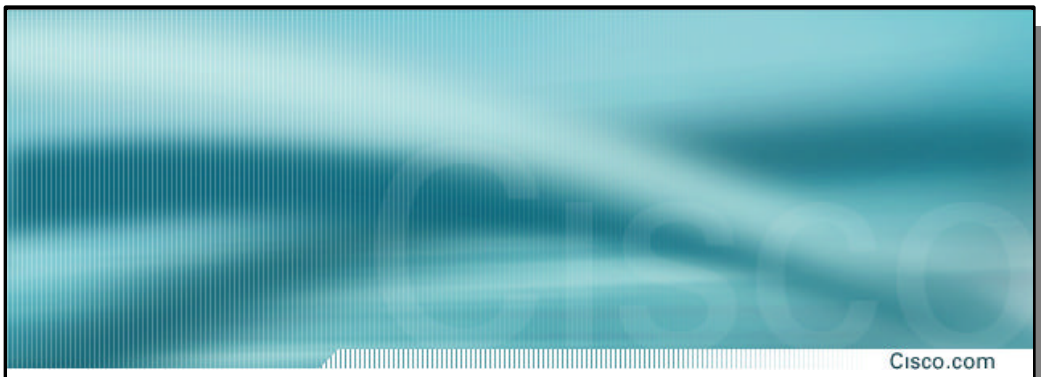


PIM Protocol Extensions

Module 12

Objectives

Cisco.com

- **Upon completion of this module, you will be able to perform the following tasks:**
  - **Describe the principles of Source-Specific Multicast and configure it.**
  - **Describe the principles of Bidir-PIM and configure it.**

Agenda

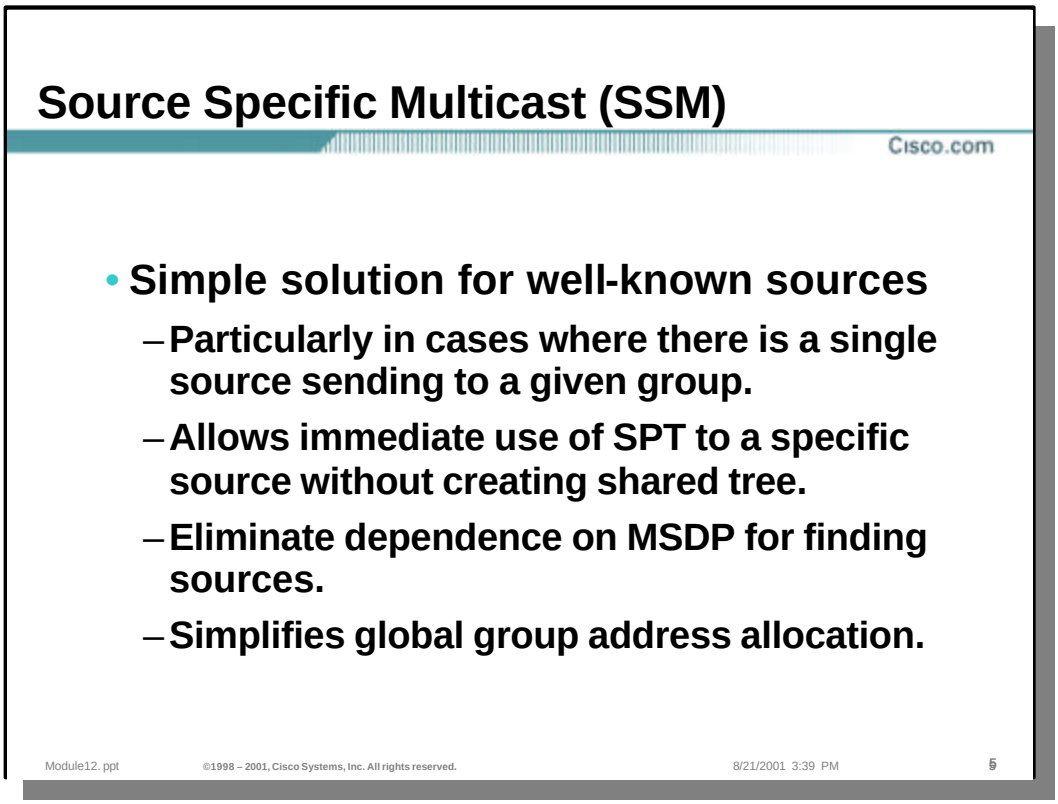
Cisco.com

- **Source Specific Multicast**
- **Bidirectional (Bidir) PIM**

Barriers to Inter-domain Multicast Deployment

Cisco.com

- **Multicast Address Allocation**
  - **Dynamic Address Allocation**
    - **No adequate dynamic address allocation methods exist**
    - **SDR – Doesn't scale**
    - **MASC – Long ways off!**
  - **Static Address Allocation (GLOP)**
    - **Based on AS number.**
    - **Insufficient address space for large Content Providers.**
- **Multicast Content "Jammers"**
  - **Undesirable sources on a multicast group.**
    - **"Capt. Midnight" sources bogus data/noise to group.**
    - **Can cause DoS attack by congesting low speed links.**



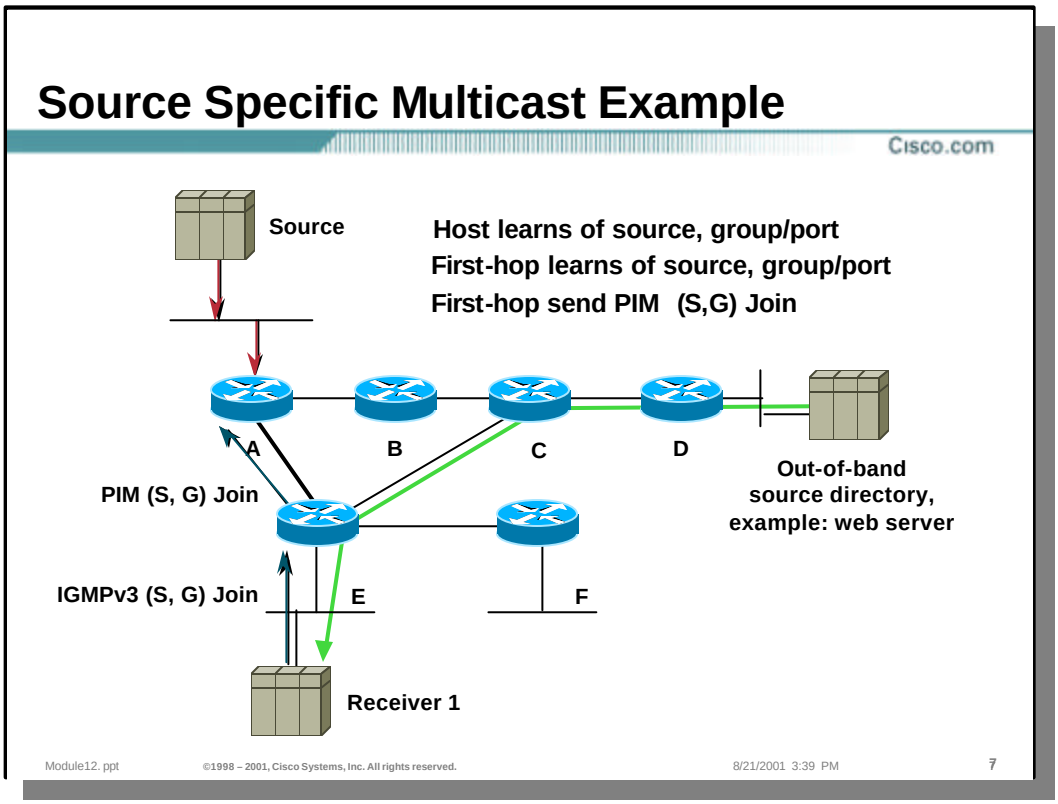
- **Source Specific Multicast**
  - Another variant of a PIM Sparse mode supports Source Specific Multicast (SSM) applications. The PIM SS (Source Specific) utilizes all the benefits of sparse mode protocols but eliminates shared trees at all and only builds source specific shortest path trees. These trees are built directly on receiving group membership reports that request a given source. The PIM SS is a draft proposal (draft-bhaskar-pim-ss-00.txt).
  - The SSM is suitable for well known sources within a domain or in another domain. The Multicast Source Discovery Protocol (MSDP) which is needed for interdomain multicast routing when regular PIM Sparse Mode is used within a domain is no longer needed for SSM.
  - A dedicated multicast group address range 232/8 is used exclusively for shortest-path trees for SSM. Routers are prevented to build a shared tree for any of the groups from this address range. The address range 232/8 is assigned for global well-known sources.
  - Source specific multicast (SSM) is a datagram delivery model that best supports one-to-many applications, also known as broadcast applications.

SSM Overview

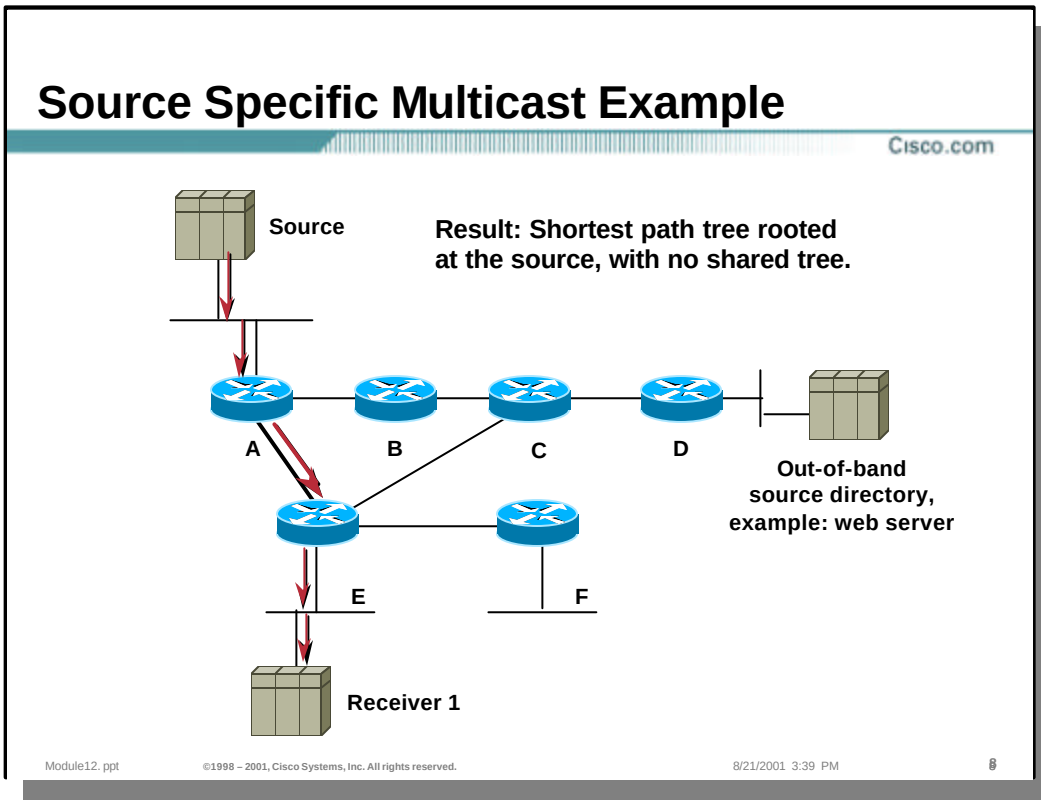
Cisco.com

- **Hosts initiate join requests for a specific source(s) within a group.**
- **Last-hop router sends (S,G) join toward source without joining/creating shared tree.**
- **Content identified by both source and group address instead of group address alone.**
- **Eliminates shared tree, simplifying address allocation.**
  - **Dissimilar content sources can use same group without fear of interfering with each other.**

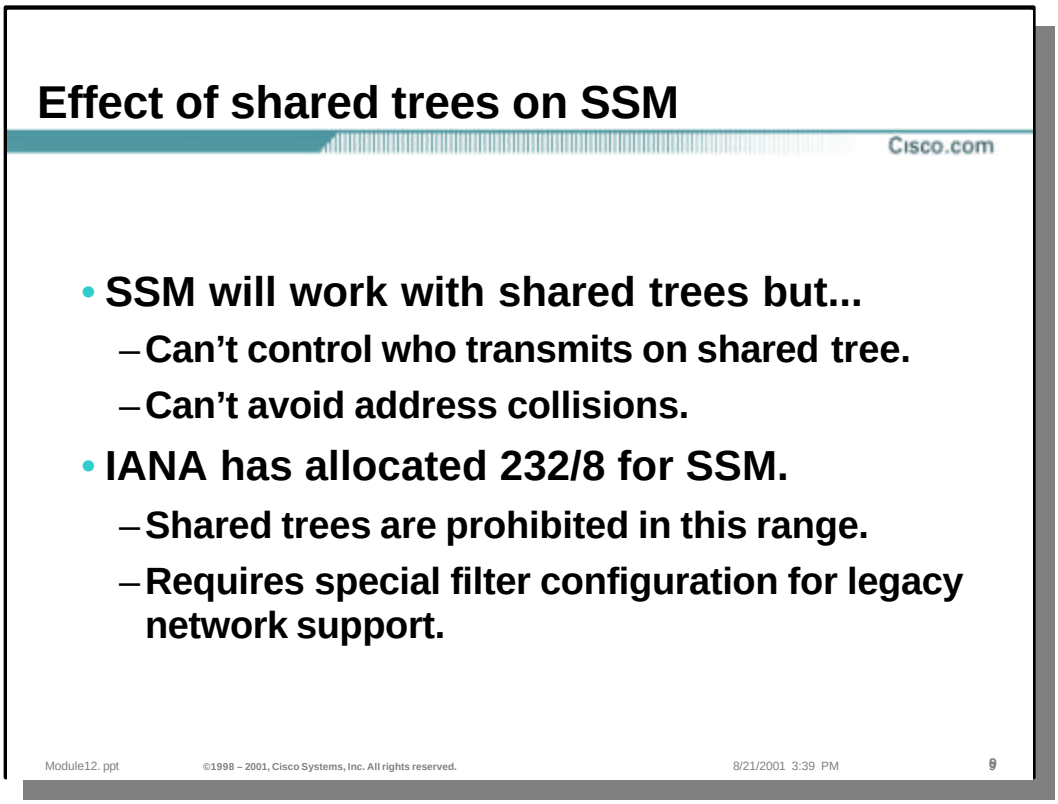
- **SSM: For Well-know Sources**
  - The Source Specific Multicast allows last-hop router to immediately send (S,G) Join towards the source. Thus the PIM Sparse Mode (*,G) Join towards the RP is eliminated at all and first-hop routers start forwarding the multicast traffic donw the shortest-path tree (SPT) from the very beginning - as soon as the SPT is built by receiving first (S,G) Join.
  - The assigned address range 232/8 also simplifies the address allocation problems since the range is a global range for sorces that have to be well-known. Implementations in routers must not build any shared tree for those groups.
  - Source specific groups can coexist with other groups in PIM Sparse mode domains.



