Computer Networks : Protocols and Practice

Part 7 : Multicast

Olivier Bonaventure
http://inl.info.ucl.ac.be/

The motivation for non-unicast

- Principles of unicast transmission
  - Source is identified by a unique address
  - Destination is identified by a unique address
  - Packets are sent from Source towards Destination
- Motivation for a different mode than unicast
  - How to transmit same info to set of receivers ?
    - Source-based solution
      - sender sends N copies of this information, one for each receiver
      - easy to implement, but consumes a lot of bandwidth
    - Network-based solution
      - sender sends a single copy of the information
      - network (routers) takes care of distributing this information efficiently towards all interested receivers
      - more efficient, but requires specific protocols and mechanisms

General references on multicast include

B. Williamson, Developing IP Multicast Networks, Cisco Press, 2000

A presentation with the same information as in this book is available from ftp://ftpeng.cisco.com/ipmulticast/index.html

J. Crowcroft, M. Hanldey, Internetworking multimedia, UCL Press, 1998
http://www.cs.ucl.ac.uk/staff/jon/mmbook/book/book.html

A good list of references may be found in :
http://www.switch.ch/network/ipmcast/references.html

See also
M. Handley, J. Crowcroft, Internet multicast today, Internet Protocol Journal, Vol2, N4, Dec. 1999,
http://www.cisco.com/en/US/about/ac123/ac147/ac174/ac198/about_cisco_ipj_archive_article09186a00800c851e.html
For tomorrow's multicast, see :
Ian Brown, Jon Crowcroft, Mark Handley, Brad Cain, Internet Multicast Tomorrow, Internet Protocol Journal,

http://www.cisco.com/en/US/about/ac123/ac147/ac175/about_cisco_ipj_archive09186a0080132b7c.html
A short discussion of IP multicast may be found in :
J. Kurose, K. Ross, Computer Networking : a top-down approach featuring the Internet, 2nd Edition, Addison Wesley, 2002
R. Perlman, Interconnections, 2nd Edition, Addison Wesley

Dimensions to consider

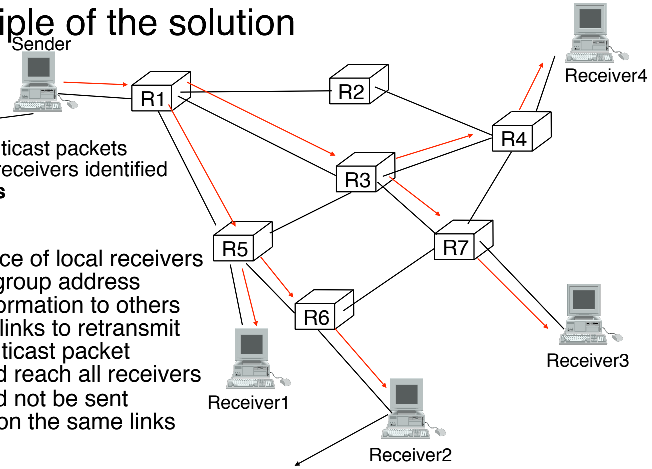
- Senders
  - Number of senders
    - How many endsystems will send packets
  - Localization of the senders
    - Where are these senders located ?
  - Volatility of the senders
    - How often does the set of senders change ?
- Receivers
  - Number of receivers
    - How many endsystems will receive packets ?
  - Localization of the receivers
    - Where are all the receivers for a given transmission ?
  - Volatility of the senders
    - How often does the set of receivers change ?

Dimensions to consider (2)

- Characteristics of the data traffic
  - Traffic volume
    - bits, kilobits, megabits per second
  - Traffic burstiness
    - one packet from time to time
    - a continuous flow of packet
    - Large bursts of packets separated by long idle times
- Lifetime of a multicast transmission
  - one second, a few minutes, several hours
- Num. of concurrent multicast transmissions
  - a few, a few tens, a few thousands, millions
- Multicast level
  - LAN-level, network level, application level

Network-level multicast

- Principle of the solution

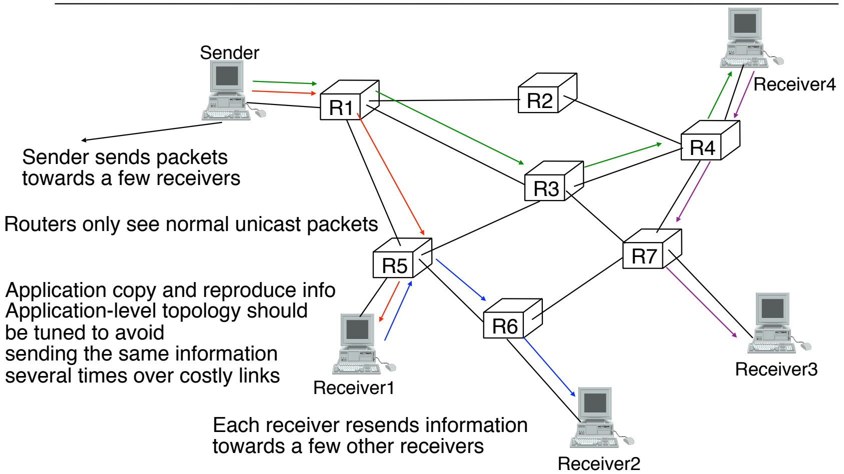


Sender sends multicast packets towards group of receivers identified by **group address**

