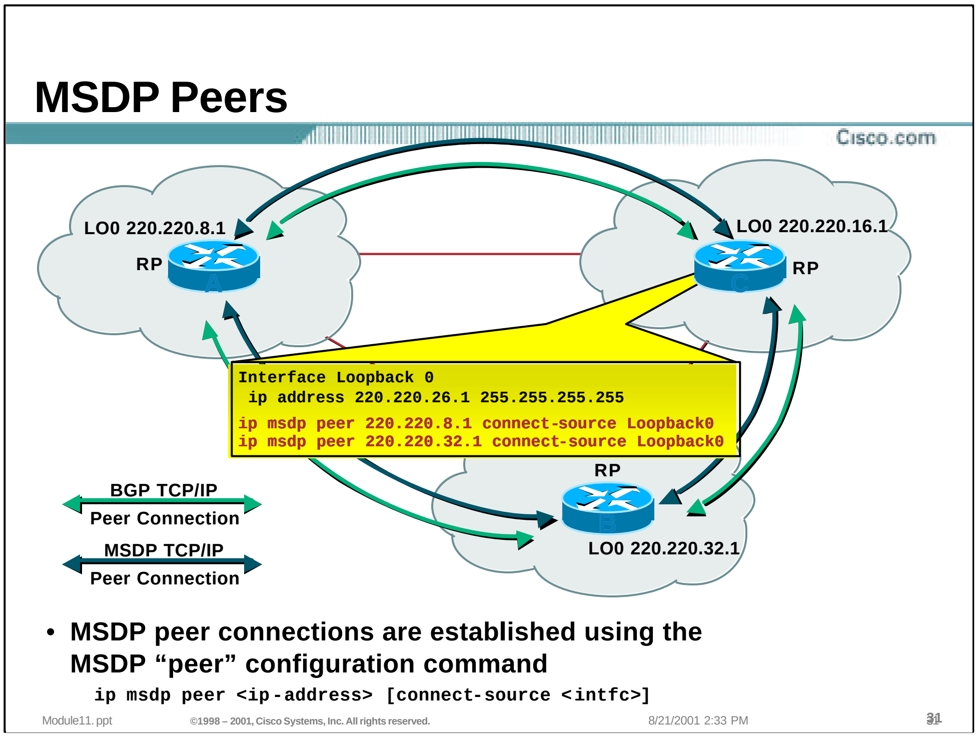
- MSDP peer connections are established using the MSDP “peer” configuration command

```
ip msdp peer <ip-address> [connect-source <intfc>]
```

- **MSDP Peers**
  - **Peer connections are establish by the use of the following IOS command:**

    ```
    ip msdp peer <ip-address> [connect-source <interface>]
    ```

  - **In the above example Router A has MSDP peer connections with both Routers B and C using their Loopback address as the connection address.**



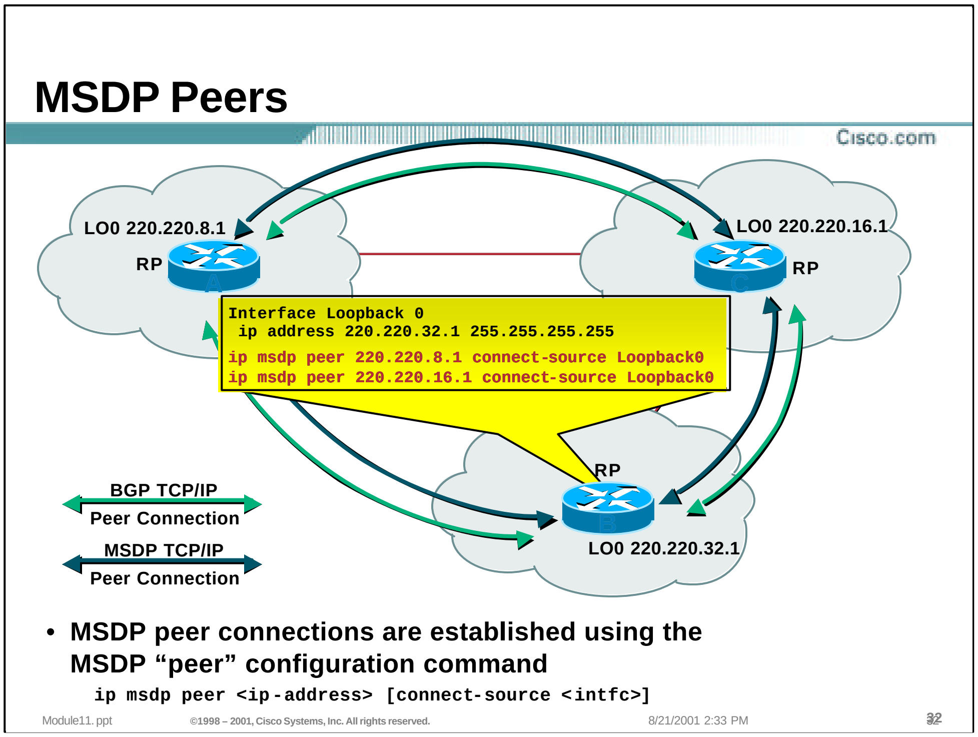
MSDP Peers

```
Interface Loopback 0
 ip address 220.220.26.1 255.255.255.255
ip msdp peer 220.220.8.1 connect-source Loopback0
ip msdp peer 220.220.32.1 connect-source Loopback0
```

- MSDP peer connections are established using the MSDP “peer” configuration command

```
ip msdp peer <ip-address> [connect-source <intfc>]
```

- **MSDP Peers**
  - in the above example Router C has MSDP peer connections with both Routers A and B using their Loopback address as the connection address.



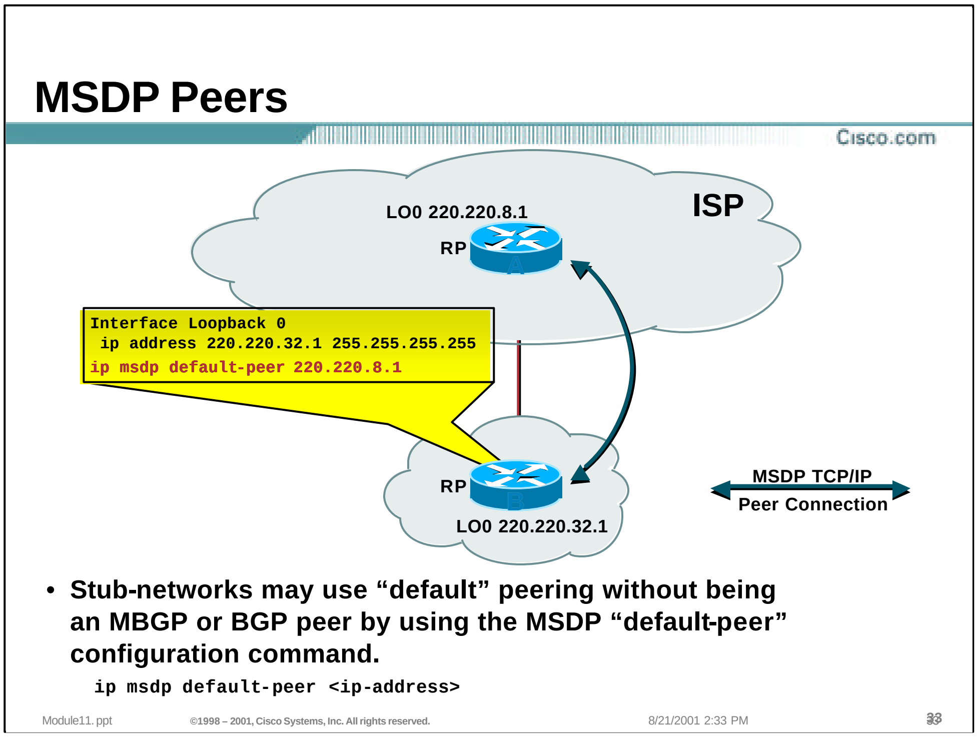
MSDP Peers

```
Interface Loopback 0
 ip address 220.220.32.1 255.255.255.255
ip msdp peer 220.220.8.1 connect-source Loopback0
ip msdp peer 220.220.16.1 connect-source Loopback0
```

- MSDP peer connections are established using the MSDP “peer” configuration command

```
ip msdp peer <ip-address> [connect-source <intfc>]
```

- **MSDP Peers**
  - in the above example Router B has MSDP peer connections with both Routers A and C using their Loopback address as the connection address.



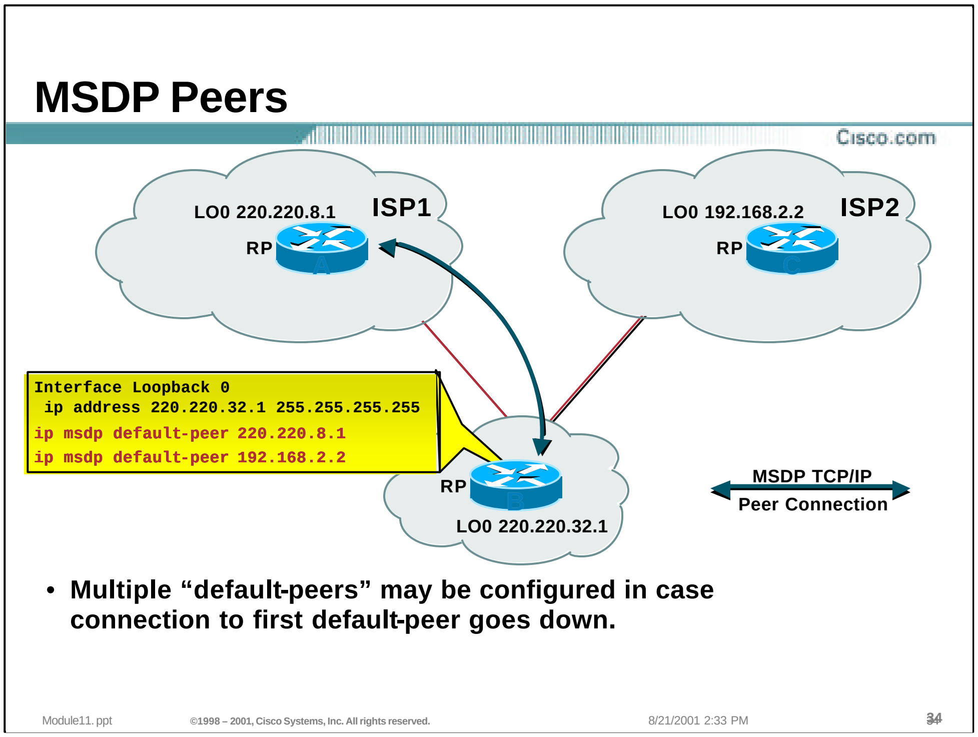
MSDP Peers

```
Interface Loopback 0
 ip address 220.220.32.1 255.255.255.255
ip msdp default-peer 220.220.8.1
```

- Stub-networks may use “default” peering without being an MBGP or BGP peer by using the MSDP “default-peer” configuration command.

```
ip msdp default-peer <ip-address>
```

- **MSDP Peers**
  - **Stub networks may use “default” peering to a single MSDP peer. This eliminates the need to run (M)BGP to have the information necessary to perform the RPF check on arriving SA messages.**
    - Note: Since there is only a single connection, there is no need to perform the RPF check since this is the only path that SA messages can take.
  - **The format of the IOS command that establishes a default peer connection is:**
    - ip msdp default-peer <ip-address>



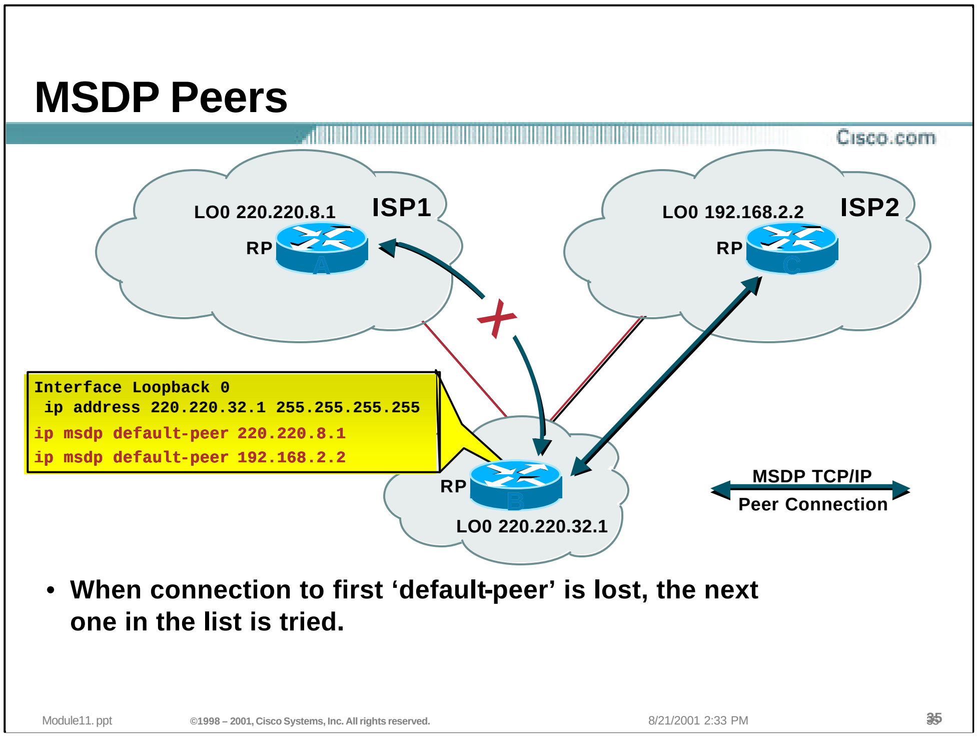
MSDP Peers

Cisco.com

```
Interface Loopback 0
 ip address 220.220.32.1 255.255.255.255
ip msdp default-peer 220.220.8.1
ip msdp default-peer 192.168.2.2
```

- Multiple “default-peers” may be configured in case connection to first default-peer goes down.

- **MSDP Peers**
  - Stub networks may configure additional secondary “default” peer connections to provide some redundancy in case the primary “default” peer goes down.
  - In the above example, the primary “default” peer connection is to Router A (220.220.8.1). The secondary “default” peer connection is to Router C. This connection will not be activated unless the connect to Router A is lost.

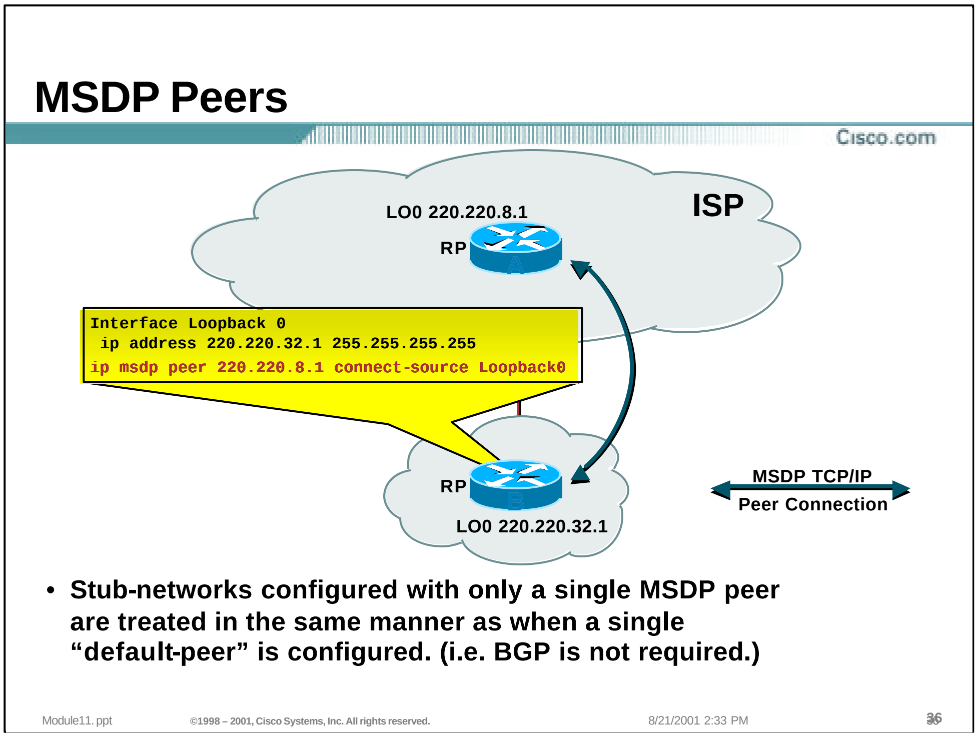


MSDP Peers

```
Interface Loopback 0
 ip address 220.220.32.1 255.255.255.255
ip msdp default-peer 220.220.8.1
ip msdp default-peer 192.168.2.2
```

- When connection to first ‘default-peer’ is lost, the next one in the list is tried.

- **MSDP Peers**
  - **Continuing with the previous example, the primary “default” peer connection is to Router A (220.220.8.1) has gone down. The secondary “default” peer connection is to Router C is now activated.**

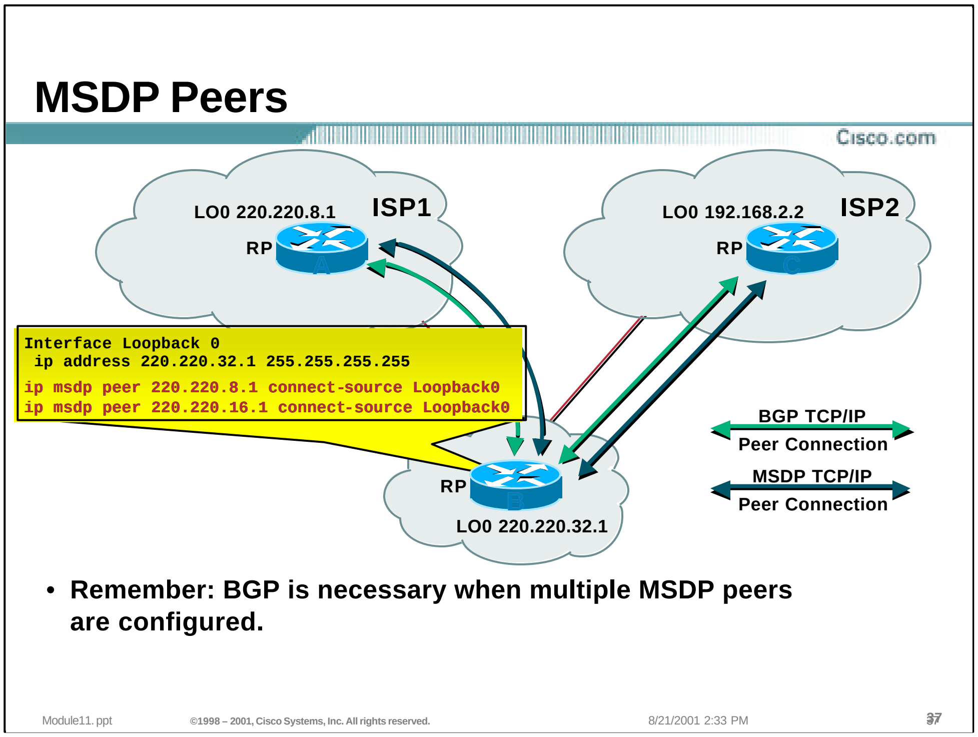


MSDP Peers

```
Interface Loopback 0
 ip address 220.220.32.1 255.255.255.255
ip msdp peer 220.220.8.1 connect-source Loopback0
```

- **Stub-networks configured with only a single MSDP peer are treated in the same manner as when a single “default-peer” is configured. (i.e. BGP is not required.)**

- **MSDP Peers**
  - **Stub networks may configure a single MSDP peer using the normal ‘ip msdp peer’ IOS command. When only a single MSDP peer is configured in this manner, it is treated in the same manner as a “default” peering. This eliminates the need to run (M)BGP to have the information necessary to perform the RPF check on arriving SA messages.**
    - Note: Since there is only a single connection, there is no need to perform the RPF check since this is the only path that SA messages can take.

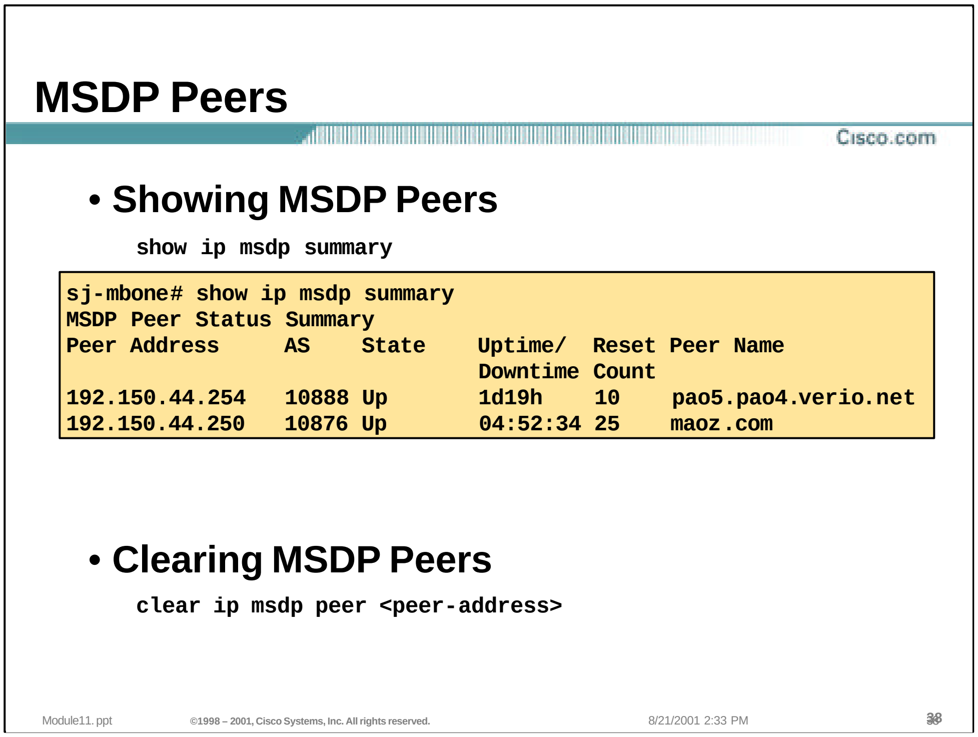


MSDP Peers

```
Interface Loopback 0
 ip address 220.220.32.1 255.255.255.255
ip msdp peer 220.220.8.1 connect-source Loopback0
ip msdp peer 220.220.16.1 connect-source Loopback0
```

- Remember: BGP is necessary when multiple MSDP peers are configured.

- **MSDP Peers**
  - **If more than one MSDP peer is configured using the ‘ip msdp peer’ command, (M)BGP must also be configured.**
  - **In the example above, Router B has active MSDP peering sessions with both Router A and Router C. In this case, (M)BGP must also be configured so that Router B has the necessary AS-PATH information to properly RPF Check arriving SA messages.**
    - Note: The only exception to this rule is if all three routers are in an MSDP Mesh Group. (Mesh Groups are discussed in a later section.)



MSDP Peers

- **Showing MSDP Peers**

```
show ip msdp summary
```

```
sj-mbone# show ip msdp summary
MSDP Peer Status Summary
Peer Address     AS     State    Uptime/  Reset Peer Name
                                 Downtime Count
192.150.44.254   10888 Up        1d19h    10    pao5.pao4.verio.net
192.150.44.250   10876 Up        04:52:34 25    maoz.com
```

- **Clearing MSDP Peers**

```
clear ip msdp peer <peer-address>
```

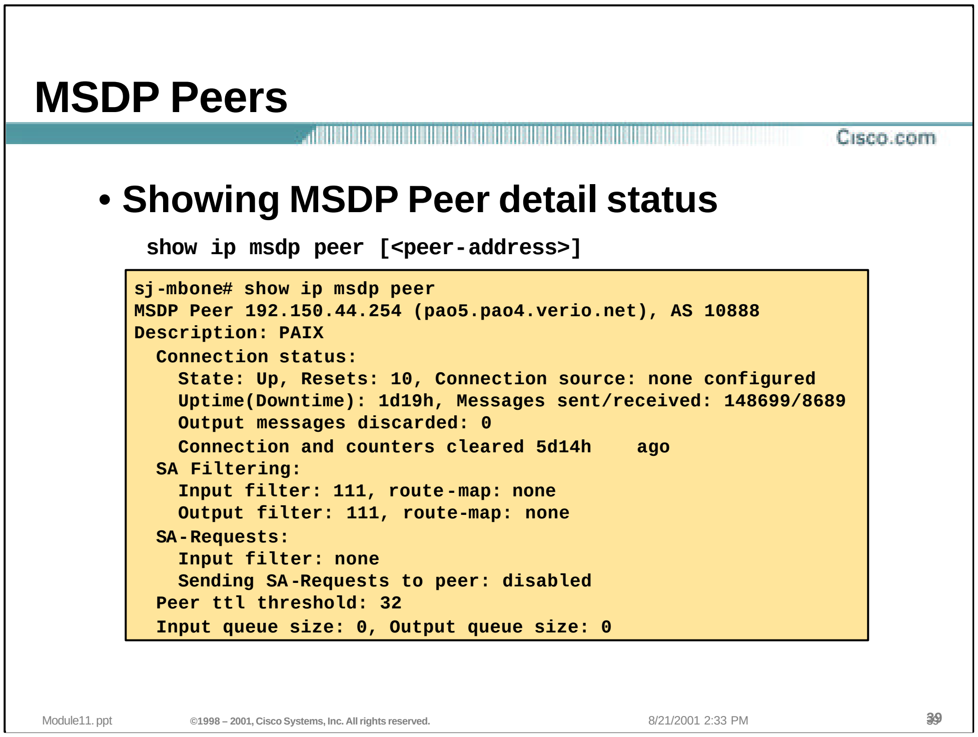
Module11. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 38

- **MSDP Peers**
  - **Summary information on a router's MSDP peer connections can be displayed using the following command:**

    ```
    show ip msdp summary
    ```

  - **An MSDP connection can be reset by using the following command:**

    ```
    clear ip msdp peer
    ```



MSDP Peers

Cisco.com

- **Showing MSDP Peer detail status**

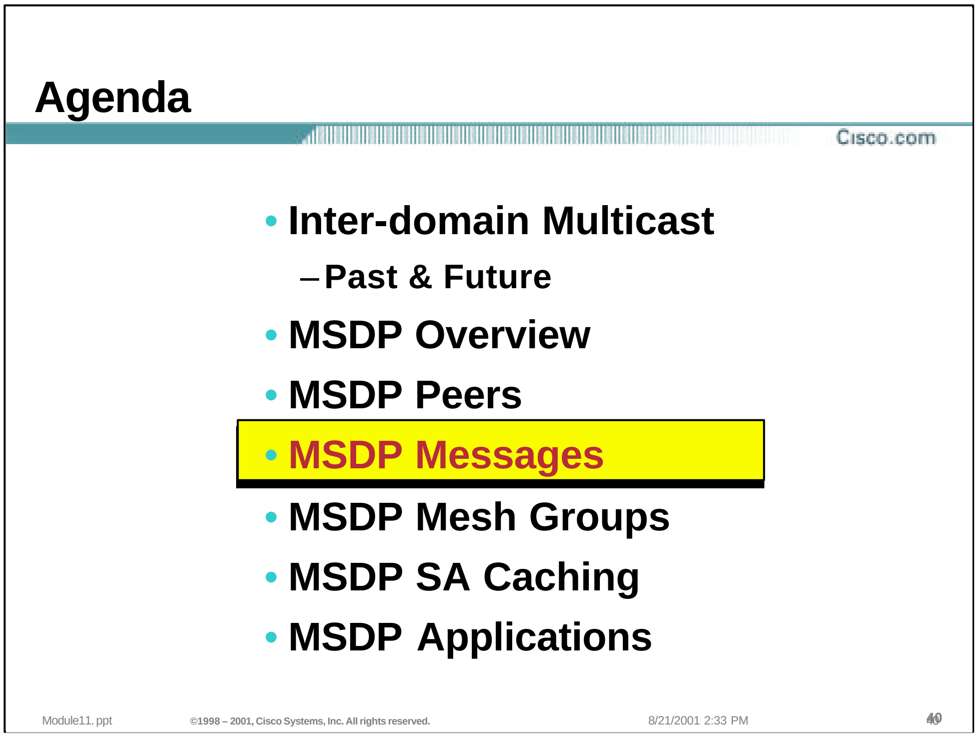
  `show ip msdp peer [<peer-address>]`

```
sj-mbone# show ip msdp peer
MSDP Peer 192.150.44.254 (pao5.pao4.verio.net), AS 10888
Description: PAIX
  Connection status:
    State: Up, Resets: 10, Connection source: none configured
    Uptime(Downtime): 1d19h, Messages sent/received: 148699/8689
    Output messages discarded: 0
    Connection and counters cleared 5d14h    ago
  SA Filtering:
    Input filter: 111, route-map: none
    Output filter: 111, route-map: none
  SA-Requests:
    Input filter: none
    Sending SA-Requests to peer: disabled
  Peer ttl threshold: 32
  Input queue size: 0, Output queue size: 0
```

Module11. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 39

- **MSDP Peers**
  - **Detailed information on a router's MSDP peer connections can be displayed using the following command:**

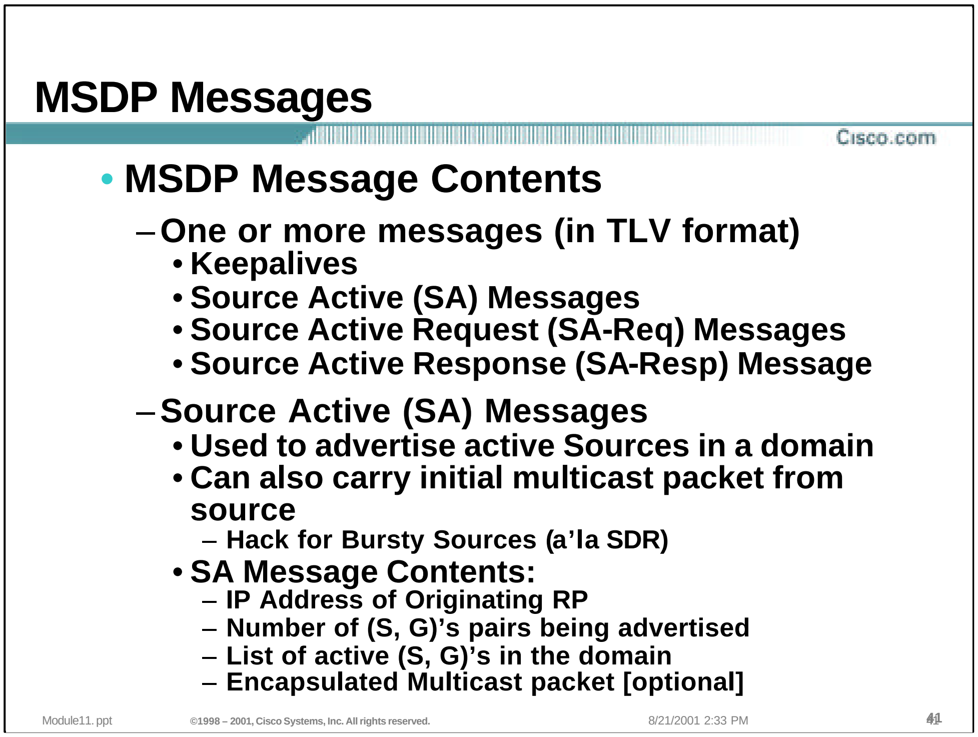
    `show ip msdp peer [<`*`peer-address`*`>]`



Agenda

Cisco.com

- Inter-domain Multicast
  - Past & Future
- MSDP Overview
- MSDP Peers
- MSDP Messages
- MSDP Mesh Groups
- MSDP SA Caching
- MSDP Applications



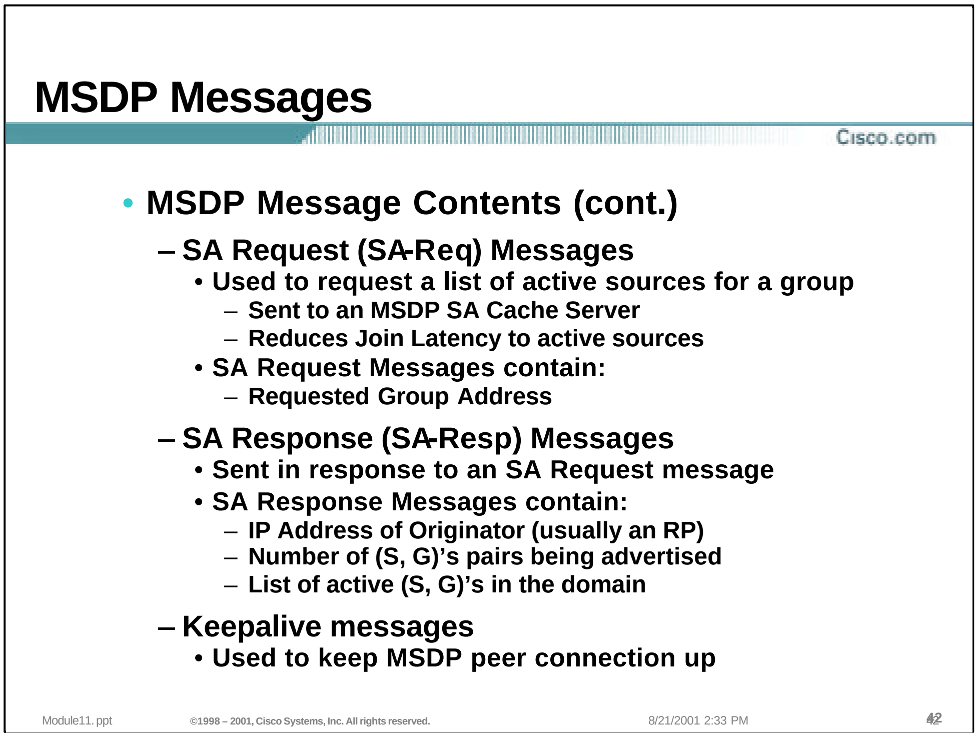
MSDP Messages

Cisco.com

- MSDP Message Contents
  - One or more messages (in TLV format)
    - Keepalives
    - Source Active (SA) Messages
    - Source Active Request (SA-Req) Messages
    - Source Active Response (SA-Resp) Message
  - Source Active (SA) Messages
    - Used to advertise active Sources in a domain
    - Can also carry initial multicast packet from source
      - Hack for Bursty Sources (a'la SDR)
    - SA Message Contents:
      - IP Address of Originating RP
      - Number of (S, G)'s pairs being advertised
      - List of active (S, G)'s in the domain
      - Encapsulated Multicast packet [optional]

Module11. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 41

- **MSDP Message Contents**
  - **There are four basic MSDP message types, each encoded in their own TLV format.**
    - Keepalives
    - Source Active (SA)
    - Source Active Request (SA-Req)
    - Source Active Response (SA-Resp)
  - **Source Active (SA) Messages**
    - These messages are used to advertise active sources in a domain. In addition, these SA messages may contain the initial multicast data packet that was sent by the source. Carrying this first data packet in the initial SA message helps to deal with the bursty source problem such as low rate SDR announcements.
    - SA Messages contain the IP address of the originating RP as well as one or more (S,G) pairs being advertised. In addition, the SA message may contain an encapsulated data packet.