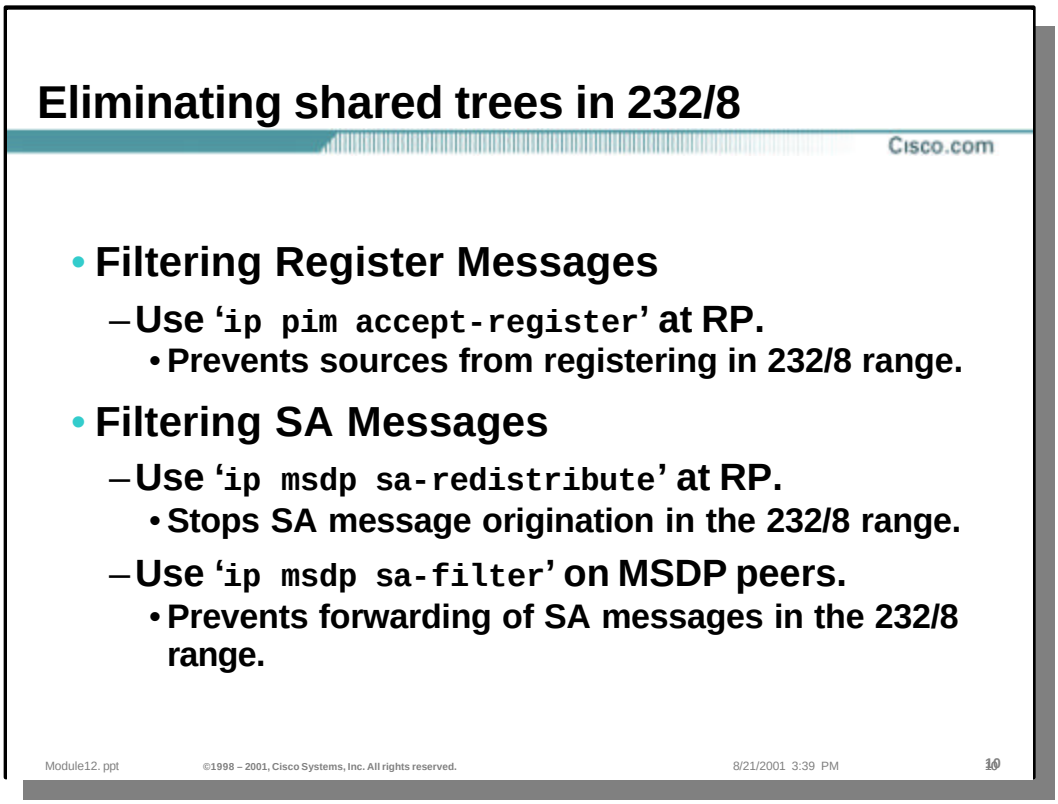
- **SSM – Example**
  - The prerequisite for SSM deployment is a mechanism that allows hosts not only to report the group they want to join but also the source for the group. This mechanism is built into emerging IGMP version3 standard. With IGMP v3 last-hop routers learn from the report for the multicast source and the group. It then simply creates (S,G) Join and forwards it directyl to the source.
  - The ways how hosts learn about existence of sources can be different – normally via some directory services (session announcements directly from sources or some out-of-band mechanisms, e.g. web pages). Most of those mechanisms distribute the information via multicast.



- **SSM – Example**
  - The result of building source-rooted tree (shortest-path tree) right from beginning is that RP mechanisms for source-specific groups are completely eliminated. The RPs for those groups are not needed any more and routers must not build shared trees for groups in the range 232/8.
  - The benefits of building shortest-path trees directly (and not via PIM Sparse mode switchover mechanism) are evident – the latency of multicast traffic is decreased and less multicast state is kept in multicast forwarding tables.
  - Another major benefit of SSM in in address management. Traditionally multicast applications had to acquire a unique IP multicast group address because traffic distribution was based only on the group address used. When two applications with different sources and receivers used the same IP multicast group address, the receivers received the traffic from both sources.
  - In SSM, traffic from each source is forwarded between routers in the network independent of traffic from other sources. Thus different sources can reuse multicast group addresses in the SSM range



- **Effect of Shared Trees on SSM**
  - Source specific multicast can coexist with shared trees but there are some caveats that affect the deployment. The assigned address range 232/2 for SSM is not recognized by older multicast implementations and if a source creates a session to any of the groups in this range legacy multicast routers could build shared trees for those groups.
  - The assignment of 232/8 address range was done under assumption that implementations will prohibit building shared trees for the range at all. With older multicast router implementations this is certainly not true which can result at least in in address collisions if not in multicast forwarding loops.



- **Eliminating Shared Trees in 232/8**
  - If some multicast routers in the network do not support Source Specific Multicast 232/8 address range an additional filtering is needed to prevent building shared trees for the address range assigned for SSM. The following 232/8 filtering mechanisms have to be in place:
    - Prevent first-hop routers to register to the RP – filter on the RP (e.g. send Register-Stop immediately)
    - Prevent last-hop routers to originate (*,G) Joins – filter on last-hop routers
    - Prevent intermediate routers to originate any (S,G) Prune with RP-bit set – filter on intermediate routers
  - In inter-domain multicast routing using MSDP prevent origination and/or forwarding information on sources that are active for groups in SSM range –filter on the RP or on a border router

SSM — Host Signalling Overview

Cisco.com

- **IETF proposed signalling: IGMPv3**
  - **draft-ietf-idmr-igmp-v3-04.txt**
  - **Proposed for IP SSM**
    - **Also for filtering in RFC1112 style IP Multicast service.**
  - **IGMPv3 will only be active ...**
    IF supported in last-hop routers
    AND IF supported in host operating systems
    AND IF supported in receiver applications

IGMPv3 — Host Signalling

Cisco.com

- (Too?) Complex protocol !
  - No report suppression by hosts.
  - Complex Set-theory Rules for Include/Exclude processing.
- Host membership reports sent to 224.0.0.22.
  - Allows for implementation of IGMPv3 snooping in less expensive switches.
- Possible memberships for hosts:
  - INCLUDE({S1,…,Sn}, G) - traffic from {S1,…,Sn} to G
  - EXCLUDE({S1,…,Sn}, G) - traffic from all sources to G, except for traffic from {S1,…,Sn}

IGMPv3 — Host Signalling

Cisco.com

- **IGMPv3 interaction inside of SSM-Range:**
  - **Routers ignore IGMPv1/v2 membership reports.**
  - **Routers ignore IGMPv3 EXCLUDE type membership reports.**
  - **Uses only IGMPv3 INCLUDE({S1,…,Sn},G) memberships!**
  - **No way to request “all sources traffic”:**
- **IGMPv3 interaction outside of SSM-Range:**
  - **Filters traffic on local wire according to Include/Exclude membership reports received from hosts.**
  - **Permits PIM to filter traffic back to source.**
    - **PIM-DM: Simple, only (S,G) state anyhow.**
    - **PIM-SM: More complex but doable. (SPT vs. RPT filtering)**
    - **Bidir-PIM: No way, there is only (*,G) state !**

SSM — Host Signalling
How to bootstrap deployment ?

Cisco.com

- **IGMPv3:**
  - **Should eventually become industry standard.**
  - **Cisco IGMPv3 implementation in IOS 12.1(3)T and 12.0(15)S.**
- **Questions:**
  - **When will host Operating Systems get IGMPv3 support?**
  - **When will applications be written to use IGMPv3 support?**
  - **Do we want to wait for all this to happen?**
- **Answer: *No!***
  - **We need the benefits of IP SSM today to:**
    - **Resolve certain multicast Security issues**
    - **Avoid address collisions**

SSM — Host Signalling
Bootstrap Solutions

Cisco.com

- Cisco IOS value added IP SSM bootstrap solutions
  - URD: (URL Rendezvous Directory)
    - Enable existing receiver applications for IP SSM via the web.
  - IGMP v3lite:
    - Provide for a partial IGMPv3 API on IGMPv1/v2 hosts.
    - Enables us to write and run IP SSM applications NOW
- Common idea of URD and IGMP v3lite:
  - Generate a (S,G) channel subscription (somehow) in addition to the IGMPv1/v2 membership that MUST already come from the kernel of the applications host.
    … and let the router figure out the right thing to do ...

SSM — Host Signalling
Bootstrap Solutions

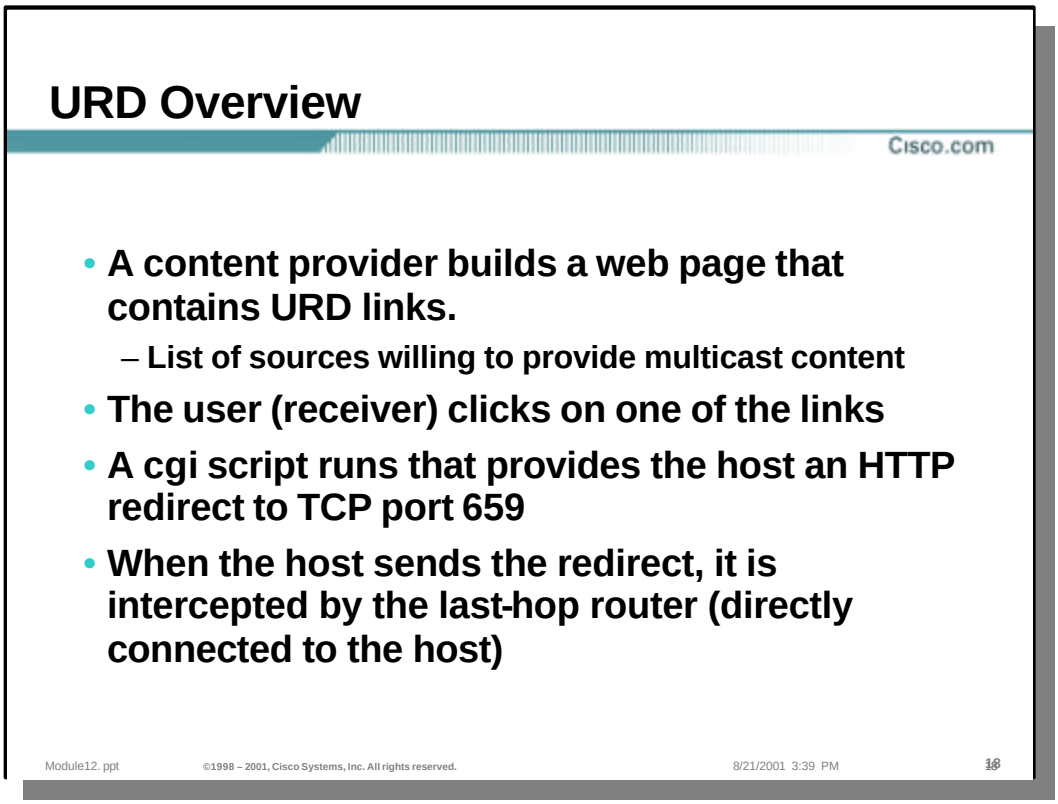
Cisco.com

- **URD and IGMP v3lite Concepts :**
  - **IGMPv1/v2 report alone has no meaning in SSM!**
    - **IGMPv1/v2 report indicates host wants to receive traffic.**
    - **(S,G) subscription tells routers from which sources!**
  - **Router begins forwarding (S,G) traffic IFF:**
    - **IGMPv1/v2 Report received AND**
    - **URD or v3lite (S,G) channel subscription received.**
  - **Router continues forwarding (S,G) traffic:**
    - **Based on IGMPv1/v2 Group membership.**
    - **No need for refreshing of (S,G) subscriptions.**
    - **Implication: *(S,G) traffic flows until last host leaves group!***

SSM — Host Signalling Bootstrap Solutions

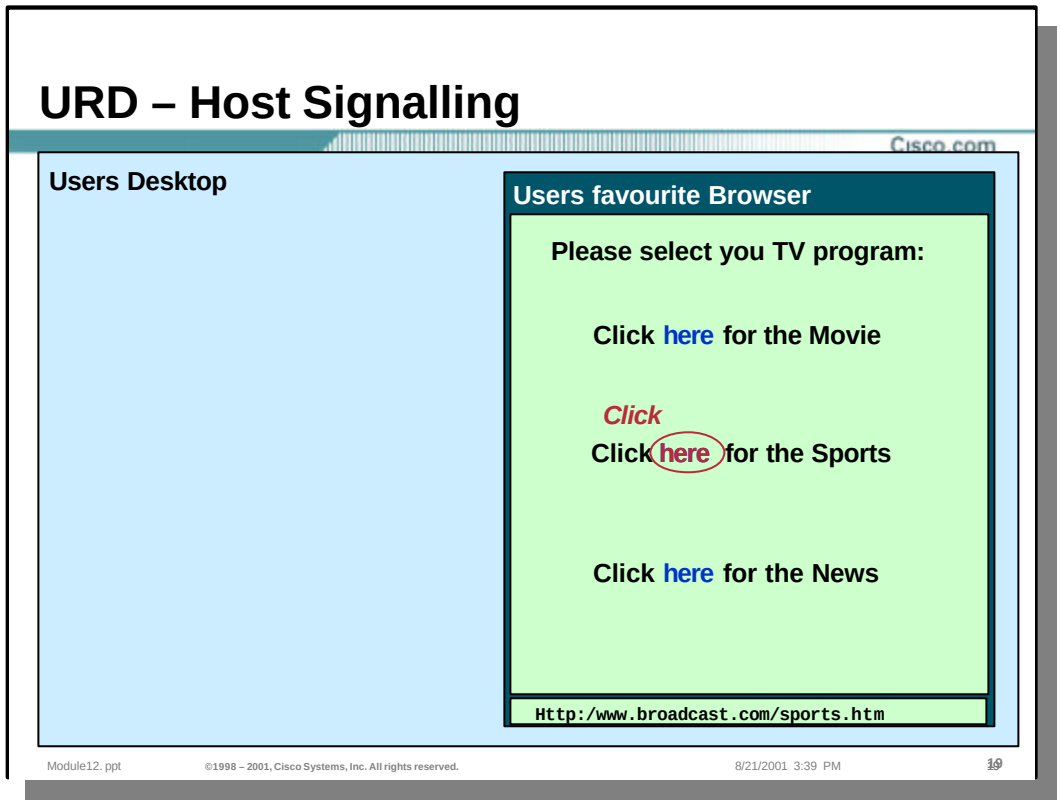
Cisco.com

- **URD and IGMP v3lite Concepts:**
  - **IGMP (v1/v2) Snooping and CGMP still work!**
    - **Group based traffic restriction due to IGMPv1/v2 reports.**
    - **IGMP Snooping will not work with full IGMPv3 reports unless supported explicitly !**
  - **Solutions don't work outside of IP SSM range.**
    - **IGMPv1/v2 reports still interpreted as (*,G) Joins!**
    - **Causes router to join the Shared Tree.**
      - **Implies "forward ALL sources in the group" !**



- **URD Overview**
  - The idea of URD as an interim solution for transition to IGMP v3 is that the content provider builds a web page that contains URD links. Those links contain information on sources that are willing to provide the multicast content for certain groups.
  - When a user clicks on such a link the browser of a host will try to open a TCP connection to the web server on port 659. If the last hop router is enabled for URD on the interface where the router receives the TCP packets from the host, it will intercept all packets for TCP connections destined to port 659 independent of the actual destination address of the TCP connection. From the information in URD the router learns about sources and groups.
  - Because normal IGMPv1/v2 group membership reports are still sent by the application, URD is compatible with IGMPv1/v2 snooping and CGMP in switches.