Routers need to :
- know the existence of local receivers for each defined group address
- distribute this information to others
- decide on which links to retransmit each received multicast packet
  - packets should reach all receivers
  - packets should not be sent several times on the same links

Receivers must indicate their interest in receiving packets destined to specific group addresses

Application-level multicast



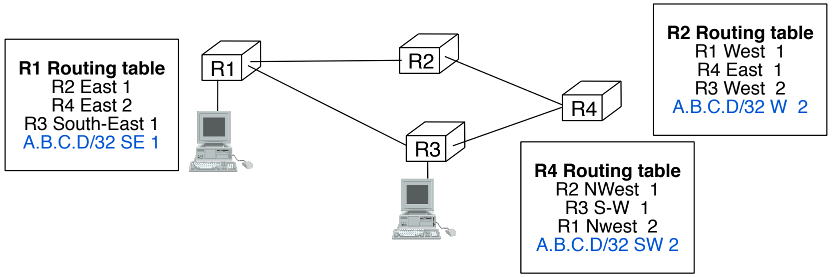
- Examples : Usenet news, IRC, some P2P apps, ...

Sample non-unicast applications

- Localization applications
  - a server announces its availability
  - a client tries to locate the closest server
- Distribution of real-time information
  - Multiparty conferences
  - Distributed games
  - Real-time Monitoring
    - Temperature
    - stock prices
- Reliable distribution of information
  - software updates
  - news information

Localization of services

- How to find the closest server for service Y ?
  - Network-based anycast service
- Network-level solution
  - IP address A.B.C.D reserved for service Y
  - each server uses A.B.C.D as logical address
  - routers distribute route towards A.B.C.D/32



The anycast service is defined in :

Patridge, C., Mendez, T. and W. Milliken, "Host Anycasting Service", RFC 1546, November 1993.

This type of service is used notably for some DNS servers. For example, several ISPs place several copies of their DNS servers at various locations and each of those servers uses the same IP address. This allows the ISP customers to utilize the same IP addresses in their configuration independently of their location.

Another example is the utilization of anycast for one of the root servers, see :
http://www.as112.net/

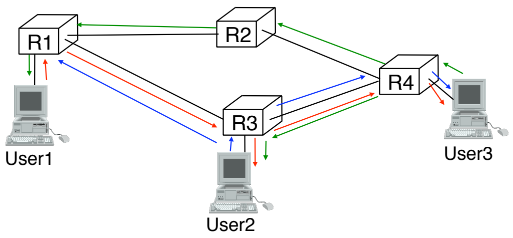
http://www.ripe.net/ripe/draft-documents/k-root-anycasting.html

Distribution of realtime information

- Example : N-way videoconference with RTP
- Network-level solution
  - One multicast group for audio
  - One multicast group for video
  - One multicast group for shared whiteboard
  - All participants send data to everybody

Number of senders : a few
Number of receivers : a few
Data traffic
- continuous for video
- may be bursty for audio
- bursty for whiteboard

Lifetime : long
Volaticity : stable
Concurrent sessions : can be large



CNPP/2008.7.

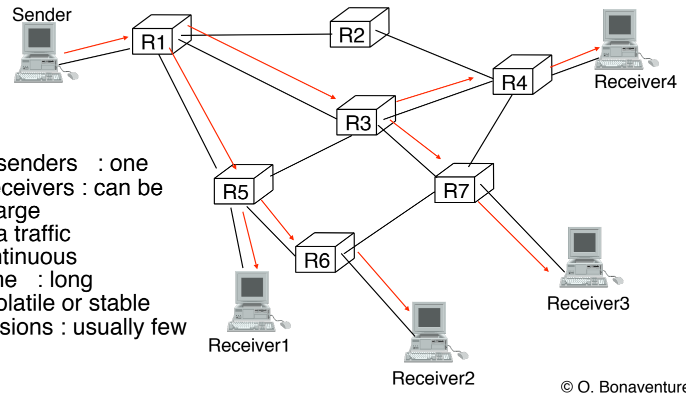
A distributed game would be different

For example realtime applications supporting multicast, see :

http://www-mice.cs.ucl.ac.uk/multimedia/software/

Distribution of realtime information (2)

- Live audio/video
  - one sender transmits streaming audio/video to large number of receivers



Number of senders : one
Number of receivers : can be large
Data traffic continuous
Lifetime : long
Volatility : volatile or stable
Concurrent sessions : usually few

Reliable distribution of information

- Information is delivered reliably and efficiently to all destinations

- Application level solution
  - Usenet News, Mailing lists, software updates
  - Application relays placed at key network locations
    - is widely deployed today

- Network-level solution
  - Information distributed by multicast packets
  - Issue
    - Multicast transport protocol are required for reliability
      - Those protocols are deployed in some enterprise networks, but not widely available in the global Internet

Concerning the application-level solutions, several researchers are currently working on the definition of generic multicast overlays that could be used by applications.