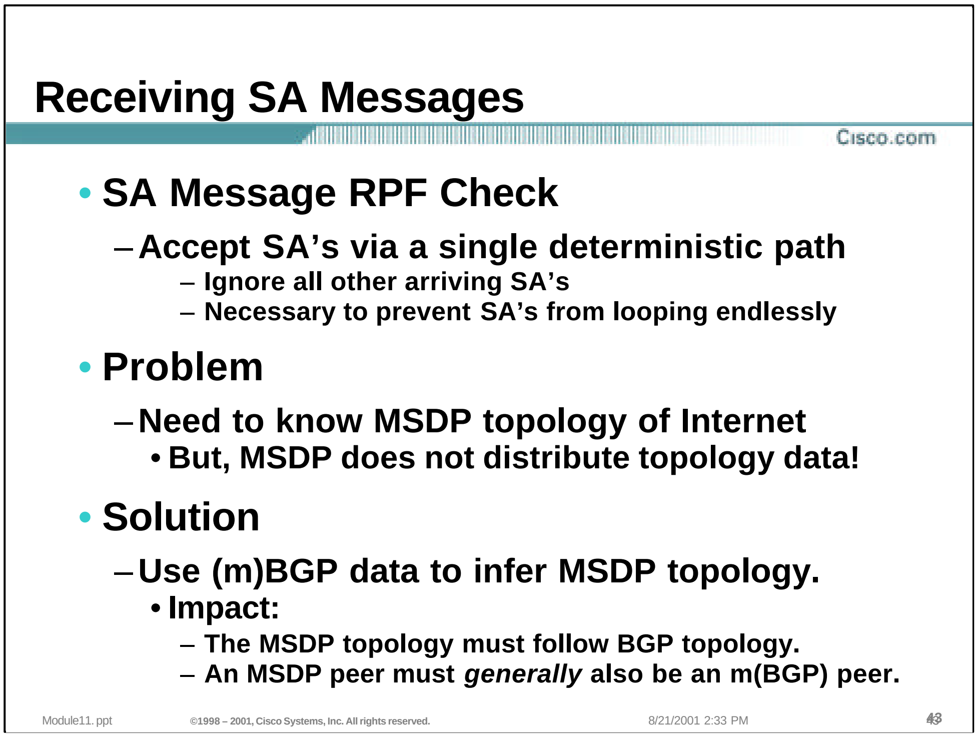
MSDP Messages

Cisco.com

- MSDP Message Contents (cont.)
  - SA Request (SA-Req) Messages
    - Used to request a list of active sources for a group
      - Sent to an MSDP SA Cache Server
      - Reduces Join Latency to active sources
    - SA Request Messages contain:
      - Requested Group Address
  - SA Response (SA-Resp) Messages
    - Sent in response to an SA Request message
    - SA Response Messages contain:
      - IP Address of Originator (usually an RP)
      - Number of (S, G)'s pairs being advertised
      - List of active (S, G)'s in the domain
  - Keepalive messages
    - Used to keep MSDP peer connection up

Module11. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 42

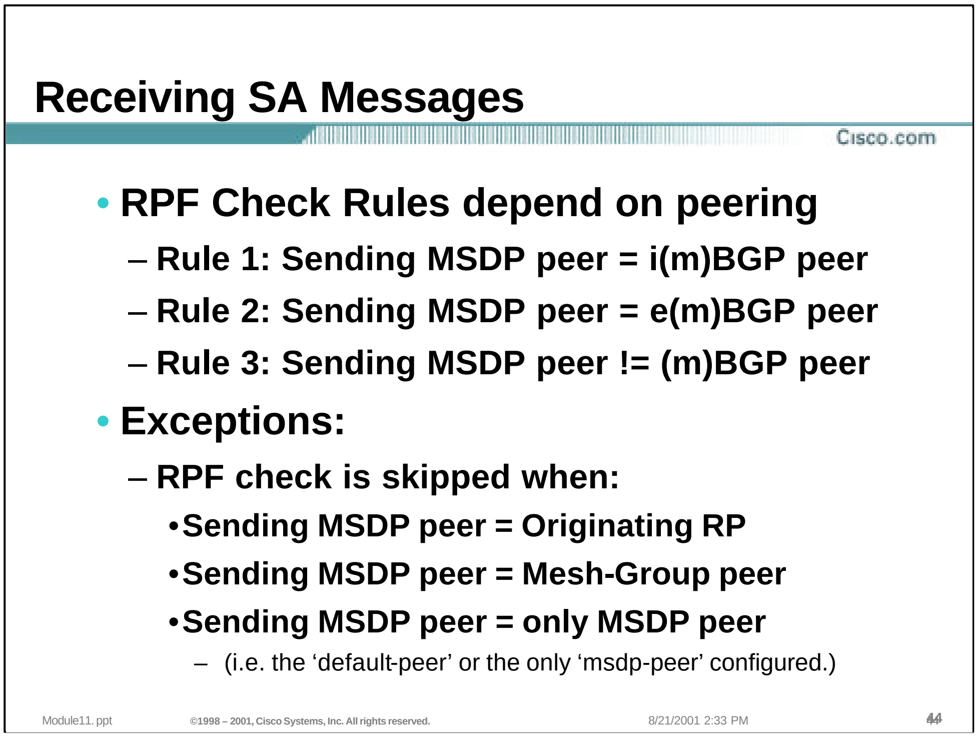
- **MSDP Message Contents**
  - **Source Active Request (SA-Req) Messages**
    - These messages are used to request a list of active sources for a specific group. These messages are sent to an MSDP SA Cache Server that is maintaining a list of active (S, G) pairs in its SA cache.
    - Join latency can be reduced by using this technique to request the list of active sources for a group instead of having to wait up to 60 seconds for all active sources in the group to be readvertised by the originating RP(s).
  - **Source Active Response (SA-Resp) Messages**
    - These messages are sent by the MSDP SA Cache Server in response to an SA-Req message.
    - These SA-Resp message contains the IP address of the originating RP as well as one or more (S, G) pairs of the active sources in the originating RP's domain.
  - **Keepalive Messages**
    - These messages are sent every 60 seconds in order to keep the MSDP session active. If no Keepalives or SA messages are received for 75 seconds, the MSDP session is cleared and reopened.



Receiving SA Messages

- SA Message RPF Check
  - Accept SA’s via a single deterministic path
    - Ignore all other arriving SA’s
    - Necessary to prevent SA’s from looping endlessly
- Problem
  - Need to know MSDP topology of Internet
    - But, MSDP does not distribute topology data!
- Solution
  - Use (m)BGP data to infer MSDP topology.
    - Impact:
      - The MSDP topology must follow BGP topology.
      - An MSDP peer must *generally* also be an m(BGP) peer.

- **SA Message RPF Check**
  - **SA messages must only be accepted from the MSDP RPF peer that is in the best path back towards the originator. The same SA message arriving from other MSDP peers must be ignored or SA loops can occur.**
  - **Deterministically selecting the MSDP RPF peer for an arriving SA message requires knowledge of the MSDP topology. However, MSDP does not distribute topology information in the form of routing updates. This means that the MSDP topology must be *inferred* via some other means.**
  - **The solution is to use the (m)BGP routing data as the best approximation of the MSDP topology for the SA RPF check mechanism. This has the following implications:**
    - The MSDP topology must follow the same general topology as the BGP peer topology.
    - This means that with a couple of exceptions, an MSDP peer generally should also be an m(BGP) peer.

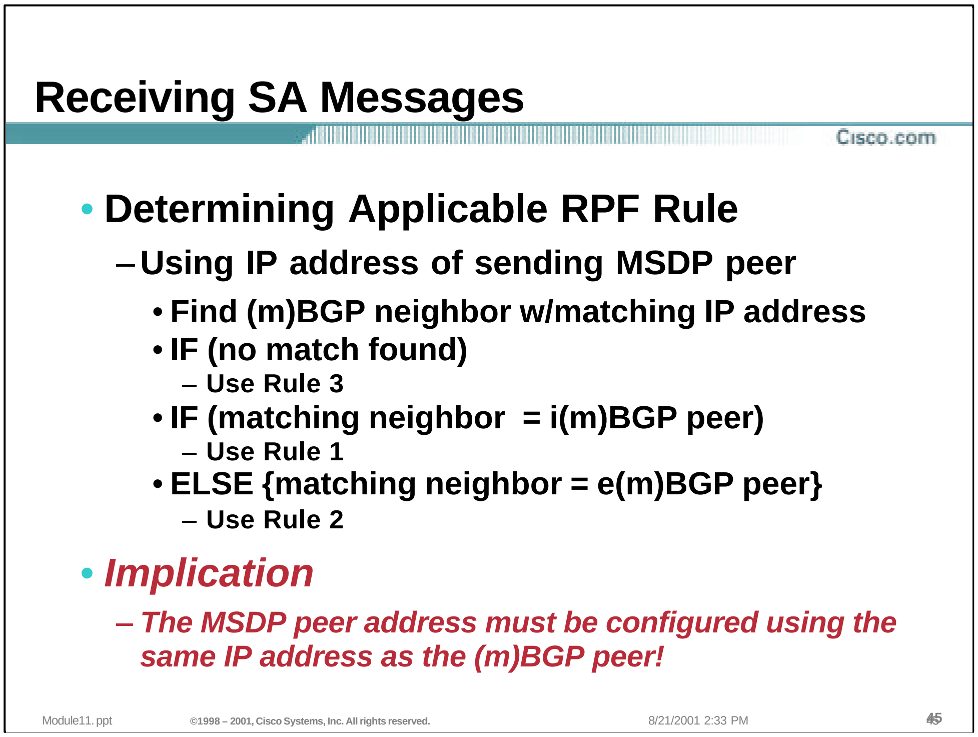


Receiving SA Messages

Cisco.com

- RPF Check Rules depend on peering
  - Rule 1: Sending MSDP peer = i(m)BGP peer
  - Rule 2: Sending MSDP peer = e(m)BGP peer
  - Rule 3: Sending MSDP peer != (m)BGP peer
- Exceptions:
  - RPF check is skipped when:
    - Sending MSDP peer = Originating RP
    - Sending MSDP peer = Mesh-Group peer
    - Sending MSDP peer = only MSDP peer
      - (i.e. the ‘default-peer’ or the only ‘msdp-peer’ configured.)

- **Receiving SA Messages**
  - **RPF Check rules depend on the BGP peering between the MSDP peers.**
    - Rule 1: Applied when the sending MSDP peer is also an i(m)BGP peer.
    - Rule 2: Applied when the sending MSDP peer is also an e(m)BGP peer.
    - Rule 3: Applied when the sending MSDP peer is not an (m)BGP peer.
  - **RPF Checks are not done in the following cases:**
    - If the sending MSDP peer is the only MSDP Peer. This would be the case if a single ‘msdp-peer’ command is configured or if only the ‘default-peer’ command is used.
    - If the sending MSDP peer is a Mesh-Group peer.
    - If the sending MSDP peer address is the RP address contained in the SA message.



Receiving SA Messages

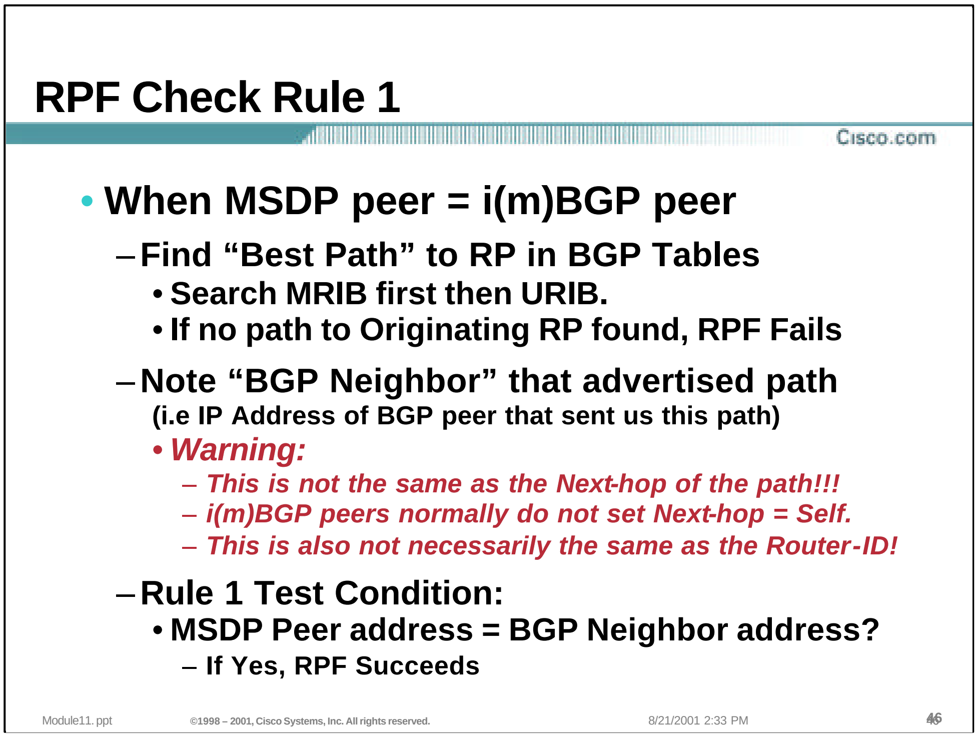
- Determining Applicable RPF Rule
  - Using IP address of sending MSDP peer
    - Find (m)BGP neighbor w/matching IP address
    - IF (no match found)
      - Use Rule 3
    - IF (matching neighbor = i(m)BGP peer)
      - Use Rule 1
    - ELSE {matching neighbor = e(m)BGP peer}
      - Use Rule 2
- *Implication*
  - *The MSDP peer address must be configured using the same IP address as the (m)BGP peer!*

- **Determining the Applicable RPF Rule**

  Cisco IOS uses the following logic to determine which RPF rule will be applied:

  - Find the (m)BGP neighbor that has the same IP address as the sending MSDP peer.
  - IF no match is found
    - Apply Rule 3
  - IF the matching (m)BGP neighbor is an internal BGP peer
    - Apply Rule 1
  - IF the matching (m)BGP neighbor is an external BGP peer
    - Apply Rule 2

  - **The implication of the above rule selection logic is the following:**
    - The IP address used to configure an MSDP peer on a router must match the IP address used to configure the (m)BGP peer on the same router.



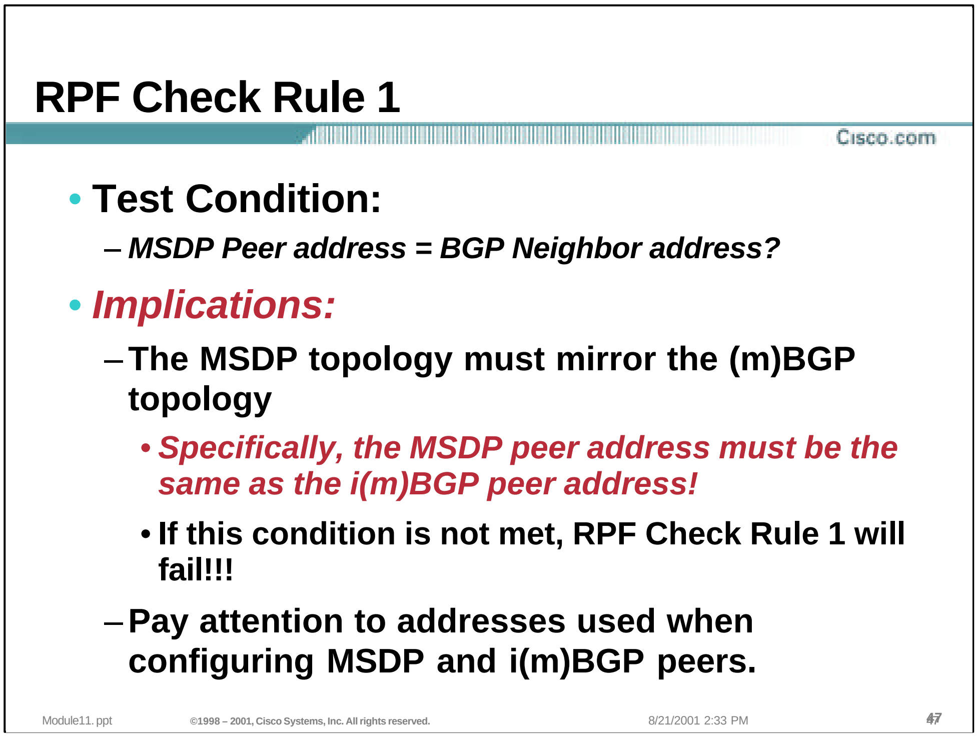
RPF Check Rule 1

- When MSDP peer = i(m)BGP peer
  - Find “Best Path” to RP in BGP Tables
    - Search MRIB first then URIB.
    - If no path to Originating RP found, RPF Fails
  - Note “BGP Neighbor” that advertised path (i.e IP Address of BGP peer that sent us this path)
    - *Warning:*
      - *This is not the same as the Next-hop of the path!!!*
      - *i(m)BGP peers normally do not set Next-hop = Self.*
      - *This is also not necessarily the same as the Router-ID!*
  - Rule 1 Test Condition:
    - MSDP Peer address = BGP Neighbor address?
      - If Yes, RPF Succeeds

- **RPF Check Rule 1**

  Applied when the sending MSDP peer is also an i(m)BGP peer.

  - **Search the BGP MRIB for the “best-path” to the RP that originated the SA message. If a path is not found in the MRIB, search the URIB. If a path is still not found, the RPF check fails.**
  - **Determine the address of the “BGP Neighbor” for this path. (This is the address of the BGP neighbor the sent us this path in a BGP Update message.)**
    - Be careful not to assume the “BGP Neighbor” address is the same as the Next-Hop address in the path. Since i(m)BGP peers do not update the Next-Hop attribute of a path, it is usually the case that the Next-Hop address is ***not*** the same as the address of the BGP peer that sent us the path.
    - The “BGP Neighbor” address is also not necessarily the same as the BGP “Router-Id” of the peer that sent us the path.
  - **Rule 1 Test:**
    - If the IP address of the sending MSDP peer is the same as the BGP Neighbor address (i.e. the address of the BGP peer that sent us the path), then the RPF check succeeds; otherwise it fails.



RPF Check Rule 1

Cisco.com

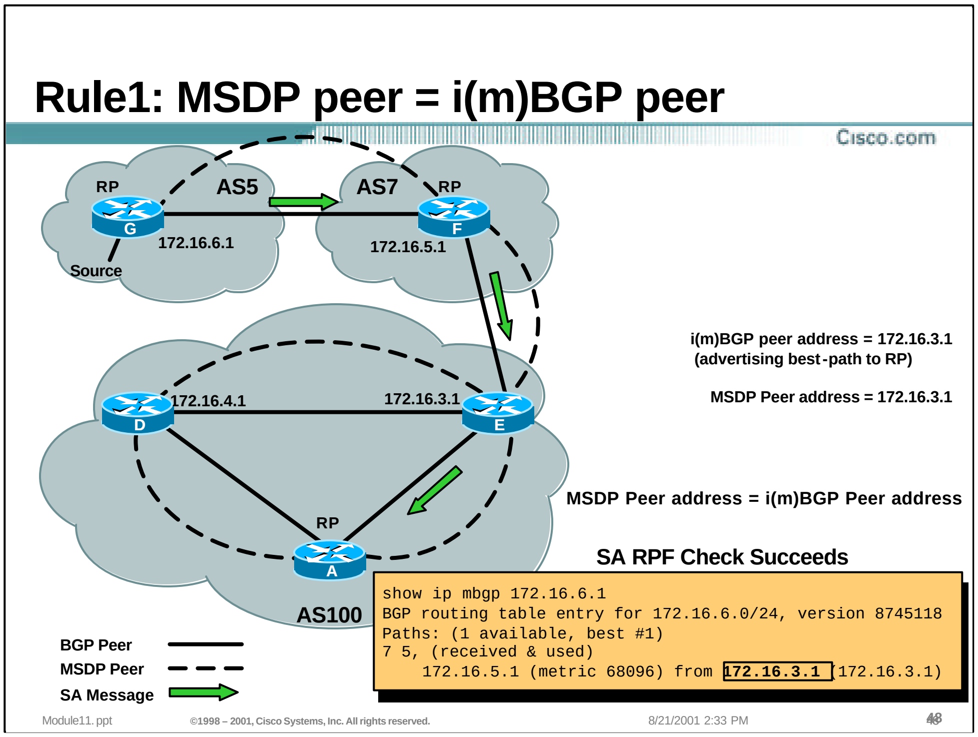
- **Test Condition:**
  - ***MSDP Peer address = BGP Neighbor address?***
- ***Implications:***
  - **The MSDP topology must mirror the (m)BGP topology**
    - ***Specifically, the MSDP peer address must be the same as the i(m)BGP peer address!***
    - **If this condition is not met, RPF Check Rule 1 will fail!!!**
  - **Pay attention to addresses used when configuring MSDP and i(m)BGP peers.**

Module11. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 47

- **RPF Check Rule 1 Implications**
  - **The MSDP topology must mirror the (m)BGP topology.**
    - Generally speaking, this means that wherever you have an i(m)BGP peer connection between two routers, you should configure an MSDP peer connection.
    - More specifically, the IP address of the far end MSDP peer connection ***must be the same*** as the far end i(m)BGP peer connection.
    - The reason for this is that BGP topology between i(m)BGP peers inside an AS is not described by the AS path.

      If it were always the case that i(m)BGP peers updated the Next-Hop address in the path when sending an update to another i(m)BGP peer, then we could rely on the Next-Hop address to describe the i(m)BGP topology (and hence the MSDP topology). However, this is not the case since the default is for i(m)BGP peers to ***not*** update the Next-Hop address.

      Instead, we must use the address of the i(m)BGP peer that sent us the path to describe the i(m)BGP/MSDP topology inside the AS. (Fortunately, BGP keeps track of the "sending (m)BGP peer address" information so it is easy for it to use this information for this MSDP RPF Check rule.
  - **Care must be taken when configuring the MSDP peer addresses to make sure that the same address is used as was used when configuring the i(m)BGP peer addresses.**



Rule1: MSDP peer = i(m)BGP peer

i(m)BGP peer address = 172.16.3.1
(advertising best-path to RP)

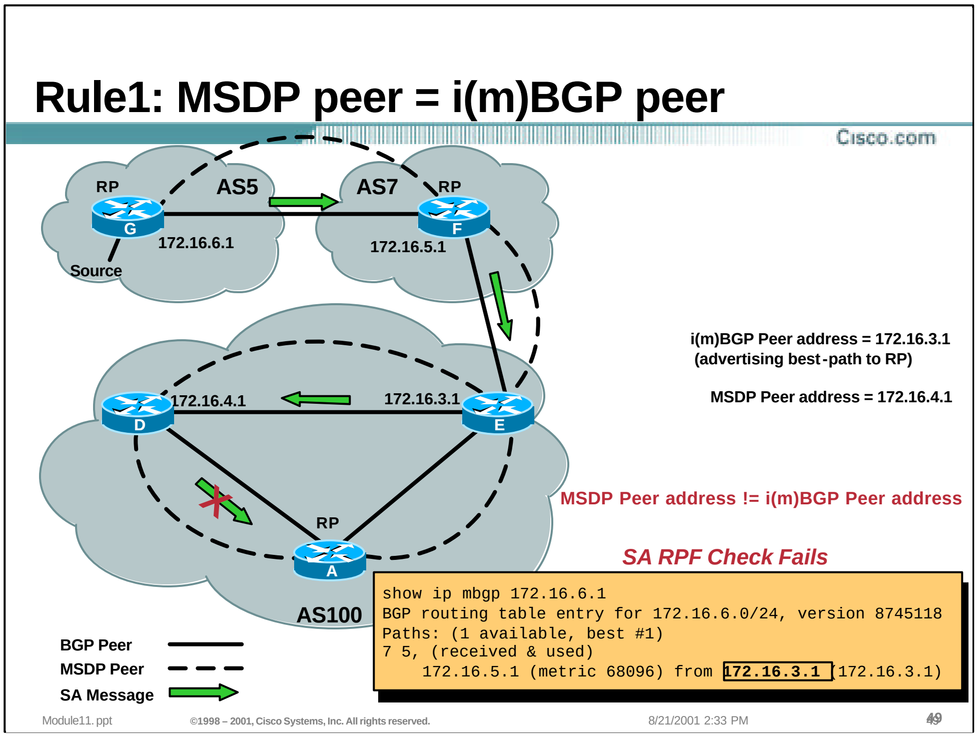
MSDP Peer address = 172.16.3.1

MSDP Peer address = i(m)BGP Peer address

SA RPF Check Succeeds

```
show ip mbgp 172.16.6.1
BGP routing table entry for 172.16.6.0/24, version 8745118
Paths: (1 available, best #1)
7 5, (received & used)
    172.16.5.1 (metric 68096) from 172.16.3.1 (172.16.3.1)
```

- **Rule 1 Example 1**
  - **In this example, router A receives an SA message originated by router G from router E which is an i(m)BGP peer.**
  - **Applying Rule 1, the following occurs:**
    - The best path in the BGP M-RIB for 172.16.6.1 (the originating RP) is located.
    - This best path was received from i(m)BGP peer, 172.16.3.1
    - The sending MSDP peer address is also 172.16.3.1
    - Therefore the RPF check Rule 1 succeeds.



Rule1: MSDP peer = i(m)BGP peer

i(m)BGP Peer address = 172.16.3.1
(advertising best-path to RP)

MSDP Peer address = 172.16.4.1

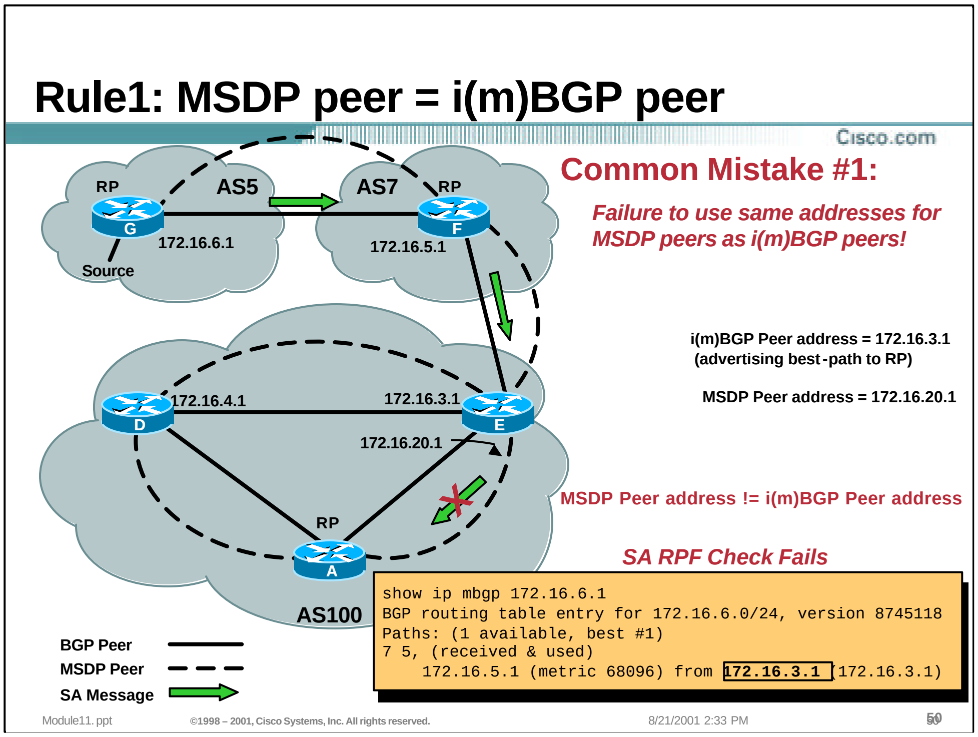
MSDP Peer address != i(m)BGP Peer address

SA RPF Check Fails

```
show ip mbgp 172.16.6.1
BGP routing table entry for 172.16.6.0/24, version 8745118
Paths: (1 available, best #1)
7 5, (received & used)
    172.16.5.1 (metric 68096) from 172.16.3.1 (172.16.3.1)
```

Module11.ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 49

- **Rule 1 Example 2**
  - **In this example, router A receives the same SA message (originated by router G) from router D which is an i(m)BGP peer.**
  - **Applying Rule 1, the following occurs:**
    - The best path in the BGP M-RIB for 172.16.6.1 (the originating RP) is located.
    - This best path was received from i(m)BGP peer, 172.16.3.1
    - The sending MSDP peer address is 172.16.4.1
    - Therefore RPF check Rule 1 fails.



Rule1: MSDP peer = i(m)BGP peer

Cisco.com

Common Mistake #1:

Failure to use same addresses for MSDP peers as i(m)BGP peers!

i(m)BGP Peer address = 172.16.3.1
(advertising best-path to RP)

MSDP Peer address = 172.16.20.1

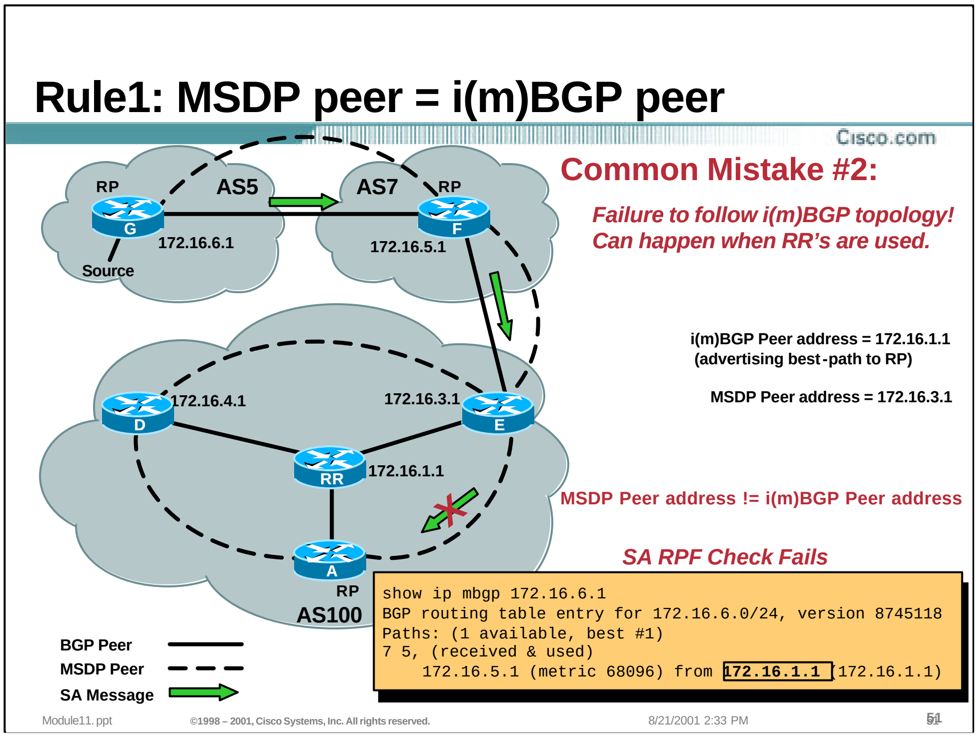
MSDP Peer address != i(m)BGP Peer address

SA RPF Check Fails

```
show ip mbgp 172.16.6.1
BGP routing table entry for 172.16.6.0/24, version 8745118
Paths: (1 available, best #1)
7 5, (received & used)
    172.16.5.1 (metric 68096) from 172.16.3.1 (172.16.3.1)
```

Module11.ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 50

- **Rule 1 Common Mistake 1**
  - **The most common mistake is for the MSDP and (m)BGP peering sessions to the same router to use different IP addresses. In this example,**
    - The MSDP peer address to router C is 172.16.20.1
    - The (m)BGP peer address to router C is 172.16.3.1
  - **Router A receives the SA message (originated by router G) from router E which is an i(m)BGP peer.**
  - **Applying Rule 1, the following occurs:**
    - The best path in the BGP M-RIB for 172.16.6.1 (the originating RP) is located.
    - This best path was received from router C which is the i(m)BGP peer with an IP address of 172.16.3.1
    - However, the sending MSDP peer (also router C) address is 172.16.20.1
    - Therefore RPF check Rule 1 fails. (Even though the SA message arrived via the correct path.)



Rule1: MSDP peer = i(m)BGP peer

Common Mistake #2:

Failure to follow i(m)BGP topology!
Can happen when RR's are used.

i(m)BGP Peer address = 172.16.1.1
(advertising best-path to RP)

MSDP Peer address = 172.16.3.1

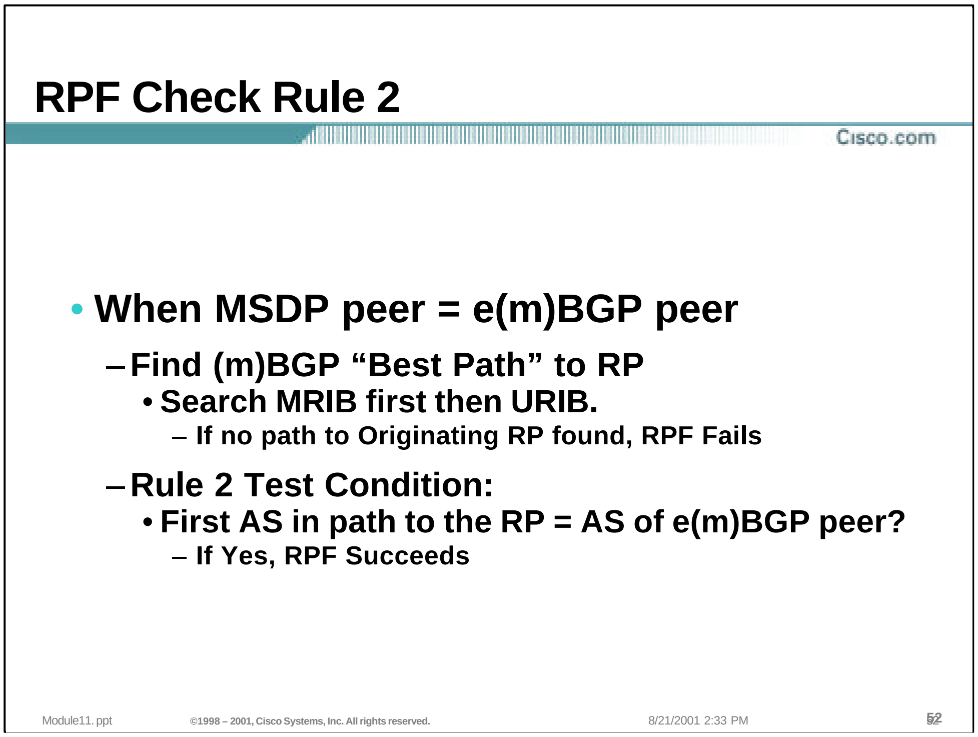
MSDP Peer address != i(m)BGP Peer address

SA RPF Check Fails

```
show ip mbgp 172.16.6.1
BGP routing table entry for 172.16.6.0/24, version 8745118
Paths: (1 available, best #1)
7 5, (received & used)
    172.16.5.1 (metric 68096) from 172.16.1.1 (172.16.1.1)
```

Module11.ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 51

- **Rule 1 Common Mistake 2**
  - **The other common mistake is to failure of the MSDP topology to follow the i(m)BGP peering topology. This can happen when Route Reflectors are used. In this example,**
    - The MSDP peer address of router E is 172.16.3.1
    - The i(m)BGP peer router is the Route Reflector "RR" whose peer address is 172.16.1.1
  - **Router A receives an SA message (originated by router G) from router E which is the MSDP peer.**
  - **Applying Rule 1, the following occurs:**
    - The best path in the BGP M-RIB for 172.16.6.1 (the originating RP) is located.
    - This best path was received from the Route Reflector which is the i(m)BGP peer with an IP address of 172.16.1.1
    - However, the sending MSDP peer is router E whose address is 172.16.20.1
    - Therefore RPF check Rule 1 fails.



RPF Check Rule 2

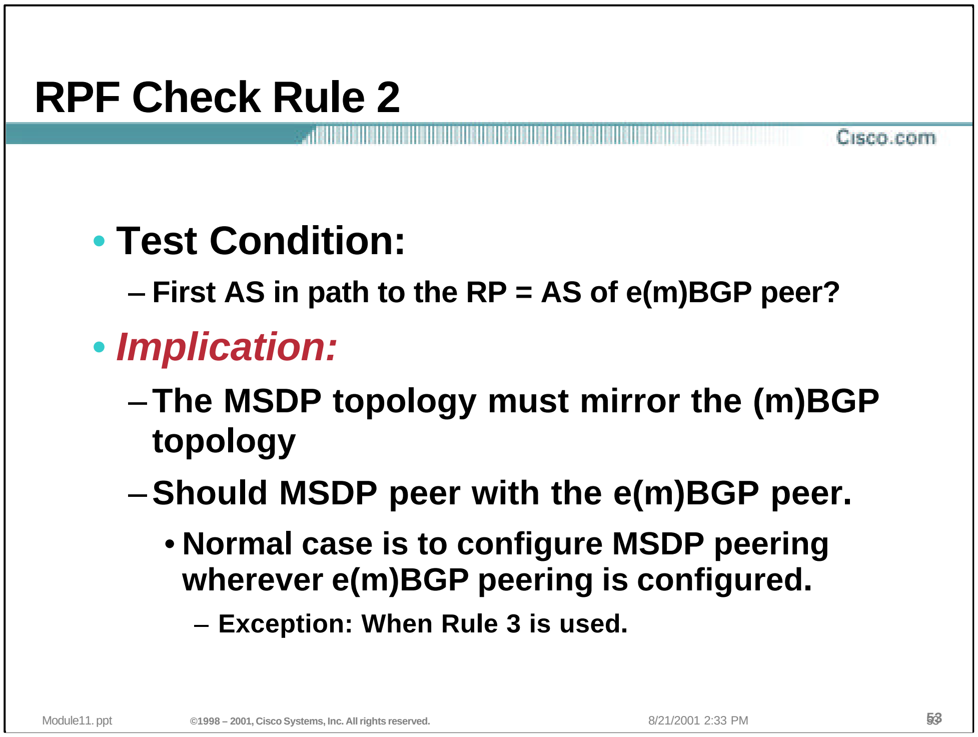
- When MSDP peer = e(m)BGP peer
  - Find (m)BGP "Best Path" to RP
    - Search MRIB first then URIB.
      - If no path to Originating RP found, RPF Fails
  - Rule 2 Test Condition:
    - First AS in path to the RP = AS of e(m)BGP peer?
      - If Yes, RPF Succeeds

Module11. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 52

- **RPF Check Rule 2**

  Applied when the sending MSDP peer is also an e(m)BGP peer.