URD – Host Signalling



URD – Host Signalling

Cisco.com

Users Desktop

Users favourite Browser

Thank you for choosing this Sports channel

Currently showing Euro 2000 Soccer live from Brussels

England : Germany 3 : 1

Min 89:00

URD – Host Signalling

Cisco.com

Users Desktop

Old streaming video receiver application. Does IP Multicast, but not IP SSM

Works fine if we don't try to run it in the SSM-Range

Users favourite Browser

Thank you for choosing this Sports channel

Currently showing Euro 2000 Soccer live from Brussels

England : Germany 3 : 1

Min 89:00

URD – Host Signalling

Cisco.com

Users Desktop

Old streaming video receiver application. Does IP Multicast, but not IP SSM

That is..
Unless some unwanted traffic disturbs the reception, maybe some DoS attack...

Users favourite Browser

Thank you for choosing this Sports channel

Currently showing
Euro 2000 Soccer
live from Brussels

England : Germany
3 : 1

Min 89:00

URD – Host Signalling

Cisco.com

Users Desktop

Old streaming video receiver application. Does IP Multicast, but not IP SSM

Running the application on an SSM-Range alone does not help: The application will receive nothing!

Users favourite Browser

Thank you for choosing this Sports channel

Currently showing
Euro 2000 Soccer
live from Brussels

England : Germany
3 : 1

Min 89:00

URD – Host Signalling

Cisco.com

Users Desktop

Old streaming video receiver application. Does IP Multicast, but not IP SSM

But thanks to URD, the old application can run on an address in the SSM-Range and will only receive traffic from the right source!

Users favourite Browser

Thank you for choosing this Sports channel

Currently showing
Euro 2000 Soccer
live from Brussels

England : Germany
3 : 1

Min 89:00

Retrieved URL String successfully

URD – Host Signalling

Cisco.com

Users Desktop

Let's repeat this in Slow Motion...

Users favourite Browser

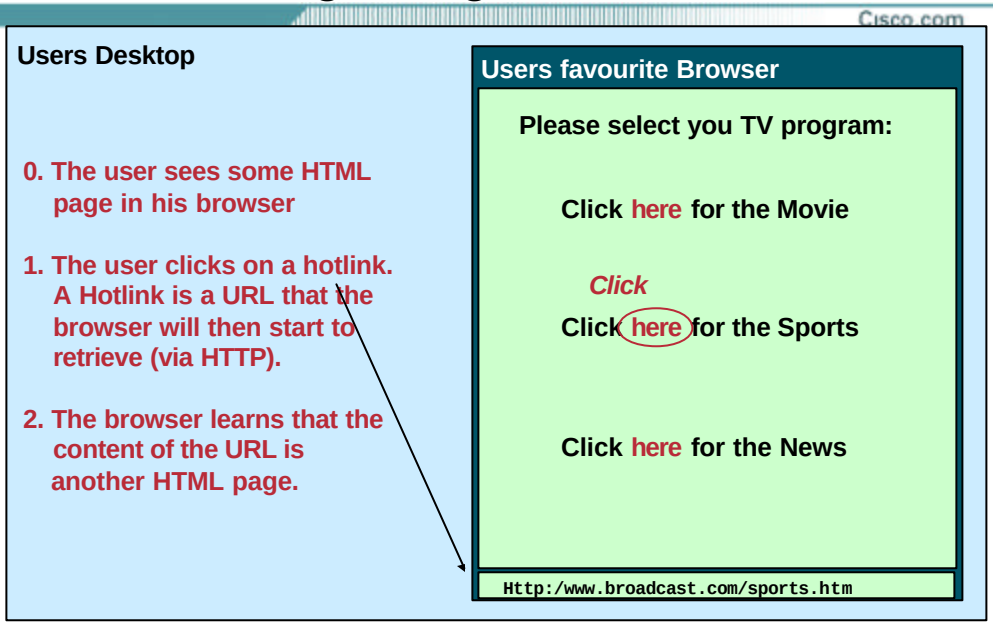
Please select you TV program:

Click here for the Movie

Click here for the Sports

Click here for the News

URD – Host Signalling



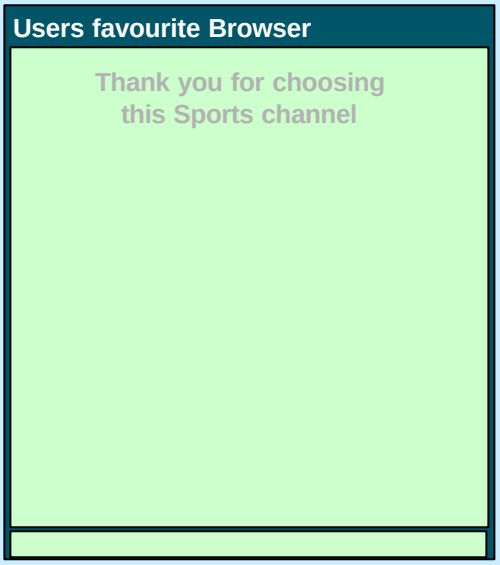
URD – Host Signalling

Cisco.com

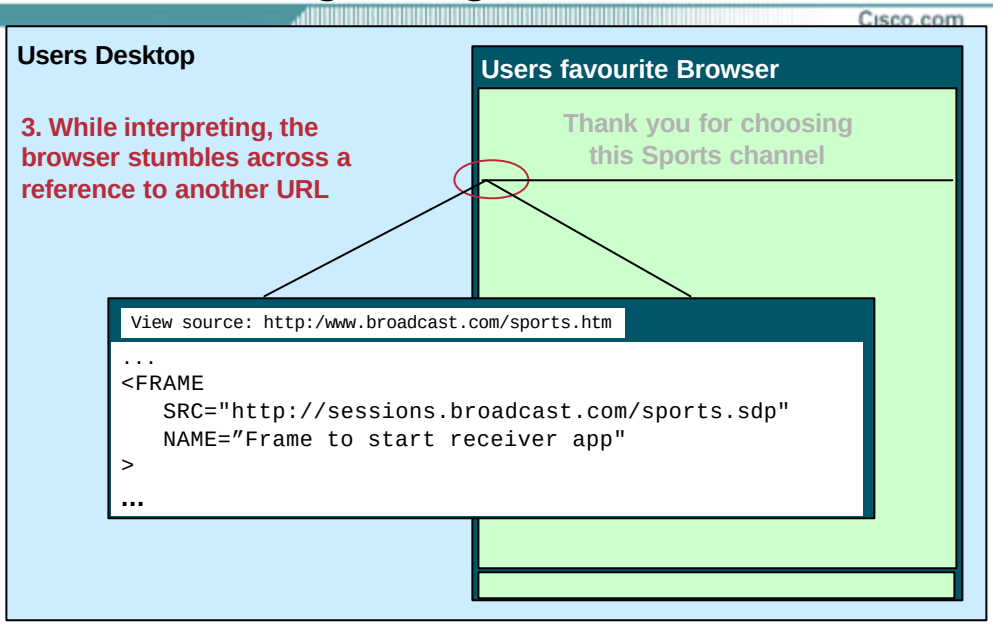
Users Desktop

3. The browser will clear the display to start "*painting*" this new HTML page.

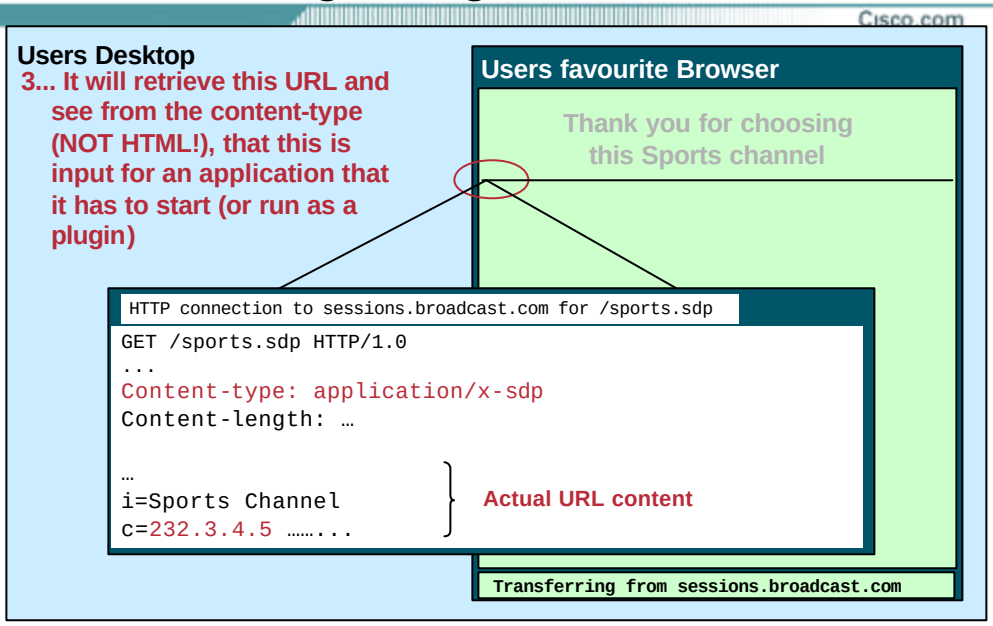
The browser then starts reading and interpreting that HTML page.