Several reliable multicast protocols have been proposed during th e last years.

RFC 2887: The Reliable Multicast Design Space for Bulk Data Transfer
M. Handley, S. Floyd, B. Whetten, R. Kermode, L. Vicisano, M. Luby, August 2000

RFC 3208: PGM Reliable Transport Protocol Specification
T. Speakman, J. Crowcroft, J. Gemmell, D. Farinacci, S. Lin, D. Leshchiner, M. Luby, T. Montgomery, L. Rizzo, A. Tweedly, N. Bhaskar, R. Edmonstone, R. Sumanasekera, L. Vicisano, December 2001

See also :
http://www.ietf.org/html.charters/rmt-charter.html

The benefits and costs of network-level multicast

- Benefits
  - Multicast allows a more efficient utilization of the available bandwidth
  - no unnecessary packet duplication
- But, to support IP multicast, routers must
  - implement special multicast routing protocols
    - implementation and configuration cost
  - maintain some state information for multicast
    - memory consumption
  - IP multicast packets require special treatment
    - CPU consumption

→ The bandwidth saved should be compared to the router costs!

The Any-Source IP Multicast (ASM) service model

- Principles
  - A fraction of the IP address space is reserved exclusively for IP multicast
    - 224.0.0.0 -> 239.255.255.255
  - A multicast address identifies a group of receivers
    - receivers can join and leave a group at any time
    - members of a group can be located anywhere
  - Any source host can transmit IP packets towards any IP multicast address
    - a sender does not need to be a member of a group
  - Routers take care of the efficient transmission of multicast packets towards their destinations
    - Packet forwarding based on destination address

The ASM model was defined in :

S. Deering, Host extensions for IP multicasting. RFC1112, 1-1989.

IP Multicast addresses

- Initial Allocation of IP addresses
  - Well-known addresses
    - 224.0.0.1 : all hosts
    - 224.0.0.2 : all multicast routers
    - 224.0.0.5 : all OSPF routers
    - Other addresses like 224.0.1.11
      - Reserved addresses to broadcast IETF meetings
  - Dynamic multicast address assignment
    - `sdr` application listens to special multicast group
      - this group is used to broadcast announcements of the creation of new IP multicast groups
    - When an application needs to create a multicast group, it requests `sdr` to select one address at random
      - `sdr` checks that it did not heard anything about this IP address
    - Suitable for research experiments, ...
      more difficult in a commercial Internet

The list of well-known IP multicast addresses may be found in :
http://www.iana.org/assignments/multicast-addresses

IP Multicast addresses (2)

- Multicast Address allocation
  - Evolution is to allocate addresses hierarchically
    - Administratively scoped addresses
      - 239.0.0.0 - 239.255.255.255
      - reserved for utilization inside private domains
      - Should not appear on the global Internet
    - Static group address assignment (GLOP)
      - 233.0.0.0 - 233.255.255.255
      - each Autonomous System receives 256 addresses in this range and can allocate those addresses to its users
    - MASC : Multicast Address Set-Claim
      - Dynamic Method to allocate blocks of IP multicast addresses to ISPs under development
      - Multicast Addresses allocated for some time only
      - far from being deployed

The assignment of 233.0.0.0/8 is defined in
RFC 3180: GLOP Addressing in 233/8 , D. Meyer, P. Lothberg, September 2001

Work on the allocation of IP multicast addresses is being performed in :
http://www.ietf.org/html.charters/malloc-charter.html

Evaluation of the ASM model

- Advantages
  - Any source can send packets to any group
  - Any receiver may join any group

- Drawbacks
  - How do we allocate IP multicast addresses ?
    - Random allocation is not really a long term solution
    - Risk of collision is severe
  - Lack of access control
    - Any source can send packets to any group
      - Consider a videoconference between professors where students could easily send packets to interrupt the conference...
    - Any receiver may join any group
      - Encryption can be used to protect packet contents, but ...
  - Inefficient handling of well-known sources
    - CNN uses 226.0.0.10 and IraqiTV selects same address...

The Source-Specific IP Multicast (SSM) service model

- Objective
  - Provide a deployable one-to-many service
- Principles
  - SSM multicast address = channel from a sender
    - A single source can use as many channels as there are SSM addresses
      - Several senders can use same channel id without problems
    - A fraction of the IP address space is reserved exclusively for SSM multicast addresses
      - 232.0.0.0 -> 232.255.255.255
  - Receivers subscribe to channels from senders
    - Receiver knows from which sender packets will arrive
  - Routers retransmit multicast packets from the sender to all channel subscribers
    - Packet forwarding based on both packet source and destination addresses (Channel)

Source-Specific Multicast for IP , H. Holbrook, B. Cain, RFC4602, 2006

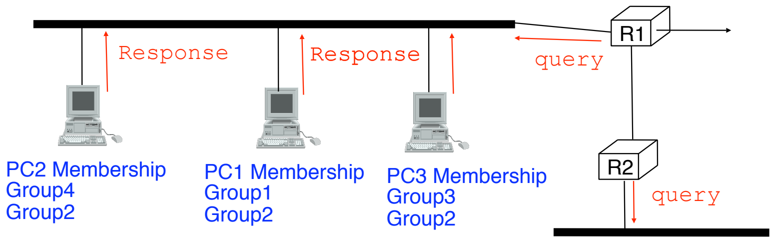
An Overview of Source-Specific Multicast(SSM) Deployment , S. Bhattacharyya et al., RFC3569, 2003