  - **Search the BGP MRIB for the "best-path" to the RP that originated the SA message. If a path is not found in the MRIB, search the URIB. If a path is still not found, the RPF check fails.**
  - **Rule 2 Test:**
    - If the first AS in the 'best-path' to the RP is the same as the AS of the e(m)BGP peer (which is also the sending MSDP peer), then the RPF check succeeds; otherwise it fails.



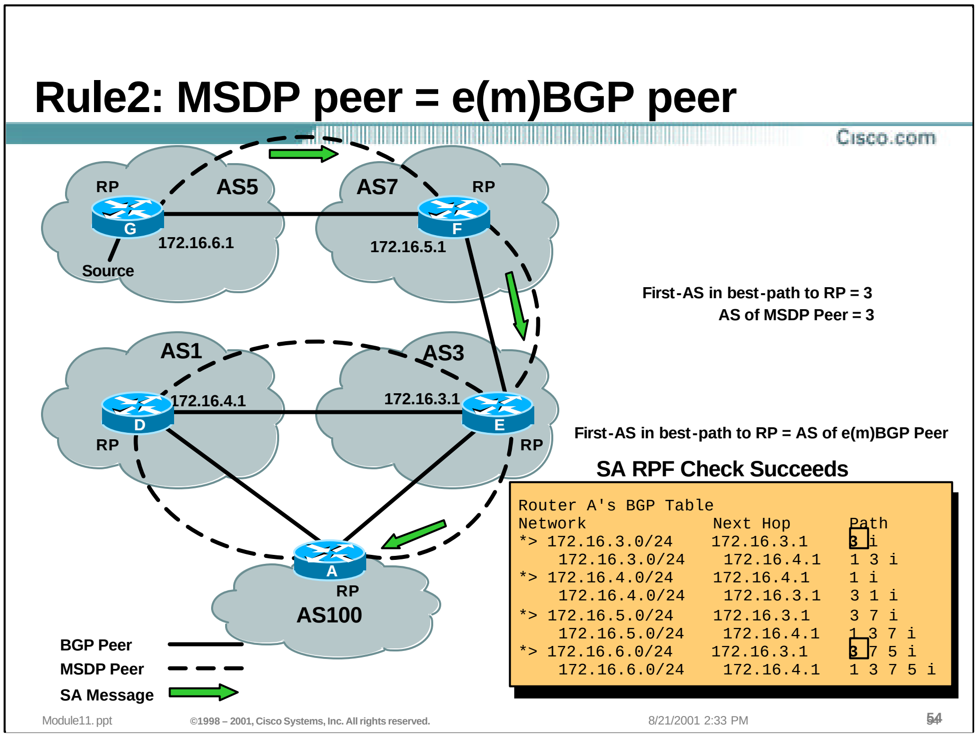
RPF Check Rule 2

- **Test Condition:**
  - **First AS in path to the RP = AS of e(m)BGP peer?**
- ***Implication:***
  - **The MSDP topology must mirror the (m)BGP topology**
  - **Should MSDP peer with the e(m)BGP peer.**
    - **Normal case is to configure MSDP peering wherever e(m)BGP peering is configured.**
      - **Exception: When Rule 3 is used.**

- **RPF Check Rule 2 Implications**
  - **The MSDP topology must mirror the (m)BGP topology.**
    - Generally speaking, this means that wherever you have an e(m)BGP peer connection between two routers, you should configure an MSDP peer connection.
    - In this case, the IP address of the far end MSDP peer connection ***does not have to be the same*** as the far end e(m)BGP peer connection.
    - The reason that the addresses do not have to be identical is that BGP topology between two e(m)BGP peers is not described by the AS path.

      If it were always the case that i(m)BGP peers updated the Next-Hop address in the path when sending an update to another i(m)BGP peer, then we could rely on the Next-Hop address to describe the i(m)BGP topology (and hence the MSDP topology). However, this is not the case since the default is for i(m)BGP peers to ***not*** update the Next-Hop address.

      Instead, we must use the address of the i(m)BGP peer that sent us the path to describe the i(m)BGP/MSDP topology inside the AS. (Fortunately, BGP keeps track of the "sending (m)BGP peer address" information so it is easy for it to use this information for this MSDP RPF Check rule.
  - **Care must be taken when configuring the MSDP peer addresses to make sure that the same address is used as was used when configuring the i(m)BGP peer addresses.**



Rule2: MSDP peer = e(m)BGP peer

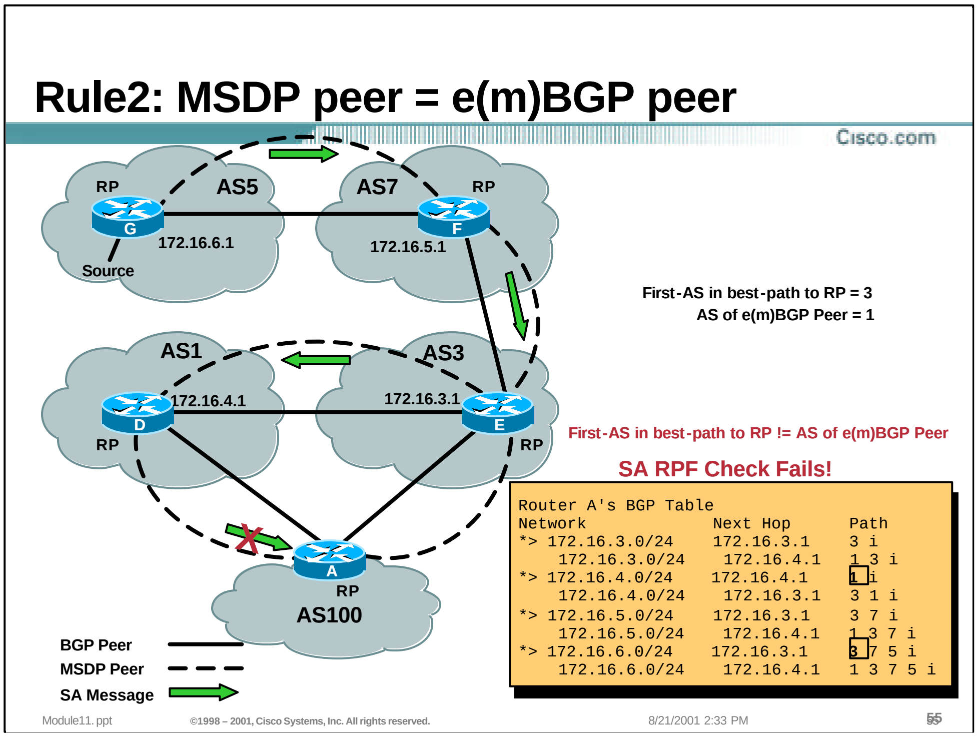
First-AS in best-path to RP = 3
AS of MSDP Peer = 3

First-AS in best-path to RP = AS of e(m)BGP Peer

SA RPF Check Succeeds

```
Router A's BGP Table
Network               Next Hop          Path
*> 172.16.3.0/24     172.16.3.1        3 i
     172.16.3.0/24     172.16.4.1      1 3 i
*> 172.16.4.0/24     172.16.4.1        1 i
     172.16.4.0/24     172.16.3.1      3 1 i
*> 172.16.5.0/24     172.16.3.1        3 7 i
     172.16.5.0/24     172.16.4.1      1 3 7 i
*> 172.16.6.0/24     172.16.3.1        3 7 5 i
     172.16.6.0/24     172.16.4.1      1 3 7 5 i
```

- **Rule 2 Example 1**
  - **In this example, router A receives an SA message originated by router G via router E which is an e(m)BGP peer.**
  - **Applying Rule 2, the following occurs:**
    - The best path in the BGP M-RIB for 172.16.6.1 (the originating RP) is located.
    - The first-hop AS in the best path to the originating RP is AS3.
    - The origin AS of the sending MSDP peer (172.16.3.1) is also AS3. (This is determined by locating the best-path to the MSDP peer and then finding the last AS in the AS-Path list.)
    - Therefore the RPF check Rule 2 succeeds.



Rule2: MSDP peer = e(m)BGP peer

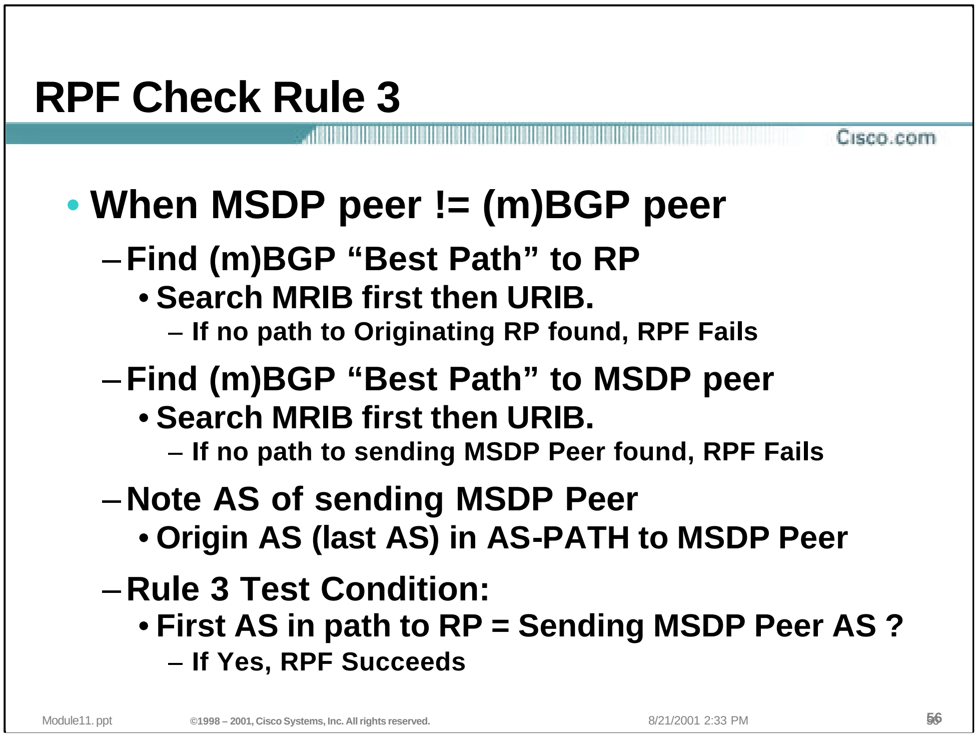
First-AS in best-path to RP = 3
AS of e(m)BGP Peer = 1

First-AS in best-path to RP != AS of e(m)BGP Peer

SA RPF Check Fails!

```
Router A's BGP Table
Network             Next Hop      Path
*> 172.16.3.0/24    172.16.3.1    3 i
     172.16.3.0/24    172.16.4.1  1 3 i
*> 172.16.4.0/24    172.16.4.1    1 i
     172.16.4.0/24    172.16.3.1  3 1 i
*> 172.16.5.0/24    172.16.3.1    3 7 i
     172.16.5.0/24    172.16.4.1  1 3 7 i
*> 172.16.6.0/24    172.16.3.1    3 7 5 i
     172.16.6.0/24    172.16.4.1  1 3 7 5 i
```

- **Rule 2 Example 2**
  - **In this example, router A receives the same SA message (originated by router G) via router D which is also an e(m)BGP peer.**
  - **Applying Rule 2, the following occurs:**
    - The best path in the BGP M-RIB for 172.16.6.1 (the originating RP) is located.
    - The first-hop AS in the best path to the originating RP is AS3.
    - The origin AS of the sending MSDP peer (172.16.4.1) is ***not*** AS3, it is AS1. (This is determined by locating the best-path to the MSDP peer and then finding the last AS in the AS-Path list.)
    - Therefore the RPF check Rule 2 fails.



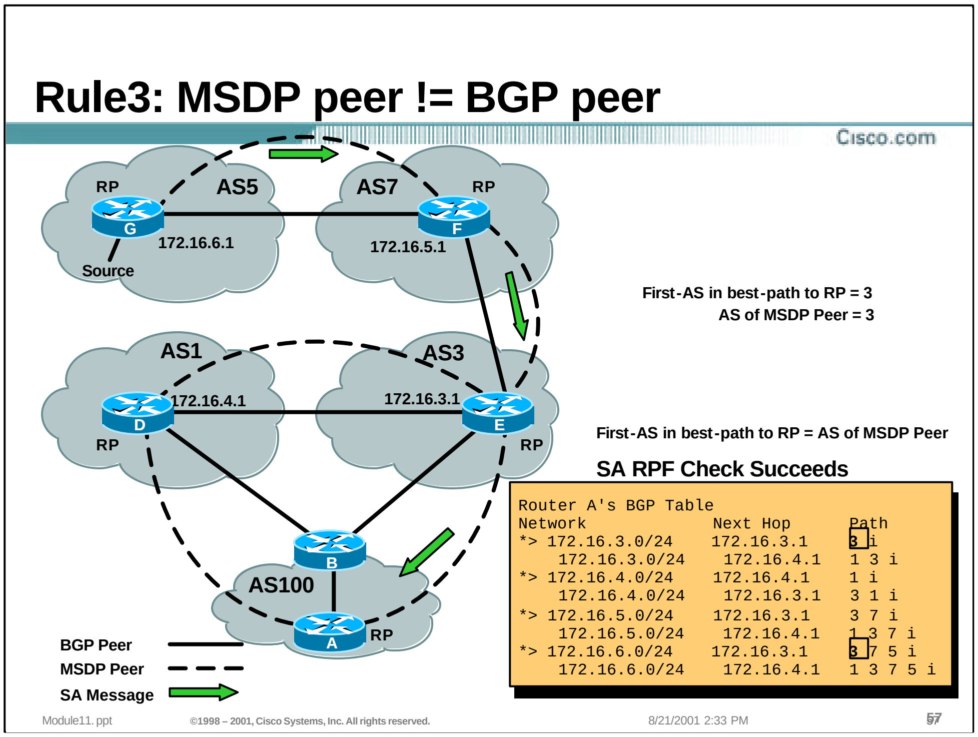
RPF Check Rule 3

- When MSDP peer != (m)BGP peer
  - Find (m)BGP “Best Path” to RP
    - Search MRIB first then URIB.
      - If no path to Originating RP found, RPF Fails
  - Find (m)BGP “Best Path” to MSDP peer
    - Search MRIB first then URIB.
      - If no path to sending MSDP Peer found, RPF Fails
  - Note AS of sending MSDP Peer
    - Origin AS (last AS) in AS-PATH to MSDP Peer
  - Rule 3 Test Condition:
    - First AS in path to RP = Sending MSDP Peer AS ?
      - If Yes, RPF Succeeds

- **RPF Check Rule 3**

  Applied when the sending MSDP peer is not an (m)BGP peer at all.

  - **Search the BGP MRIB for the “best-path” to the RP that originated the SA message. If a path is not found in the MRIB, search the URIB. If a path is still not found, the RPF check fails.**
  - **Search the BGP MRIB for the “best-path” to the MSDP peer that sent us the SA message. If a path is not found in the MRIB, search the URIB. If a path is still not found, the RPF check fails.**
  - **Note the AS of MSDP peer that sent us the SA. (This is the “origin AS” which is the last AS in the AS-PATH to the MSDP peer.)**
  - **Rule 3 Test:**
    - If the first AS in the ‘best-path’ to the RP is the same as the AS of the sending MSDP peer, then the RPF check succeeds; otherwise it fails.



Rule3: MSDP peer != BGP peer

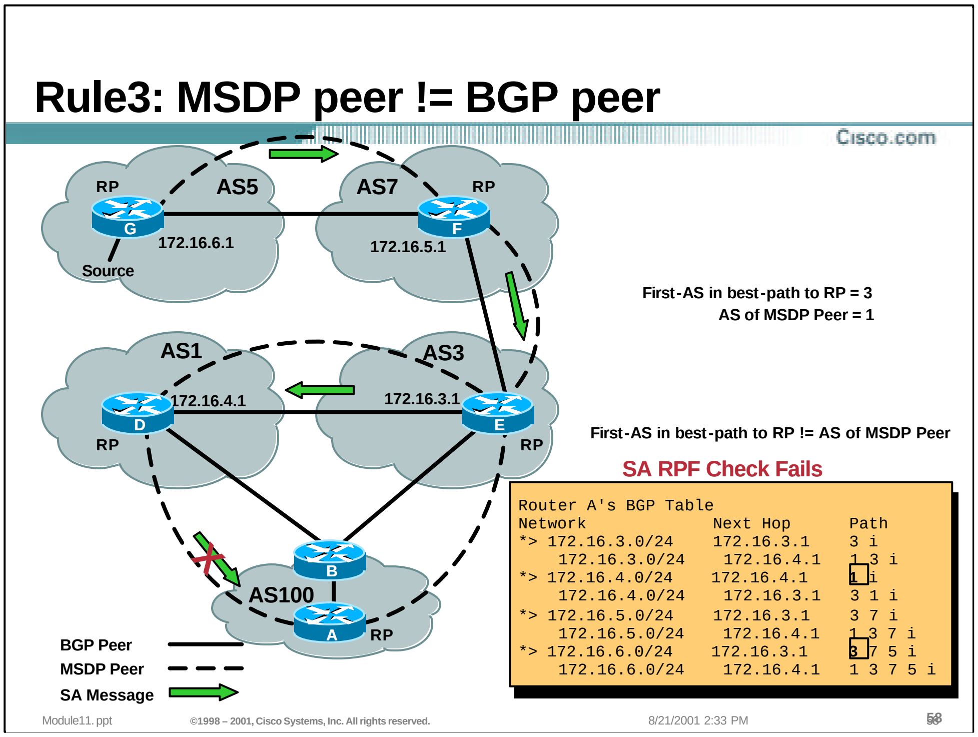
First-AS in best-path to RP = 3
AS of MSDP Peer = 3

First-AS in best-path to RP = AS of MSDP Peer

SA RPF Check Succeeds

```
Router A's BGP Table
Network              Next Hop       Path
*> 172.16.3.0/24    172.16.3.1     3 i
     172.16.3.0/24    172.16.4.1   1 3 i
*> 172.16.4.0/24    172.16.4.1     1 i
     172.16.4.0/24    172.16.3.1   3 1 i
*> 172.16.5.0/24    172.16.3.1     3 7 i
     172.16.5.0/24    172.16.4.1   1 3 7 i
*> 172.16.6.0/24    172.16.3.1     3 7 5 i
     172.16.6.0/24    172.16.4.1   1 3 7 5 i
```

- **Rule 3 Example 1**
  - **In this example, router A receives an SA message originated by router G via router E which is neither an i(m)BGP peer nor an e(m)BGP peer.**
  - **Applying Rule 3, the following occurs:**
    - The best path in the BGP M-RIB for 172.16.6.1 (the originating RP) is located.
    - The first-hop AS in the best path to the originating RP is AS3.
    - The origin AS of the sending MSDP peer (172.16.3.1) is also AS3. (This is determined by locating the best-path to the MSDP peer and then finding the last AS in the AS-Path list.)
    - Therefore the RPF check Rule 3 succeeds.



Rule3: MSDP peer != BGP peer

First-AS in best-path to RP = 3
AS of MSDP Peer = 1

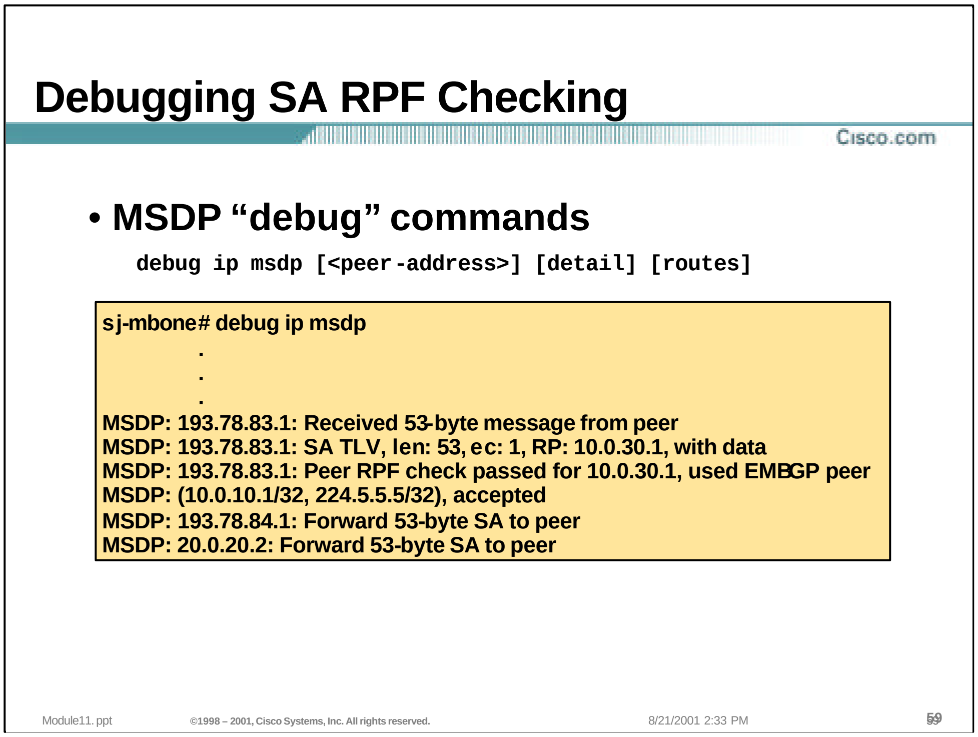
First-AS in best-path to RP != AS of MSDP Peer

SA RPF Check Fails

```
Router A's BGP Table
Network              Next Hop       Path
*> 172.16.3.0/24     172.16.3.1     3 i
     172.16.3.0/24     172.16.4.1   1 3 i
*> 172.16.4.0/24     172.16.4.1     1 i
     172.16.4.0/24     172.16.3.1   3 1 i
*> 172.16.5.0/24     172.16.3.1     3 7 i
     172.16.5.0/24     172.16.4.1   1 3 7 i
*> 172.16.6.0/24     172.16.3.1     3 7 5 i
     172.16.6.0/24     172.16.4.1   1 3 7 5 i
```

Module11.ppt 8/21/2001 2:33 PM 58

- **Rule 3 Example 2**
  - **In this example, router A receives the same SA message (originated by router G) via router D which is neither an i(m)BGP peer nor an e(m)BGP peer.**
  - **Applying Rule 3, the following occurs:**
    - The best path in the BGP M-RIB for 172.16.6.1 (the originating RP) is located.
    - The first-hop AS in the best path to the originating RP is AS3.
    - The origin AS of the sending MSDP peer (172.16.4.1) is AS1. (This is determined by locating the best-path to the MSDP peer and then finding the last AS in the AS-Path list.)
    - Therefore the RPF check Rule 3 fails.



Debugging SA RPF Checking

- **MSDP “debug” commands**

```
debug ip msdp [<peer-address>] [detail] [routes]
```

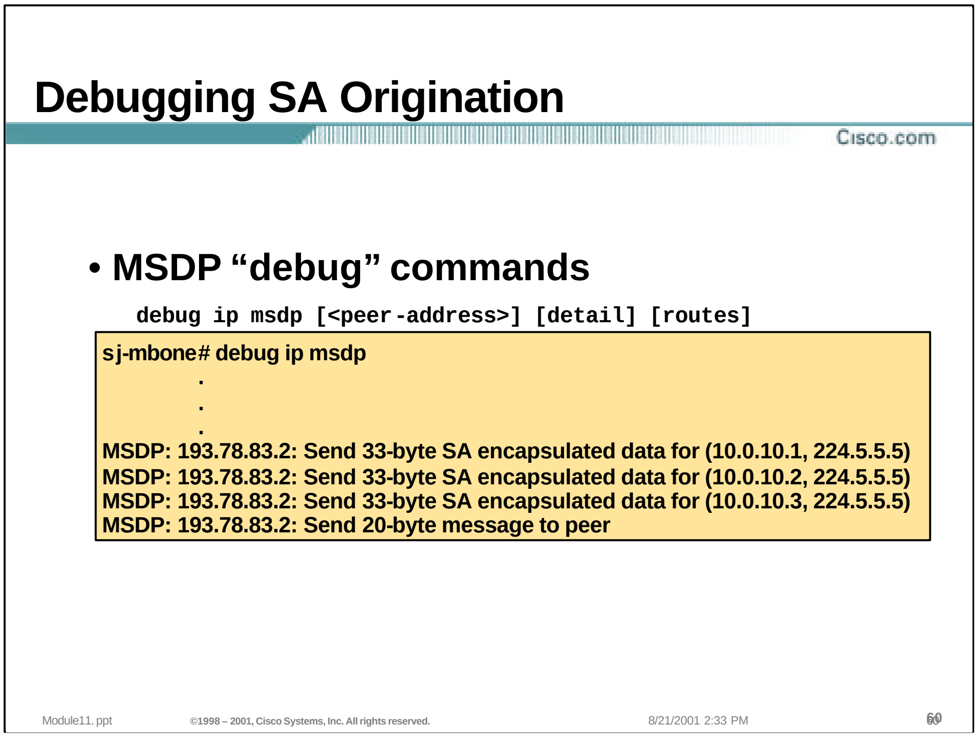
```
sj-mbone# debug ip msdp
            .
            .
            .
MSDP: 193.78.83.1: Received 53-byte message from peer
MSDP: 193.78.83.1: SA TLV, len: 53, ec: 1, RP: 10.0.30.1, with data
MSDP: 193.78.83.1: Peer RPF check passed for 10.0.30.1, used EMBGP peer
MSDP: (10.0.10.1/32, 224.5.5.5/32), accepted
MSDP: 193.78.84.1: Forward 53-byte SA to peer
MSDP: 20.0.20.2: Forward 53-byte SA to peer
```

Module11.ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 59

- **Debugging SA RPF Checking**
  - **Use the following IOS command to debug the exchange of MSDP messages and to see the results of the RPF check as each SA message arrives.**

    ```
    debug ip msdp [<peer-address>] [detail] [routes]
    ```

  - **The MSDP debug lines in the above example indicates that:**
    - A 53 byte MSDP message was received from MSDP peer 193.78.83.1
    - This MSDP message was an SA message that was originated by the RP whose address is 10.0.30.1. It also contains an encapsulated multicast data packet.
    - The RPF check succeeded on this message and the rule that was applied was the “MSDP Peer = External (M)BGP Neighbor” rule.
    - The contents of the SA message contained a single (S, G) source advertisement for (10.0.10.1, 224.5.5.5).
    - The SA message was forwarded to MSDP Peer 193.78.84.1
    - The SA message was forwarded to MSDP Peer 20.0.20.2



Debugging SA Origination

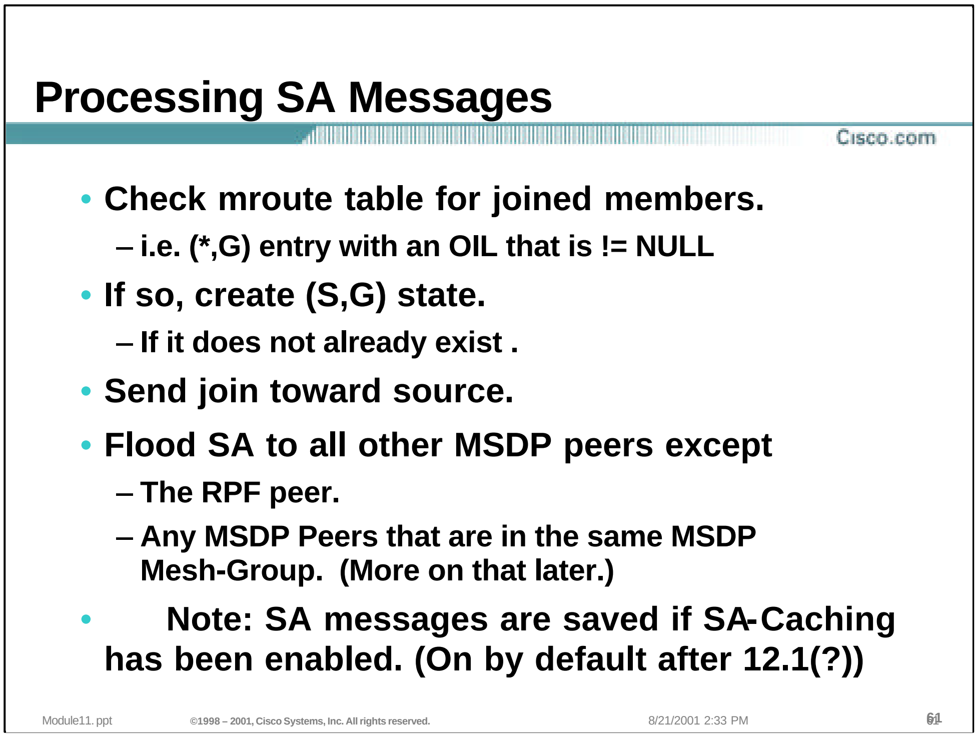
Cisco.com

- MSDP “debug” commands

```
debug ip msdp [<peer-address>] [detail] [routes]
```

```
sj-mbone# debug ip msdp
            .
            .
            .
MSDP: 193.78.83.2: Send 33-byte SA encapsulated data for (10.0.10.1, 224.5.5.5)
MSDP: 193.78.83.2: Send 33-byte SA encapsulated data for (10.0.10.2, 224.5.5.5)
MSDP: 193.78.83.2: Send 33-byte SA encapsulated data for (10.0.10.3, 224.5.5.5)
MSDP: 193.78.83.2: Send 20-byte message to peer
```

- **Debugging SA Origination on RP’s**
  - **The ‘debug ip msdp’ command can also be used to debug the origination of SA Messages by RP’s.**
  - **The MSDP debug lines in the above example indicates that the router (which is an RP) originated SA messages for the following active sources in its local PIM-SM domain:**
    - The router originated an SA message for local source (10.0.10.1, 224.5.5.5) to MSDP neighbor 193.78.83.2
    - The router originated an SA message for local source (10.0.10.2, 224.5.5.5) to MSDP neighbor 193.78.83.2
    - The router originated an SA message for local source (10.0.10.2, 224.5.5.5) to MSDP neighbor 193.78.83.2
    - The router sent a 20 byte MSDP message to MSDP peer 193.78.83.2.



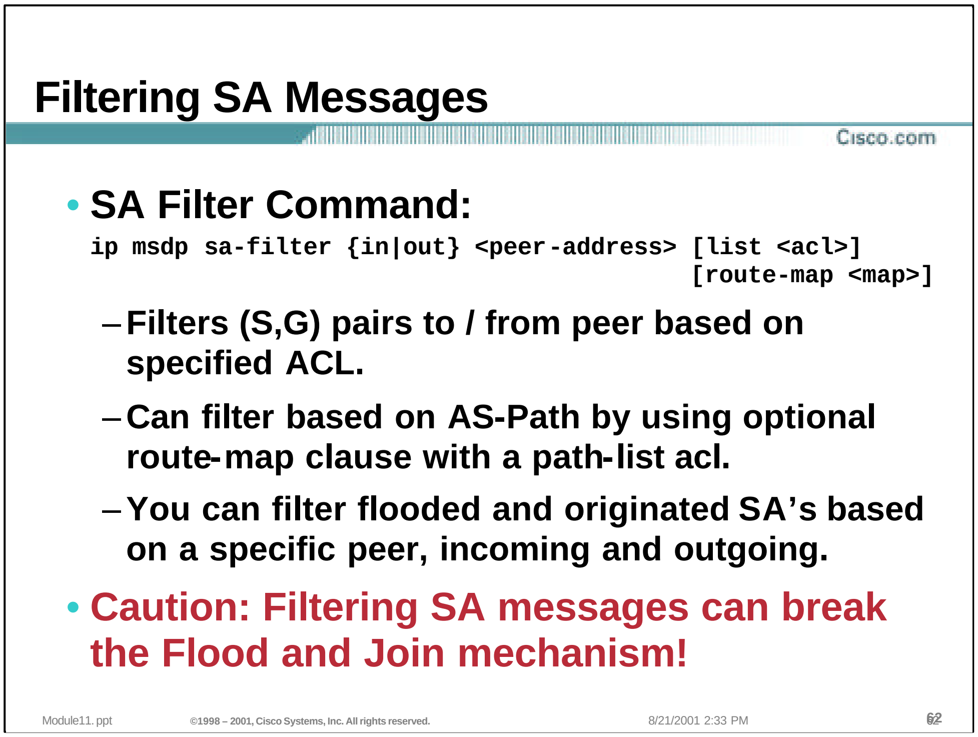
Processing SA Messages

Cisco.com

- Check mroute table for joined members.
  - i.e. (*,G) entry with an OIL that is != NULL
- If so, create (S,G) state.
  - If it does not already exist .
- Send join toward source.
- Flood SA to all other MSDP peers except
  - The RPF peer.
  - Any MSDP Peers that are in the same MSDP Mesh-Group. (More on that later.)
- Note: SA messages are saved if SA-Caching has been enabled. (On by default after 12.1(?))

Module11. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 61

- **Processing SA Messages**
  - **The following steps are taken by a router whenever it processes an SA message:**
  - **Using group address "G" of the (S, G) pair in the SA message, locate the associated (*, G) entry in the mroute table.**
  - **If the (*, G) entry is found AND it's outgoing interface list isnot Null, then there are active receivers in the PIM-SM domain for the source advertised in the SA message.**
    - Create an (S, G) entry for the advertised source.
    - If the (S, G) entry did not already exist, immediately trigger an (S, G) Join toward the source in order to join the source tree.
  - **Flood the SA message to all other MSDP peers with the exception of:**
    - The MSDP peer from which the SA message was received
    - Any MSDP peers that are in the same MSDP Mesh Group as this router. (More on MSDP Mesh Groups later.)
  - **Note: SA messages are not stored locally by the router unless SA-Caching has been enabled on the router. (In most cases, Network Administrators enable SA-Caching in order to improve network debugging capabilities.)**



Filtering SA Messages

Cisco.com

- **SA Filter Command:**

```
ip msdp sa-filter {in|out} <peer-address> [list <acl>]
                                          [route-map <map>]
```

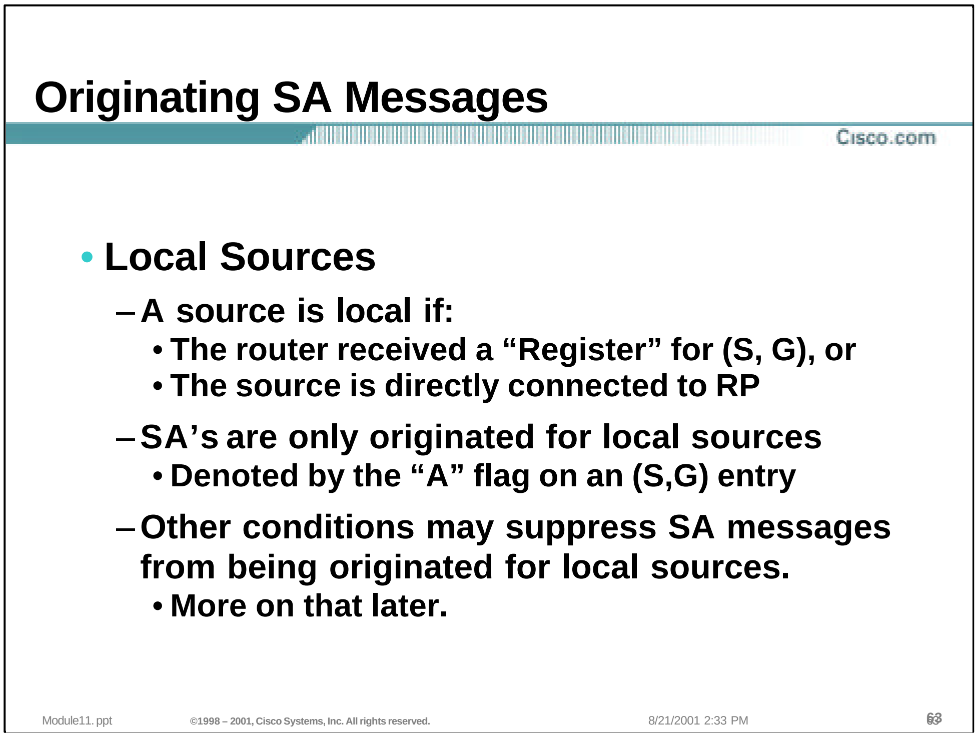
  - **Filters (S,G) pairs to / from peer based on specified ACL.**
  - **Can filter based on AS-Path by using optional route-map clause with a path-list acl.**
  - **You can filter flooded and originated SA's based on a specific peer, incoming and outgoing.**

- **Caution: Filtering SA messages can break the Flood and Join mechanism!**

- **SA Filtering**
  - **SA Filtering can be configured by the use of the following IOS command:**

```
ip msdp sa-filter {in|out} <peer-address> [list <acl>] [route-map <map>]
```

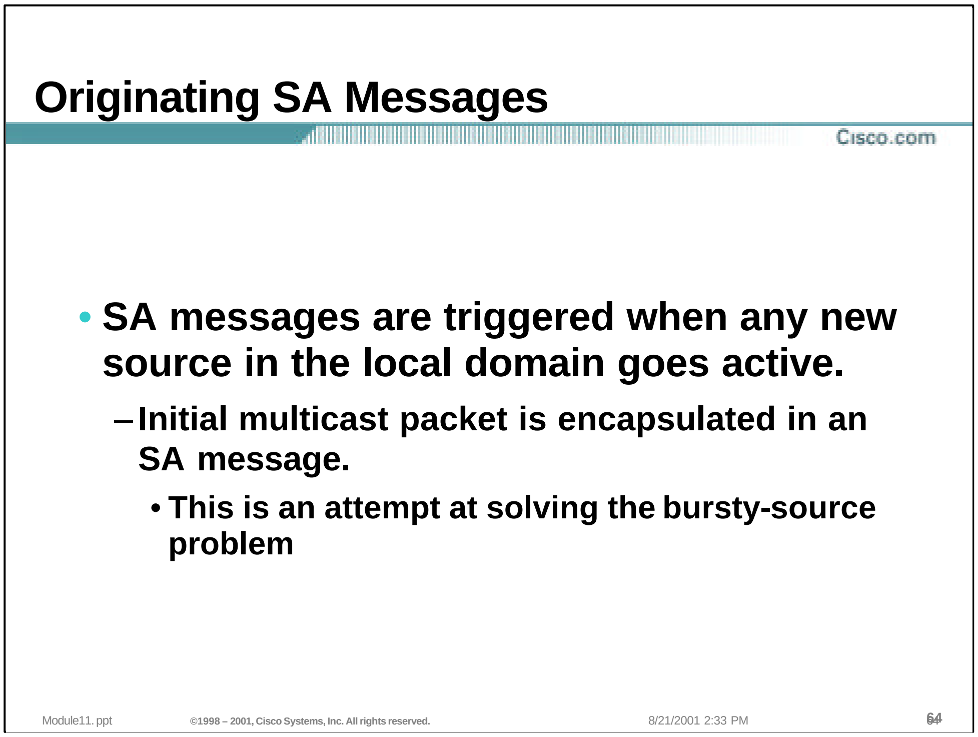
    - The above command may be used to filter incoming or outgoing SA Messages based on the (S, G) pairs specified in the **`list <acl>`** clause.
    - The above command may also be used to filter incoming or outgoing SA Messages based AS-PATH using the route map specified by the **`route-map <map>`** clause.
  - **Caution: Arbitrary filtering of SA Messages can result in downstream MSDP Peers from being starved of SA Messages for legitimate active sources. Care should be used when using these sorts of filters so that this does not occur. (Normally, these filters are only used to reject Bogons such as sources in network 10.0.0.0, etc.)**



Originating SA Messages

- Local Sources
  - A source is local if:
    - The router received a “Register” for (S, G), or
    - The source is directly connected to RP
  - SA’s are only originated for local sources
    - Denoted by the “A” flag on an (S,G) entry
  - Other conditions may suppress SA messages from being originated for local sources.
    - More on that later.