URD – Host Signalling

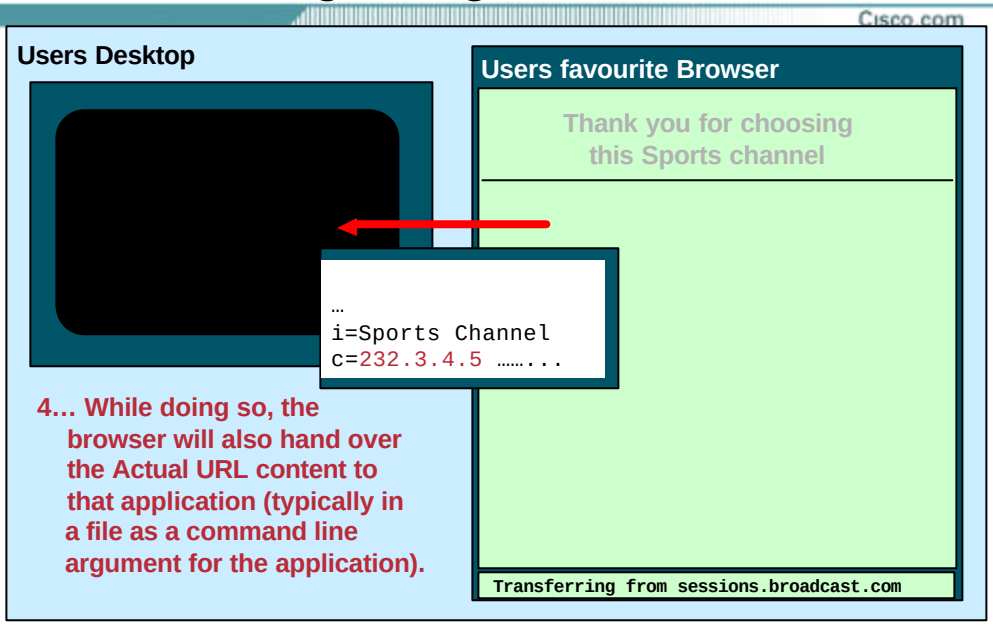


URD – Host Signalling

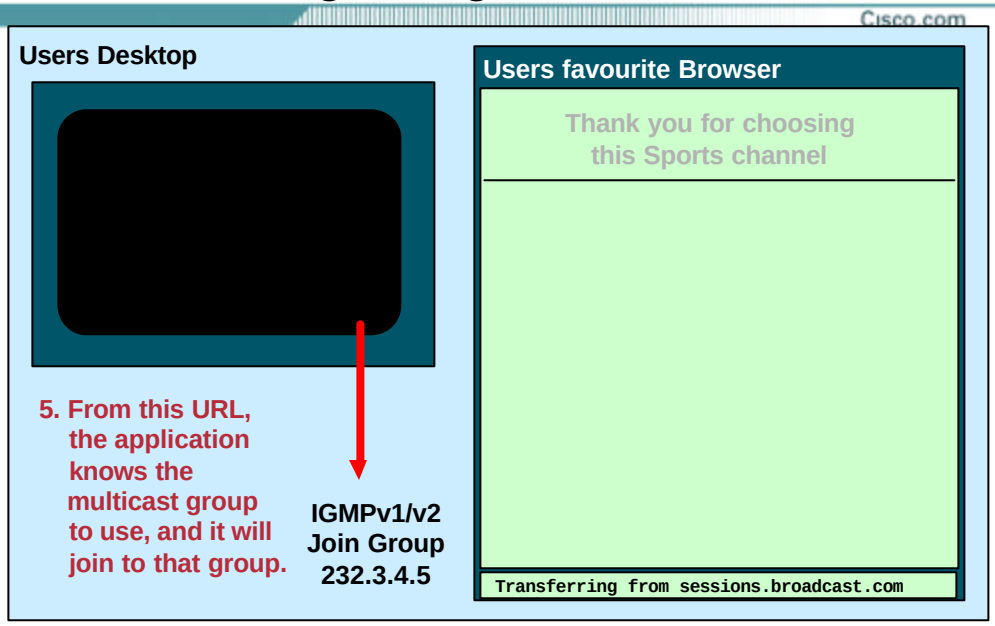


URD – Host Signalling

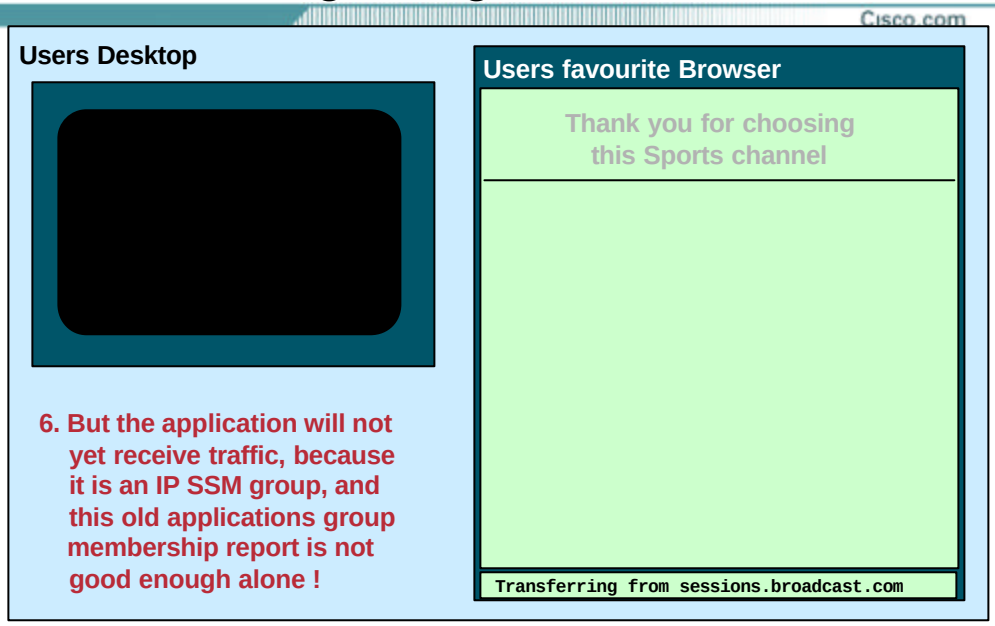
URD – Host Signalling



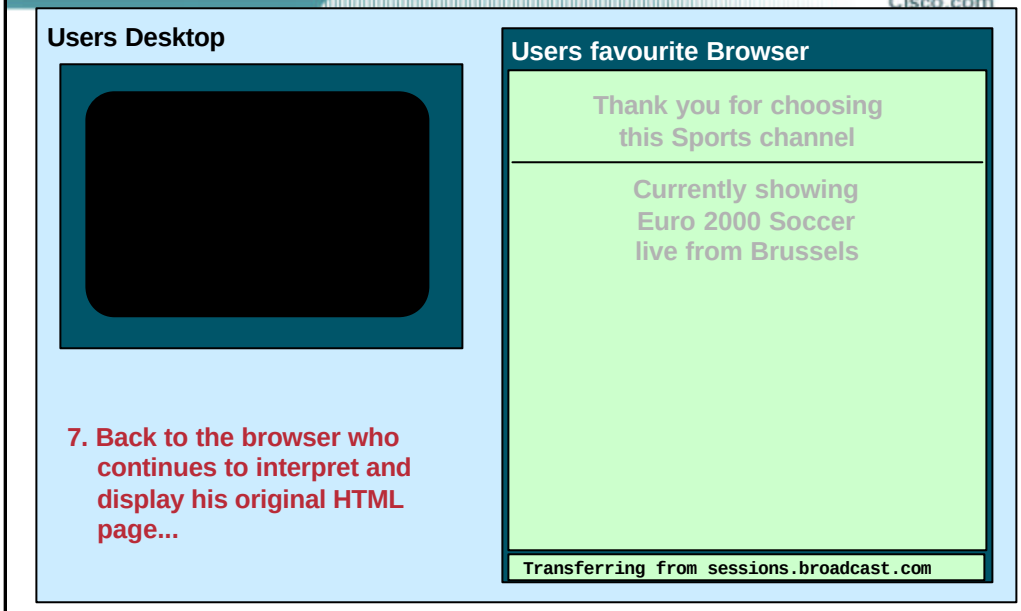
URD – Host Signalling



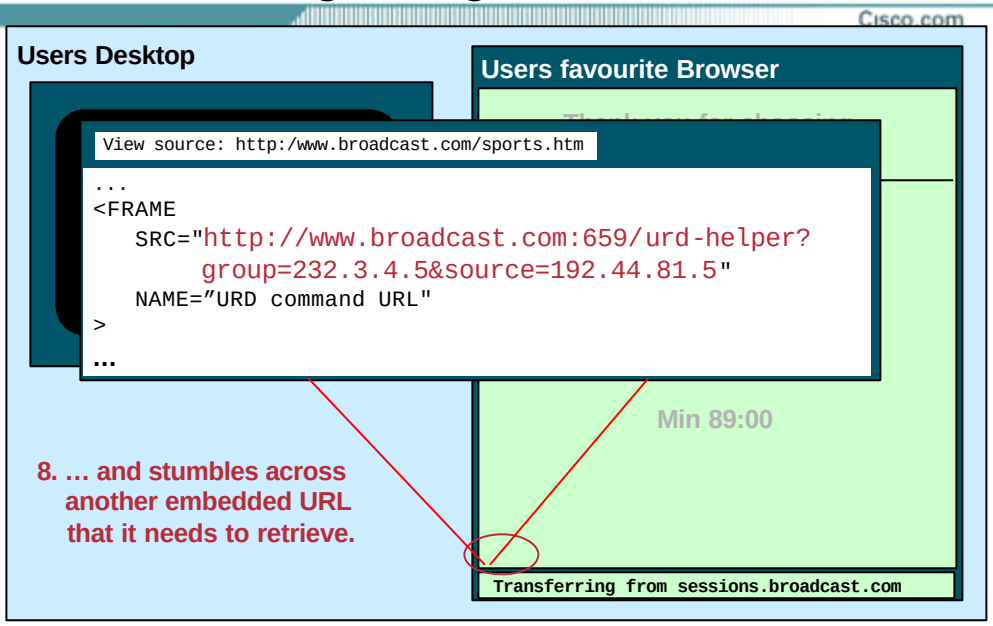
URD – Host Signalling



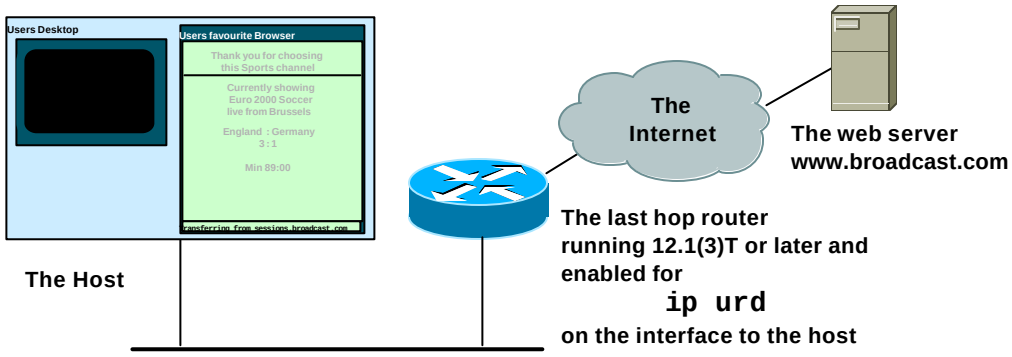
URD – Host Signalling



URD – Host Signalling

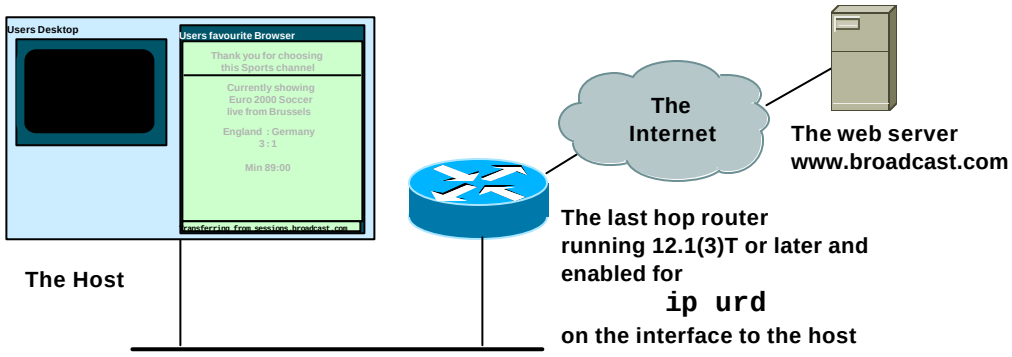


URD – Host Signalling



Let's zoom out a bit...

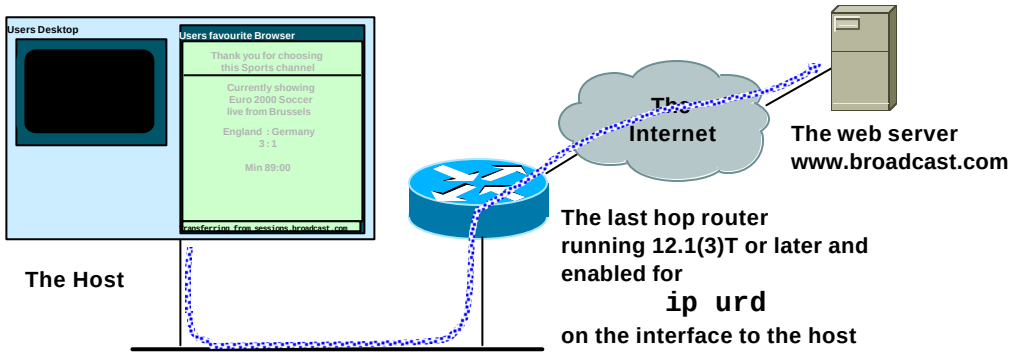
URD – Host Signalling



If the browser tries to retrieve the URL

```
http://www.broadcast.com:659/urd-
helper?group=232.3.4.5&source=192.44.81.5
```

URD – Host Signalling

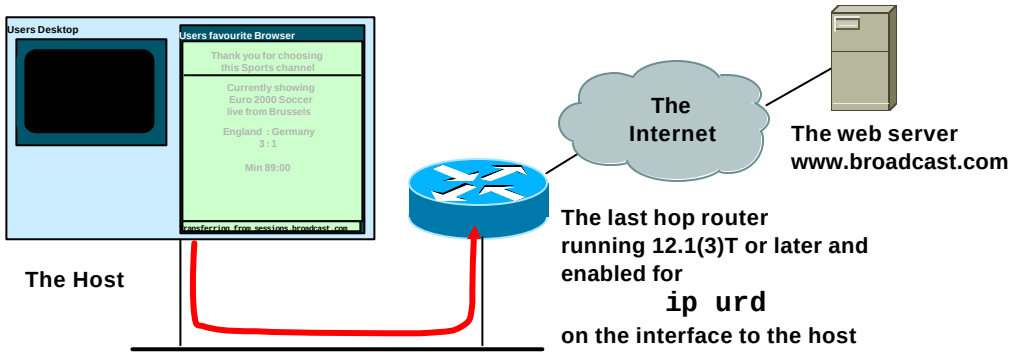


If the browser tries to retrieve the URL

```
http://www.broadcast.com:659/urd-
helper?group=232.3.4.5&source=192.44.81.5
```

Then it wants to open a TCP connection to www.broadercast.com, port 659

URD – Host Signalling

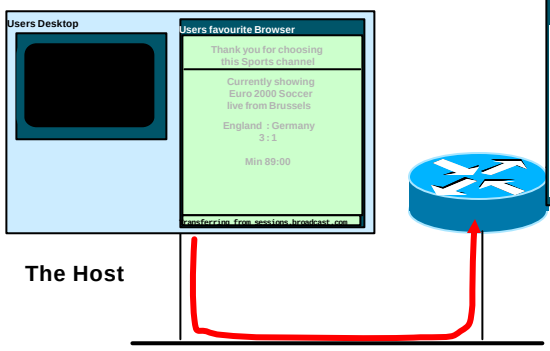


If the browser tries to retrieve the URL

```
http://www.broadcast.com:659/urd-
helper?group=232.3.4.5&source=192.44.81.5
```

Then it wants to open a TCP connection to www.broadercast.com, port 659
But it only gets up to the first-hop router, who intercepts all TCP connections to port 659, whatever destination address they are for !

URD – Host Signalling



The Host

```
Watch TCP connection
GET /urd-helper?
    group=232.3.4.5&
    source=192.44.81.5
    HTTP/1.0
```

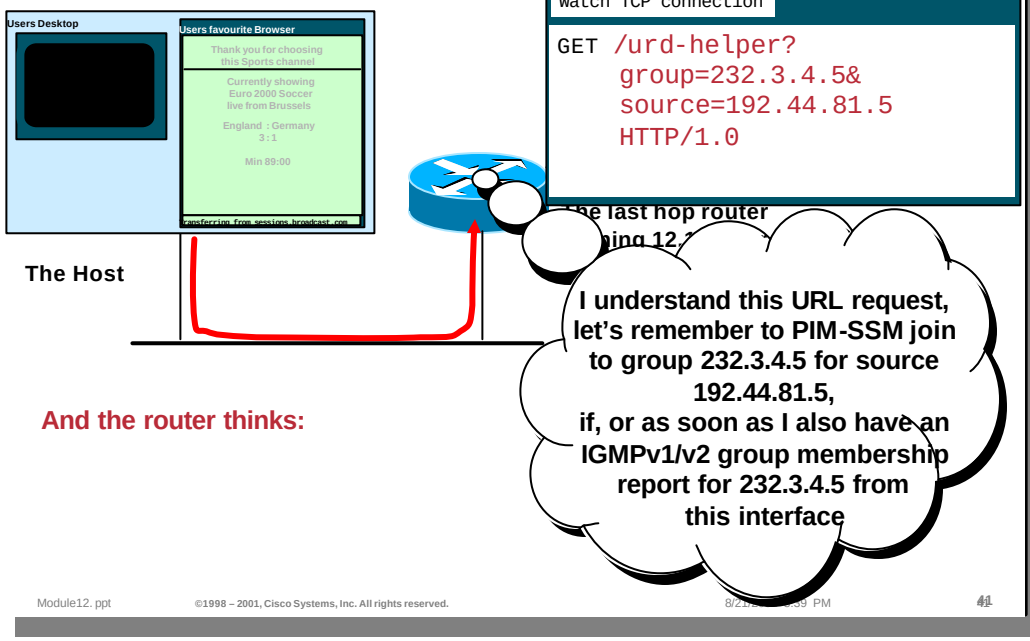
The last hop router running 12.1(3)T or later and enabled for

`ip urd`

on the interface to the host

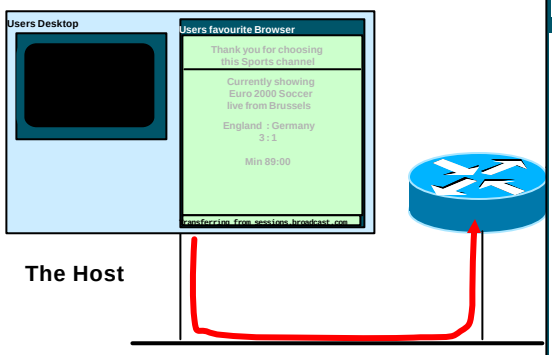
The router disguises itself as a web server and listens to what the host want to have.

URD – Host Signalling



And the router thinks:

URD – Host Signalling



The Host

Watch TCP connection

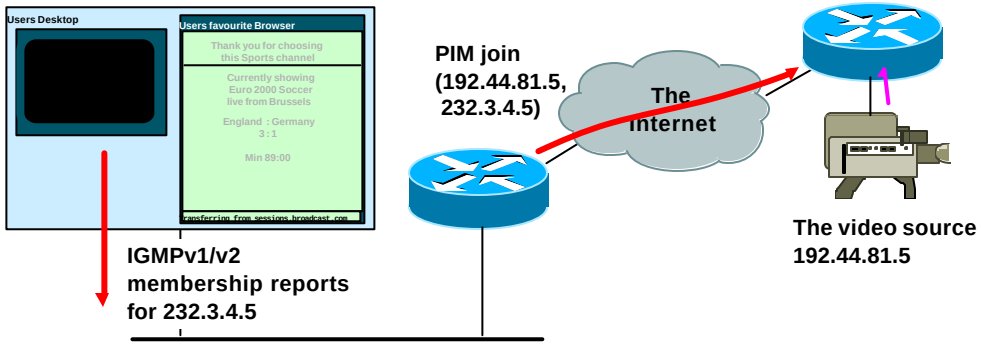
```
GET /urd-helper?
    group=232.3.4.5&
    source=192.44.81.5
    HTTP/1.0

HTTP/1.1 200 OK
Server: cisco IOS
Content-Type: text/html
<html>
<body>
Retrieved URL string
successfully
</body>
</html>
```

And so the router answers:

And closes the TCP connection.