Source-Specific Protocol Independent Multicast in 232/8, D. Meyer, R. Rockell, G. Shepherd, RFC4608, 2006

Issues for network-level multicast

1. How to localize all the members of each multicast group ?
   - Internet Group Management Protocol
2. How to efficiently distribute multicast packets from one or more sender(s) to a group of receivers ?
   - A loop-free graph must be built to connect the sender(s) to all the interested receivers
     - source tree
     - unidirectional shared tree
     - bi-directional shared tree

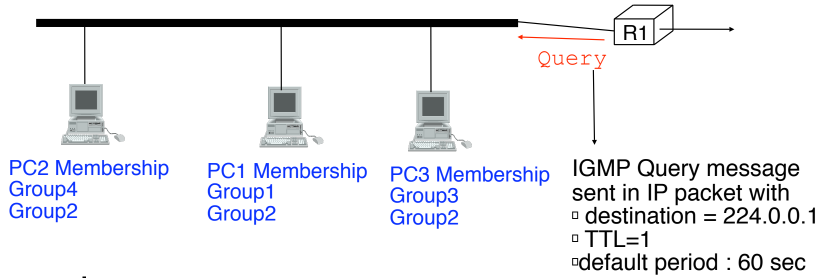
Localization of receivers



- Principle
  - Routers query their attached LANs to identify active groups
  - Hosts respond to router queries
  - Routers redistribute the localization of the receivers towards other routers

IGMPv1

- IGMP : Internet Group Management Protocol



- Issues
  - Queries must be received by all IP multicast hosts
  - Responses should not overload router

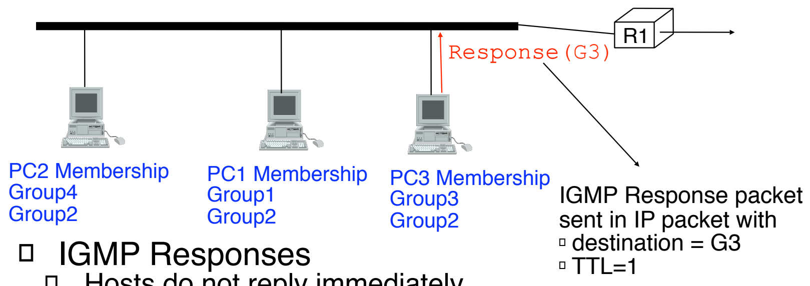
CNPP/2008.7.

All IP Multicast enabled hosts should always listen to IP address 224.0.0.1

IGMPv1 is defined in :

RFC 1112 Deering, S., "Host Extensions for IP Multicasting", STD 5, RFC 1112, August 1989.

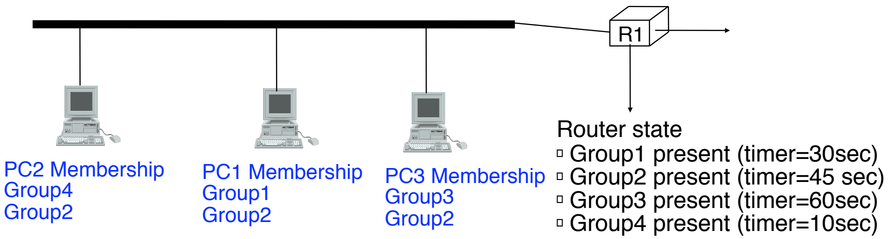
IGMPv1 (2)



- IGMP Responses
  - Hosts do not reply immediately
    - Random delay (0 - 10 seconds)
  - One IGMP response per group
  - Response sent to multicast address of the group
    - other members of the group will suppress their response
- Consequences
  - a single response packet for each group on each LAN
  - router must listen to all multicast groups to receive IGMPv1 responses

As an optimization, it is possible for a host to send an IGMP response message without receiving a Query when they join a group. This allows the router to be immediately informed about new group members.

IGMPv1 (3)



- How to leave a group ?
- Soft-state solution
  - timer associated to router state for each group
  - if timer expires before the reception of a group response, group is considered to be absent on the LAN
    - timer should be longer than interval between IGMP queries