- **Originating SA Messages for Local Sources**
  - **A local source is is defined as a source for which the RP:**
    - Has received a Register message for the source, or
    - The sources is directly connected to the RP.
  - **An RP “originates” SA Messages only for local sources in its PIM-SM domain.**
    - A local source is denoted by the “A” flag being set in the (S, G) mroute entry on the RP. This indicates that the source is a “candidate” for advertisement by the RP to other MSDP peers.
    - NOTE: In some IOS versions, the key in the “show ip mroute” command states that the “A” flag indicates that the (S, G) *IS* being announced via MSDP. This in fact is not correct as other factors such as filters may block this (S, G) from actually being advertised.



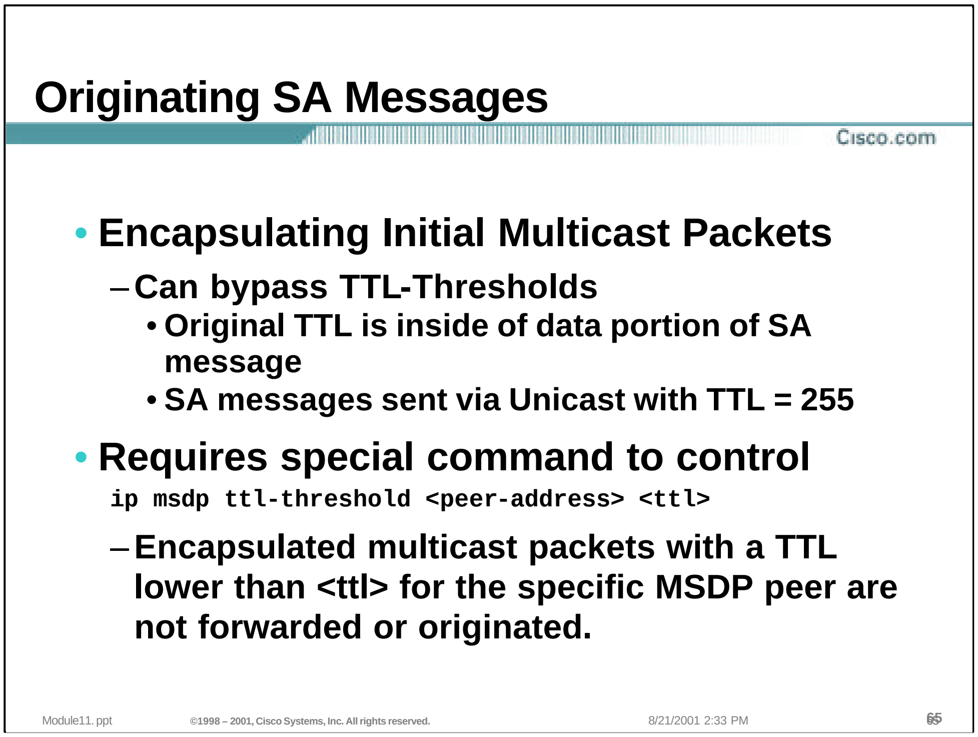
Originating SA Messages

Cisco.com

- SA messages are triggered when any new source in the local domain goes active.
  - Initial multicast packet is encapsulated in an SA message.
    - This is an attempt at solving the bursty-source problem

Module11. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 64

- **Originating SA Messages**
  - **SA messages are triggered by an RP (assuming MSDP is configured) when any new source goes active within the local PIM-SM domain.**
  - **When a source in the local PIM-SM domain initially goes active, it causes the creation of (S, G) state in the RP. New sources are detected by the RP by:**
    - The receipt of a Register message or
    - The arrival of the first (S, G) packet from a directly connected source.
  - **The initial multicast packet sent by the source (either encapsulated in the Register message or received from a directly connected source) is encapsulated in the initial SA message in an attempt to solve the problem of bursty sources.**



Originating SA Messages

Cisco.com

- **Encapsulating Initial Multicast Packets**
  - **Can bypass TTL-Thresholds**
    - **Original TTL is inside of data portion of SA message**
    - **SA messages sent via Unicast with TTL = 255**
- **Requires special command to control**

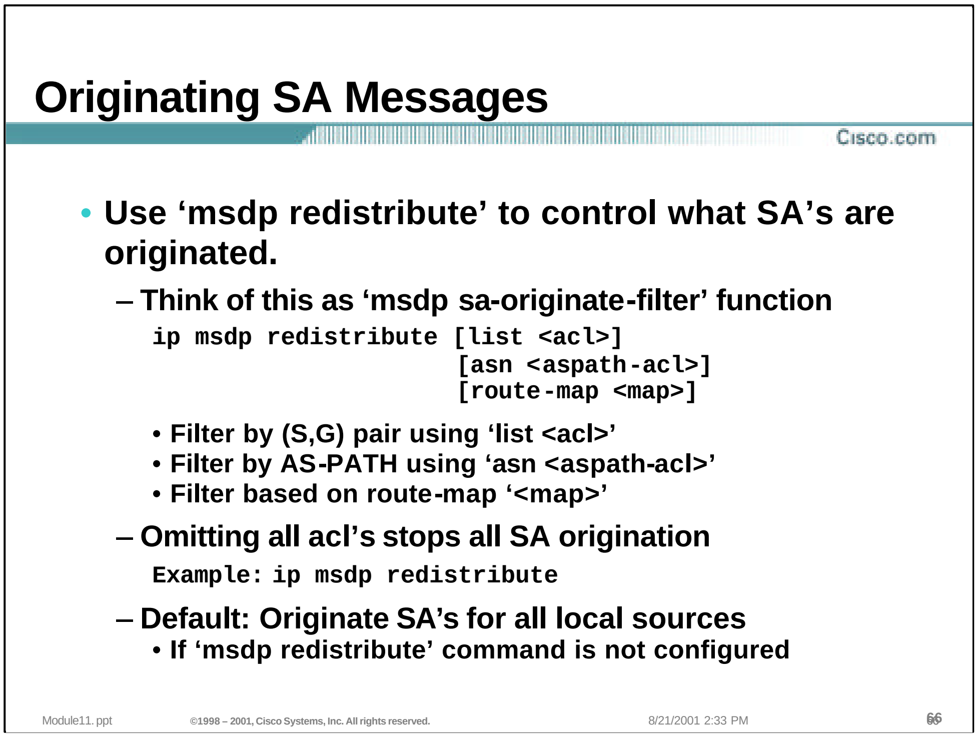
```
ip msdp ttl-threshold <peer-address> <ttl>
```

  - **Encapsulated multicast packets with a TTL lower than <ttl> for the specific MSDP peer are not forwarded or originated.**

- **Originating SA Messages**
  - **A TTL-Threshold problem can be introduced by the encapsulation of a source's initial multicast packet in an SA Message. Because the multicast packet is encapsulated inside of the unicast SA Message (whose TTL = 255), its TTL is not decremented as the SA message travels to the MSDP peer. Furthermore, the total number of hops that the SA message traverses can be drastically different than a normal multicast packet. This is because multicast and unicast traffic may follow completely different paths to the MSDP peer and hence the remote PIM-SM domain. This can result in TTL-Thresholds being violated by this encapsulated packet.**
  - **The solution to this problem is to configure a TTL Threshold that is associated with any multicast packet that is encapsulated in an SA message sent to a particular MSDP peer. This can be accomplished by configuring the following IOS command:**

```
ip msdp ttl-threshold <peer-address> <ttl>
```

    - The above command prevents any multicast packet whose TTL is below ***<ttl>*** from being encapsulated in an SA message sent to the MSDP peer whose IP address is ***<peer-address>***.



Originating SA Messages

- **Use ‘msdp redistribute’ to control what SA’s are originated.**
  - **Think of this as ‘msdp sa-originate-filter’ function**

```
ip msdp redistribute [list <acl>]
                     [asn <aspath-acl>]
                     [route-map <map>]
```

  - **Filter by (S,G) pair using ‘list <acl>’**
  - **Filter by AS-PATH using ‘asn <aspath-acl>’**
  - **Filter based on route-map ‘<map>’**
- **Omitting all acl’s stops all SA origination**

```
Example: ip msdp redistribute
```

- **Default: Originate SA’s for all local sources**
  - **If ‘msdp redistribute’ command is not configured**

Module11.ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 66

- **Originating SA Messages**
  - **By default, an RP configured to run MSDP will originate SA messages for any and all local sources for which it is the RP. In some cases this may not be desirable. (Example: If a sources inside the PIM-SM domain are using private addresses such as network 10.0.0.0/8, it is generally not a good idea to advertise these to other MSDP peers in the Internet.)**
  - **Control over which local sources should be advertised in SA Message can be accomplished using the following IOS command on the RP:**

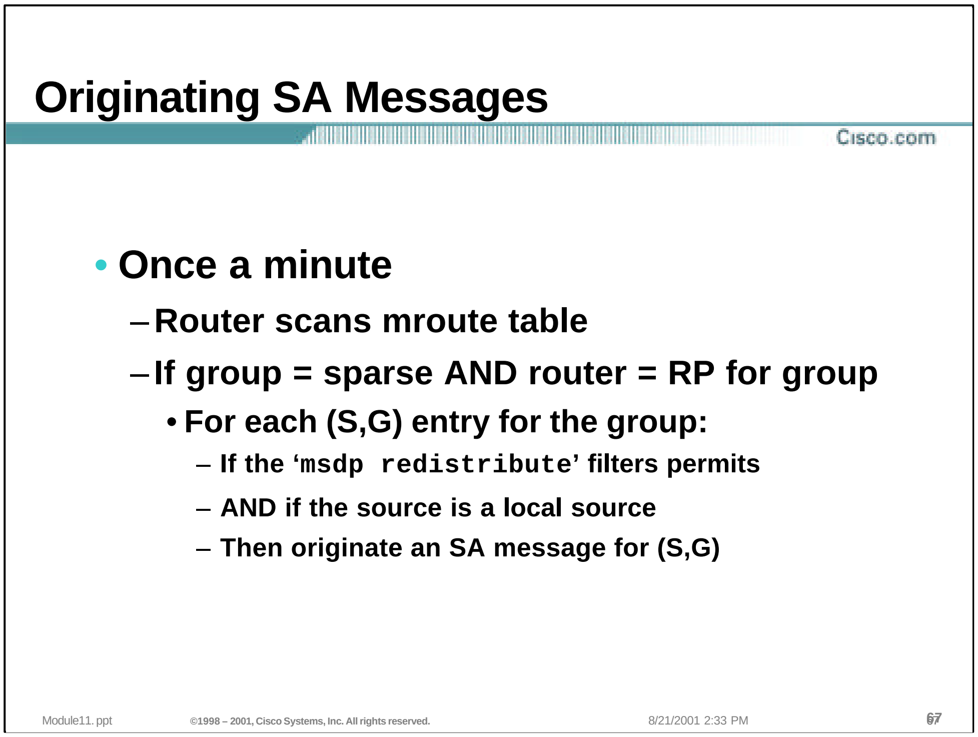
```
ip msdp redistribute [list <acl>]
                     [asn <aspath-acl>]
                     [route-map <map>]
```

This command permits filtering of the SA messages that are “originated” by the RP based on:

- (S, G) pair using the **`list <acl>`** clause.
- AS-PATH using the **`asn <aspath-acl>`** clause.
- Other criteria using the **`route-map <map>`** clause.

Configuring this command without any of the acl or router-map clauses causes all SA origination by this RP to be stopped. (Note: The router will still forward SA messages from other MSDP peers in the normal fashion. It will just not “originate” any of its own.

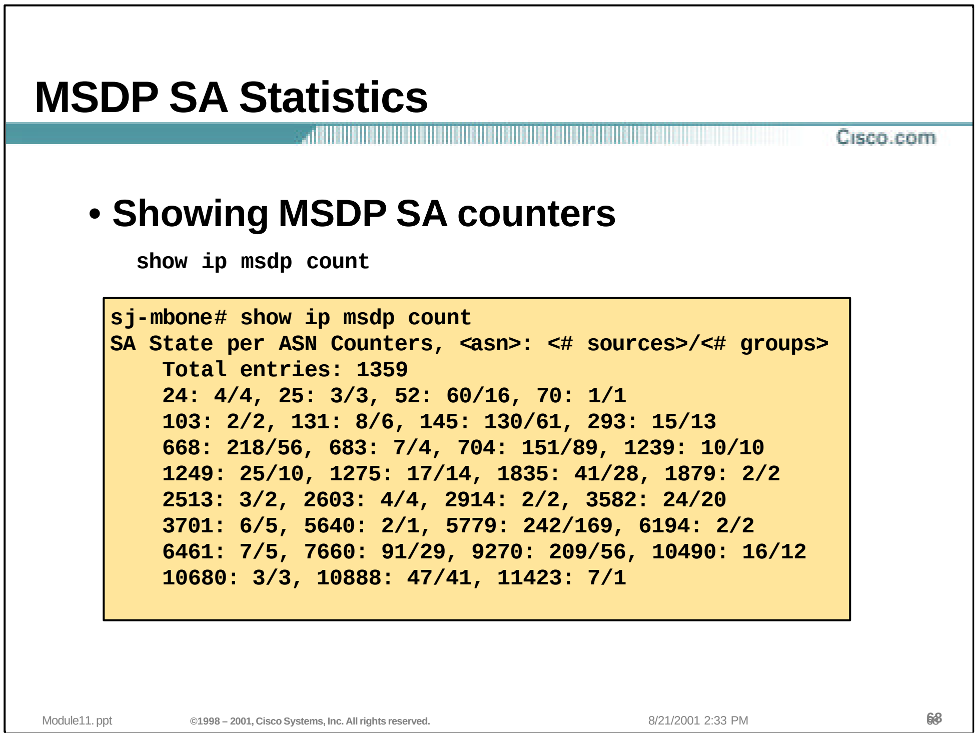
Author’s Note: ***The choice of syntax for this command is a bit confusing and could have been better chosen. The term* redistribute *typically implies some other operation in IOS such as route redistribution. I prefer to mentally translate the syntax of this command into* ‘ip msdp sa-originate-filter’ *because it is more descriptive of what the command actually does.***



Originating SA Messages

- Once a minute
  - Router scans mroute table
  - If group = sparse AND router = RP for group
    - For each (S,G) entry for the group:
      - If the '`msdp redistribute`' filters permits
      - AND if the source is a local source
      - Then originate an SA message for (S,G)

- **Originating SA Messages**
  - **The RP continues to periodically (every 60 seconds) originate SA messages for all active sources in the local PIM-SM domain for which it is functioning as the RP. The details of this mechanism that is performed every minute is as follows:**
  - **For all Sparse mode (*, G) entries in the mroute table for which the router is functioning as the RP, originate an SA message for each subordinate (S, G) entry that meets the following conditions:**
    - The entry must be "permitted" by any 'msdp redistribute' filters AND
    - The source is a local source. (Denoted by the "A" flag.)



MSDP SA Statistics

Cisco.com

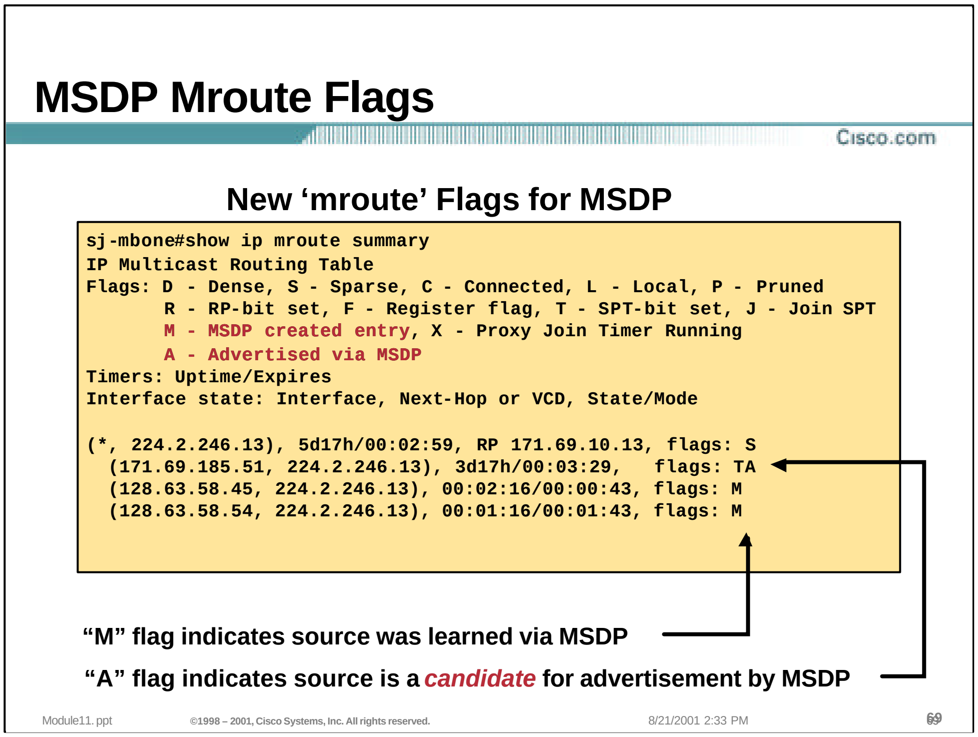
- **Showing MSDP SA counters**

  `show ip msdp count`

```
sj-mbone# show ip msdp count
SA State per ASN Counters, <asn>: <# sources>/<# groups>
    Total entries: 1359
    24: 4/4, 25: 3/3, 52: 60/16, 70: 1/1
    103: 2/2, 131: 8/6, 145: 130/61, 293: 15/13
    668: 218/56, 683: 7/4, 704: 151/89, 1239: 10/10
    1249: 25/10, 1275: 17/14, 1835: 41/28, 1879: 2/2
    2513: 3/2, 2603: 4/4, 2914: 2/2, 3582: 24/20
    3701: 6/5, 5640: 2/1, 5779: 242/169, 6194: 2/2
    6461: 7/5, 7660: 91/29, 9270: 209/56, 10490: 16/12
    10680: 3/3, 10888: 47/41, 11423: 7/1
```

Module11. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 68

- **MSDP Statistics**
  - **Information on the number of sources and groups being advertised on an AS basis can be obtain by the use of the following IOS command:**
    - show ip msdp count
  - **The example ‘show ip msdp count’ shown above indicates that:**
    - There are a total of 1359 sources being advertised via MSDP
    - AS 24 is advertising 4 sources and 4 groups
    - AS 25 is advertising 3 sources and 3 groups
    - AS 52 is advertising 60 sources and 16 groups
    - etc, etc.



MSDP Mroute Flags

Cisco.com

New ‘mroute’ Flags for MSDP

```
sj-mbone#show ip mroute summary
IP Multicast Routing Table
Flags: D - Dense, S - Sparse, C - Connected, L - Local, P - Pruned
       R - RP-bit set, F - Register flag, T - SPT-bit set, J - Join SPT
       M - MSDP created entry, X - Proxy Join Timer Running
       A - Advertised via MSDP
Timers: Uptime/Expires
Interface state: Interface, Next-Hop or VCD, State/Mode

(*, 224.2.246.13), 5d17h/00:02:59, RP 171.69.10.13, flags: S
  (171.69.185.51, 224.2.246.13), 3d17h/00:03:29,   flags: TA
  (128.63.58.45, 224.2.246.13), 00:02:16/00:00:43, flags: M
  (128.63.58.54, 224.2.246.13), 00:01:16/00:01:43, flags: M
```

“M” flag indicates source was learned via MSDP

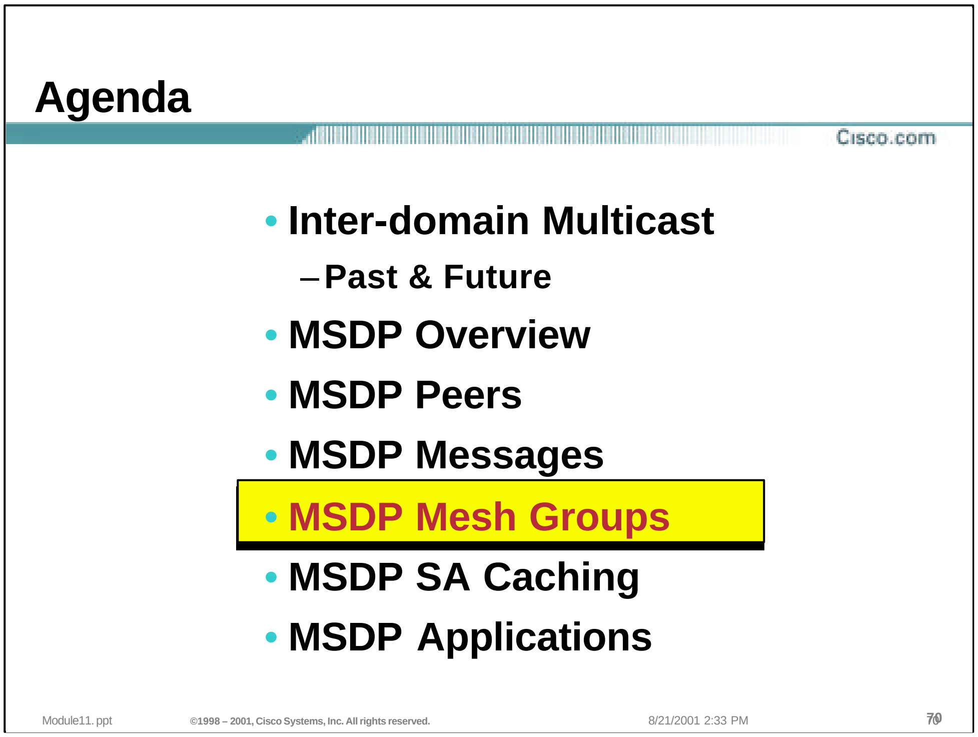
“A” flag indicates source is a *candidate* for advertisement by MSDP

Module11. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 69

- **MSDP Mroute Flags**
  - **MSDP has added two new flags to the complement of flags that may appear on (S, G) entries in the Mroute Table on the RP. These new flags are as follows:**
    - “M” - Indicates that this source was learned via an MSDP SA message.
    - “A” - Indicates that this source is a “candidate” for advertisement in an SA message.

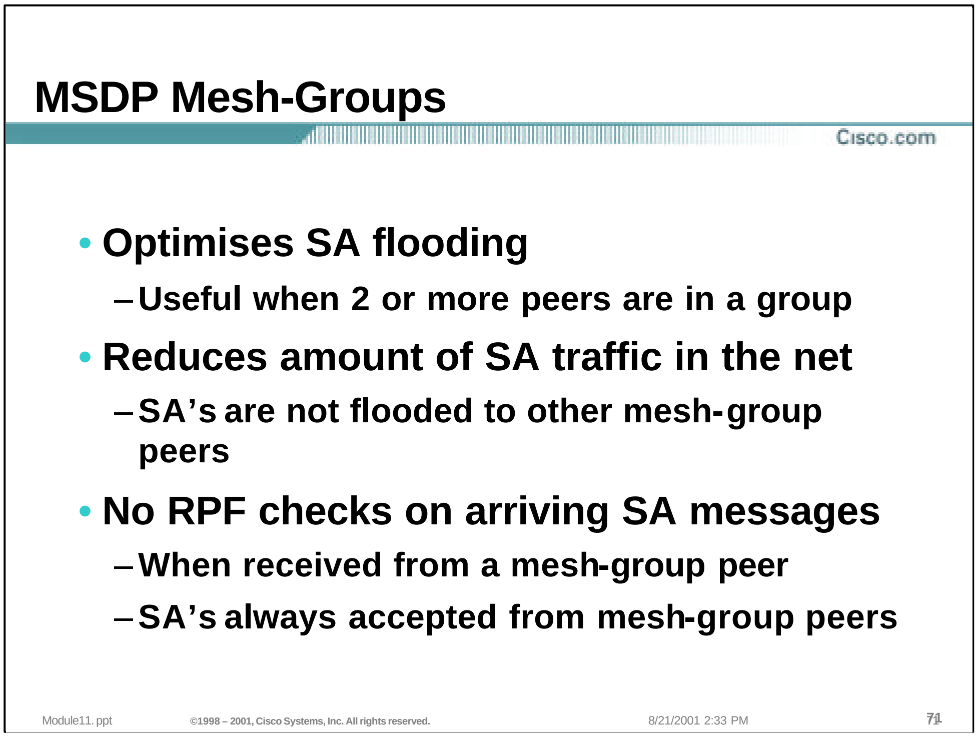
  Note that the flag key is not 100% correct as it is possible for the “A” flag to be set without this source being advertised via MSDP by this router. This could be the case if an ‘msdp redistribute’ filter were in use which “denied” a particular source from being advertised.

  Author’s Note: ***I prefer to think of the “A” flag as an indicator that the source is a local source in the PIM-SM domain and thus is a “candidate” for the RP to advertise via MSDP. Unfortunately, there is currently no flag that positively indicates that a source actually is being advertised by MSDP. About the only way to determine that is to turn on ‘debug ip msdp’.***



Agenda

- Inter-domain Multicast
  - Past & Future
- MSDP Overview
- MSDP Peers
- MSDP Messages
- **MSDP Mesh Groups**
- MSDP SA Caching
- MSDP Applications

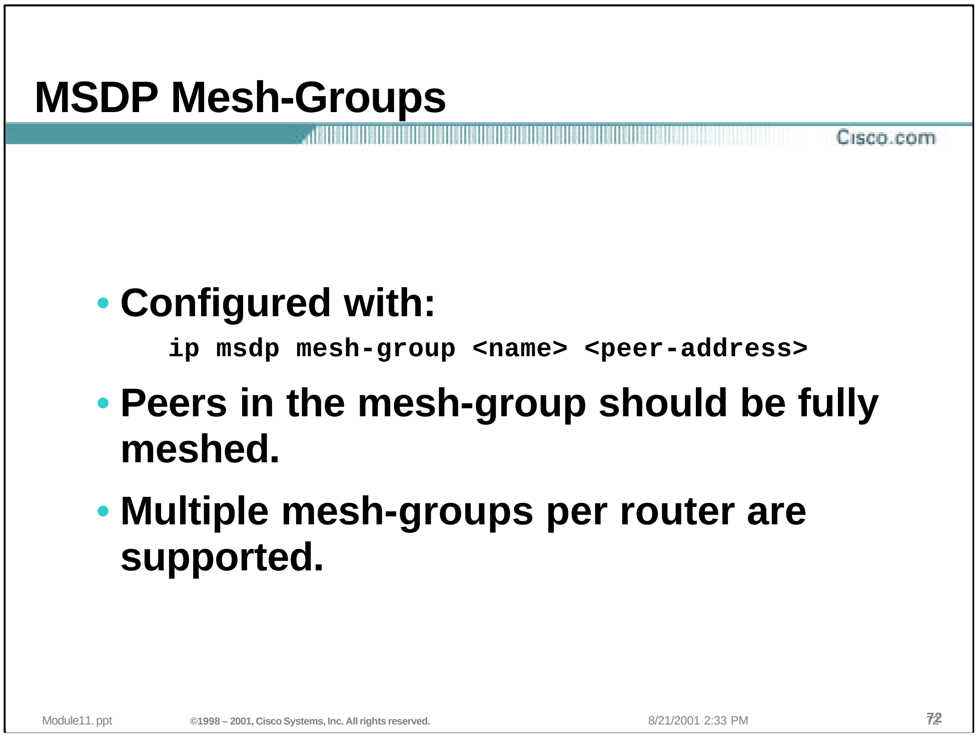


MSDP Mesh-Groups

- Optimises SA flooding
  - –Useful when 2 or more peers are in a group
- Reduces amount of SA traffic in the net
  - –SA’s are not flooded to other mesh-group peers
- No RPF checks on arriving SA messages
  - –When received from a mesh-group peer
  - –SA’s always accepted from mesh-group peers

Module11. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 71

- MSDP Mesh-Groups
  - An MSDP Mesh-Groups can be configured on a group of MSDP peers that are fully meshed. (In other words, each of the MSDP peers in the group has an MSDP connection to every other MSDP peer in the group.
  - When an MSDP Mesh-Group is configured between a group of MSDP peers, SA flooding is reduced. This is because when an MSDP peer in the group receives an SA message from another MDP peer in the group, it can assume that this SA message was sent to all the other MSDP peers in the group. As a result, it is not necessary nor desirable for the receiving MSDP peer to flood the SA message to the other MSDP peers in the group.
  - MSDP Mesh-Groups may also be used to eliminate the need to run (M)BGP to do RPF checks on arriving SA messages. This is because SA messages are never flooded to other MSDP peers in the mesh-group. As a result, it is not necessary to perform the RPF check on arriving SA messages.



MSDP Mesh-Groups

Cisco.com

- **Configured with:**

```
ip msdp mesh-group <name> <peer-address>
```

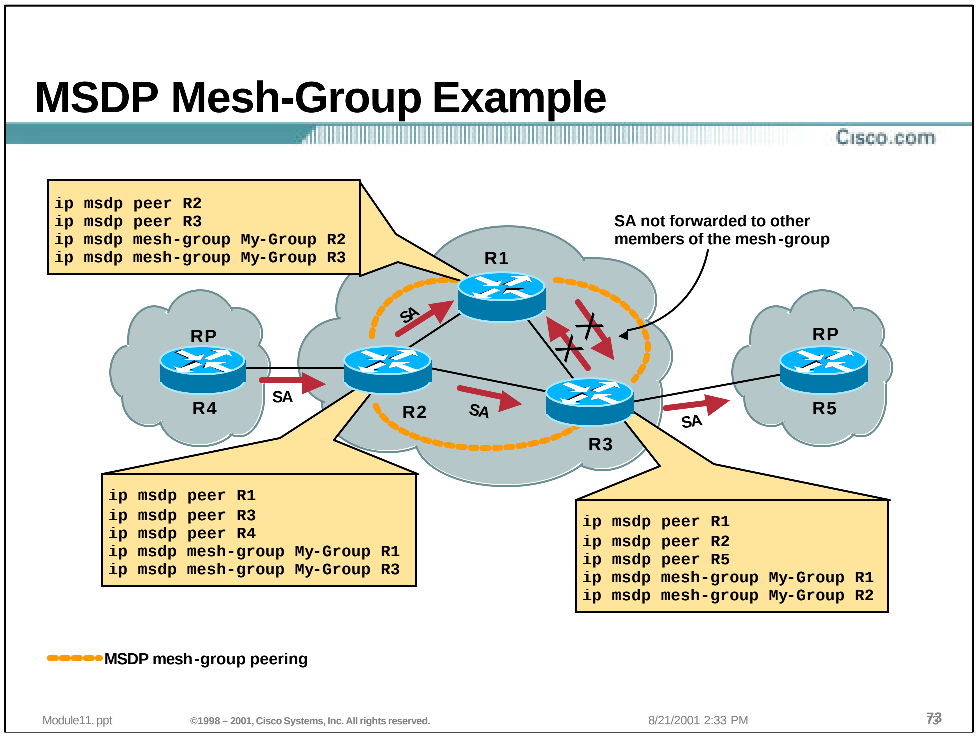
- **Peers in the mesh-group should be fully meshed.**
- **Multiple mesh-groups per router are supported.**

Module11. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 72

- **MSDP Mesh-Groups**
  - **A MSDP Mesh-Group may be configured by using the following IOS configuration command:**

    ```
    ip msdp mesh-group <name> <peer-address>
    ```

    This command configures the router as a member of the mesh-group *<name>* **for which MSDP peer** *<peer-address>* **is also a member.**
  - **All MSDP peers in the mesh-group must be fully-meshed with all other MSDP peers in the group. This means that each router must be configured with 'ip msdp peer' and 'ip msdp mesh-group' commands for each member of the mesh-group.**
  - **Routers may be members of multiple Mesh-Groups.**



MSDP Mesh-Group Example

```
ip msdp peer R2
ip msdp peer R3
ip msdp mesh-group My-Group R2
ip msdp mesh-group My-Group R3
```

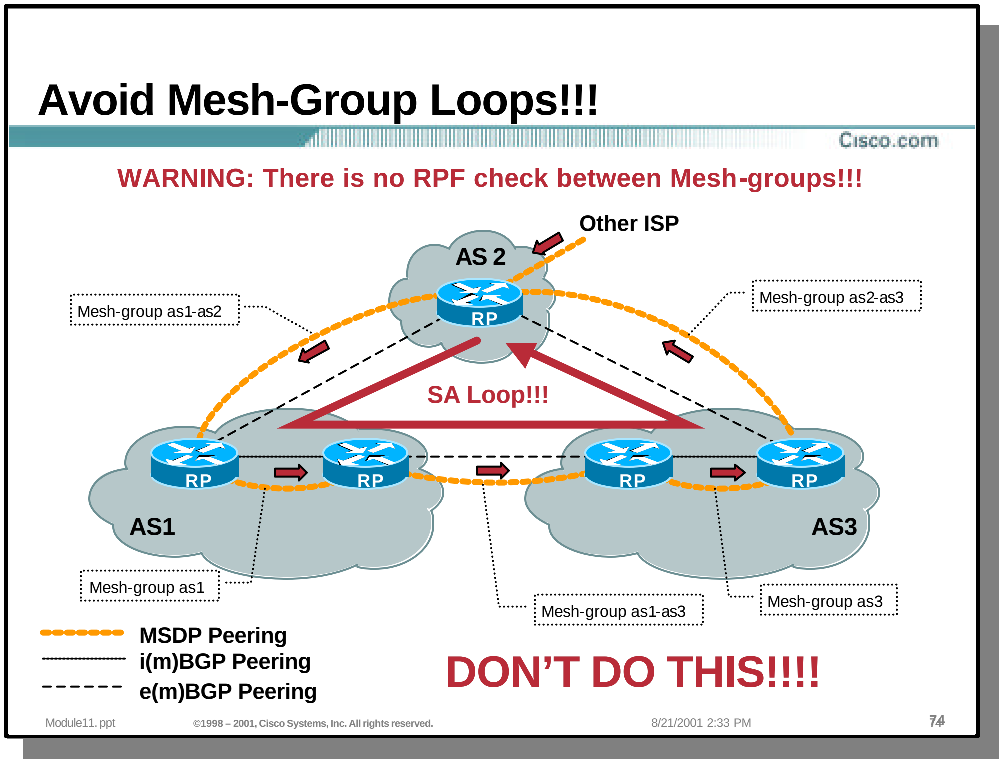
SA not forwarded to other members of the mesh-group

```
ip msdp peer R1
ip msdp peer R3
ip msdp peer R4
ip msdp mesh-group My-Group R1
ip msdp mesh-group My-Group R3
```

```
ip msdp peer R1
ip msdp peer R2
ip msdp peer R5
ip msdp mesh-group My-Group R1
ip msdp mesh-group My-Group R2
```

MSDP mesh-group peering

- **MSDP Mesh-Group Example.**
  - **In the above example, routers R1, R2 and R3 are all configured as members of the same MSDP mesh-group. In addition, router R1 is also MSDP peering with router R4 and router R3 is MSDP peering with router R5. Neither R4 nor R5 are members of the MSDP mesh-group.**
  - **Assume router R4 originates an SA message for a source in its local PIM-SM domain. This message is sent to route R2 as shown in the drawing above.**
  - **When router R2 receives this SA message, it must perform an RPF check on the message because it was received from an MSDP peer that is not a member of the mesh-group.**
  - **In this case the RPF check is successful and router R2 floods the SA message (received from a non-mesh-group member) to all other members of the mesh-group.**
  - **When routers R1 and R3 receive the SA message from mesh-group member R2, they do not have to perform an RPF check on the arriving message nor do they flood the SA message to each other since they are both members of the mesh-group. (They know that the other members of the mesh-group will have received a copy directly from R2 and therefore they do not have to forward the SA message to each other. This is why a full mesh between mesh-group members is required.)**
  - **Finally, router R3 floods the SA message to all of its MSDP peers that are not members of the mesh-group. In this case, the SA message is flooded to router R5 to continue the flow of the SA message downstream away from the RP.**



Avoid Mesh-Group Loops!!!