URD – Host Signalling

Cisco.com

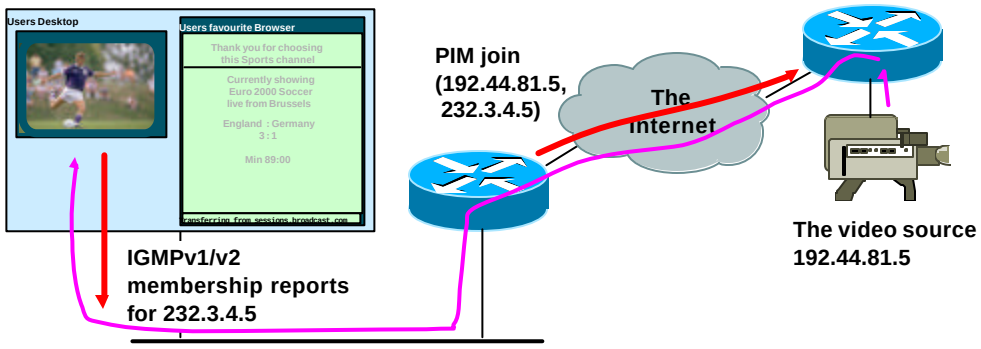


And once it sees the first IGMPv1/v2 report for the group (from the application), the router will join to the source via PIM-SS and continue as long as the IGMPv1/v2 group reports come in.

Note: The URL request from the browser and the first IGMPv1/2 report from the application may arrive in any order within ~ 1 minute

URD – Host Signalling

Cisco.com



And finally the picture arrives and is being forwarded as long as the application runs and sends the IGMPv1/v2 membership reports

URD – Host Signalling

Cisco.com

Users Desktop

Old streaming video receiver application. Does IP Multicast, but not IP SSM

And all the user could notice, is the string returned by the router (may be hidden)!

Users favourite Browser

Thank you for choosing this Sports channel

Currently showing
Euro 2000 Soccer
live from Brussels

England : Germany
3 : 1

Min 89:00

Retrieved URL String successfully

Module12. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 3:39 PM 45

URD – Configuration

Cisco.com

- **Enable IP SSM for existing applications**
  - **Works with every browser that supports frames (or one click more for those without)**
- **No plugin's required**
  - **Complete host platform independence**
- **Nothing to configure on the host**
- **URL easily added to WWW server HTML pages**
  - **No additional CGI scripts required.**

URD – Cisco's Implementation

Cisco.com

- Supported in IOS 12.1(3)T, 12.0(15)S and later.
- Supported in the process, fast and CEF paths
- Intercepting solely based on TCP port 659
  - If first hop router is not URD enabled, www-server may want to reply to HTTP on that port too (error discovery)
- Port 659 assigned by IANA for Cisco URD.
- URD - URL Rendezvous Directory
  - Name still carries the idea that it is also quite simple to write a CGI-Script to completely emulate an RP, i.e.: add web pages, where you would click onto if you are a source, and the script would then create the URD command URLs for the receivers.

IGMP v3lite Overview

Cisco.com

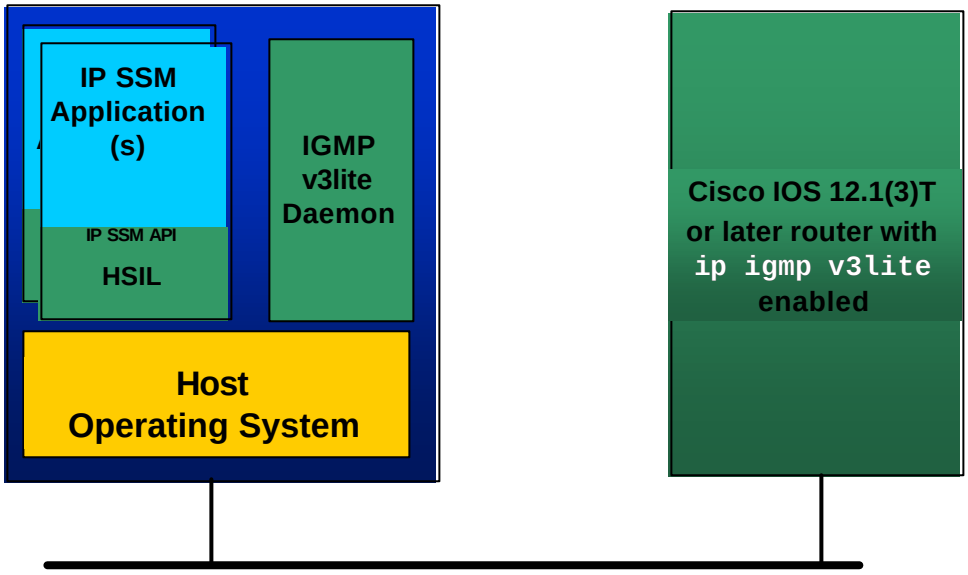
- **Source side:**
  - **No application changes required!**
- **Receiver side IGMPv3 API:**
  - **draft-ietf-idmr-msf-api-00.txt**
    - *Socket Interface Extensions for Multicast Source Filter*
    - **Supports all memberships possible with IGMPv3: Group membership with INCLUDE or EXCLUDE list of sources.**
    - **Different subsets of the API defined (one for IP SSM)**
    - **Kernel implementation of this will also filter out any unwanted received traffic still forwarded to host.**
    - **Layer2 of hosts is not IP SSM aware, input filtering on group only.**

IGMP v3lite Overview

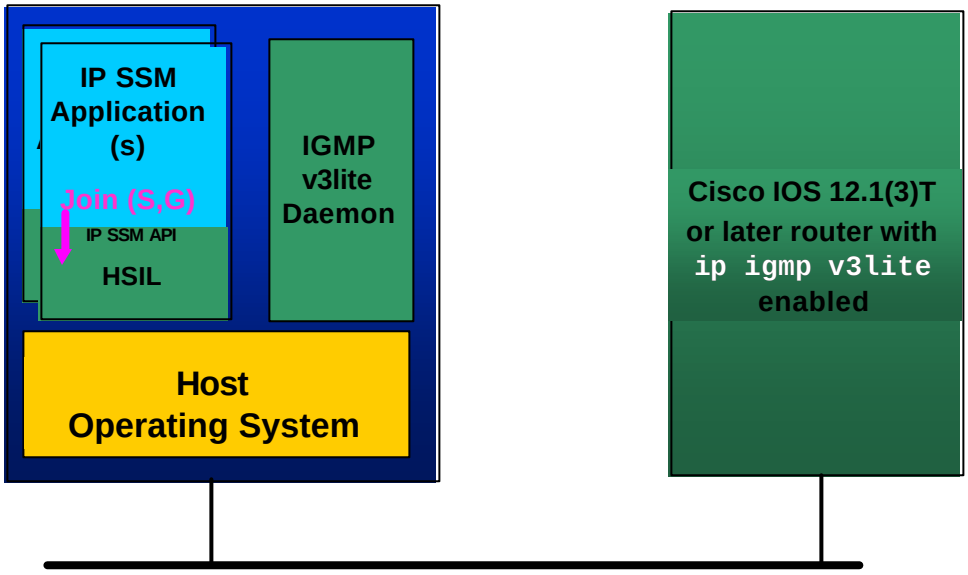
Cisco.com

- **IGMP v3lite HSIL** *(Host Side IGMP Library)*
  - **Provides for the IP SSM subset of IGMPv3 API**
  - **Applications must still filter out unwanted traffic.**
  - **Forward compatible with OS supported IGMPv3:**
    - **Recompile of application without HSIL**
    - **HSIL may also be able to detect and support host native IGMPv3 if available.**

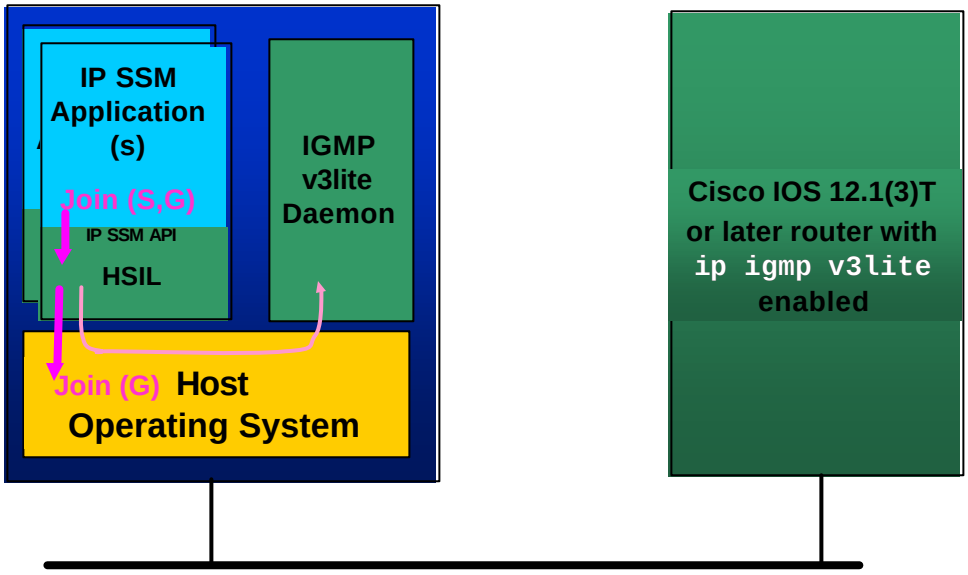
IGMP v3lite – Host Signalling



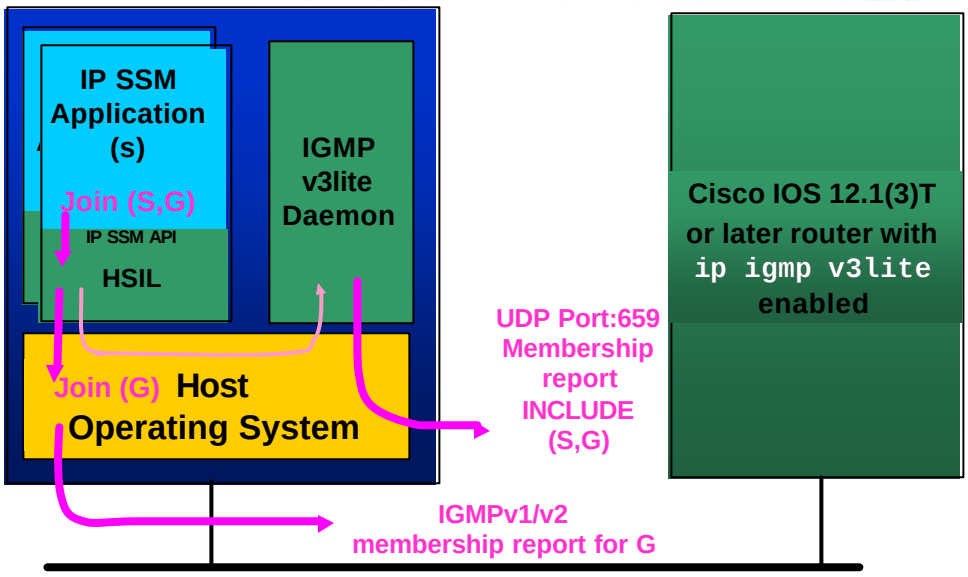
IGMP v3lite – Host Signalling



IGMP v3lite – Host Signalling



IGMP v3lite – Host Signalling



IGMP v3lite – Cisco's Implementation

Cisco.com

- Solution to start developing and deploying IP SSM applications with an IGMPv3 API subset.
  - Router side supported in IOS 12.1(3)T and later
  - Host side written by Whitebarn for Cisco
  - Supported for typical OS's (Windows, Unix, Linux)
  - Host side binaries will be freely downloadable from `www.whitebarn.com`
  - Supported ONLY at the IP SSM API in the host
    - i.e.: Do not try to write your own HSIL or Daemon and expect IOS to do the right thing.

SSM – Summary

Cisco.com

- **Solves multicast address allocation problems.**
  - **Flows differentiated by both source and group.**
    - **Not just by group.**
  - **Content providers can use same group ranges.**
    - **Since each (S,G) flow is unique.**
- **Helps prevent certain DoS attacks**
  - **"Bogus" source traffic:**
    - **Can't consume network bandwidth.**
    - **Not received by host application.**

Agenda

Cisco.com

- Source Specific Multicast
- Bidirectional (Bidir) PIM

Multicast Application Categories

Cisco.com

- One-to-Many Applications
  - Video, TV, Radio, Concerts, Stock Ticker, etc.
- Few-to-Few Applications
  - Small (<10 member) Video/Audio Conferences
- Few-to-Many Applications
  - TIBCO RV Servers (Publishing)
- Many-to-Many Applications
  - Stock Trading Floors, Gaming
- Many-to-Few Applications
  - TIBCO RV Clients (Subscriptions)

Multicast Application Categories

PIM-SM (S, G) State

Cisco.com

- One-to-Many Applications
  - Single (S,G) entry
- Few-to-Few Applications
  - Few (<10 typical) (S,G) entries
- Few-to-Many Applications
  - Few (<10 typical) (S,G) entries
- Many-to-Many Applications
  - *Unlimited (S,G) entries*
- Many-to-Few Applications
  - *Unlimited (S,G) entries*

Many-to-Any State Problem

Cisco.com

- Creates huge amounts of (S,G) state
  - State maintenance workloads skyrocket
    - High OIL fanouts make the problem worse
  - Router performance begins to suffer
- Using Shared-Trees only.
  - Provides some (S,G) state reduction
    - Results in (S,G) state only along SPT to RP
    - Frequently still too much (S,G) state
    - Need a solution that only uses (*,G) state

Eliminating (S,G) State
Solution 1

Cisco.com

- **Register-Encapsulate all data to RP**
  - **Easy to implement**
    - **RP never bothers to send a Register-Stop**
  - **Effectively IP-IP tunneling traffic to RP**
  - **Still results in (S,G) state in:**
    - **The RP**
    - **The first-hop routers**
  - **Each packet must be de-encapsulated**
    - **Process-switched**
    - **Only feasible if data rates are very low.**

Eliminating (S,G) State
Solution 2

Cisco.com

- **Bidirectional Shared-Trees**
  - **Allows data to flow *up* the Shared Tree**
    - **Source traffic follows Shared Tree to get to the RP and all other receivers on the Shared Tree**
  - **Cannot use current (*,G) RPF rules**
    - **Care must be taken to avoid multicast loops**
  - **Requires a Designated Forwarder (DF)**
    - **Responsible for forwarding traffic up Shared Tree**
      - **DF's will accept data on the interfaces in their OIL.**
      - **Then send it out all other interfaces. (Including the IIF.)**