IGMP v2

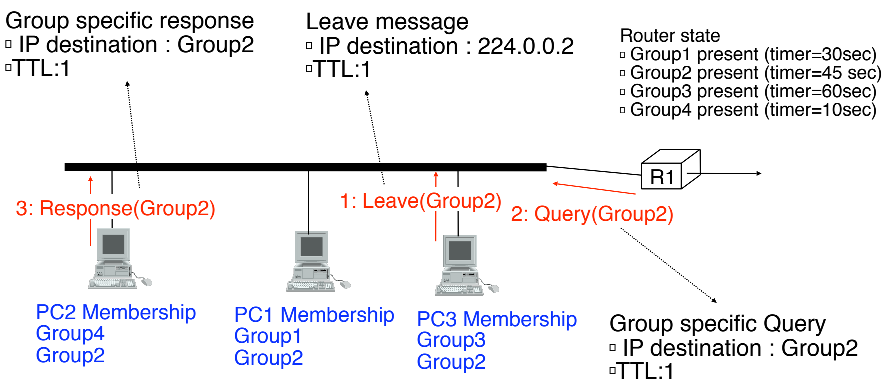
- Main additions in IGMPv2
  - Explicit Leave Group message
    - a host may send an explicit leave message to indicate that it is no more interested by a given group
      - leave message is sent to All-Multicast-Routers group (224.0.0.2)
  - Group specific Query
    - A router may send a specific query to a single group to verify whether there are still members of the group
      - A group specific query is sent to the group address

IGMPv2 is defined in :

RFC 2236: Internet Group Management Protocol, Version 2 , W. Fenner, November 1997

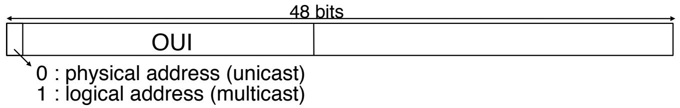
IGMPv2 (2)

- Example leave



LAN-level multicast

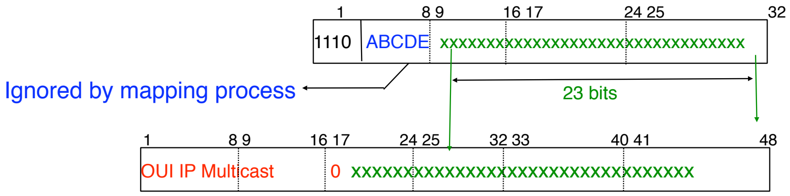
- Principle
  - Two types of destination Ethernet addresses
    - Physical addresses
      - identifies one Ethernet adapter
    - Logical addresses
      - identifies a logical group of Ethernet destinations



- Transmission of multicast frame
  - sender transmits frame with multicast destination addr.
- Reception of multicast frames
  - Ethernet adapters can be configured to capture frames whose destination address is
    - Their unicast address
    - One of a set of multicast addresses

IP Multicast over Ethernet

- IP Multicast over Ethernet
  - Algorithmic mapping of IP multicast addresses
  - special OUI reserved for IP multicast [0x00005e]



- The low order 23 bits of the IP multicast address are copied in the low order 23 bits of the Ethernet address
- 32 IP multicast address are mapped on the same Ethernet logical address

CNPP/2008.7.

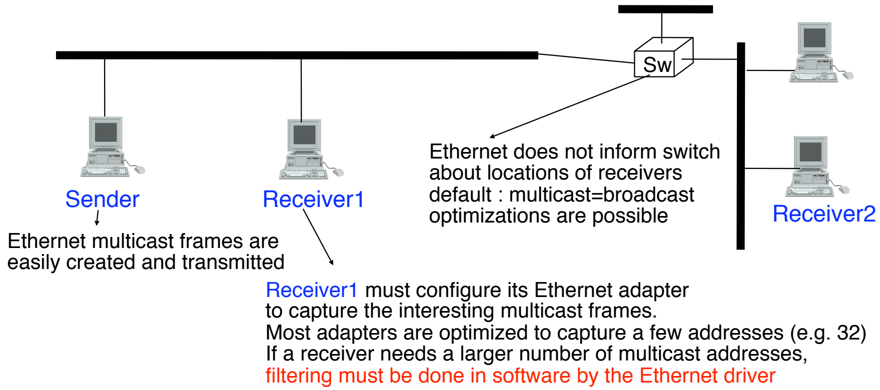
Issue :

32 IP addresses are mapped on the same Ethernet address. For example, Ethernet address 01-00-5e-0a-00-0a is used to send IP packets destined to the following IP multicast addresses :

224.10.0.1, 225.10.0.1, 226.10.0.1, 227.10.0.1,
228.10.0.1, 229.10.0.1, 230.10.0.1, 231.10.0.1,
232.10.0.1, 233.10.0.1, 234.10.0.1, 235.10.0.1,
236.10.0.1, 237.10.0.1, 238.10.0.1, 239.10.0.1,
224.138.0.1, 225.138.0.1, 226.138.0.1, 227.138.0.1,
228.138.0.1, 229.138.0.1, 230.138.0.1, 231.138.0.1,
232.138.0.1, 233.138.0.1, 234.138.0.1, 235.138.0.1,
236.138.0.1, 237.138.0.1, 238.138.0.1, 239.138.0.1,

The hidden cost of LAN multicast

- Multicast on a single Ethernet
  - group of receivers identified by special address
  - receivers join group by configuring their adapter

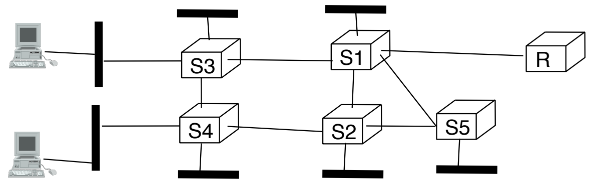


CNPP/2008.7.