Cisco.com

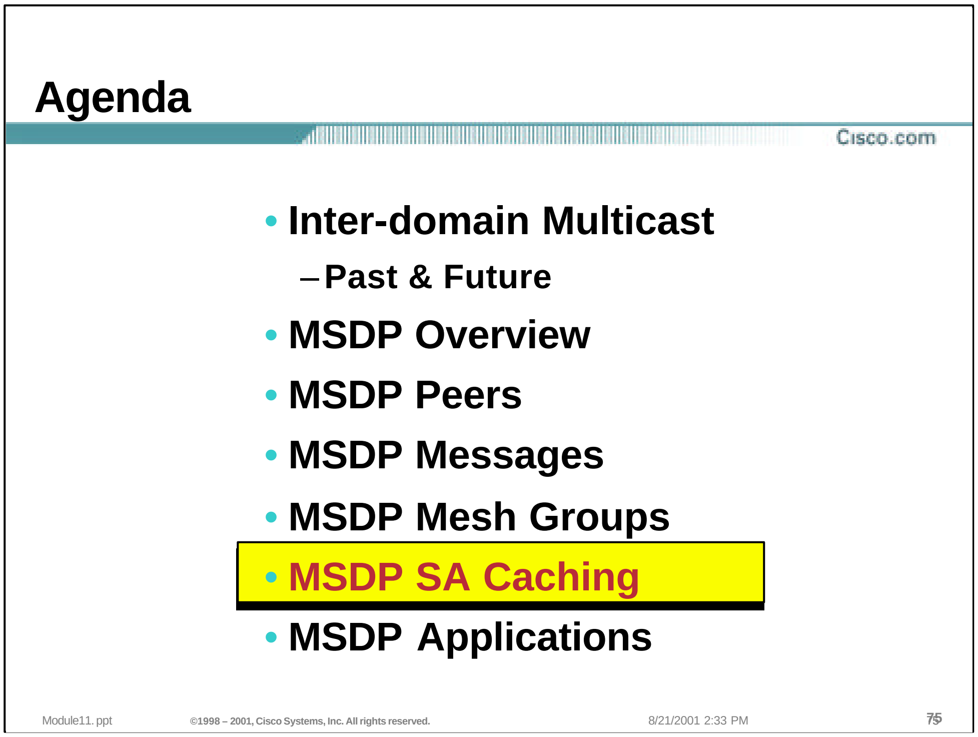
WARNING: There is no RPF check between Mesh-groups!!!

MSDP Peering
i(m)BGP Peering
e(m)BGP Peering

DON’T DO THIS!!!!

Module11. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 74

- **Avoid Mesh-Group Loops!!!**
  - **There is no RPF checking between Mesh-groups.**
  - **Insure that you aren’t creating a continuous loop of Mesh-Groups.**



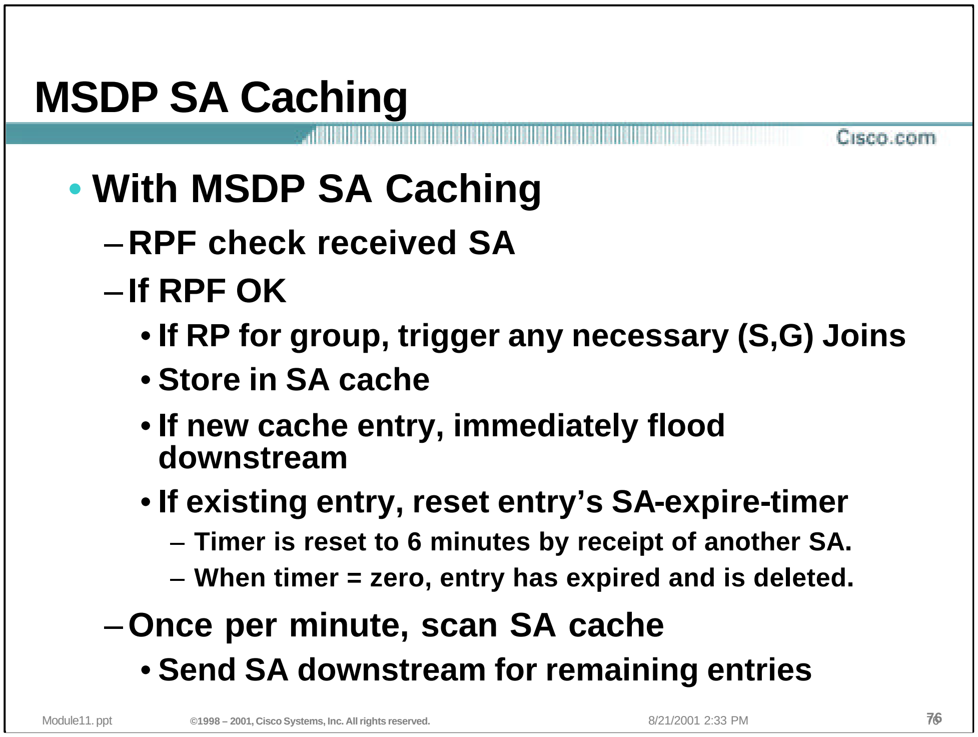
Agenda

Cisco.com

- Inter-domain Multicast
  - Past & Future
- MSDP Overview
- MSDP Peers
- MSDP Messages
- MSDP Mesh Groups
- **MSDP SA Caching**
- MSDP Applications

Module11.ppt

8/21/2001 2:33 PM



MSDP SA Caching

- With MSDP SA Caching
  - RPF check received SA
  - If RPF OK
    - If RP for group, trigger any necessary (S,G) Joins
    - Store in SA cache
    - If new cache entry, immediately flood downstream
    - If existing entry, reset entry's SA-expire-timer
      - Timer is reset to 6 minutes by receipt of another SA.
      - When timer = zero, entry has expired and is deleted.
  - Once per minute, scan SA cache
    - Send SA downstream for remaining entries

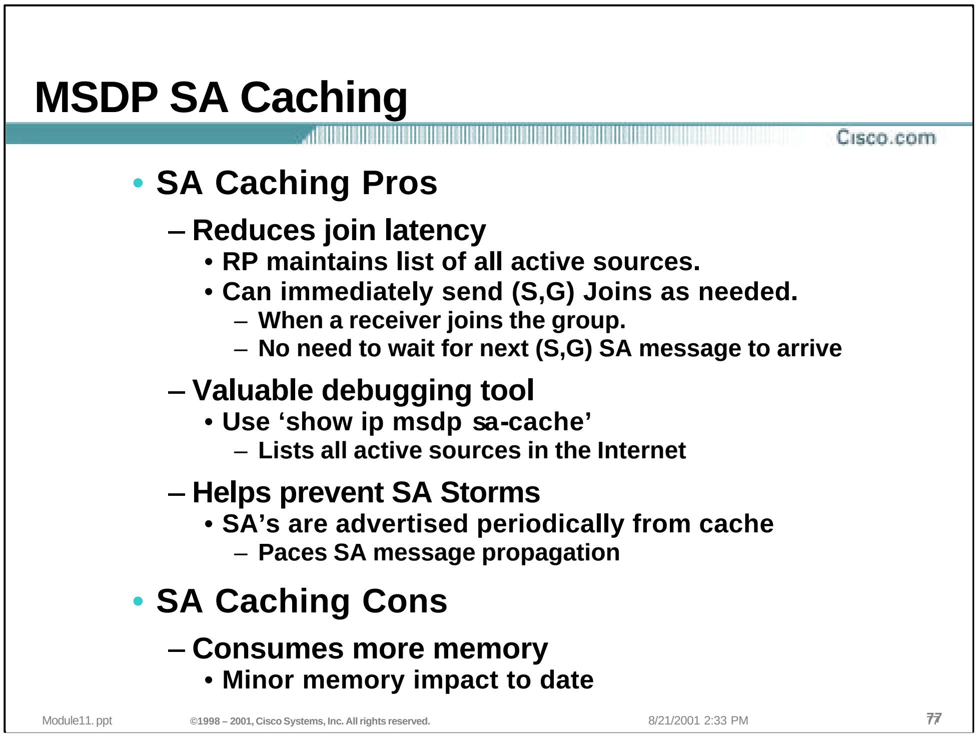
Module11. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 76

- **With MSDP SA Caching**
  - **Arriving SA messages are RPF checked in the normal fashion.**

    Note: Arriving SA messages must pass any incoming SA filters that have been configured.

  - **If the RPF check succeeds:**
    - An (S,G) join is triggered toward this source if the router is the active RP for this group AND it has a non-null (*,G) OIL (which indicates that there are members for the group on the shared tree).
    - The router then stores the SA message in the SA cache. If this is a new entry, the SA message is immediately flooded downstream to all MSDP neighbors. Next, the SA-expire-timer for the entry (new or existing) is reset to 6 minutes. Each time an (S,G) SA message is received, this timer in the cache entry is reset to 6 minutes. If the timer ever counts down to zero, then the entry is expired and it is removed from the SA cache.
  - **Once per minute**
    - The router scans the SA cache and sends an SA message downstream for each unexpired cache entry.

      Note: The SA messages must pass any outgoing SA filters that have been configured before it is sent.



MSDP SA Caching

- SA Caching Pros
  - Reduces join latency
    - RP maintains list of all active sources.
    - Can immediately send (S,G) Joins as needed.
      - When a receiver joins the group.
      - No need to wait for next (S,G) SA message to arrive
  - Valuable debugging tool
    - Use ‘show ip msdp sa-cache’
      - Lists all active sources in the Internet
  - Helps prevent SA Storms
    - SA’s are advertised periodically from cache
      - Paces SA message propagation
- SA Caching Cons
  - Consumes more memory
    - Minor memory impact to date

Module11.ppt 8/21/2001 2:33 PM 77

- **SA Caching Pros**
  - **Reduces Join latency**

    When this SA caching is configured, the router will begin caching all (S,G) pairs received in SA messages. This reduces join latency as the RP maintains a list of all active sources. Therefore, when the first receiver joins the group, the RP doesn’t have to wait 60 seconds for the next SA message before sending out the (S,G) Join.

  - **Valuable debugging tool**

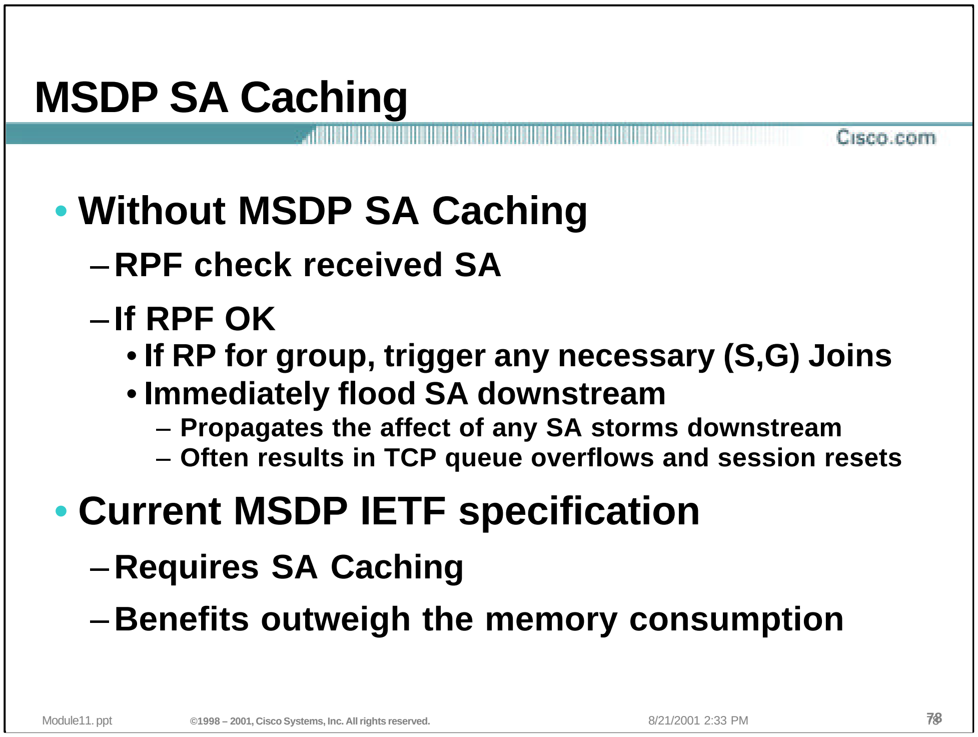
    The contents of the SA cache is a valuable source of MSDP debugging information. The ‘**show ip msdp sa-cache**’ command will list its contents and display all active sources in the Internet along with information on what AS they reside and what RP advertised it.

  - **Helps prevent SA Storms**

    Since SA’s are advertised periodically from cache (instead of as soon as they are received from an upstream neighbor), the propagation of SA messages through out the Internet are better paced. This helps to avoid overrunning TCP input queues on the MSDP peers which results in session resets and instability of MSDP.

- **SA Caching Cons**
  - **Memory consumption**

    The memory impact of turning on SA Caching in most RP’s is, in general, very small.

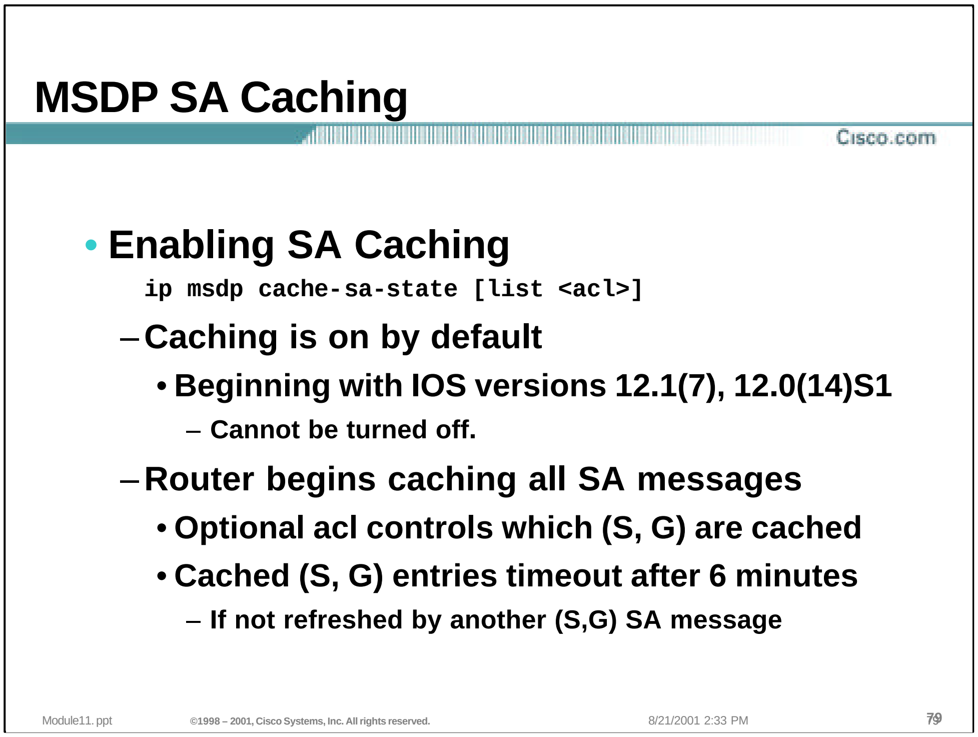


MSDP SA Caching

- Without MSDP SA Caching
  - –RPF check received SA
  - –If RPF OK
    - If RP for group, trigger any necessary (S,G) Joins
    - Immediately flood SA downstream
      - – Propagates the affect of any SA storms downstream
      - – Often results in TCP queue overflows and session resets
- Current MSDP IETF specification
  - –Requires SA Caching
  - –Benefits outweigh the memory consumption

- **Without MSDP SA Caching**
  - **Arriving SA messages are RPF checked in the normal fashion.**

    Note: Arriving SA messages must pass any incoming SA filters that have been configured.
  - **If the RPF check succeeds:**
    - An (S,G) join is triggered toward this source if the router is the active RP for this group AND it has a non-null (*,G) OIL (which indicates that there are members for the group on the shared tree).
    - The router then immediately forwards the SA message to all downstream MSDP peers. This behavior does nothing to prevent SA Storms since SA's are flooded to downstream peers as fast as they arrive.
- **Current MSDP IETF specification**
  - **The current draft of the MSDP specification *requires* an implementation to use SA caching and to pace the transmission of SA messages to downstream neighbors. This was added to the specification to help prevent SA Storms which had occurred in the early days of MSDP usage in the Internet.**



MSDP SA Caching

- Enabling SA Caching

  `ip msdp cache-sa-state [list <acl>]`

  – Caching is on by default

    - Beginning with IOS versions 12.1(7), 12.0(14)S1

      – Cannot be turned off.

  – Router begins caching all SA messages

    - Optional acl controls which (S, G) are cached
    - Cached (S, G) entries timeout after 6 minutes

      – If not refreshed by another (S,G) SA message

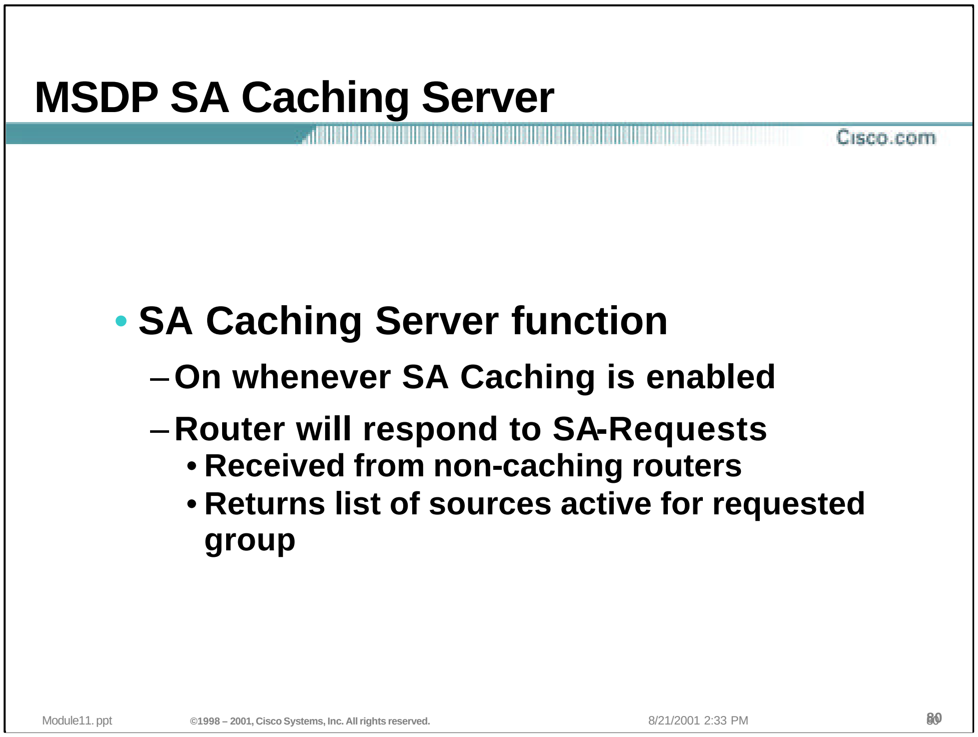
Module11. ppt 8/21/2001 2:33 PM 79

- **Enabling SA Caching**
  - **SA caching can be enabled by the use of the following IOS command :**

    ```
    ip msdp cache-sa-state [list <acl>]
    ```

    When this command is configured, the router will begin caching all (S,G) pairs received in SA messages. This reduces join latency as the RP maintains a list of all active sources. (The memory impact of turning on SA Caching in most RP's is, in general, very small. The other benefit is that additional information becomes available for network debugging if SA Caching is enabled.)

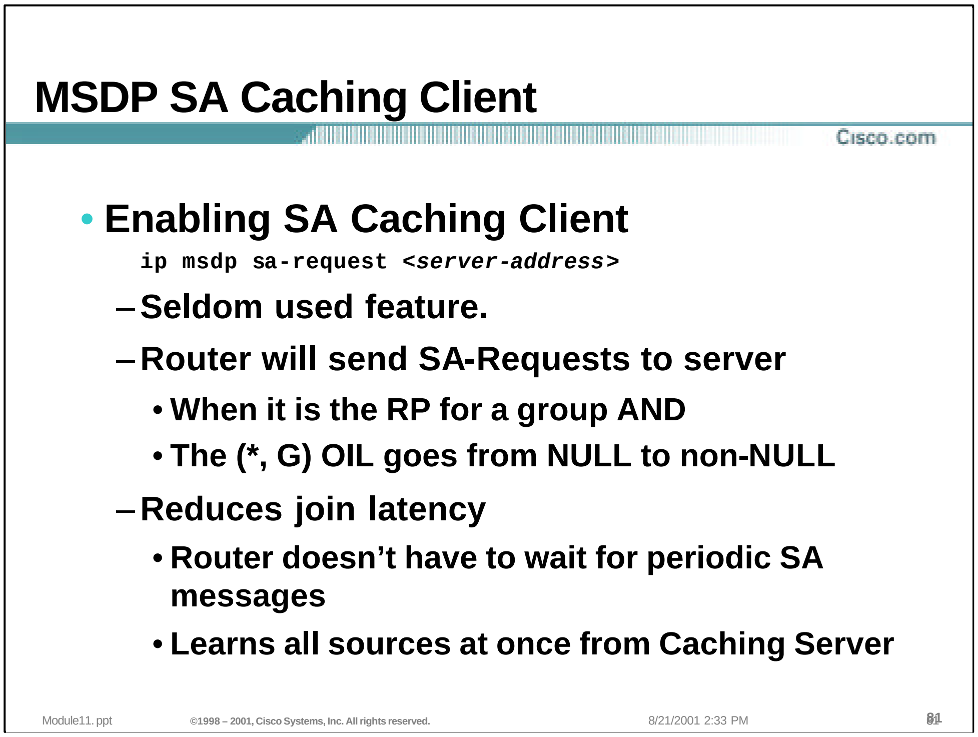
    - The optional **`list <acl>`** clause may be used to control which (S,G) pairs are cached. If this optional clause is not specified, all (S,G) pairs are cached.
    - (S,G) pairs in the SA cache have a 6 minute timeout. If a new SA message is not received with this (S,G) pair in that period, the entry is removed from the cache.



MSDP SA Caching Server

- SA Caching Server function
  - On whenever SA Caching is enabled
  - Router will respond to SA-Requests
    - Received from non-caching routers
    - Returns list of sources active for requested group

- **MSDP SA Caching Server**
  - **If SA Caching has been enabled on a router, it will respond to SA Request messages received from other routers. When an SA Request for group “G” is received by a router with SA Caching enabled, the router responds by sending back SA Response messages containing a list of all active sources for the requested group “G”.**



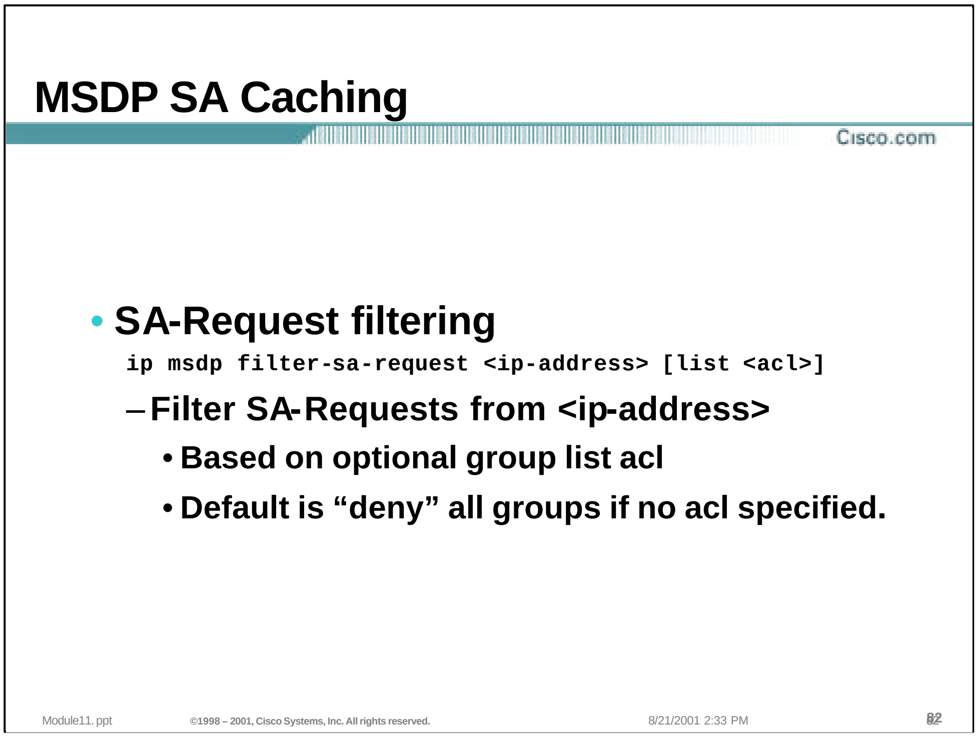
MSDP SA Caching Client

- **Enabling SA Caching Client**

```
ip msdp sa-request <server-address>
```

  - **Seldom used feature.**
  - **Router will send SA-Requests to server**
    - **When it is the RP for a group AND**
    - **The (*, G) OIL goes from NULL to non-NULL**
  - **Reduces join latency**
    - **Router doesn’t have to wait for periodic SA messages**
    - **Learns all sources at once from Caching Server**

- **MSDP SA Caching**
  - **Routers that do not have SA Caching enabled can benefit from other routers that are SA Caching enabled by sending SA Request messages whenever receivers first join a group. In order to enable the sending of SA Requests, the following IOS command must be configured:**
    - ip msdp sa-request <server-address>
  - **This command will cause the router to send an SA-Request message to the SA Caching router whose IP address is <server-address> under the following conditions:**
    - When the outgoing interface list of a (*,G) entry goes from Null to non-Null AND
    - The router is the RP for group “G”.



MSDP SA Caching

- SA-Request filtering

  `ip msdp filter-sa-request <ip-address> [list <acl>]`

  – Filter SA-Requests from <ip-address>

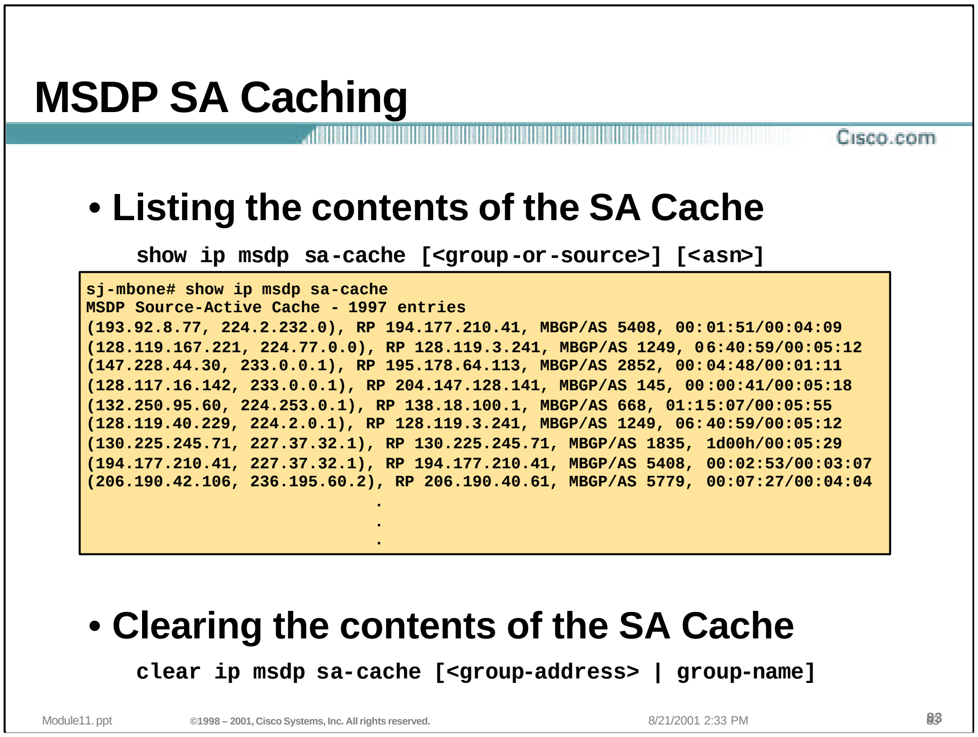
    - Based on optional group list acl
    - Default is “deny” all groups if no acl specified.

- **SA Request Filtering**

  – **In some situations, a router that has SA Caching enabled may not wish to honor all received SA Request messages. If this is desired, an SA Request filter may be configured using the following IOS command :**

```
ip msdp filter-sa-request <ip-address> [list <acl>]
```

This command will cause the router to filter all SA-Requests received from the router whose IP address is *<ip-address>* **based on the optional** `list` *<acl>* **clause. If the optional** `list` *<acl>* **clause is not specified, the default behavior is to not respond to SA Requests from this router.**



MSDP SA Caching

Cisco.com

- **Listing the contents of the SA Cache**

```
show ip msdp sa-cache [<group-or-source>] [<asn>]
```

```
sj-mbone# show ip msdp sa-cache
MSDP Source-Active Cache - 1997 entries
(193.92.8.77, 224.2.232.0), RP 194.177.210.41, MBGP/AS 5408, 00:01:51/00:04:09
(128.119.167.221, 224.77.0.0), RP 128.119.3.241, MBGP/AS 1249, 06:40:59/00:05:12
(147.228.44.30, 233.0.0.1), RP 195.178.64.113, MBGP/AS 2852, 00:04:48/00:01:11
(128.117.16.142, 233.0.0.1), RP 204.147.128.141, MBGP/AS 145, 00:00:41/00:05:18
(132.250.95.60, 224.253.0.1), RP 138.18.100.1, MBGP/AS 668, 01:15:07/00:05:55
(128.119.40.229, 224.2.0.1), RP 128.119.3.241, MBGP/AS 1249, 06:40:59/00:05:12
(130.225.245.71, 227.37.32.1), RP 130.225.245.71, MBGP/AS 1835, 1d00h/00:05:29
(194.177.210.41, 227.37.32.1), RP 194.177.210.41, MBGP/AS 5408, 00:02:53/00:03:07
(206.190.42.106, 236.195.60.2), RP 206.190.40.61, MBGP/AS 5779, 00:07:27/00:04:04
                             .
                             .
                             .
```

- **Clearing the contents of the SA Cache**

```
clear ip msdp sa-cache [<group-address> | group-name]
```

- **MSDP SA Caching**
  - **The contents of the SA Cache can be very helpful to debugging MSDP problems in the network. (This is why most network administrators enable SA Caching on all MSDP routers.) The following IOS command can be used to display the contents of the SA Cache:**

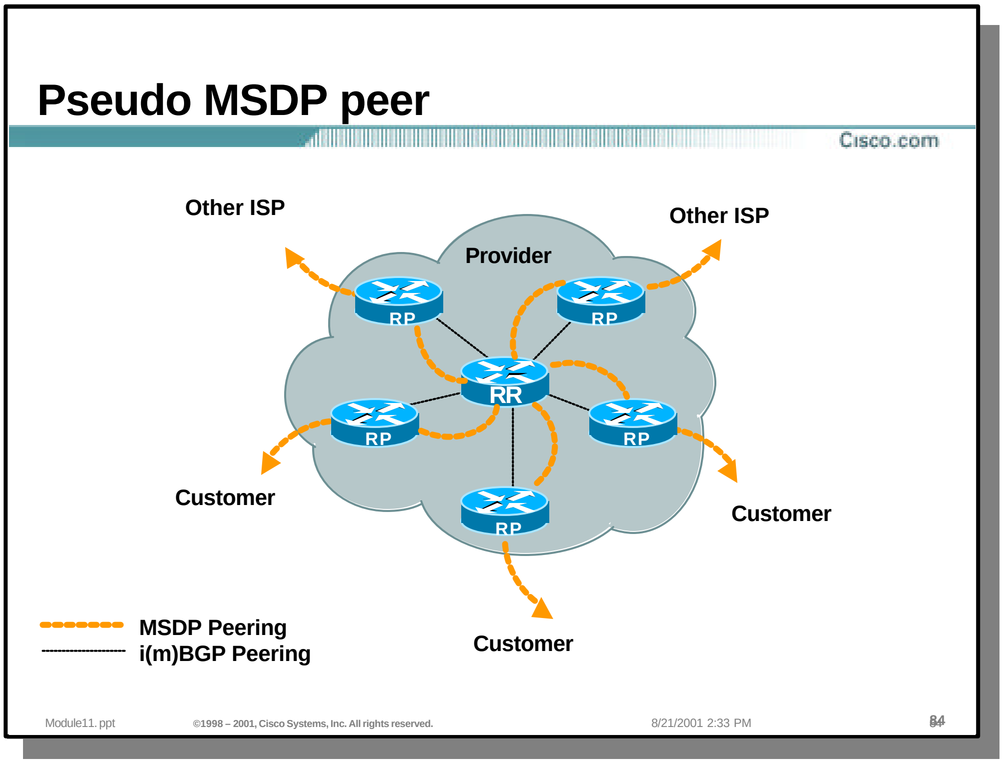
    `show ip msdp sa-cache [<`*`group-or-source`*`>] [<`*`asn`*`>]`

    The above command lists the contents of the SA Cache. The optional *`<group-or-source>`* **and** *`<asn>`* **qualifiers may be used to limited the displayed output to only those desired entries.**
  - **In the above example of this command we see that there are 1997 entries in the SA Cache. The information on the first entry is as follows:**

• (192.92.8.77, 224.2.232.0)	= Active source/group information
• RP 194.177.210.41	= IP address of the originating RP.
• MBGP/AS 5408	= The RP resides in AS 5408
• 00:01:51/00:04:09	= The source has been active for 1 min, 51 sec and will expire in 4 min, 9 sec.