Bidirectional (Bidir) PIM

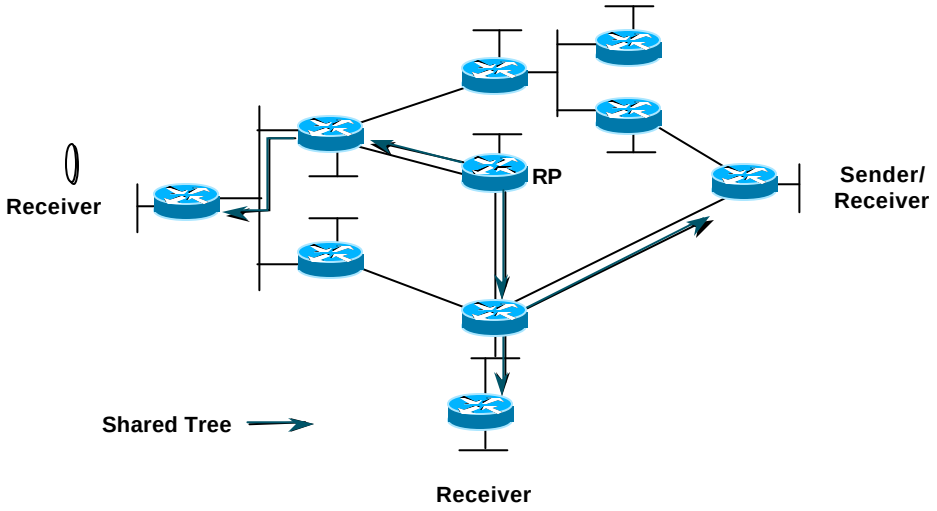
Cisco.com

- **Idea:**
  - **Use the same tree for traffic from sources towards RP and from RP to receivers**
- **Benefits:**
  - **Less state in routers**
    - **Only (*, G) state is used**
    - **Source traffic follows the Shared Tree**
      - **Flows up the Shared Tree to reach the RP.**
      - **Flows down the Shared Tree to reach all other receivers.**

- **Bidir PIM**
  - PIM Sparse Mode in its native form is unidirectional – the traffic from sources to the RP initially flows encapsulated in Register messages wich presents a significant burden due to encapsulation / decapsulation mechanisms. Additionally, shortest path tree is built between the RP and the source (initiated by the RP) which results in (*,G) and (S,G) entries at least on the way between the RP and the source.
  - Several multicast applications use many-to-many model where each participant is receiver and sender as well. In such an environment (*,G) and (S,G) entries appear everywhere along the path from participants and the associated RP in a PIM Sparse Mode domain resulting in increased memory and protocol overhead. It is also possible that the path from the source to the RP and the opposite path (from the RP to the source which is a receiver as well) are incongruent.
  - Bi-directional PIM dispenses with both encapsulation and source state by allowing packets to be natively forwarded from a source to the RP using shared tree state only. This ensures that only (*,G) entries will appear in multicast forwarding tables and that the path taken by packets flowing from the participant (source and/or receiver) to the RP and vice versa will be the same.

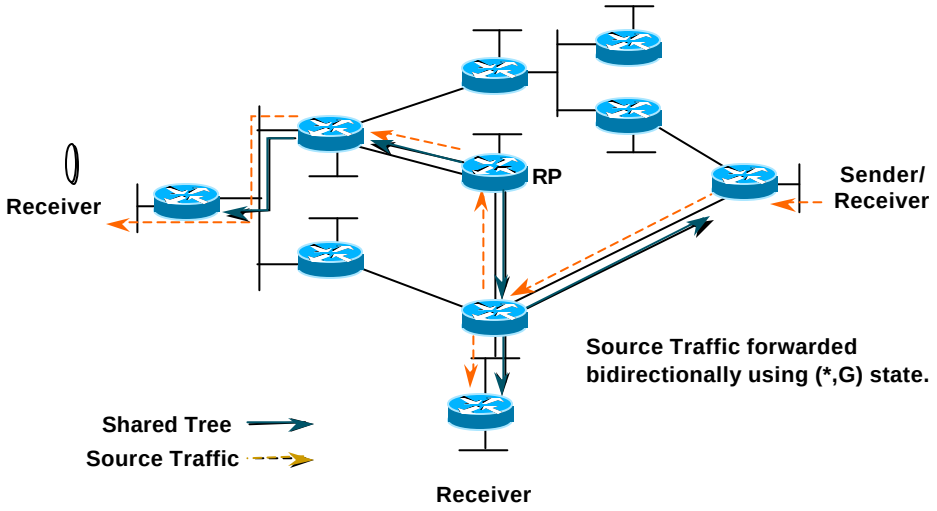
Bidirectional PIM – Overview

Cisco.com



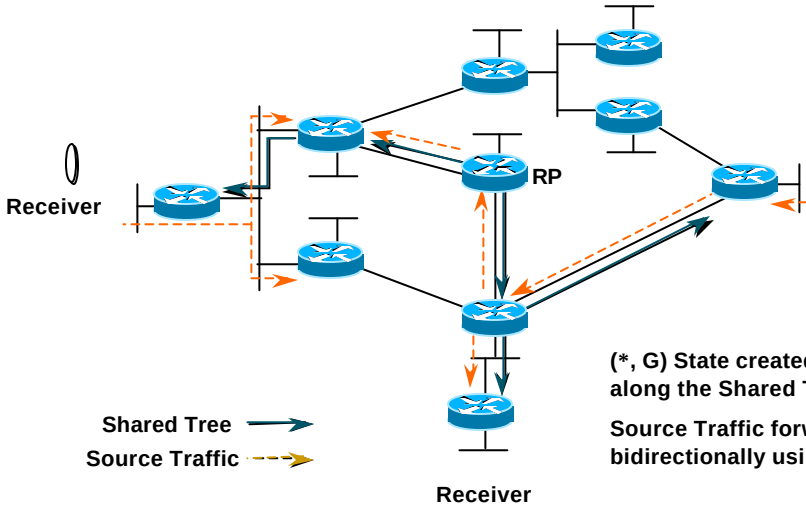
Bidirectional PIM – Overview

Cisco.com



Bidirectional PIM – Overview

Cisco.com



(*, G) State created only along the Shared Tree.

Source Traffic forwarded bidirectionally using (*,G) state.

PIM Modifications for Bidir Operation

Cisco.com

- **Designated Forwarders (DF)**
  - **On each link the router with the best path to the RP is elected to be the DF.**
    - **Note: Designated Routers (DR) are not used for bidir groups.**
  - **The DF is responsible for forwarding traffic upstream towards the RP.**
  - **No special treatment is required for local sources.**

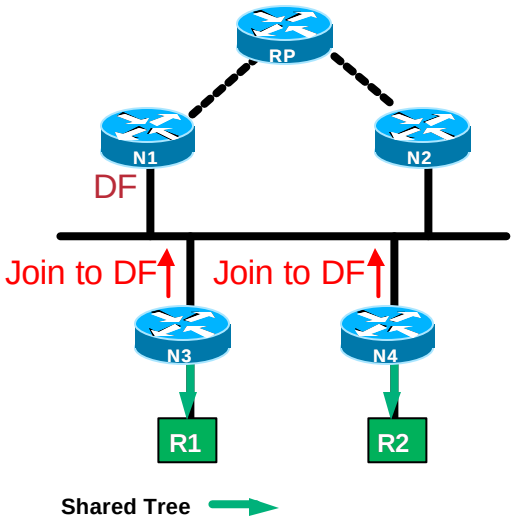
Module12. ppt 8/21/2001 3:39 PM 66

- **PIM Modifications for Bidir Operation**
  - The major modification of PIM Sparse Mode to support bidirectional mode is an addition of a Designated Forwarder, which takes over the role of a Designated Router (DR) and has the following responsibilities:
    - It is the only router that forwards packets travelling downstream (towards receiver segments) onto the link
    - It is the only router that picks-up upstream traveling packets (away from the source) off the link and forwards them towards the RP
  - There is one DF per RP for bidirectional group(s) on each link. One and only one election is performed at RP discovery time. There is no constant control traffic and control messages appear only on changes. The election is robust and enforces consistent view on all routers on link. The router with the best unicast route to the RP is elected as a DF.
  - There is no effect of this election on local sources – their traffic reaches locally attached receivers directly and special treatment is no longer required when the sources are directly connected to a router. Data from those sources will automatically be picked up by the DF and forwarded towards the RP.

Forwarding / Tree Building

Cisco.com

- **Downstream routers with receivers Join towards the DF.**



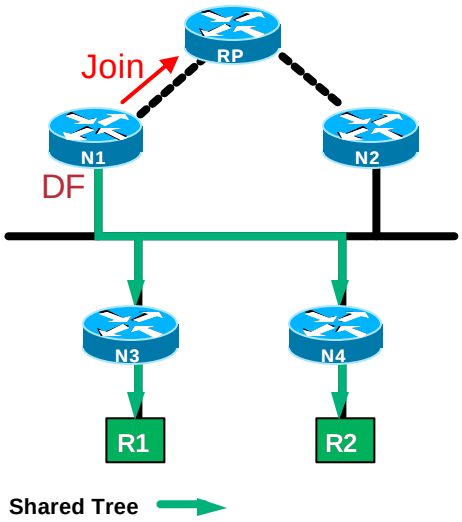
Module12. ppt 8/21/2001 3:39 PM 67

- **Forwarding / Tree Building**
  - The DF also forwards / initiates all (*,G) Joins towards the RP for the active group. Downstream routers forward their (*,G) Joins via upstream DFs. They indicate that in the Upstream Router field of a PIM Join message.

Forwarding / Tree Building

Cisco.com

- **Downstream routers with receivers Join towards the DF.**
- **DF adds link to (*,G) olist and Joins towards the RP.**



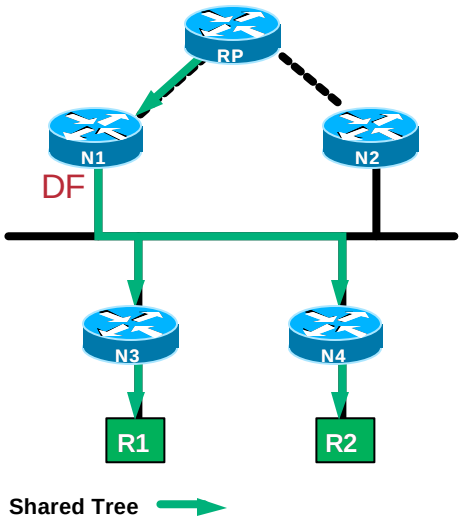
Module12. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 3:39 PM 68

- **Forwarding / Tree Building**
  - When (*,G) Join is received by the DF on the link it adds the link to the outgoing interface list (OIL) for the group. If the entry already exists the interface timer is refreshed. The (*,G) Join is then forwarded by the DF towards the RP for the group.

Forwarding / Tree Building

Cisco.com

- **Downstream routers with receivers Join towards the DF.**
- **DF adds link to (*,G) olist and Joins towards the RP.**
- **Shared Tree is now built.**



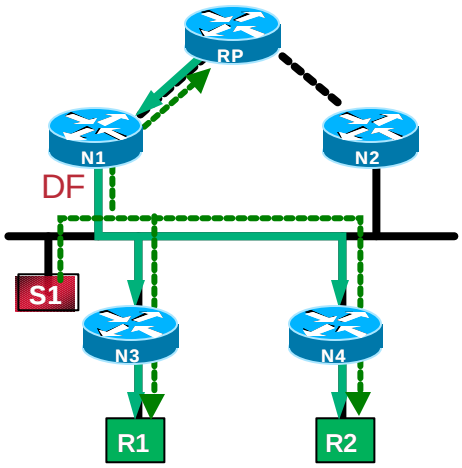
Module12. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 3:39 PM 69

- **Forwarding / Tree Building**
  - When (*,G) Join is received by the DF on the link it adds the link to the outgoing interface list (OIL) for the group. If the entry already exists the interface timer is refreshed. The (*,G) Join is then forwarded by the DF towards the RP for the group.

Forwarding / Tree Building

Cisco.com

- **Downstream routers with receivers Join towards the DF.**
- **DF adds link to (*,G) olist and Joins towards the RP.**
- **Shared Tree is now built.**
- **The DF forwards all traffic from the link upstream towards the RP. At the same time, traffic flows down the tree.**



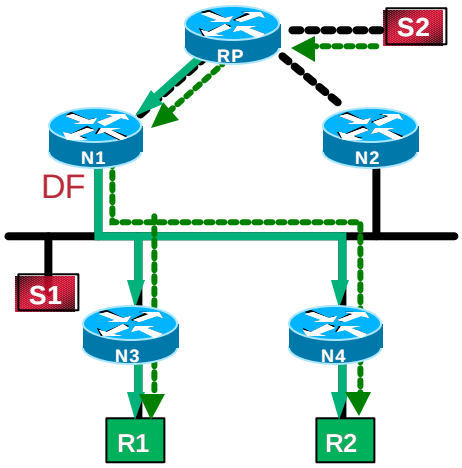
Module12. ppt 8/21/2001 3:39 PM 70

- **Forwarding/Tree Building**
  - The Designated Forwarder (DF) has all the responsibilities for forwarding multicast traffic in bidirectional PIM. It has to forward multicast traffic received on a link for which it is a DF via RPF-interface towards the RP (in addition to forward the traffic via interfaces in OIL excluding the inteface on which the traffic was received).