Some switches are able to look at the contents of IP packets to determine whether they can provide some information about the location of receivers, but this is outside the scope of this presentation.

IGMPv3

- Complete rewrite of IGMP
- Drawbacks of IGMPv1/2
  - Difficult to efficiently support IP multicast in switched LANs

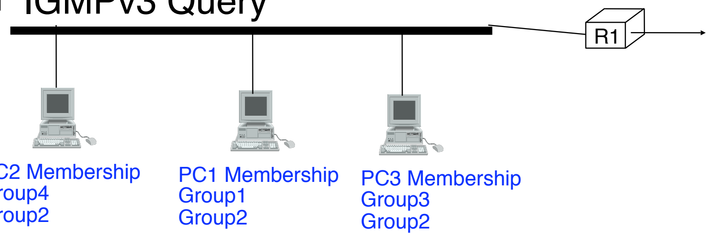


    - Routers must listen to all IP multicast addresses to receive IGMP messages
    - No support for SSM service model

Some switches are able to snoop IGMP messages and to determine automatically the location of IP multicast receivers. However, this implies that those switches need to inspect the content of the receivers multicast frames carrying IP packets.

IGMPv3 (2)

- IGMPv3 Query



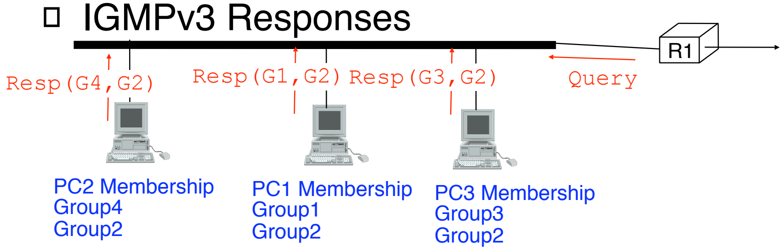
- Three types of IGM Queries
  - General Query
    - similar to IGMPv1/2, Query sent to IP address 224.0.0.1
  - Group Specific Query
    - similar to Group Specific queries in IGMPv2, sent to group addr.
  - Group-Source Specific Query
    - router asks responses from specific sources and groups
    - query sent to group address includes list of sources

IGMPv3 is defined in :

RFC 3376: Internet Group Management Protocol, Version 3 , B. Cain, S. Deering, I. Kouvelas, B. Fenner, A . Thyagarajan, October 2002

IGMPv3 (3)

- IGMPv3 Responses

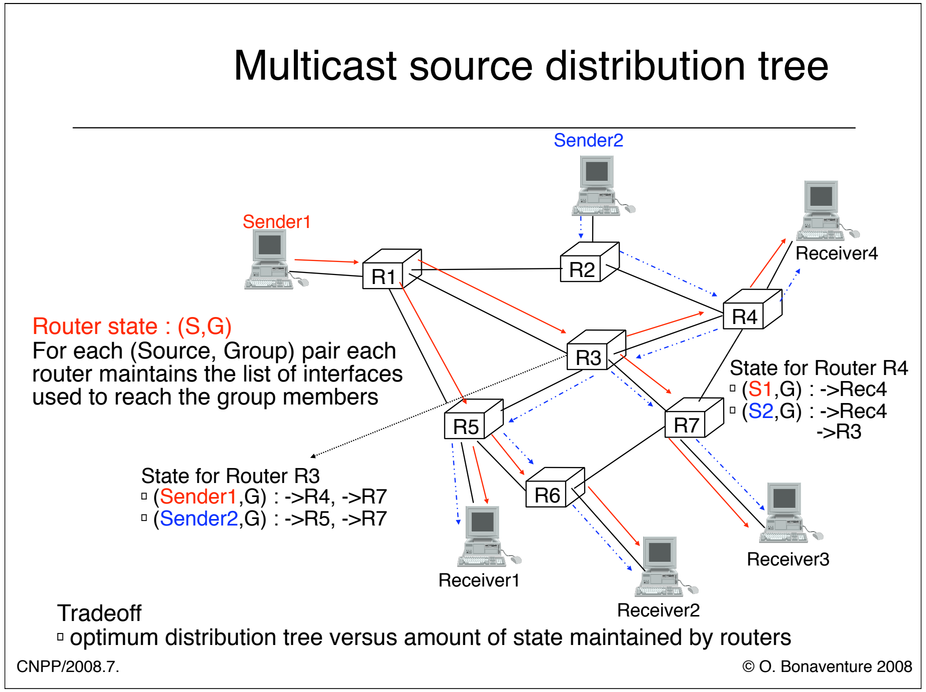


- each host sends an IGMPv3 Response
  - response contains the entire list of groups for the host
  - for each group, the host may optionally indicate that its interest is only for some specified senders (for SSM)
  - response is sent in packets with destination=224.0.0.22
  - a random timer is used before sending response to avoid overloading the network or the routers

Issues for network-level multicast

1. How to localize all the members of each multicast group ?
   - Internet Group Management Protocol

2. How to efficiently distribute multicast packets from one or more sender(s) to a group of receivers ?
   - A loop-free graph must be built to connect the sender(s) to all the interested receivers
     - source tree
     - unidirectional shared tree
     - bi-directional shared tree