  - **The contents of the SA Cache can be cleared by the use of the following IOS command:**

    `clear ip msdp sa-cache [<`*`group-address`*`> | <`*`group-name`*`>]`
  - **The optional** *`<group-address>`* **and** *`<group-name>`* **qualifiers may be specified to limit which entries are to be cleared from the cache.**

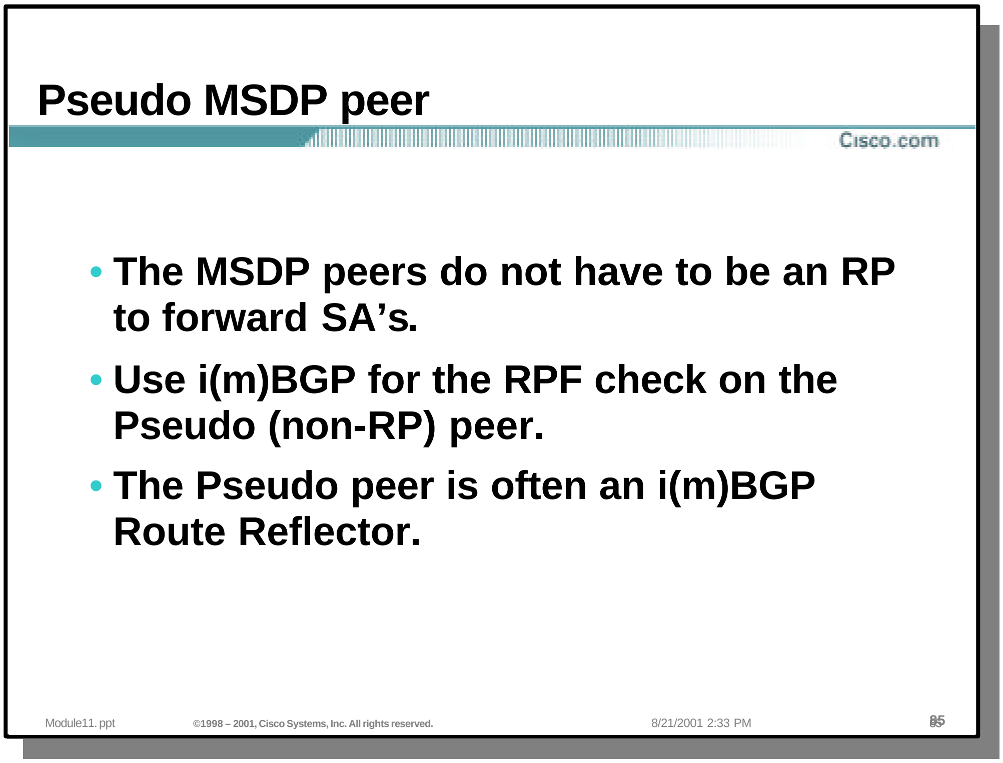


Pseudo MSDP peer

Cisco.com

Module11. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 84

- **Pseudo MSDP Peer**
  - **MSDP peers do not have to be an RP. They will still forward any SA's (that pass the RPF check) to other MSDP peers.**
  - **i(m)BGP must be used in parallel between the Pseudo (non-RP) MSDP peer and all other MSDP peers. This is necessary so that SA messages can be RPF checked.**
  - **The Pseudo MSDP peer concept is often used on a (m)BGP Route Reflector.**



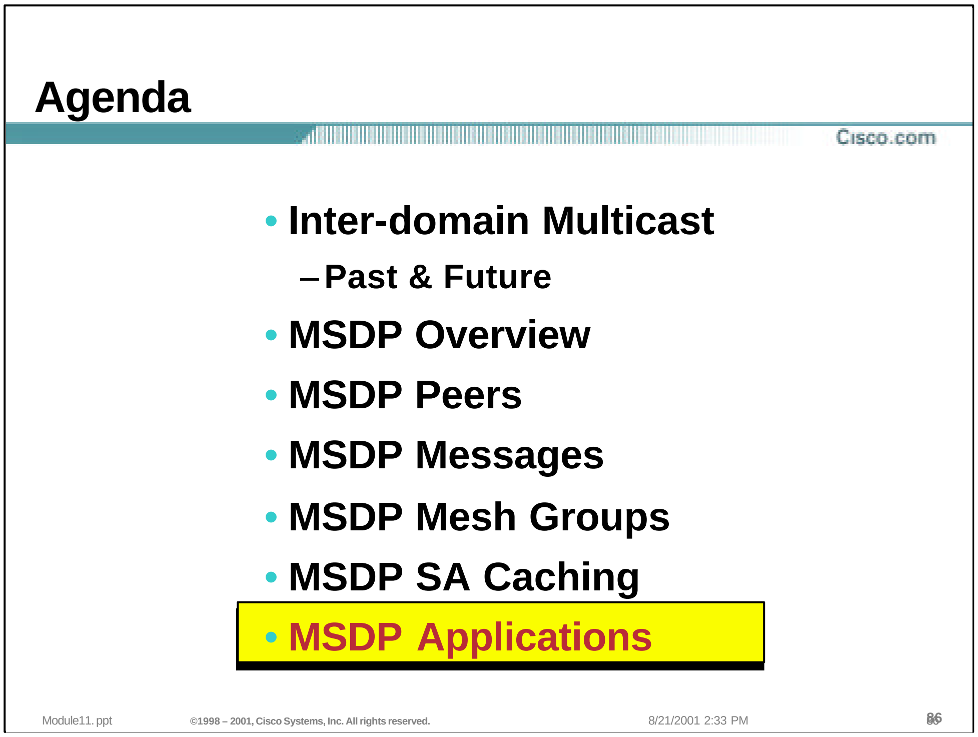
Pseudo MSDP peer

Cisco.com

- The MSDP peers do not have to be an RP to forward SA’s.
- Use i(m)BGP for the RPF check on the Pseudo (non-RP) peer.
- The Pseudo peer is often an i(m)BGP Route Reflector.

Module11. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 85

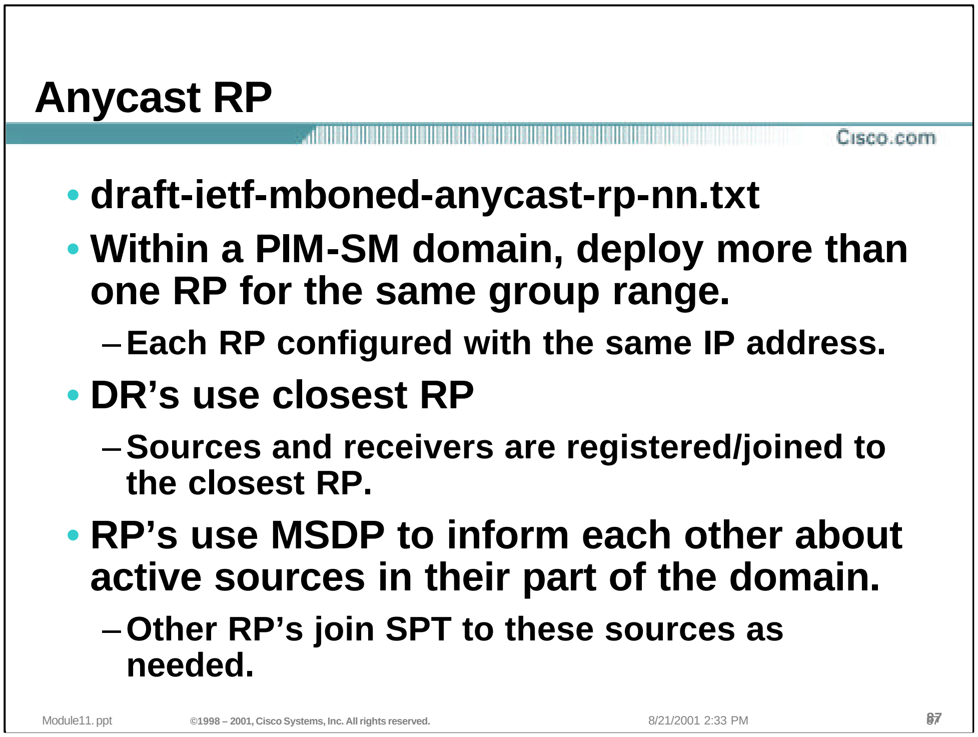
- **Pseudo MSDP Peer**
  - MSDP peers do not have to be an RP. They will still forward any SA’s (that pass the RPF check) to other MSDP peers.
  - i(m)BGP must be used in parallel between the Pseudo (non-RP) MSDP peer and all other MSDP peers. This is necessary so that SA messages can be RPF checked.
  - The Pseudo MSDP peer concept is often used on a (m)BGP Route Reflector.



Agenda

Cisco.com

- Inter-domain Multicast
  - Past & Future
- MSDP Overview
- MSDP Peers
- MSDP Messages
- MSDP Mesh Groups
- MSDP SA Caching
- MSDP Applications



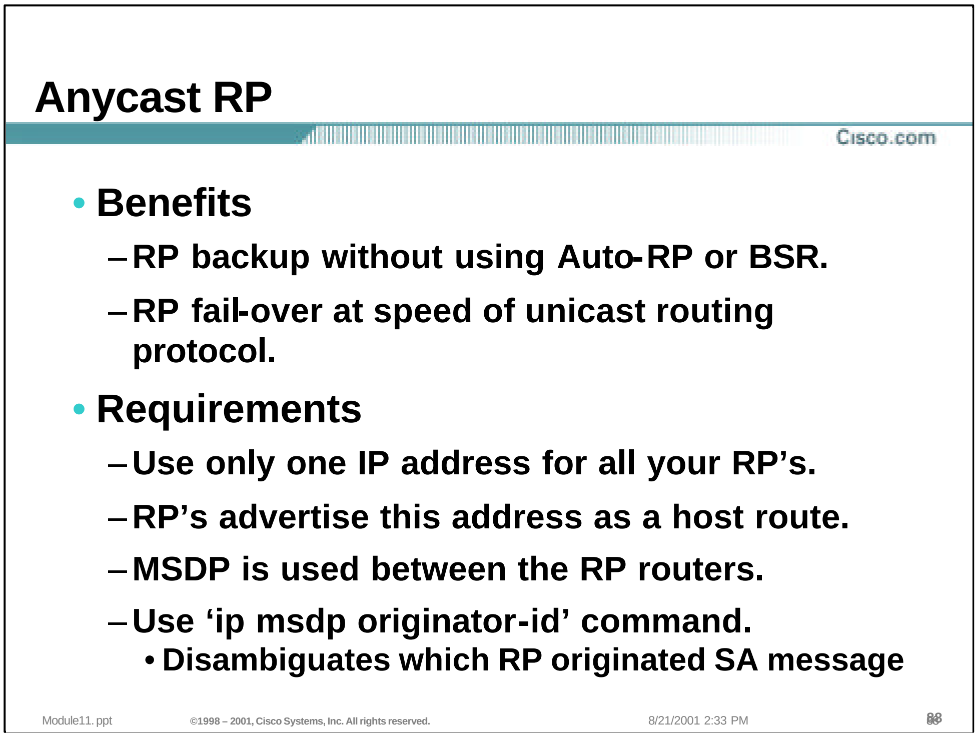
Anycast RP

Cisco.com

- draft-ietf-mboned-anycast-rp-nn.txt
- Within a PIM-SM domain, deploy more than one RP for the same group range.
  - Each RP configured with the same IP address.
- DR's use closest RP
  - Sources and receivers are registered/joined to the closest RP.
- RP's use MSDP to inform each other about active sources in their part of the domain.
  - Other RP's join SPT to these sources as needed.

Module11.ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 87

- **Anycast RP**
  - **MSDP may be used to implement the concept of Anycast RP's within a PIM-SM domain to provide RP redundancy, rapid RP fail-over and RP load-balancing. This concept was first documented in the following IETF draft:**
    - draft-ietf-mboned-anycast-rp-nn.txt
  - **The Anycast RP mechanism works as follows:**
    - Two or more routers are configured as active RP for the same group range at the same time. (This is normally is a configuration error that would partition the PIM -SM domain. However, MSDP is used to prevent this from happening.)
    - Each RP is assigned the same RP Address. (This is usually accomplished using a Loopback interface and private address space.) Each router advertises its RP address as a host route in the unicast routing protocol.
    - Sources and receivers (more specifically, their DR's) will use the closest RP based on their unicast routing table.
    - The Anycast RP's are all connected via MSDP. This allows each RP to learn which sources have been registered with the other Anycast RP's in the domain.
    - The normal PIM -SM RP behavior will result in the RP's joining the source tree of active sources in the other parts of the network as necessary.



Anycast RP

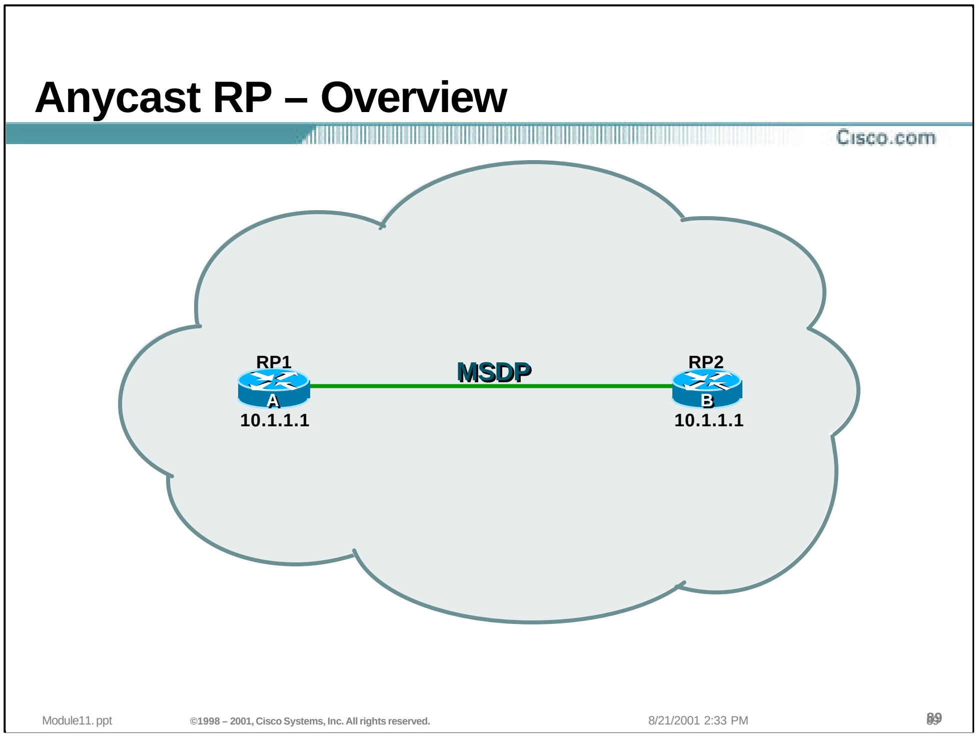
Cisco.com

- Benefits
  - RP backup without using Auto-RP or BSR.
  - RP fail-over at speed of unicast routing protocol.
- Requirements
  - Use only one IP address for all your RP’s.
  - RP’s advertise this address as a host route.
  - MSDP is used between the RP routers.
  - Use ‘ip msdp originator-id’ command.
    - Disambiguates which RP originated SA message

Module11. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 88

- **Anycast RP Benefits**
  - **Anycast RPs provide for backup RP’s without having to use Auto-RP or BSR.**
  - **RP fail-over occurs roughly at the same speed as the unicast routing protocol converges.**
- **Anycast RP Requirements**
  - **All Anycast RP’s are configured to use the same IP address.**
  - **All Anycast RP’s advertise this IP address as a host route. This causes the DR’s in the network to see only the closest RP.**

    Note: **If there does happen to be a metric tie, the normal RPF mechanism will select the only one path back to the RP. The path selected will be the one that has the highest next hop address.**
  - **All Anycast RP’s are tied together via MSDP peering sessions.**
  - **The ‘ip msdp originator-id’ command is used to control the IP address that is sent in any SA messages that are originated by an RP.**
    - This is done to disambiguate which RP originated the SA message. If this were not done, all RP’s would originate SA messages using the same IP address.



Anycast RP – Overview

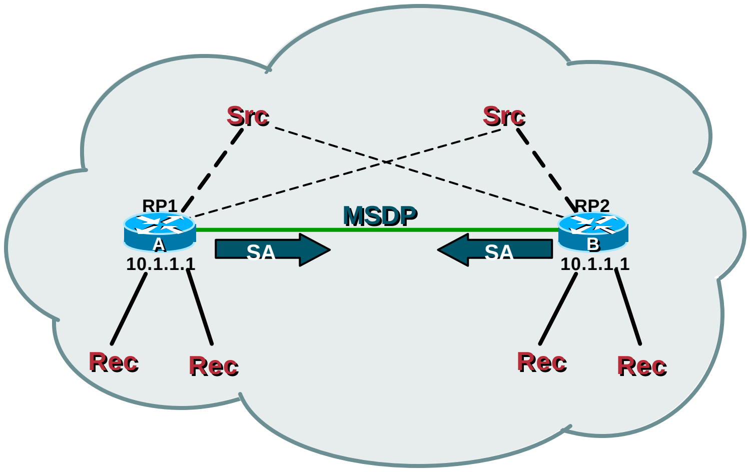
Cisco.com

Module11.ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 89

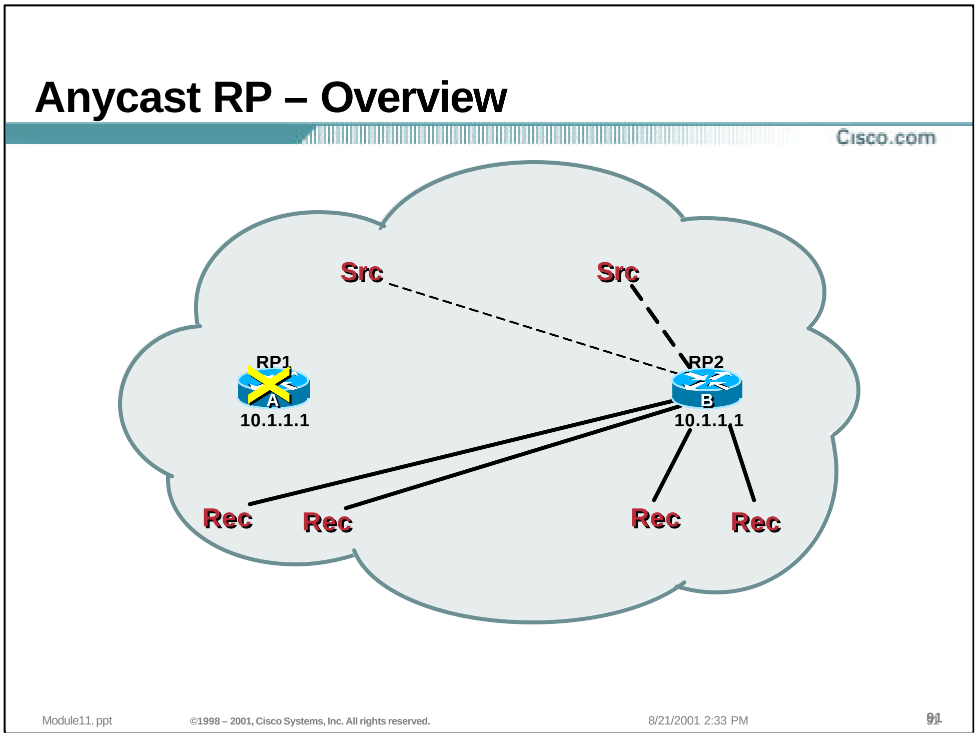
- **Anycast RP Overview**
  - **In the drawing above, two Anycast RP's are configured with the same IP address, RP1 in San Francisco and RP2 in New York. Each are connected via MSDP.**
    - (Yes, you must use some other address in the 'ip msdp peer' commands than 10.0.0.1.)

Anycast RP – Overview

Cisco.com



- **Anycast RP Overview**
  - **Notice that initially, the DR’s for the sources and receivers register to the closest RP based on their unicast routing table entry for IP address 10.0.0.1. This causes in DR’s in the eastern half of the U.S. to register/join to the RP in New York while the DR’s in the western half register/join to the RP in San Francisco.**
  - **When a new source registers with the nearest RP, that RP will send an MSDP SA message to its peer. This will cause the peer RP to join the SPT to the new source so it can pull the sources traffic to itself and then send it down the shared tree to its receivers.**

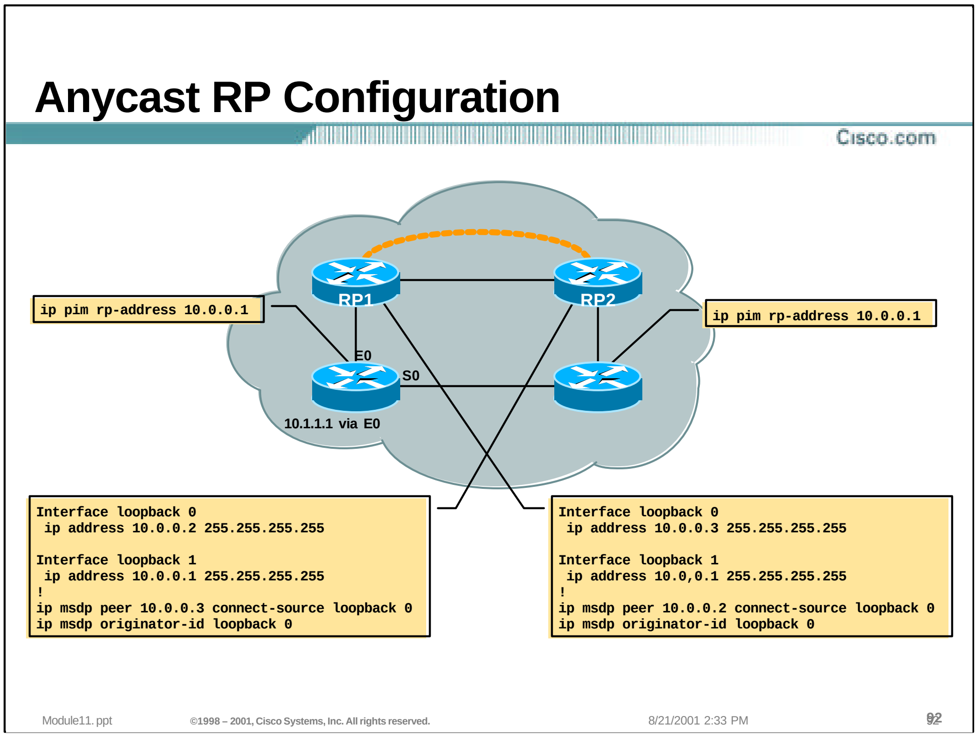


Anycast RP – Overview

Cisco.com

Module11. ppt 8/21/2001 2:33 PM 91

- **Anycast RP Overview**
  - **Continuing with our example, let’s assume that the RP in San Francisco goes down. When the unicast routing protocol reconverges, all of the DR’s in the western half of the U.S. will now see the route to IP address 10.0.0.1 points toward the the New York RP. This results in new registers/joins being sent by the DR’s in the western half of the U.S. to the RP in New York and the flow of traffic is reestablished.**



Anycast RP Configuration

```
ip pim rp-address 10.0.0.1
```

```
ip pim rp-address 10.0.0.1
```

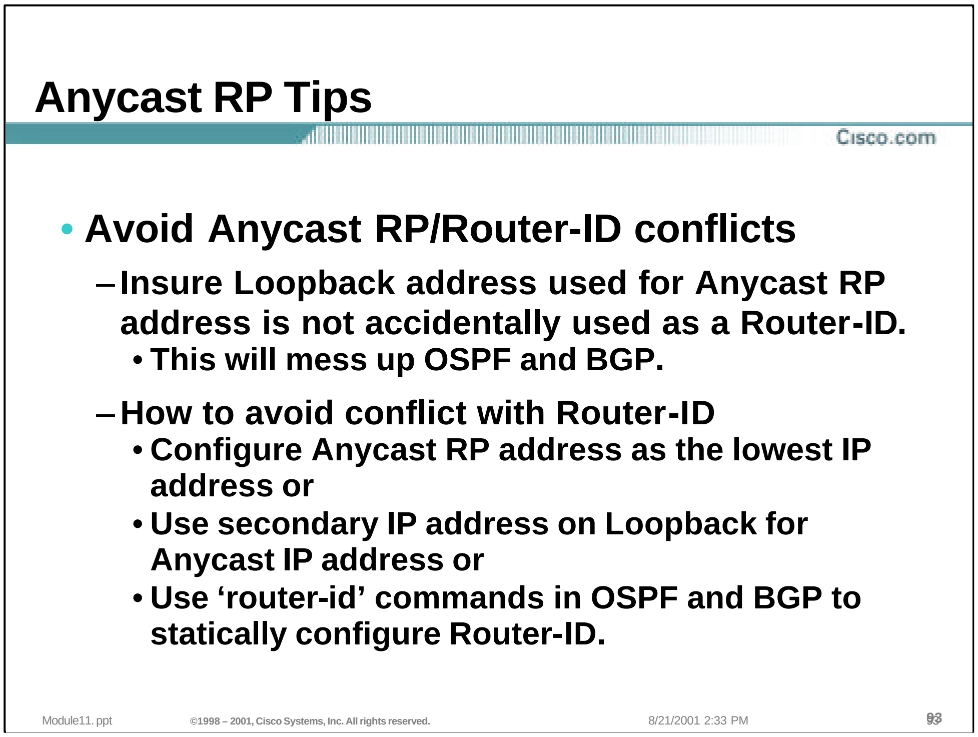
```
Interface loopback 0
 ip address 10.0.0.2 255.255.255.255

Interface loopback 1
 ip address 10.0.0.1 255.255.255.255
!
ip msdp peer 10.0.0.3 connect-source loopback 0
ip msdp originator-id loopback 0
```

```
Interface loopback 0
 ip address 10.0.0.3 255.255.255.255

Interface loopback 1
 ip address 10.0,0.1 255.255.255.255
!
ip msdp peer 10.0.0.2 connect-source loopback 0
ip msdp originator-id loopback 0
```

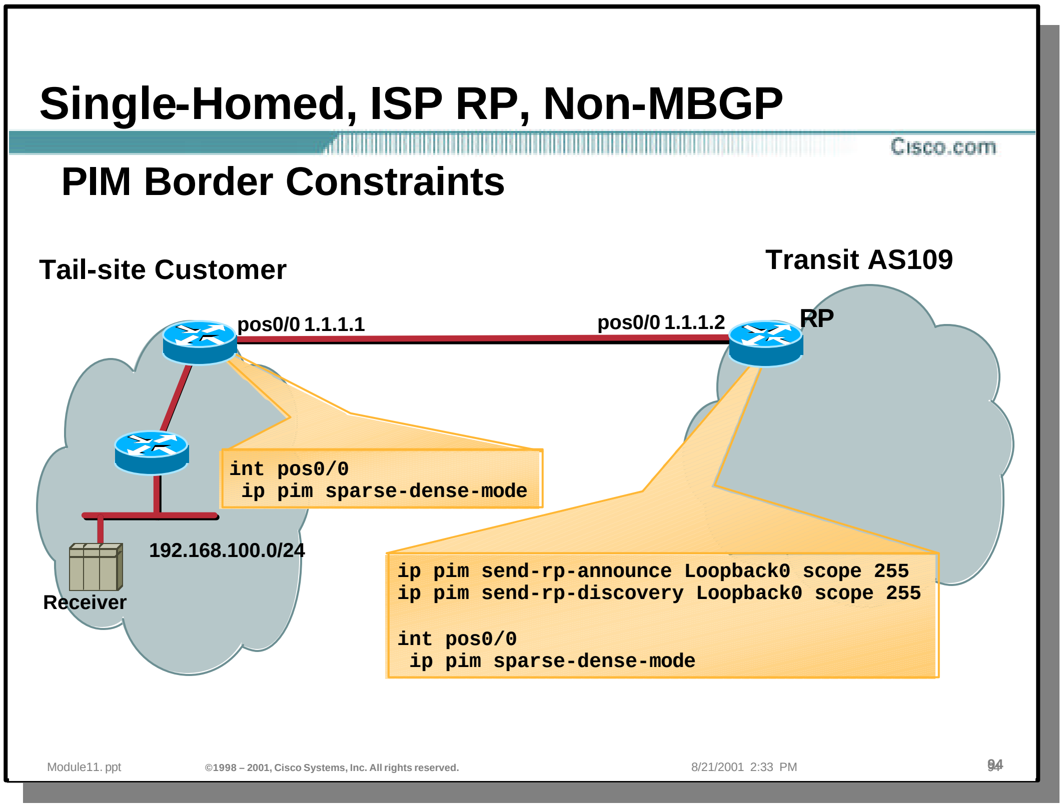
- **Anycast RP Example**
  - **In this example, two Anycast RP's are configured with the same IP address, 10.0.0.1, using Loopback 0.**
  - **Each are connected via MSDP using their Loopback 1 addresses, 10.0.0.2 and 10.0.0.3.**
    - (Yes, you must use some other address in the 'ip msdp peer' commands than 10.0.0.1.)



Anycast RP Tips

- **Avoid Anycast RP/Router-ID conflicts**
  - **Insure Loopback address used for Anycast RP address is not accidentally used as a Router-ID.**
    - **This will mess up OSPF and BGP.**
  - **How to avoid conflict with Router-ID**
    - **Configure Anycast RP address as the lowest IP address or**
    - **Use secondary IP address on Loopback for Anycast IP address or**
    - **Use ‘router-id’ commands in OSPF and BGP to statically configure Router-ID.**

- **Anycast RP Tips**
  - **Care must be taken to prevent the Loopback addresses being used for the Anycast RP’s from being accidentally used as the Router-ID for OSPF and BGP.**
    - If this occurs, there will be multiple OSPF/BGP routers in the network with the same Router-ID. (Can you say, “My network is broken? Sure, I knew you could.)
  - **Avoiding the Router-ID conflict:**
    - Configure the Anycast RP Loopback address using the lowest IP address in the box.
    - Configure a secondary address on the Loopback address and use this address for Anycast RP configuration.
    - Use the ‘router-id’ configuration commands to statically configure the OSPF and/or BGP Router-ids.



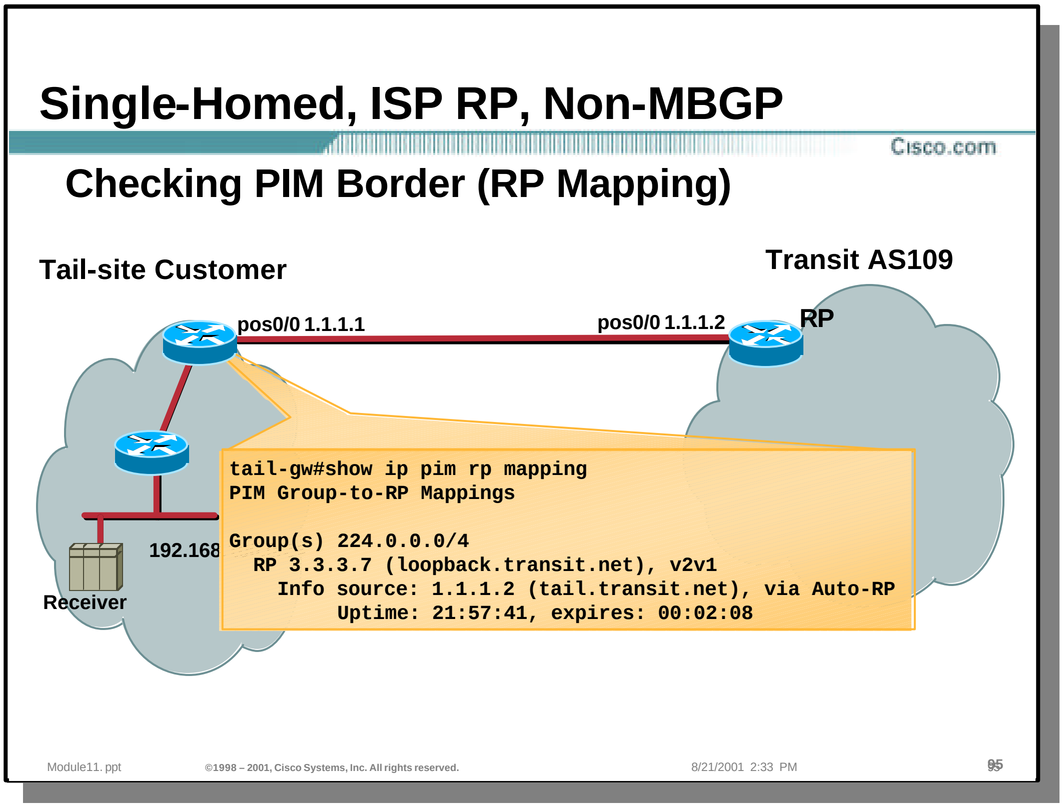
Single-Homed, ISP RP, Non-MBGP

PIM Border Constraints

```
int pos0/0
 ip pim sparse-dense-mode
```

```
ip pim send-rp-announce Loopback0 scope 255
ip pim send-rp-discovery Loopback0 scope 255

int pos0/0
 ip pim sparse-dense-mode
```

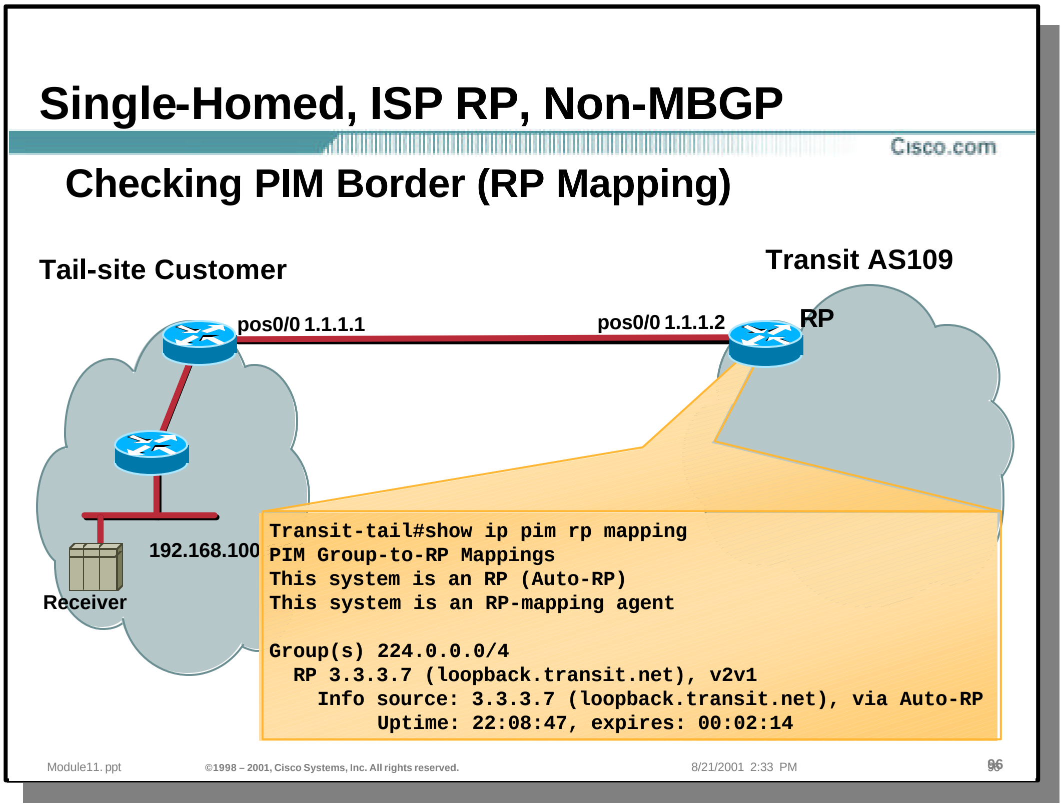


Single-Homed, ISP RP, Non-MBGP

Checking PIM Border (RP Mapping)

```
tail-gw#show ip pim rp mapping
PIM Group-to-RP Mappings

Group(s) 224.0.0.0/4
  RP 3.3.3.7 (loopback.transit.net), v2v1
    Info source: 1.1.1.2 (tail.transit.net), via Auto-RP
         Uptime: 21:57:41, expires: 00:02:08
```