Forwarding / Tree Building

Cisco.com

- **Downstream routers with receivers Join towards the DF.**
- **DF adds link to (*,G) olist and Joins towards the RP.**
- **Shared Tree is now built.**
- **The DF forwards all traffic from the link upstream towards the RP. At the same time, traffic flows down the tree.**
- **Downstream traffic is forwarded through the DF.**



Module12. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 3:39 PM 71

- **Forwarding / Tree Building**
  - The branch of the tree built via (*,G) Joins is bidirectional which means that:
    - the traffic from upstream sources follows the same (downstream) path that was built with (*,G) Joins and is forwarded to the link by the same DF
    - a single path through the DF is enforced for traffic travelling upstream to the RP

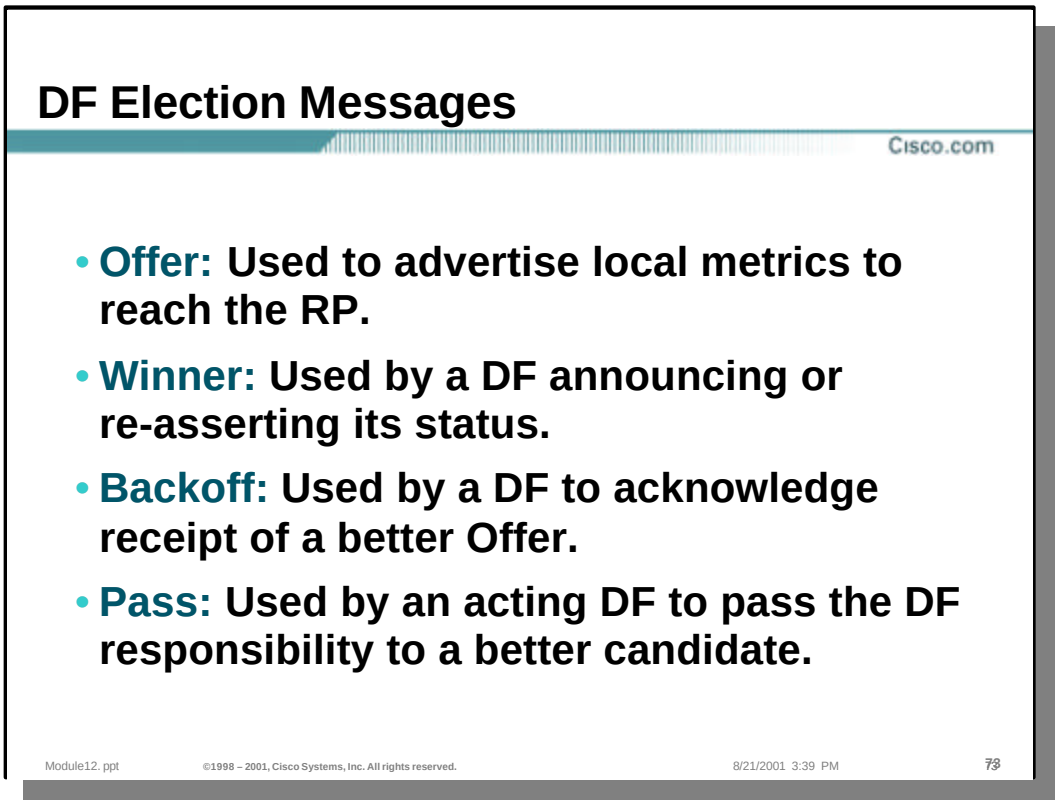
Designated Forwarder Election

Cisco.com

- **Performed as soon as a bidir RP is learned via Auto-RP or BSR.**
- **Elects the router on the link with the best path to the RP.**
- **Ensures all routers on link have a consistent view of the winner identity and metrics.**
- **Uses assert-like metric comparison rules to pick best path.**

- **DF Election**
  - The election of a Designated Forwarder on each link follows similar principles known from the Assert process in PIM Dense Mode. The mechanism ensures that all the routers on the link have consistent view of the same RP. To perform the election of the DF for a particular RP, routers on a link need to exchange their unicast routing metric information for reaching the RP.

    Note: The election of a DF is per RP and not per individual group.
  - The election process happens once only - when information on a new RP becomes available. There are however some conditions where an update to the election is needed:
    - A change in unicast metric to reach the RP for any of the routers on the link
    - The interface on which the RP is reachable changes to an interface for which the router was previously the DF
    - A new PIM neighbor on a link
    - The elected DF dies

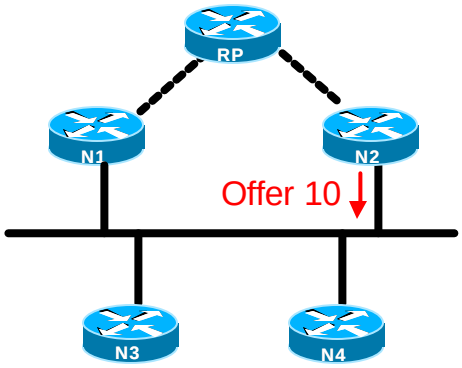


- **DF Election Messages**
  - The DF election mechanism is based on four control messages exchanged between the routers on the link.
    - The **Offer** message is used to advertise router`s unicast metric to reach the RP and is used for comparison with other routers participating in DF election.
    - The **Winner** message allows the winning router to declare to every other router on the link the identity of the winner and the metrics it is using. The message is used by the DF to reassert its status as well.
    - The **Backoff** message is used by the DF on receipt of an offer that is better than its own metric. The DF records the received information and responds with a Backoff message. This instructs the offering router to hold off for a short period of time while the unicast routing stabilizes.
    - The **Pass** message is used by the acting DF to pass its role to another router offering better metric. The old DF stops its tasks as soon as the transmission is made.

Initial Election

Cisco.com

- **On RP discovery send Offer with metric to RP.**



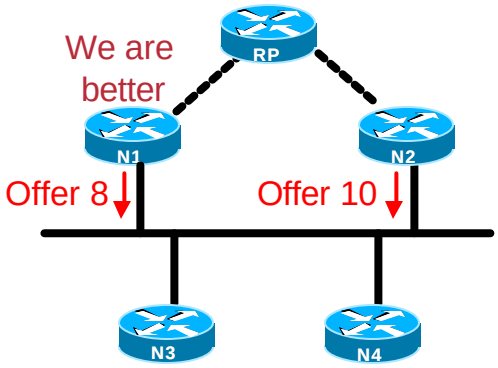
Module12. ppt 8/21/2001 3:39 PM 74

- **Initial Election**
  - When a router finds out a new RP and the DF does not exist yet it sends an Offer message. The message contains the router`s metric to reach the RP and the router`s identity. The Offer message is periodically (Offer-Interval) retransmitted.
  - If the router learns about a better metric from a neighbor it stops sending Offer messages for a period of three times the Offer-Interval. If after this period no winner is elected, the election is restarted by the router. The same happens if an Offer with a worse metric is received.
  - A router takes the role of the DF after sending three Offers without receiving any offer from any other neighbor.

Initial Election

Cisco.com

- **On RP discovery send Offer with metric to RP.**
- **Neighbors compare with own metric and send Offer only if better.**



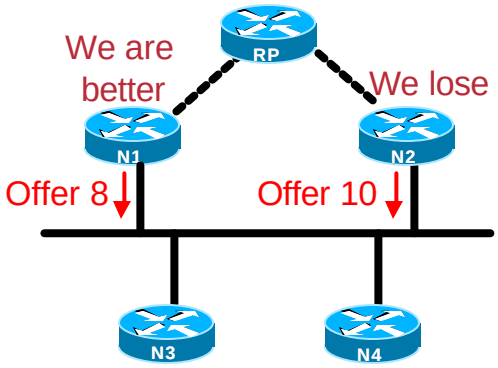
Module12. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 3:39 PM 75

- **Initial Election**
  - When neighbors hear the Offer message they compare the offered metric with their own one. If their metric is worse they back off (remain silent for three times the Offer-Interval) and thus allow the offering router to win. A timer is still running to restart offering in case election fails.
  - If the neighbor that heard the Offer has better metric it actively starts participating in the election by sending its own Offer messages including its metric to the RP and its identity.

Initial Election

Cisco.com

- **On RP discovery send Offer with metric to RP.**
- **Neighbors compare with own metric and send Offer only if better.**



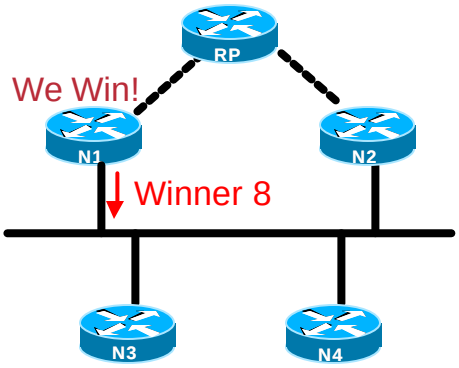
Module12. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 3:39 PM 76

- **Initial Election**
  - If the offering router hears an Offer with a better metric it assumes it lost and stops sending Offer messages for the period of three times the Offer-Interval. If after that interval the situation is not yet resolved, the election process will restart.

Initial Election

Cisco.com

- **On RP discovery send Offer with metric to RP.**
- **Neighbors compare with own metric and send Offer only if better.**
- **After repeating 3 uncontested Offers, send a Winner and assume DF role.**

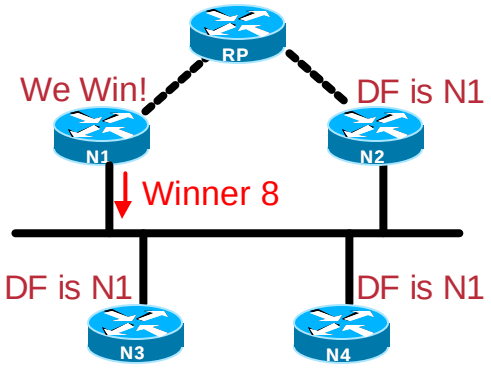


- **Initial Election**
  - The router that sent the better Offer three times (and hasn`t heard of better Offer or no Offer at all) assumes the DF role and transmits a Winner message which declares to every router on the link the identity of the winner and the metric it is using.
  - Routers hearing a Winner message stop participating in the election and record the identity and metrics of the winner.

Initial Election

Cisco.com

- **Winner message informs the other routers who is DF.**



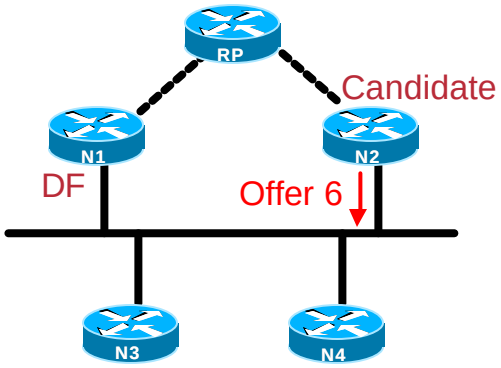
Module12. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 3:39 PM 78

- **Initial Election**
  - The router that sent the better Offer three times (and hasn`t heard of better Offer or no Offer at all) assumes the DF role and transmits a Winner message which declares to every router on the link the identity of the winner and the metric it is using.
  - Routers hearing a Winner message stop participating in the election and record the identity and metrics of the winner.

DF Preemption

Cisco.com

- **New candidate sends improved Offer.**



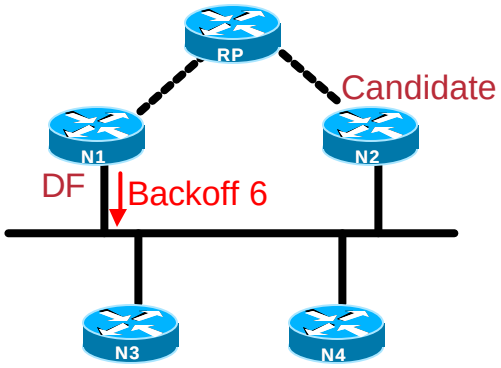
Module12. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 3:39 PM 79

- **DF Preemption**
  - Once the DF is elected the process does not restart if there are no changes in metrics, PIM neigbors, DF reachability or interfaces towards the RP. If the unicast metric to a RP changes for a non-DF router to a value that is better than that previously advertised by the DF the router sends a new Offer. A new Offer includes an improved metric and the candidate`s identity.

DF Preemption

Cisco.com

- **New candidate sends improved Offer.**
- **DF responds with Backoff instructing candidate to wait.**



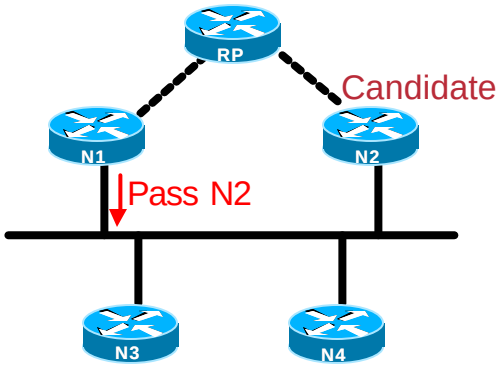
Module12. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 3:39 PM 80

- **DF Preemption**
  - Upon receipt of an Offer that is better than its current metric, the DF records the identity and metrics of the offering router and responds with a Backoff message (including the metric of the candidate that just sent the Offer).
  - The offering router will hold off for a period of time (defined in the Backoff message) while the unicast routing stabilises. All routers on the link who have pending offers with metrics equal or worse than those in the backoff message (including the original offering router) will hold further offers for the defined period.
  - If during the period someone else sends a new better Offer, the Backoff message is repeated for the new Offer and the backoff period restarted.

DF Preemption

Cisco.com

- **New candidate sends improved Offer.**
- **DF responds with Backoff instructing candidate to wait.**
- **Before backoff period expires, old DF stops forwarding and sends Pass.**

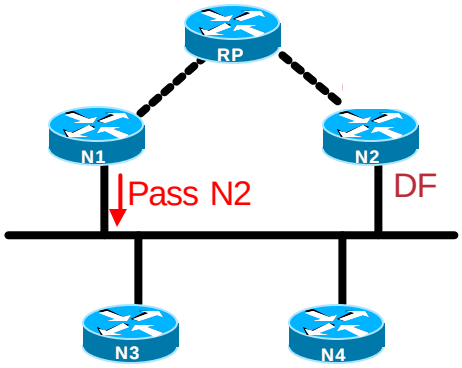


- **DF Preemption**
  - Just before the backoff period expires, the current DF declares the candidate router with the best Offer as the new DF. This is done via a Pass message which includes the IDs and metrics of both the old and new DFs.
  - The current DF stops acting as a DF soon after the Pass is transmitted. The new DF assumes the role of the DF as soon as it receives the Pass message. All other routers on the link record the identity and the metric of the newly elected DF.

DF Preemption

Cisco.com

- **New candidate sends improved Offer.**
- **DF responds with Backoff instructing candidate to wait.**
- **Before backoff period expires, old DF stops forwarding and sends Pass.**
- **On receipt candidate becomes DF.**

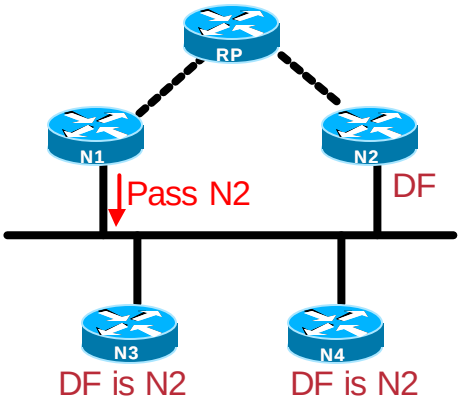


- **DF Preemption**
  - Just before the backoff period expires, the current DF declares the candidate router with the best Offer as the new DF. This is done via a Pass message which includes the IDs and metrics of both the old and new DFs.
  - The current DF stops acting as a DF soon after the Pass is transmitted. The new DF assumes the role of the DF as soon as it receives the Pass message. All other routers on the link record the identity and the metric of the newly elected DF.

DF Preemption

Cisco.com

- New candidate sends improved Offer.
- DF responds with Backoff instructing candidate to wait.
- Before backoff period expires, old DF stops forwarding and sends Pass.
- On receipt candidate becomes DF.
- Other routers hear Pass, learn N2 is now DF.