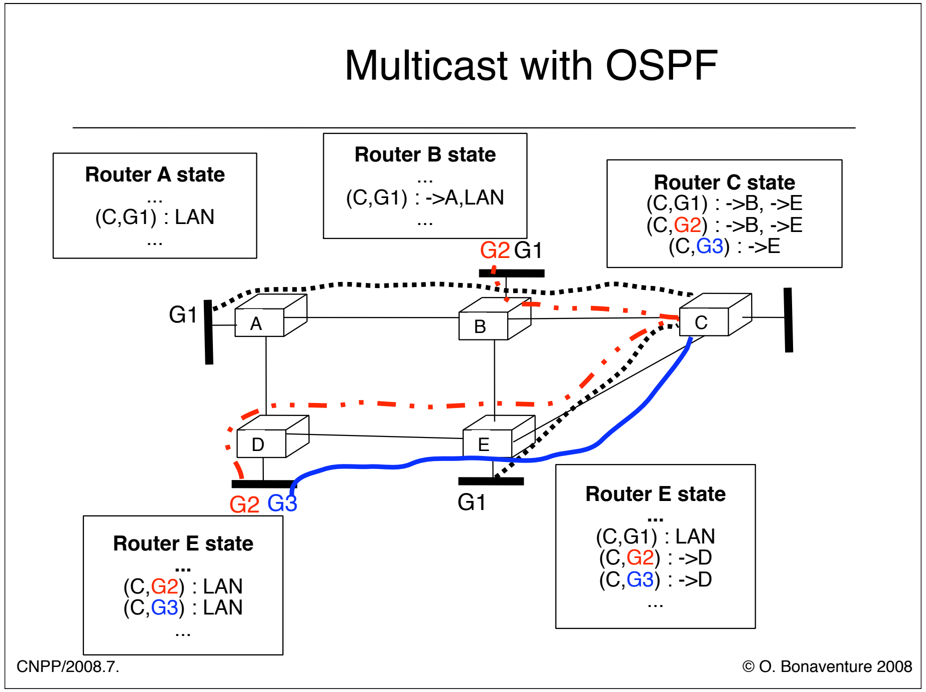
This type of source distribution tree is well suited to support the SSM service model, but can also be used to support the ASM service model.

Building the source tree

- How to build the distribution tree ?
  - Link-state solution
    - link-state routing protocol used to distribute membership
    - Each router knows the exact location of group members
    - Each router can easily compute source trees
  - Broadcast solution
    - Assume initially that group members are everywhere
    - The first packet towards a group is sent everywhere
    - If a router receives unwanted traffic, it will ask its upstream routers to stop transmitting this traffic
    - Distribution tree is built progressively

Building the source tree
Link-state solution

- Principle
  - Routers collect membership information with IGMP
  - Group membership is distributed by special type of LSPs flooded by link-state routing protocol
  - By processing those LSPs, each router knows
    - The entire network topology
    - The localization of the receivers from each multicast group
  - Based on the link state database, each router computes (S,G) spanning trees for each Source, Group pair by using the Dijkstra algorithm



The multicast extensions to OSPF are defined in

RFC1584 Multicast Extensions to OSPF. J. Moy. March 1994

RFC1585 MOSPF: Analysis and Experience. J. Moy. March 1994

Multicast with OSPF (2)

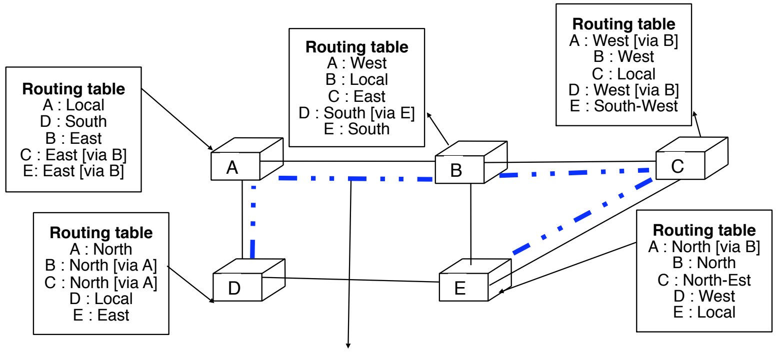
- Advantages of OSPF for IP Multicast
  - routers have complete information on
    - network topology
    - localization of group members
- Drawbacks of OSPF for IP Multicast
  - Large number of LSPs if there are many groups or many group members in the network
    - these LSPs may consume a lot of memory inside all routers whether or not there is multicast traffic
    - group membership LSPs compete with normal LSPs
  - Amount of router state depends on number of (S,G) pairs
  - Computation of (S,G) spanning trees can be costly

Building the source tree

- How to build the distribution tree ?
  - Link-state solution
    - link-state routing protocol used to distribute membership
    - Each router knows the exact location of group members
    - Each router can easily compute source trees
  - Broadcast solution
    - Assume initially that group members are everywhere
    - The first packet towards a group is sent everywhere
    - If a router receives unwanted traffic, it will ask its upstream routers to stop transmitting this traffic
    - Distribution tree is built progressively

Broadcasting IP packets

- How to broadcast packets in a large network ?