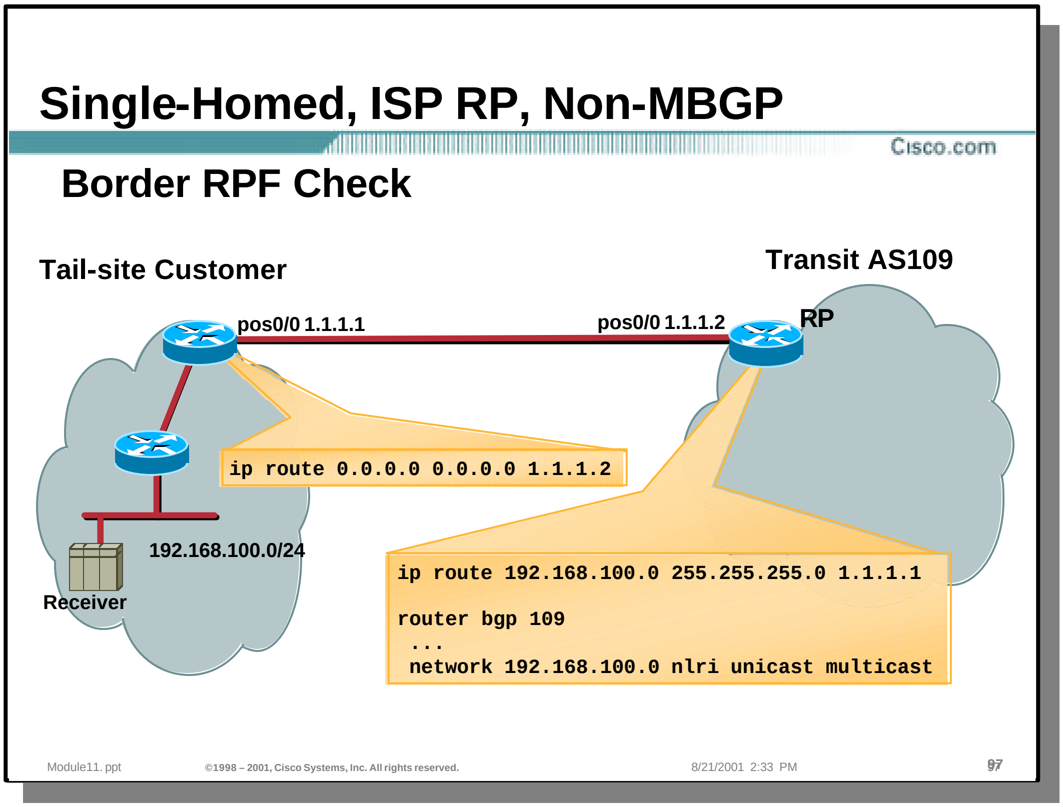


Single-Homed, ISP RP, Non-MBGP

Checking PIM Border (RP Mapping)

```
Transit-tail#show ip pim rp mapping
PIM Group-to-RP Mappings
This system is an RP (Auto-RP)
This system is an RP-mapping agent

Group(s) 224.0.0.0/4
  RP 3.3.3.7 (loopback.transit.net), v2v1
    Info source: 3.3.3.7 (loopback.transit.net), via Auto-RP
         Uptime: 22:08:47, expires: 00:02:14
```



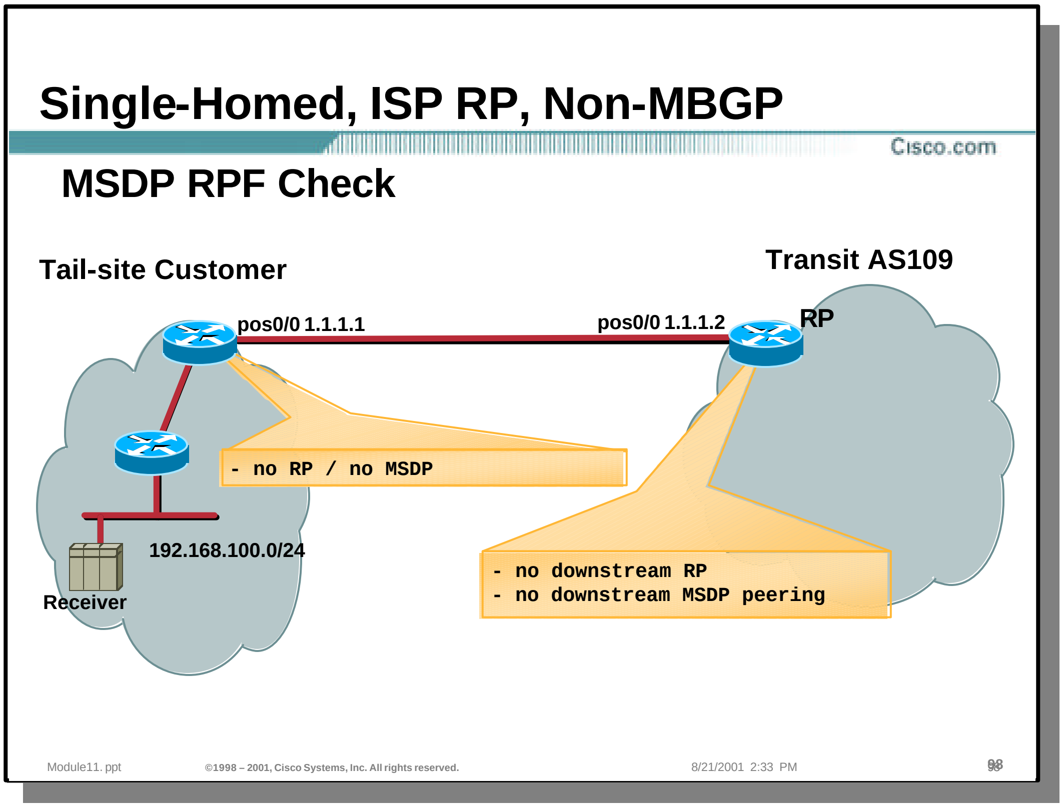
Single-Homed, ISP RP, Non-MBGP

Border RPF Check

```
ip route 0.0.0.0 0.0.0.0 1.1.1.2
```

```
ip route 192.168.100.0 255.255.255.0 1.1.1.1

router bgp 109
 ...
 network 192.168.100.0 nlri unicast multicast
```



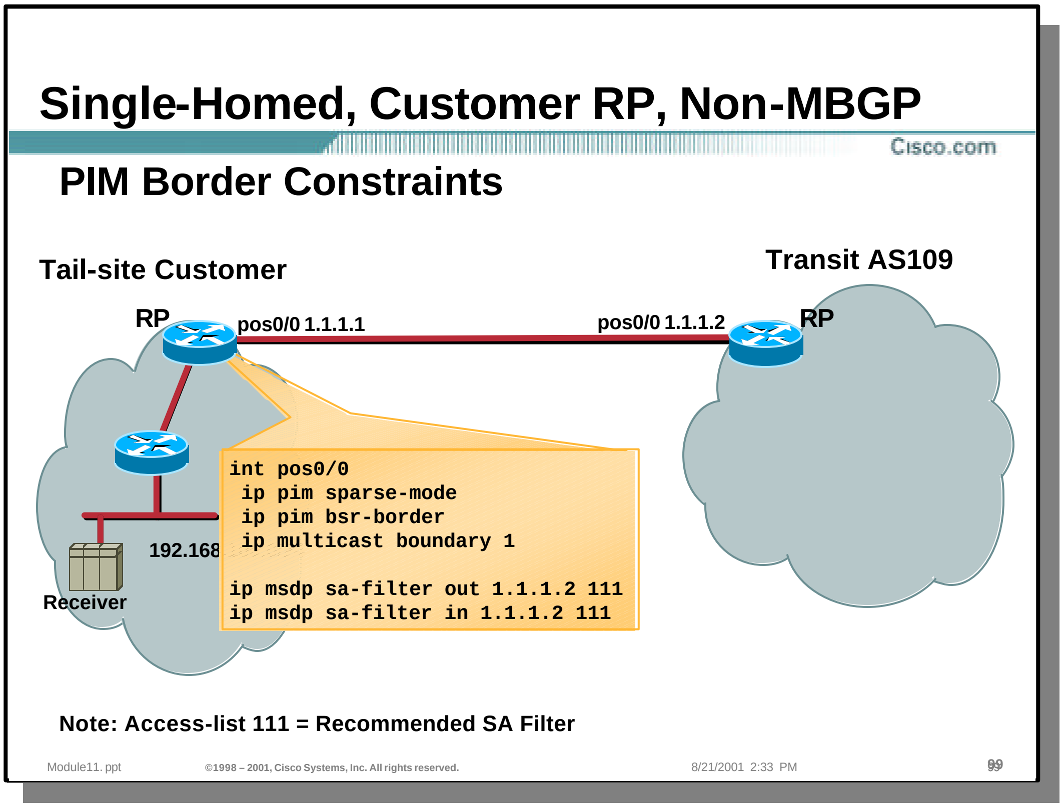
Single-Homed, ISP RP, Non-MBGP

Cisco.com

MSDP RPF Check

```
- no RP / no MSDP
```

```
- no downstream RP
- no downstream MSDP peering
```



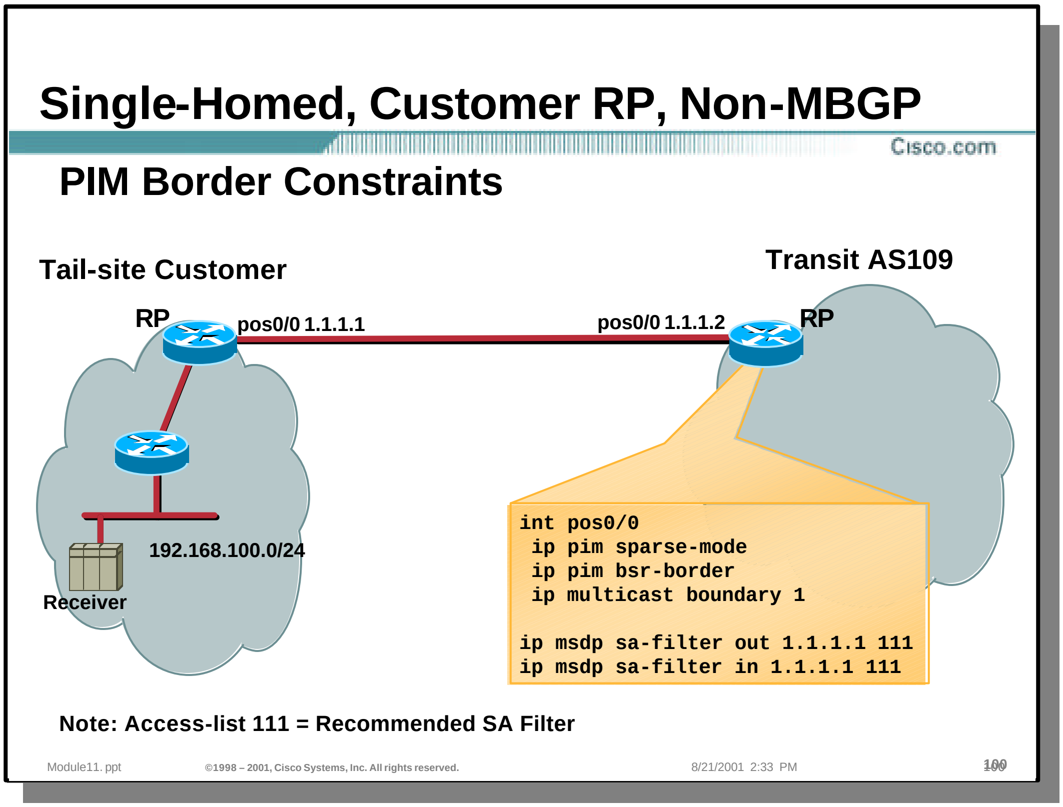
Single-Homed, Customer RP, Non-MBGP

PIM Border Constraints

```
int pos0/0
 ip pim sparse-mode
 ip pim bsr-border
 ip multicast boundary 1

ip msdp sa-filter out 1.1.1.2 111
ip msdp sa-filter in 1.1.1.2 111
```

Note: Access-list 111 = Recommended SA Filter



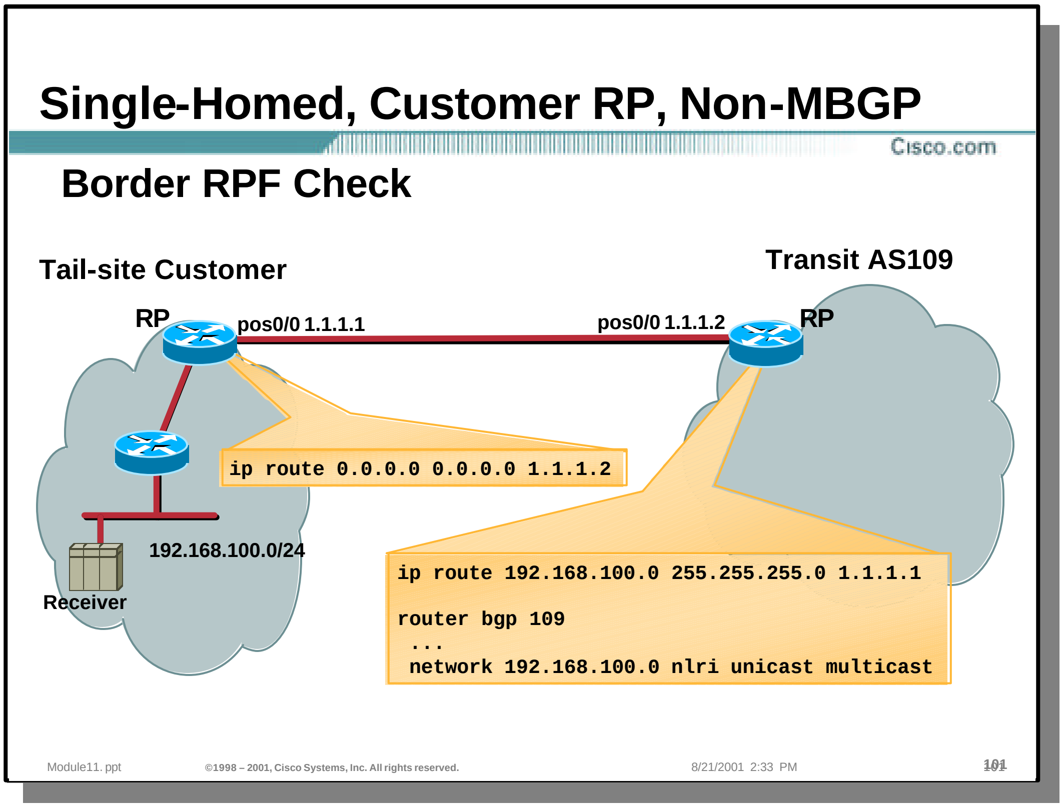
Single-Homed, Customer RP, Non-MBGP

PIM Border Constraints

```
int pos0/0
 ip pim sparse-mode
 ip pim bsr-border
 ip multicast boundary 1

ip msdp sa-filter out 1.1.1.1 111
ip msdp sa-filter in 1.1.1.1 111
```

Note: Access-list 111 = Recommended SA Filter



Single-Homed, Customer RP, Non-MBGP

Border RPF Check

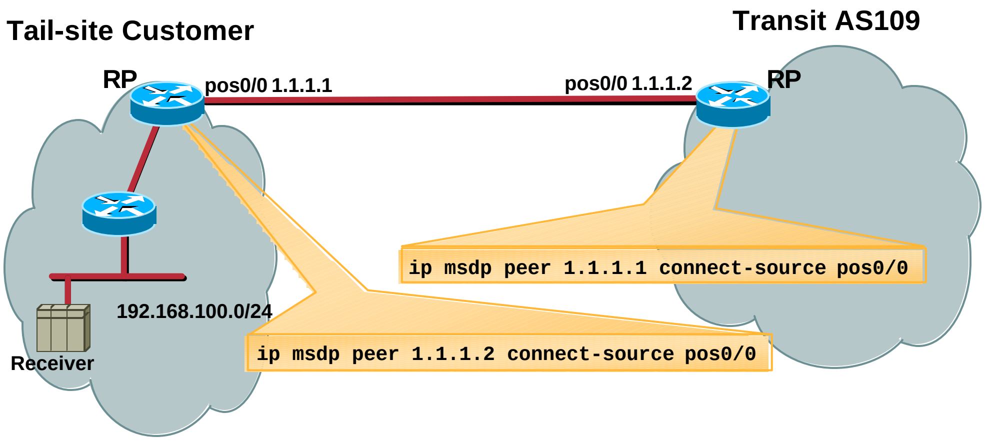
```
ip route 0.0.0.0 0.0.0.0 1.1.1.2
```

```
ip route 192.168.100.0 255.255.255.0 1.1.1.1

router bgp 109
 ...
 network 192.168.100.0 nlri unicast multicast
```

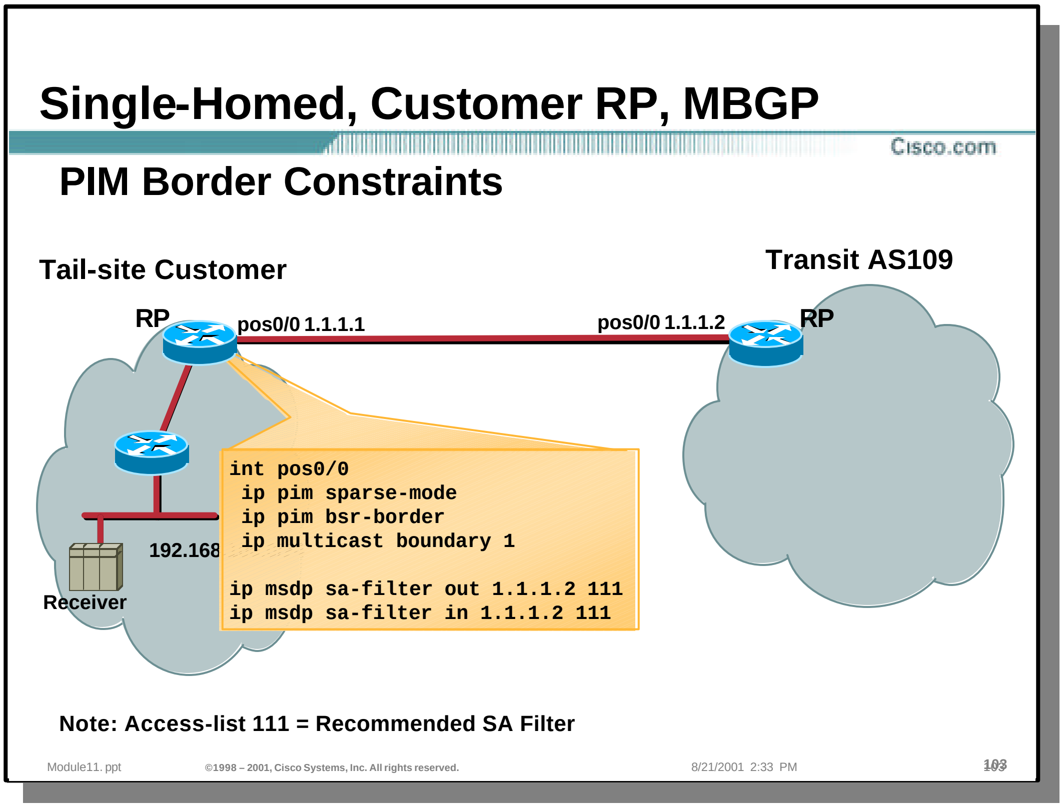
Single-Homed, Customer RP, Non-MBGP

MSDP RPF Check



```
ip msdp peer 1.1.1.1 connect-source pos0/0
```

```
ip msdp peer 1.1.1.2 connect-source pos0/0
```



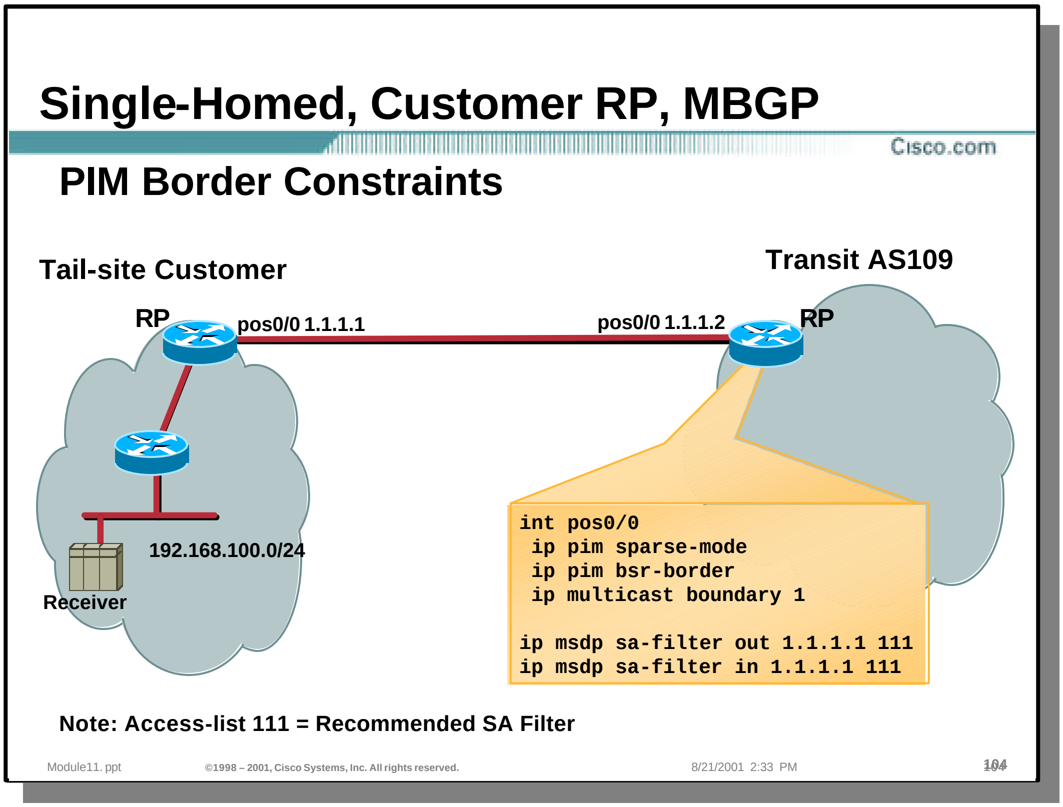
Single-Homed, Customer RP, MBGP

PIM Border Constraints

```
int pos0/0
 ip pim sparse-mode
 ip pim bsr-border
 ip multicast boundary 1

ip msdp sa-filter out 1.1.1.2 111
ip msdp sa-filter in 1.1.1.2 111
```

Note: Access-list 111 = Recommended SA Filter



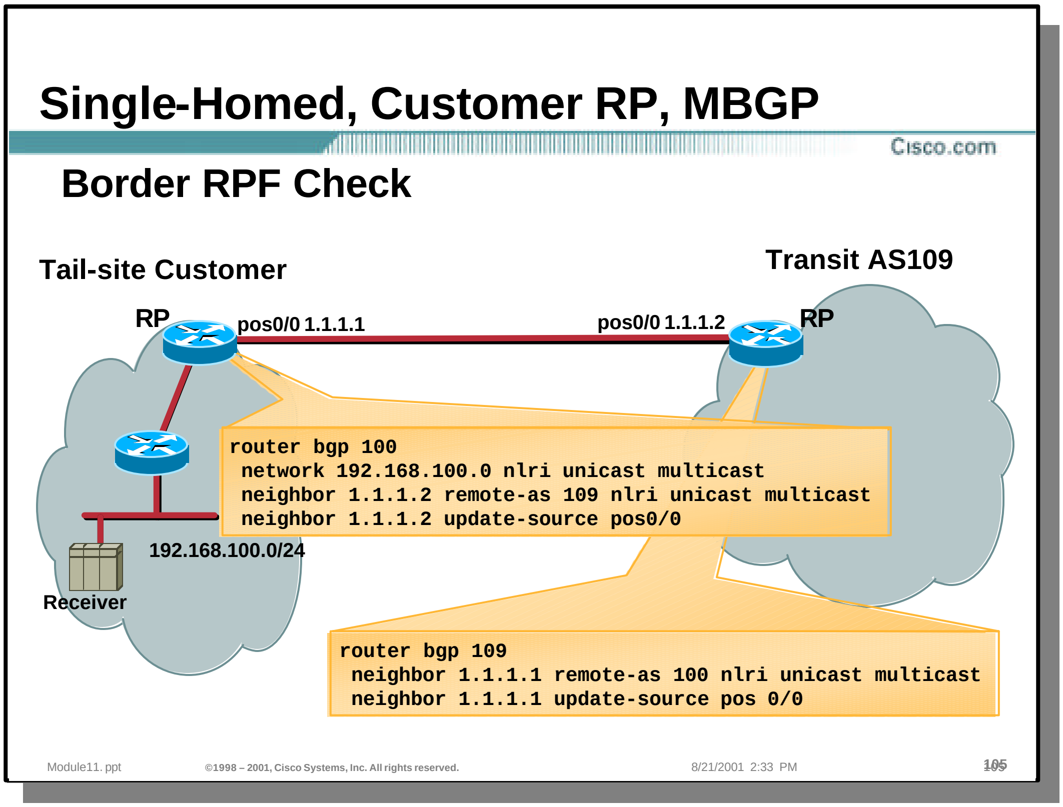
Single-Homed, Customer RP, MBGP

PIM Border Constraints

```
int pos0/0
 ip pim sparse-mode
 ip pim bsr-border
 ip multicast boundary 1

ip msdp sa-filter out 1.1.1.1 111
ip msdp sa-filter in 1.1.1.1 111
```

Note: Access-list 111 = Recommended SA Filter

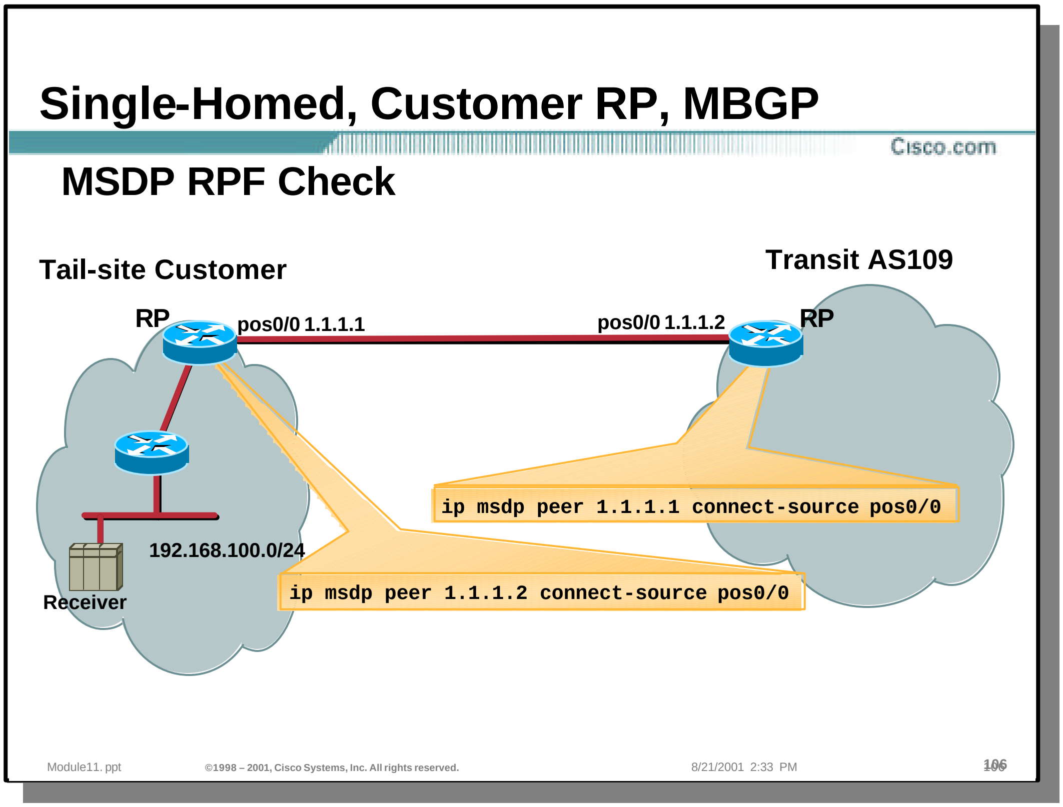


Single-Homed, Customer RP, MBGP

Border RPF Check

```
router bgp 100
 network 192.168.100.0 nlri unicast multicast
 neighbor 1.1.1.2 remote-as 109 nlri unicast multicast
 neighbor 1.1.1.2 update-source pos0/0
```

```
router bgp 109
 neighbor 1.1.1.1 remote-as 100 nlri unicast multicast
 neighbor 1.1.1.1 update-source pos 0/0
```



Single-Homed, Customer RP, MBGP

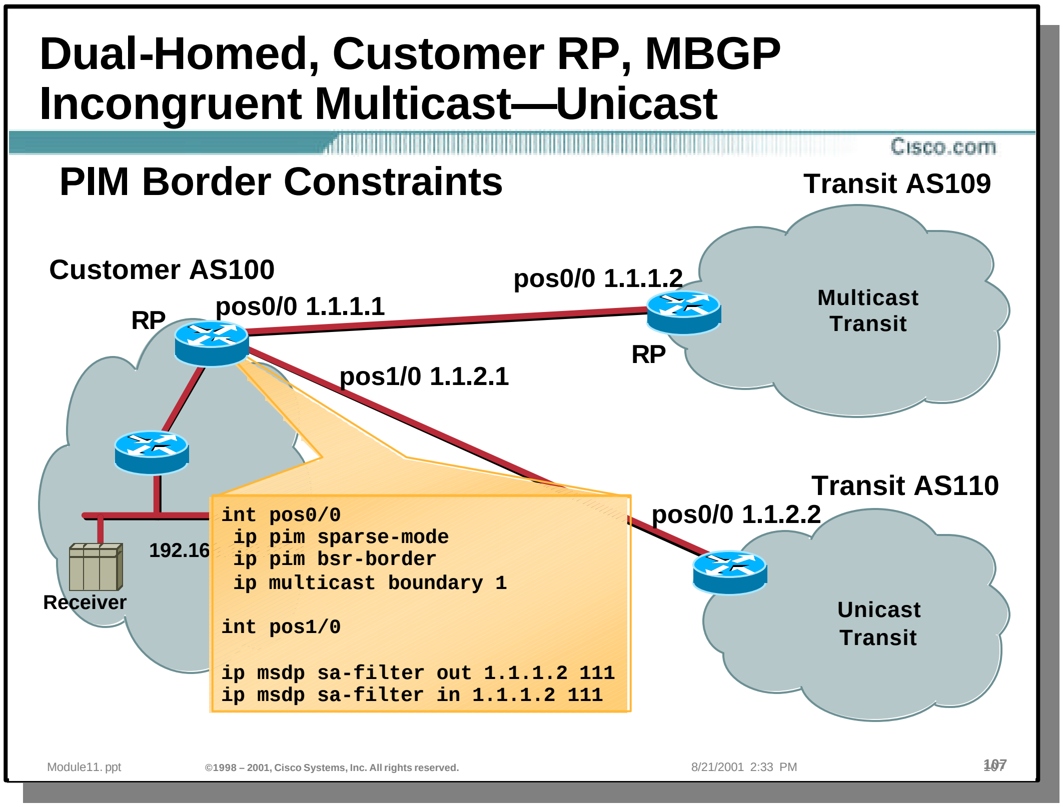
Cisco.com

MSDP RPF Check

```
ip msdp peer 1.1.1.1 connect-source pos0/0
```

```
ip msdp peer 1.1.1.2 connect-source pos0/0
```

Dual-Homed, Customer RP, MBGP Incongruent Multicast—Unicast

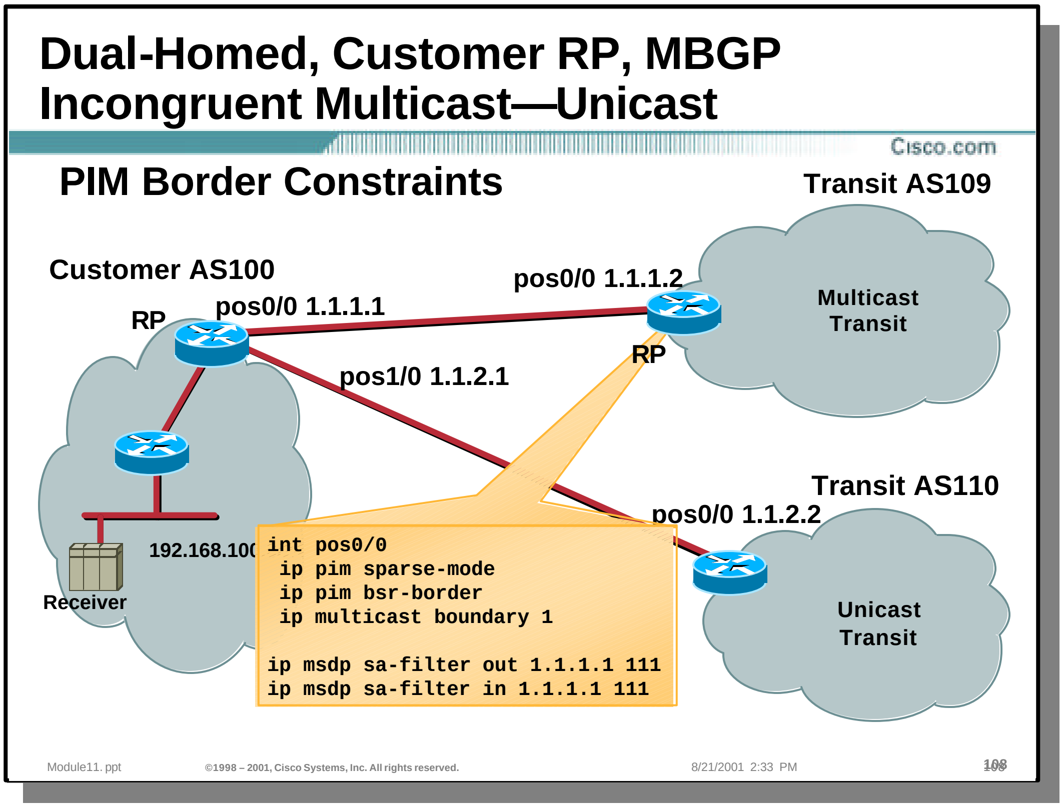


PIM Border Constraints

```
int pos0/0
 ip pim sparse-mode
 ip pim bsr-border
 ip multicast boundary 1

int pos1/0

ip msdp sa-filter out 1.1.1.2 111
ip msdp sa-filter in 1.1.1.2 111
```

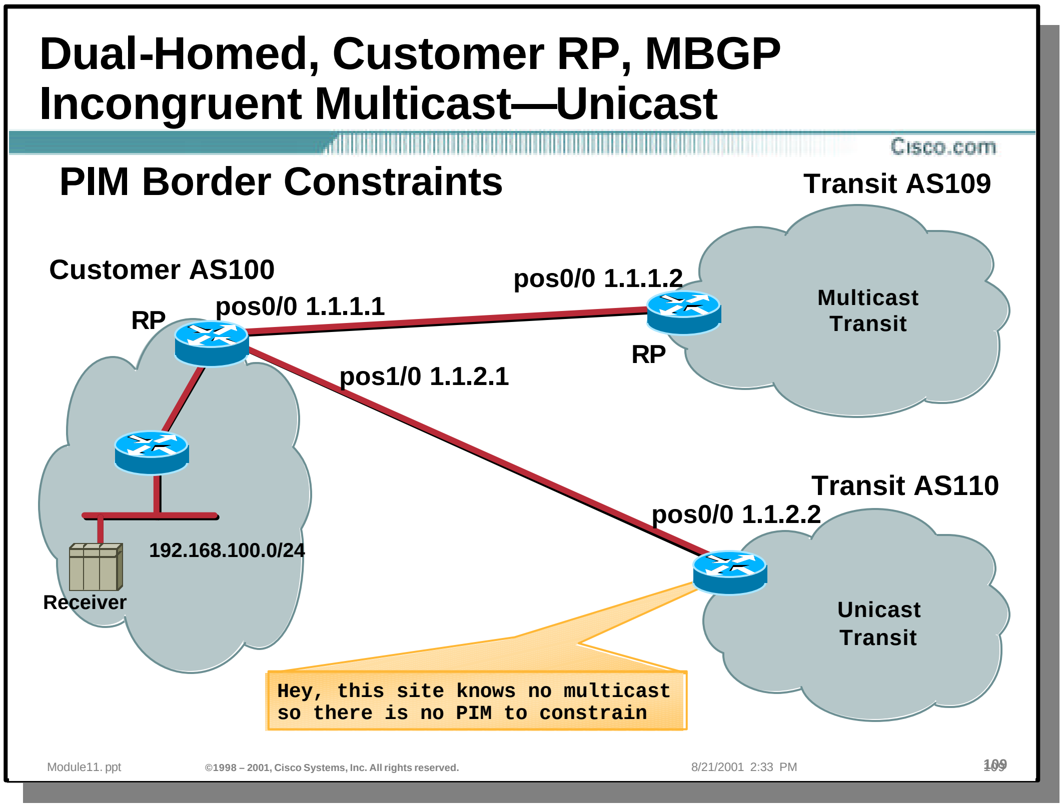


Dual-Homed, Customer RP, MBGP Incongruent Multicast—Unicast

PIM Border Constraints

```
int pos0/0
 ip pim sparse-mode
 ip pim bsr-border
 ip multicast boundary 1

ip msdp sa-filter out 1.1.1.1 111
ip msdp sa-filter in 1.1.1.1 111
```