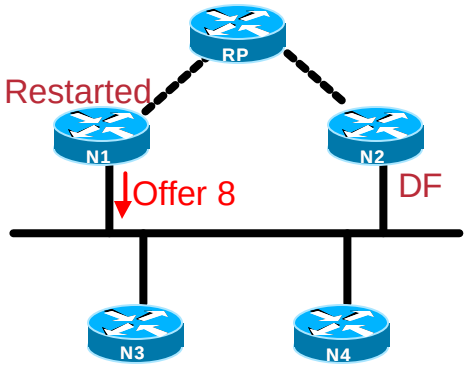


- **DF Preemption**
  - Just before the backoff period expires, the current DF declares the candidate router with the best Offer as the new DF. This is done via a Pass message which includes the IDs and metrics of both the old and new DFs.
  - The current DF stops acting as a DF soon after the Pass is transmitted. The new DF assumes the role of the DF as soon as it receives the Pass message. All other routers on the link record the identity and the metric of the newly elected DF.

PIM Neighbor Startup

Cisco.com

- **Router N1 restarts and has no knowledge of DF.**
- **On RP discovery it sends an Offer.**



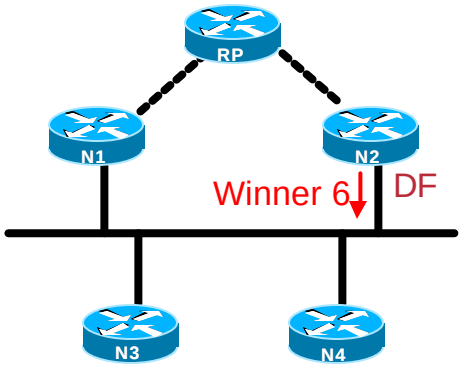
Module12. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 3:39 PM 84

- **PIM Neighbor Startup**
  - A router that started after the DF election outcome or a router that restarted in the meantime will have no knowledge of a previously elected DF. It will start advertising its metric in Offer messages on RP discovery time.

PIM Neighbor Startup

Cisco.com

- **Router N1 restarts and has no knowledge of DF.**
- **On RP discovery it sends an Offer.**
- **Acting DF responds with Winner or Backoff depending on metric comparison.**



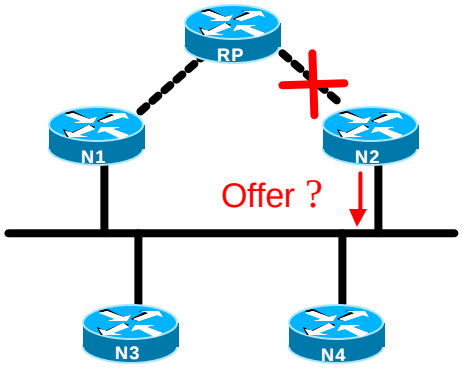
Module12. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 3:39 PM 85

- **PIM Neighbor Startup**
  - As soon as the current DF hears the Offer from the PIM neighbor that just (re)started it will respond either with a Winner or with a Backoff message depending on the metric in the Offer message. The rest of the procedure is the same as in every reelection.

DF Loses Path to the RP

Cisco.com

- **Current DF loses only path to the RP.**
- **Stops acting as the DF**
- **Sends Offer with infinite metric to trigger new DF election.**



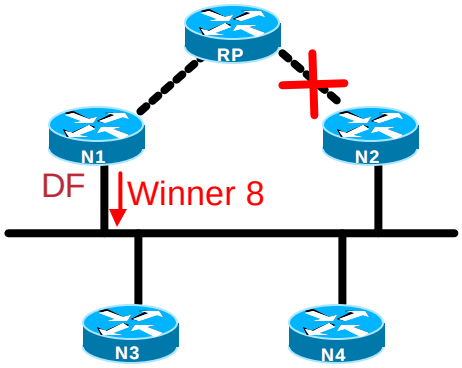
Module12. ppt ©1998 – 2001, Cisco Systems, Inc. All rights reserved. 8/21/2001 3:39 PM 86

- **DF Loses Path to the RP**
  - When the path to the RP currently used by the DF switches to be through the link for which it is the DF, then it can no longer provide forwarding services. Recall that the DF forwards the traffic (from a source) received on an interface towards the RP, but never via the interface on which the traffic was received.
  - Thus in this case the DF immediately stops being the DF and restarts the election by sending an Offer with an infinite metric. If no better Offer is received an infinite Offer is repeated periodically.

DF Loses Path to the RP

Cisco.com

- **Current DF loses only path to the RP.**
- **Stops acting as the DF**
- **Sends Offer with infinite metric to trigger new DF election.**
- **Other candidates respond with real Offers and eventually best candidate takes over with a Winner message.**



- **DF Loses Path to the RP**
  - The procedure after the router acting as a DF loses the path to the RP and its RPF-interface becomes the same as the interface for which it is the DF is similar to standard DF election procedure. Routers that hear an infinite Offer respond with their Offers and the one with the best Offer takes over the role of a new DF with the Winner message.

DF Failures

Cisco.com

- **Detecting DF Failures**
  - **Downstream Routers**
    - **RP RPF info no longer points to DF.**
  - **Non-Downstream Routers**
    - **PIM Neighbor timeout of DF.**
- **Router response to a DF Failure**
  - **Routers resend their Offer messages**
    - **Triggers new DF election**

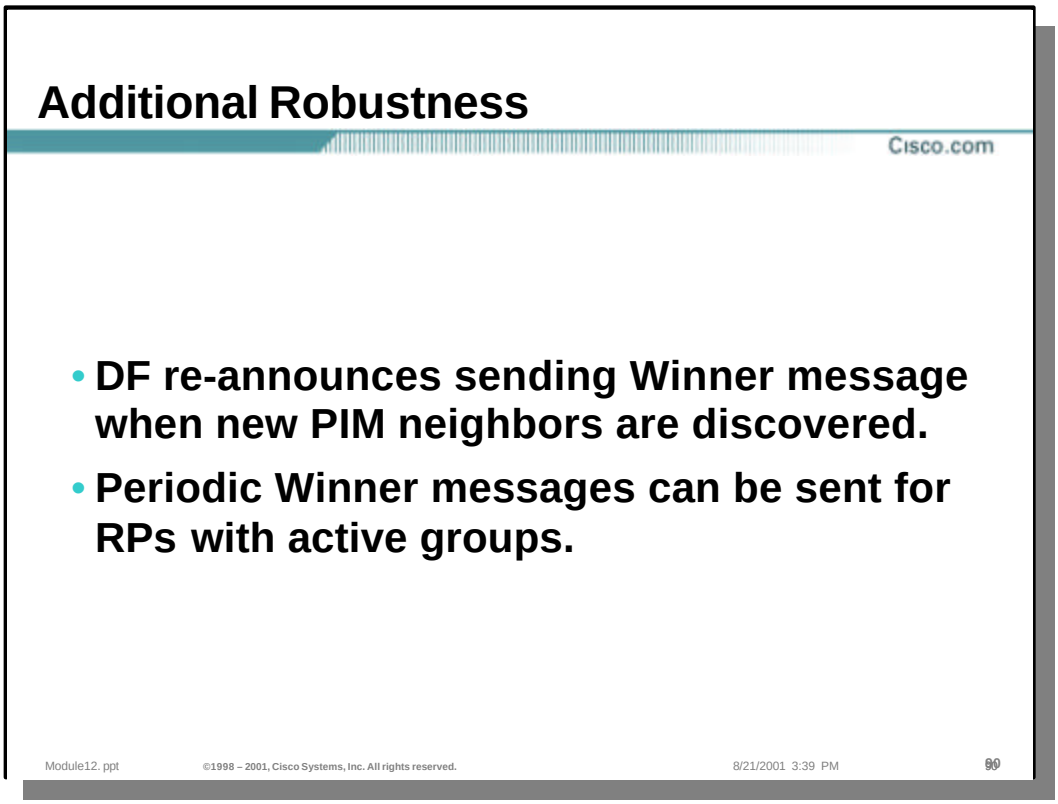
- **DF Fails**
  - The speed at which a new DF is elected after the original DF dies depends on whether there are any downstream routers on the link.
  - For downstream routers the RPF neighbor (who is the DF at the same time) will change and they will initiate the reelection by sending Offer messages. If the RP is reachable through the link via another upstream router they will use an infinite metric.
  - If no downstream routers are available the only way for other upstream routers to detect a DF failure is by the timeout of the PIM neighbor information, which will take longer.

Other Metric Changes

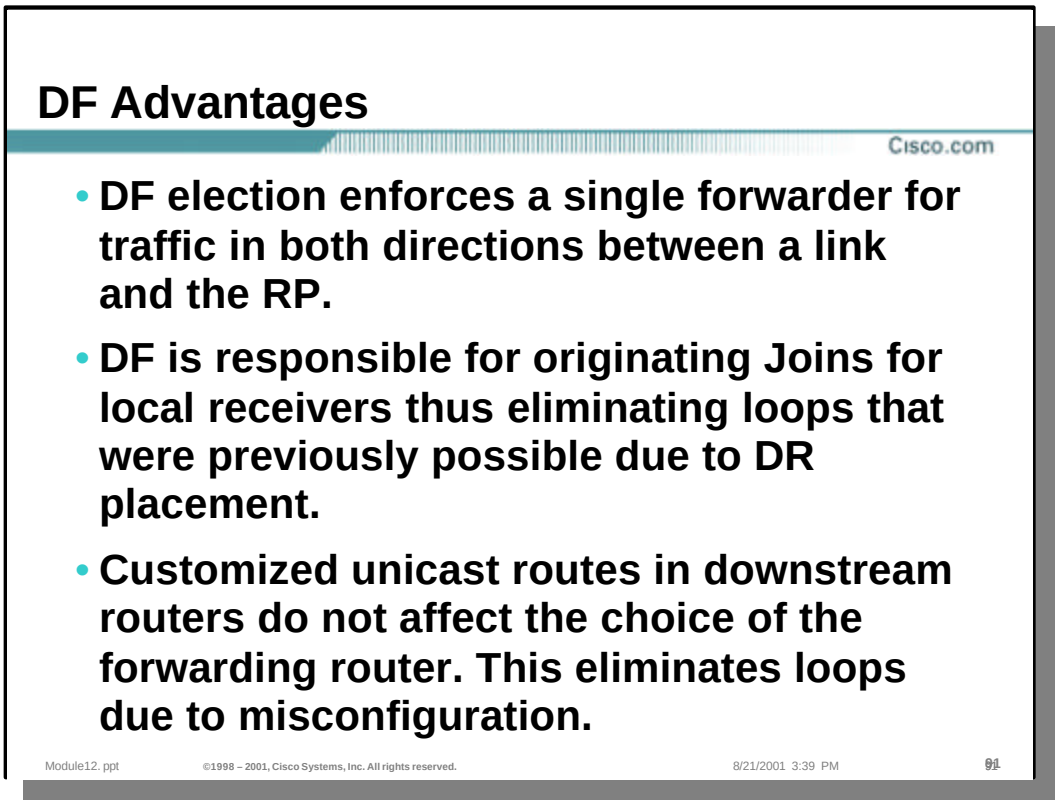
Cisco.com

- **DF metric changes:**
  - **Better metric:**
    - ***May*** **send Winner message with new metric.**
    - **Updates other routers.**
  - **Worse metric:**
    - **Sends 3 Winner messages with new metric.**
    - **Other routers can respond with a better Offer.**
- **Non-DF metric changes:**
  - **Better metric than DF:**
    - **Send new Offer to trigger DF re-election.**
  - **Worse metric than DF:**
    - **No action is taken.**

- **Other Metric Changes**
  - There are some other situations where the metric to the RP changes. When the metric of the non-DF router chenges to a value still wors than that of the current DF, no action is taken.
  - There can be changes to the metric of the current DF. If the metric becomes worse than before (assuming the DF still has a path to the RP) the DF sends a set of 3 randomly spaced Offer messages with the new metric. Routers who receive this message and have a better metric may respond with an Offer message which triggers the same procedure as follows when non-DF metric becomes better than the current DF metric. All routers assume the DF has not changed until they see a Pass or Winner message indicating the change.
  - If the routing metric at the DF changes to a better value, a single Winner message is sent advertising the new metric.



- **Additional Robustness**
  - In order to ensure an additional robustness in DF election whenever a new PIM neighbor is discovered by the current DF a Winner message is reannounced.
  - The proposal allows the DF to send periodic Winner messages for RPs serving currently active groups as well.



- **DF Advantages**
  - The implementation of PIM bidirectional mode where all the forwarding on a link is centered around the Designated Forwarder ensures highly robust PIM SM multicast networks and eliminates possible loops. All the multicast traffic from the link towards the RP and in the opposite direction passes the DF.
  - Since the role of a Designated Router (DR) is handed to a Designated Forwarder (DF) the placement of a DR is no more an issue. All (*,G) Joins are originated (forwarded) via DF which again eliminates the possibilities for forwarding loops.
  - Even if downstream routers on the link use customized unicast routes the election of a DF ensures that all those routers know who the DF is and use it for forwarding (*,G) Joins via it. This again eliminates multicast forwarding loops that were possible in regular PIM SM due to misconfigurations.

Configuring Bidir PIM
(BSR Example)

Cisco.com

- **Define Candidate RP and groups / modes it is willing to serve**

```
ip pim rp-candidate Loopback0 group-list 45 bidir
ip pim rp-candidate Loopback1 group-list 46
! Two loopbacks needed due to a nature of ACLs (permit, deny)
ip pim bsr-candidate Loopback2 4

access-list 45 permit 224.0.0.0 0.255.255.255
access-list 45 permit 227.0.0.0 0.255.255.255
! Those two groups will be PIM SM bidirectional
access-list 45 deny 225.0.0.0 0.255.255.255
! This group will be PIM DM

access-list 46 permit 226.0.0.0 0.255.255.255
! This group will be PIM SM
```

Module12. ppt 8/21/2001 3:39 PM 92

- **Configuring Bidir PIM (BSR Example)**
  - A bidirectional PIM capable router can run in bidirectional mode, sparse mode, dense mode or any combination of them. If a router is configured for bidirectional mode but does not learn of a bidirectional capable RP it will operate in sparse mode. If a bidirectional capable router learns of a bidirectional RP then the group range advertised by the RP will operate in bidirectional mode. If the RP advertises any groups with a negative prefix they will operate in dense mode.
  - By default a bidirectional RP advertises all groups as bidirectional. An access group on the RP can be used to specify a list of groups to be advertised as bidirectional. Groups with the "deny" clause will operate in dense mode.
  - A different (non bidirectional) RP address needs to be specified for groups that need to operate in sparse mode. This is because a single access-list allows only "permit" or a "deny" clause.
  - The example shows how to configure a bidirectional RP to run all 3 modes. 224/8 and 227/8 are bidirectional groups, 225/8 is dense mode and 226/8 is sparse mode. Both the bidirectional RP and the sparse mode RP are configured on one router using two different loopback interfaces.

Bidir PIM – Summary

Cisco.com

- **Drastically reduces network mroute state.**
  - **Eliminates *ALL* (S,G) state in the network.**
    - **SPT's between sources to RP eliminated.**
    - **Source traffic flows both up and down Shared Tree.**
  - **Allows Many-to-Any applications to scale.**
    - **Permits virtually an unlimited number of sources.**