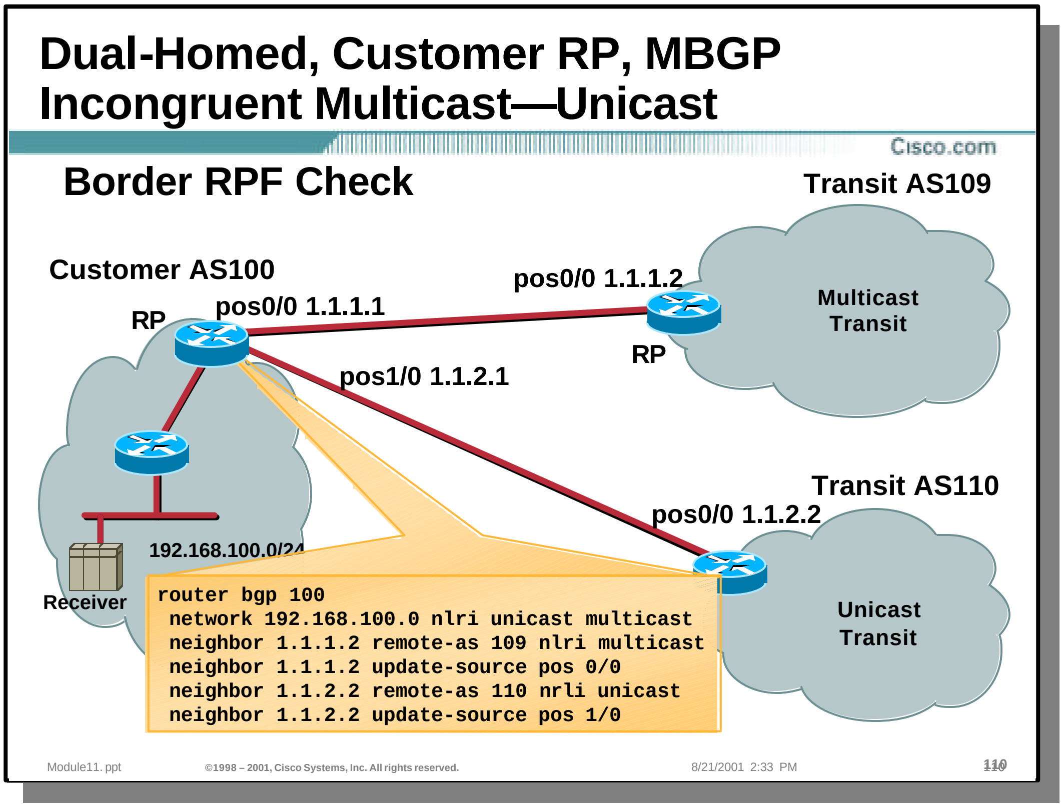


Dual-Homed, Customer RP, MBGP Incongruent Multicast—Unicast

Cisco.com

PIM Border Constraints

Hey, this site knows no multicast so there is no PIM to constrain

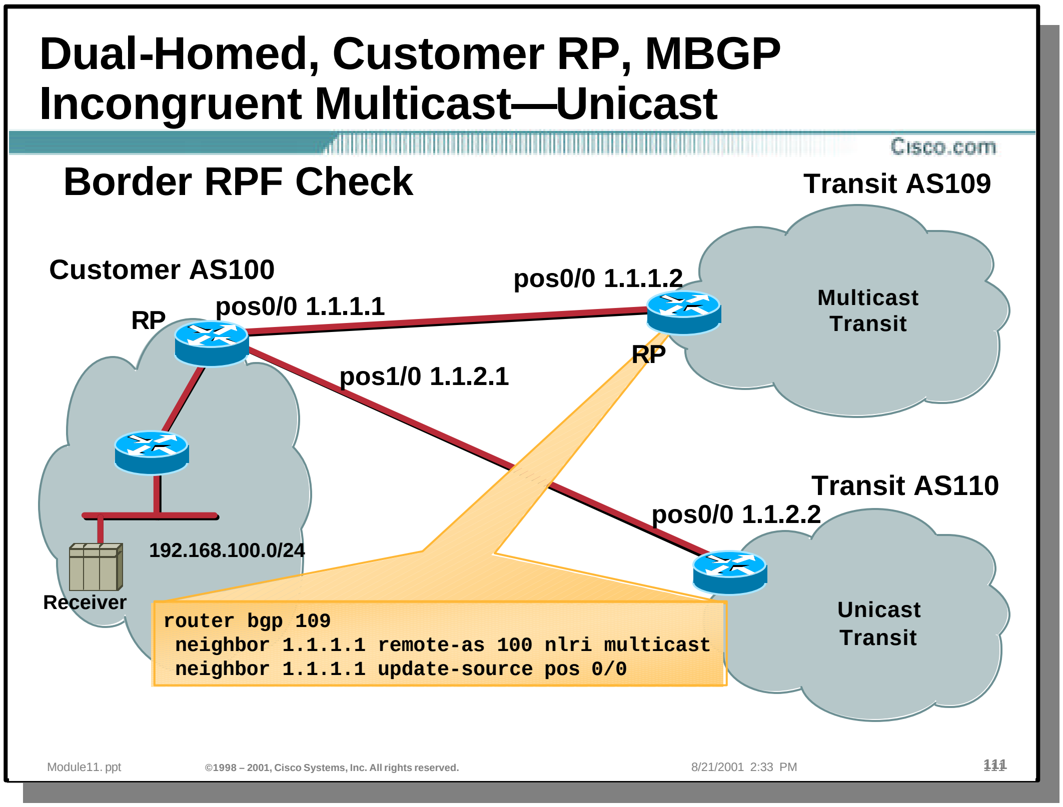


Dual-Homed, Customer RP, MBGP Incongruent Multicast—Unicast

Border RPF Check

```
router bgp 100
 network 192.168.100.0 nlri unicast multicast
 neighbor 1.1.1.2 remote-as 109 nlri multicast
 neighbor 1.1.1.2 update-source pos 0/0
 neighbor 1.1.2.2 remote-as 110 nlri unicast
 neighbor 1.1.2.2 update-source pos 1/0
```

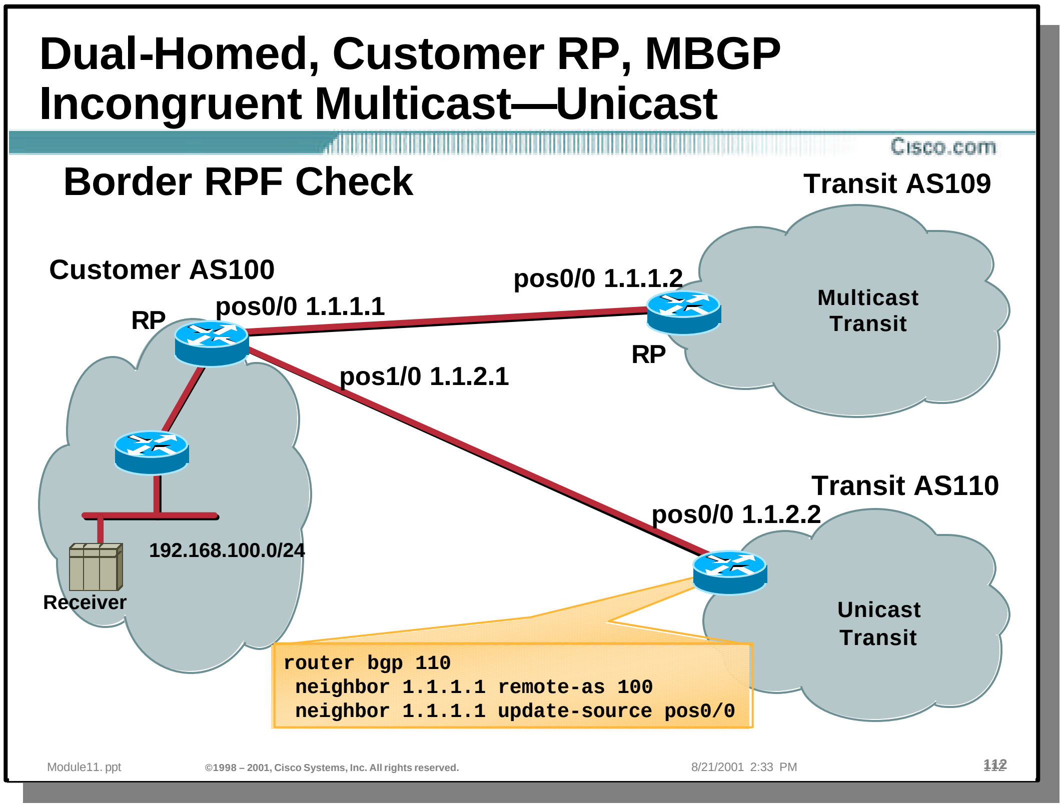
Module11. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 110



Dual-Homed, Customer RP, MBGP Incongruent Multicast—Unicast

Border RPF Check

```
router bgp 109
 neighbor 1.1.1.1 remote-as 100 nlri multicast
 neighbor 1.1.1.1 update-source pos 0/0
```



Dual-Homed, Customer RP, MBGP Incongruent Multicast—Unicast

Cisco.com

Border RPF Check

Transit AS109

Customer AS100

RP

pos0/0 1.1.1.1

pos0/0 1.1.1.2

RP

Multicast Transit

pos1/0 1.1.2.1

Transit AS110

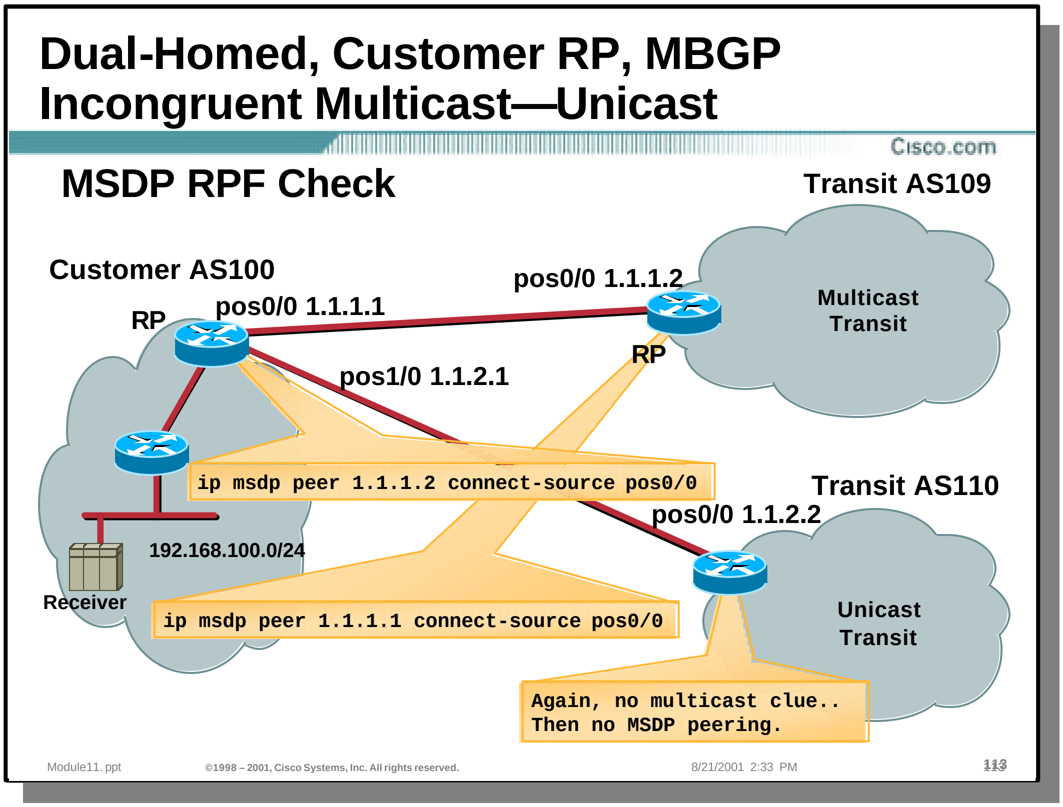
pos0/0 1.1.2.2

192.168.100.0/24

Receiver

Unicast Transit

```
router bgp 110
 neighbor 1.1.1.1 remote-as 100
 neighbor 1.1.1.1 update-source pos0/0
```



Dual-Homed, Customer RP, MBGP Incongruent Multicast—Unicast

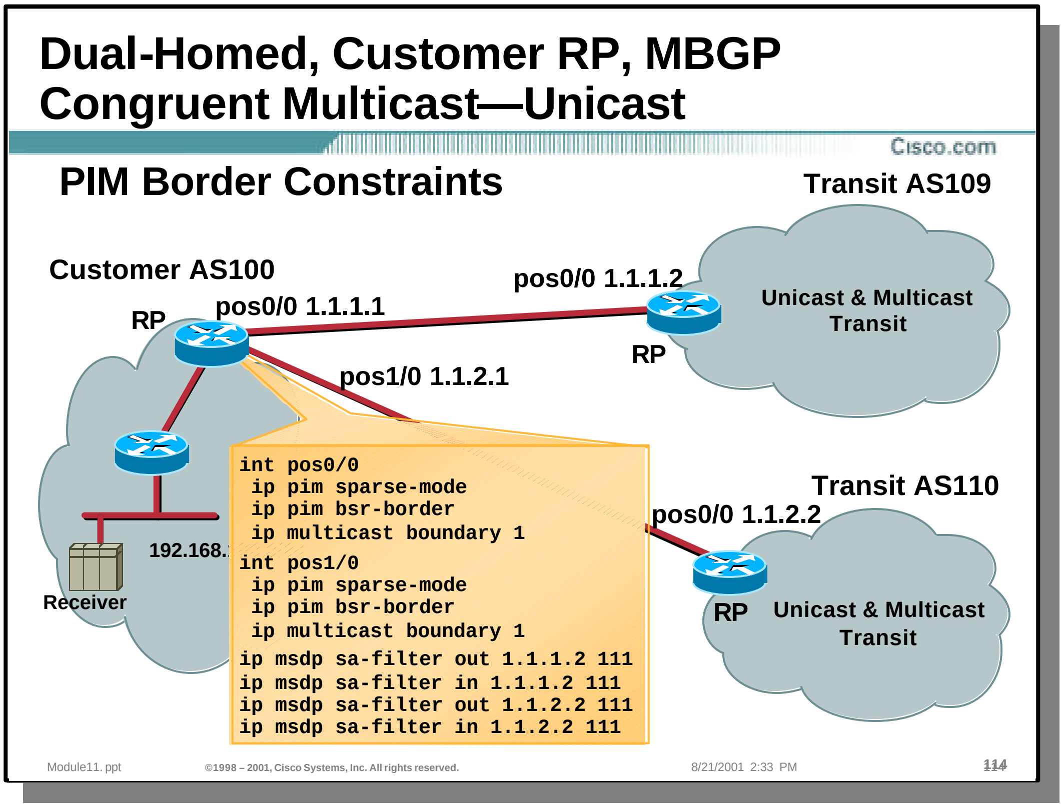
MSDP RPF Check

```
ip msdp peer 1.1.1.2 connect-source pos0/0
```

```
ip msdp peer 1.1.1.1 connect-source pos0/0
```

```
Again, no multicast clue..
Then no MSDP peering.
```

Module11. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 113

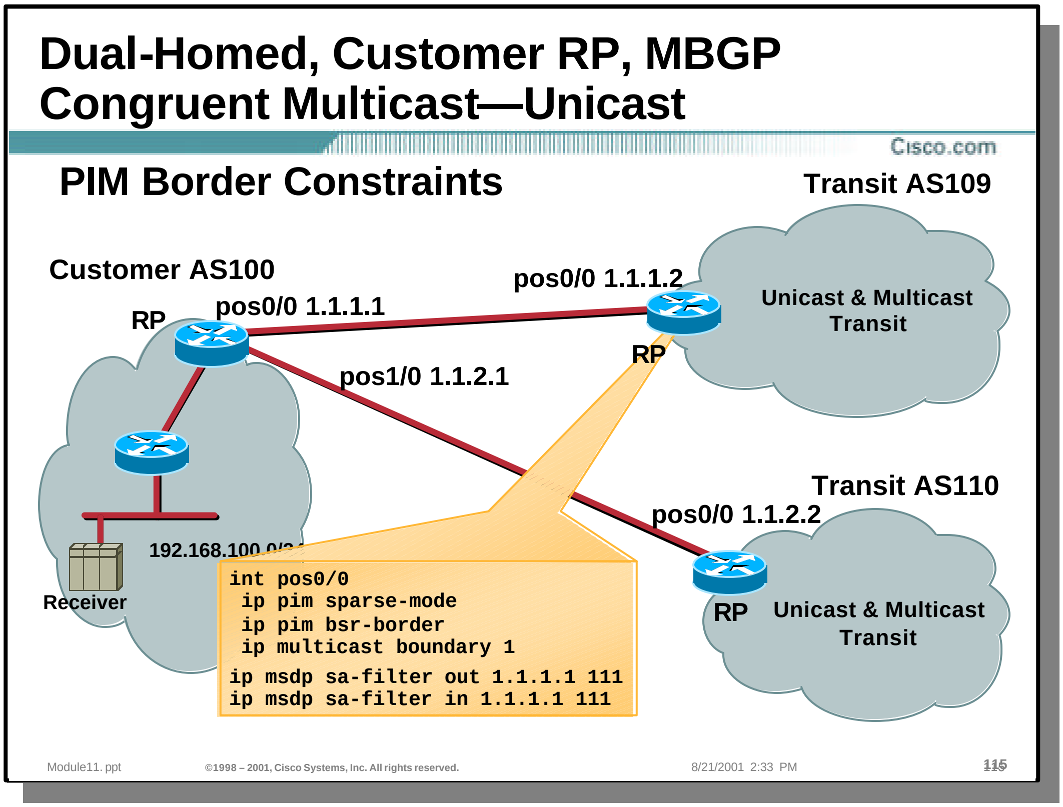


Dual-Homed, Customer RP, MBGP Congruent Multicast—Unicast

PIM Border Constraints

```
int pos0/0
 ip pim sparse-mode
 ip pim bsr-border
 ip multicast boundary 1
int pos1/0
 ip pim sparse-mode
 ip pim bsr-border
 ip multicast boundary 1
ip msdp sa-filter out 1.1.1.2 111
ip msdp sa-filter in 1.1.1.2 111
ip msdp sa-filter out 1.1.2.2 111
ip msdp sa-filter in 1.1.2.2 111
```

Module11. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 2:33 PM 114

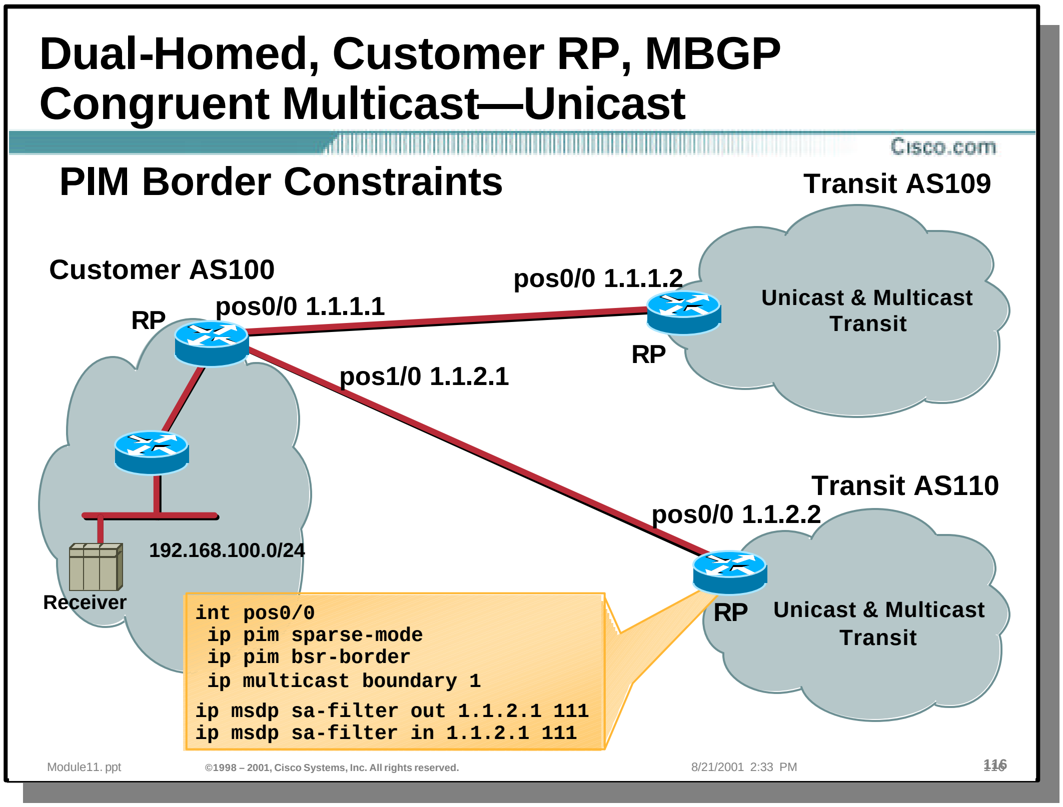


Dual-Homed, Customer RP, MBGP Congruent Multicast—Unicast

Cisco.com

PIM Border Constraints

```
int pos0/0
 ip pim sparse-mode
 ip pim bsr-border
 ip multicast boundary 1
ip msdp sa-filter out 1.1.1.1 111
ip msdp sa-filter in 1.1.1.1 111
```



Dual-Homed, Customer RP, MBGP Congruent Multicast—Unicast

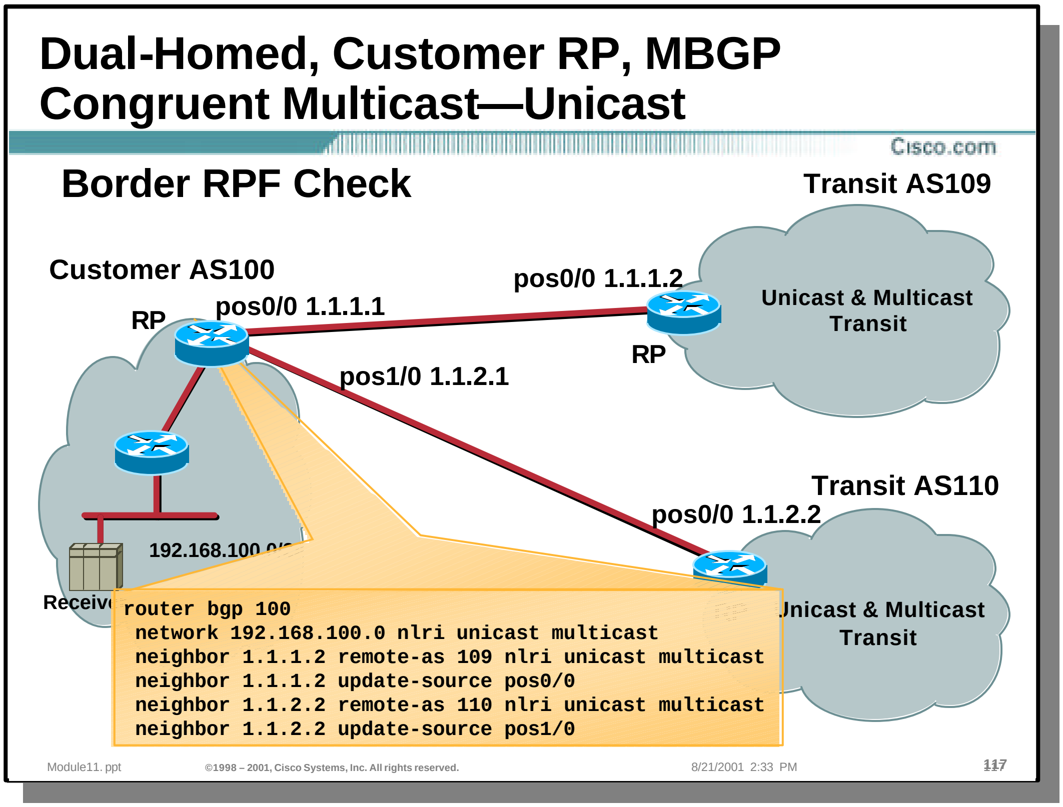
PIM Border Constraints

```
int pos0/0
 ip pim sparse-mode
 ip pim bsr-border
 ip multicast boundary 1
ip msdp sa-filter out 1.1.2.1 111
ip msdp sa-filter in 1.1.2.1 111
```

Module11. ppt

8/21/2001 2:33 PM

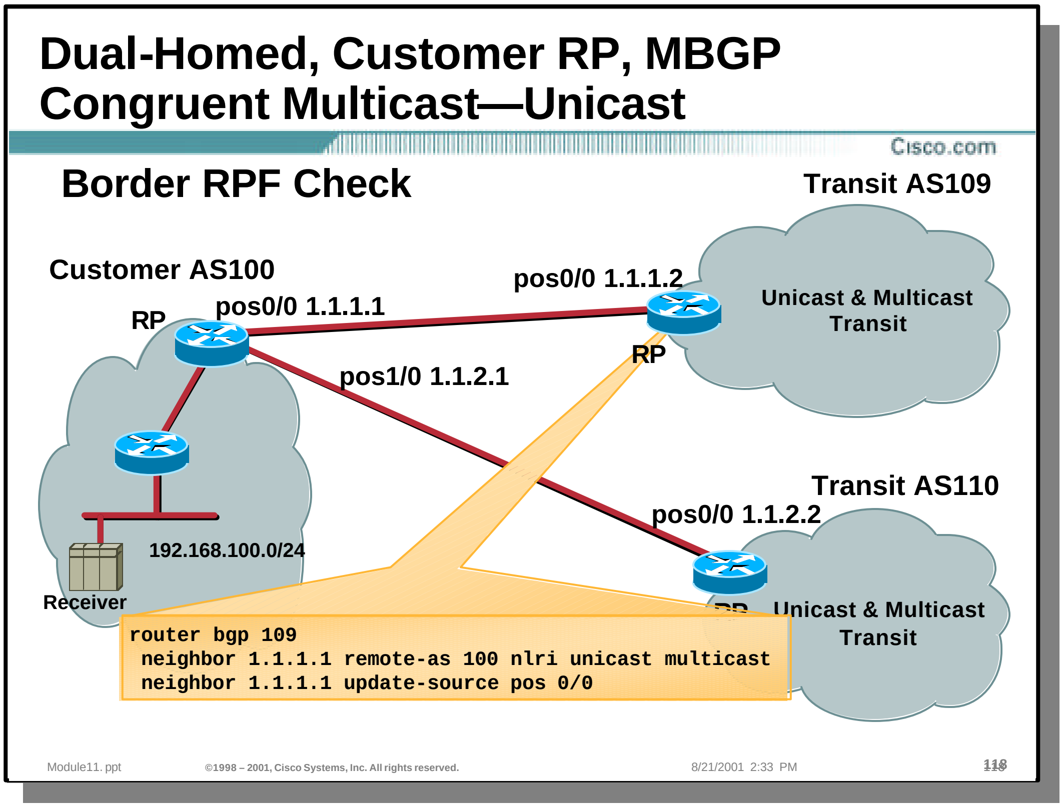
116



Dual-Homed, Customer RP, MBGP Congruent Multicast—Unicast

Border RPF Check

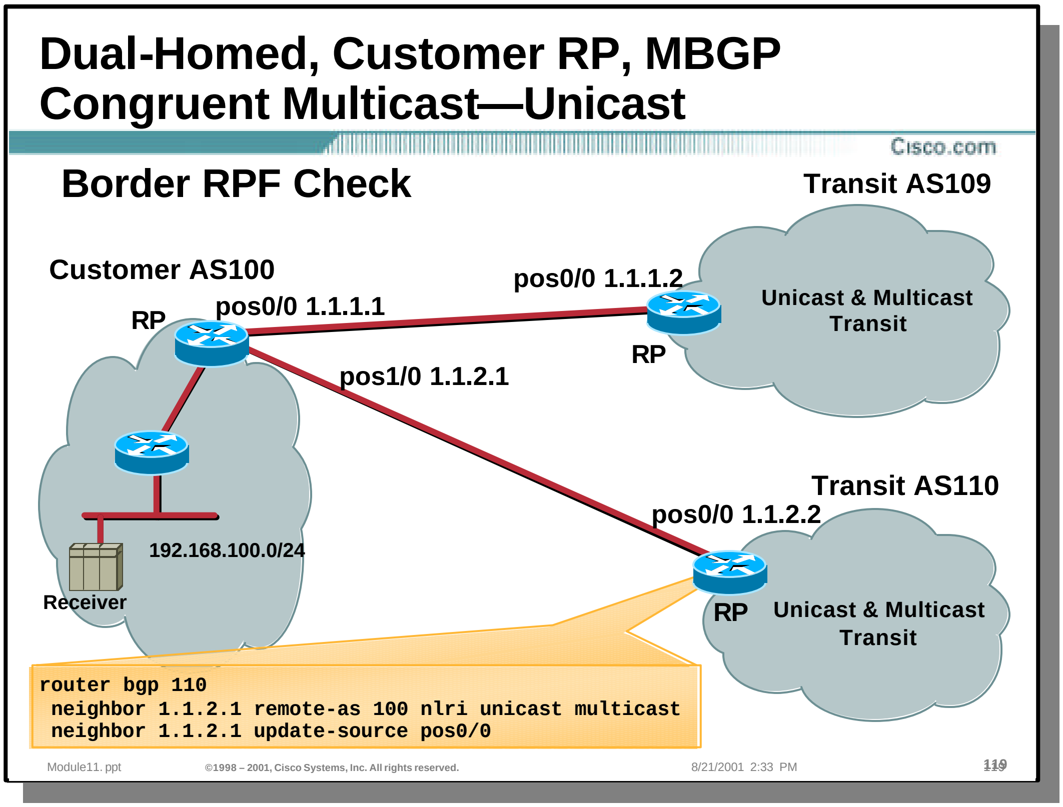
```
router bgp 100
 network 192.168.100.0 nlri unicast multicast
 neighbor 1.1.1.2 remote-as 109 nlri unicast multicast
 neighbor 1.1.1.2 update-source pos0/0
 neighbor 1.1.2.2 remote-as 110 nlri unicast multicast
 neighbor 1.1.2.2 update-source pos1/0
```



Dual-Homed, Customer RP, MBGP Congruent Multicast—Unicast

Border RPF Check

```
router bgp 109
 neighbor 1.1.1.1 remote-as 100 nlri unicast multicast
 neighbor 1.1.1.1 update-source pos 0/0
```

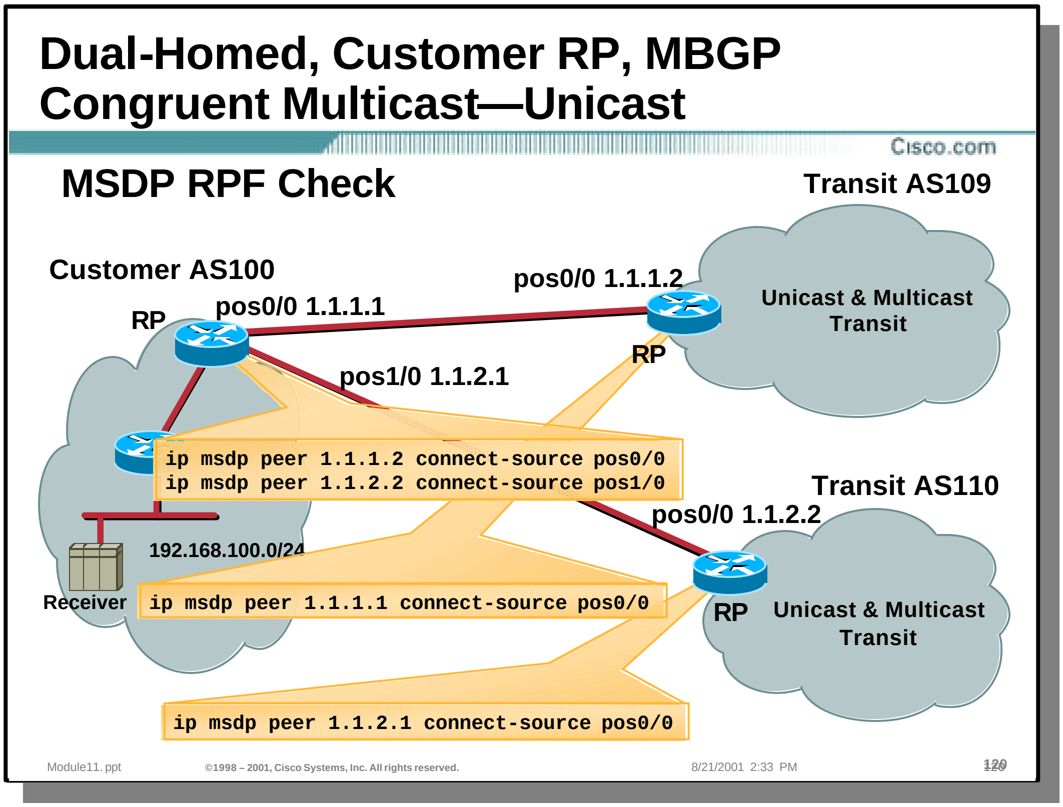


Dual-Homed, Customer RP, MBGP Congruent Multicast—Unicast

Cisco.com

Border RPF Check

```
router bgp 110
 neighbor 1.1.2.1 remote-as 100 nlri unicast multicast
 neighbor 1.1.2.1 update-source pos0/0
```



Dual-Homed, Customer RP, MBGP Congruent Multicast—Unicast

Cisco.com

MSDP RPF Check

Customer AS100

Transit AS109

Transit AS110

RP

pos0/0 1.1.1.1

pos0/0 1.1.1.2

pos1/0 1.1.2.1

RP

Unicast & Multicast Transit

```
ip msdp peer 1.1.1.2 connect-source pos0/0
ip msdp peer 1.1.2.2 connect-source pos1/0
```

pos0/0 1.1.2.2

192.168.100.0/24

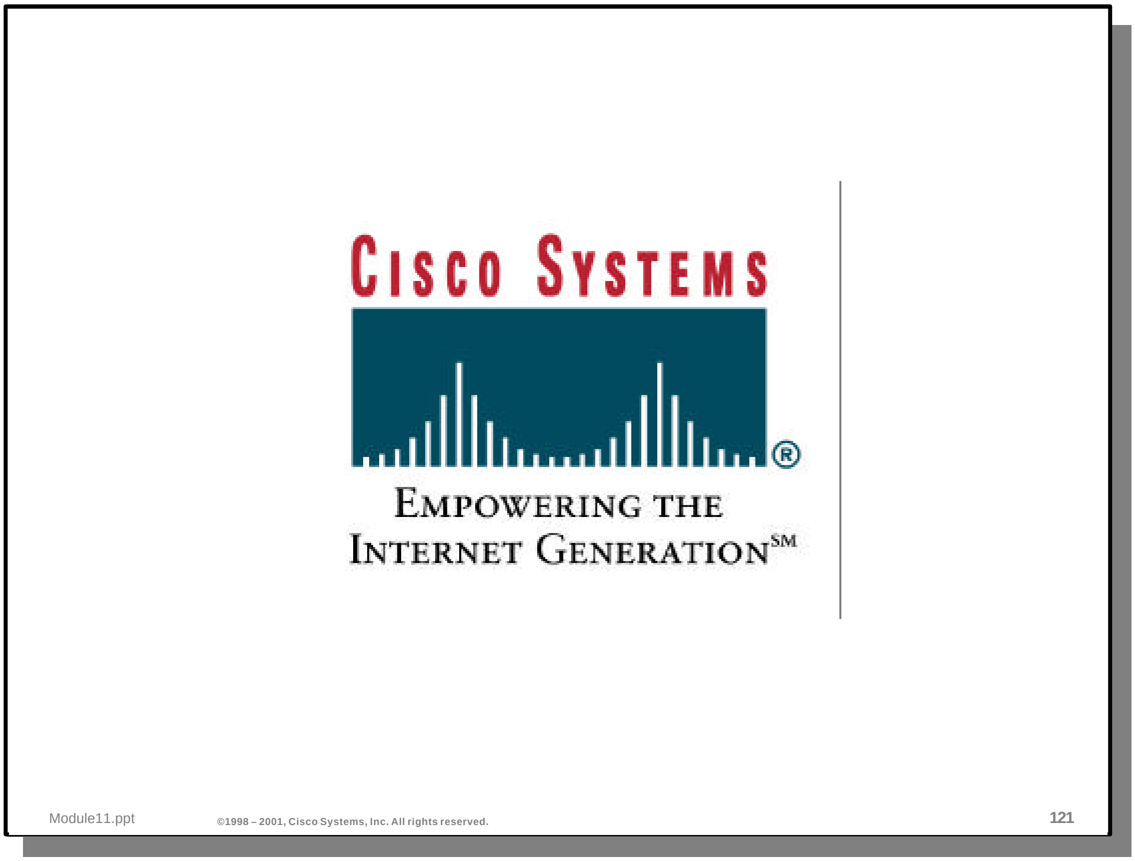
Receiver

```
ip msdp peer 1.1.1.1 connect-source pos0/0
```

RP

Unicast & Multicast Transit

```
ip msdp peer 1.1.2.1 connect-source pos0/0
```



CISCO SYSTEMS

EMPOWERING THE
INTERNET GENERATION℠