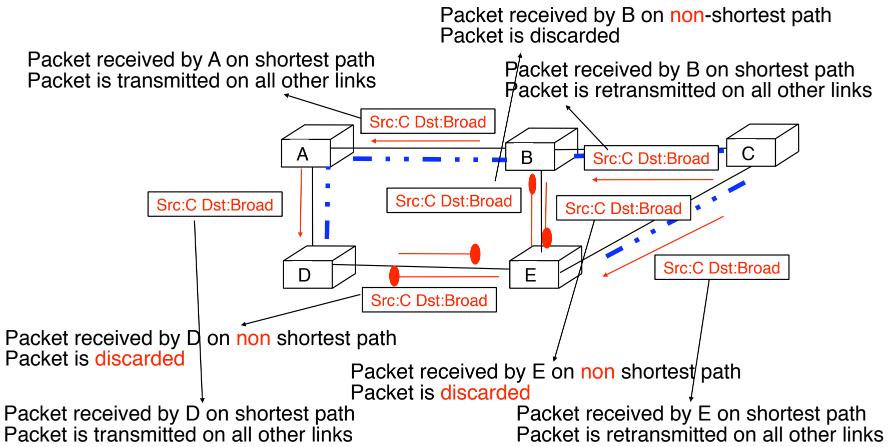
The graph containing the links from all shortest paths originating at router C creates a spanning tree rooted at C

Reverse Path Forwarding

- Principle of the solution
  - Forwarding of broadcast packets depends on
    - Multicast Destination IP address
    - Unicast Source IP address
  - Simple router forwarding algorithm

```
Broadcast packet from Source S received on interface in:
// consult routing table
// to find shortest route towards S
if (interface in == shortest_route(to S))
 for(i=0; i++; i<N)
     {
          if (i!=in) Send_packet(interface i);
     }
else
     Discard(packet);
}
```

RPF example

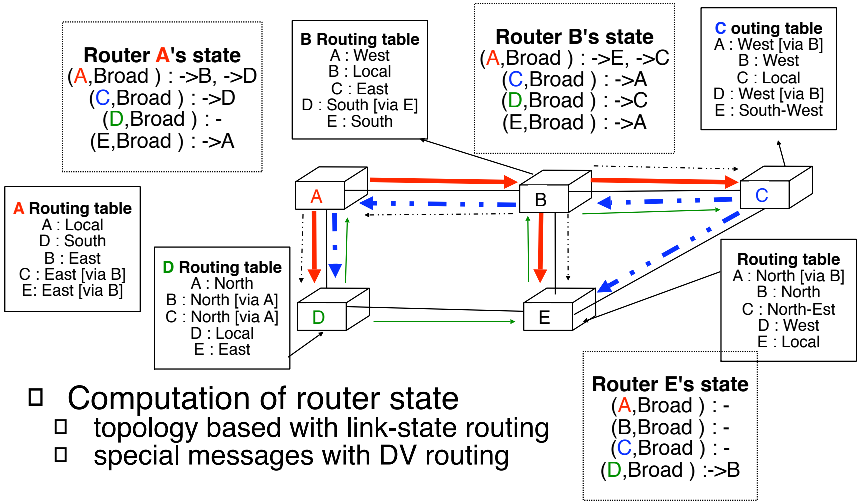


- Propagation of the broadcast packet sent by router C

Improving RPF

- How to avoid unnecessary transmissions ?
  - Sender-based solution
    - for each source, router computes list of outgoing interfaces
    - broadcast packets are only forwarded on these interfaces
    - list of outgoing interfaces is computed by including interfaces towards neighbors on shortest path from source
      - router must know shortest path chosen by neighbor router
        - possible with OSPF, requires specific messages with RIP
      - CPU intensive
  - Receiver-driven solution

Improving RPF
Sender-based solution



- Computation of router state
  - topology based with link-state routing
  - special messages with DV routing

Spanning tree rooted at router C
Spanning tree rooted at router A
Spanning tree rooted at router B
Spanning tree rooted at router D

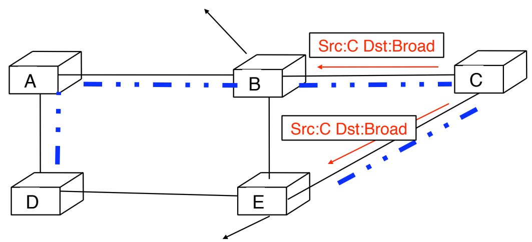
Improving RPF

- How to avoid unnecessary transmissions ?
  - Sender-based solution
  - Receiver-driven solution
    - for each source, router maintains list of outgoing interfaces
    - broadcast packets are only forwarded on these interfaces
    - Computation of list of outgoing interfaces
      - Initially list of outgoing interfaces contains all interfaces
      - when a router receives a packet over a non-shortest path, it sends a special message to the upstream router
      - the upstream router will then update its list of interfaces
    - Less CPU intensive than sender-based solution

Improving RPF
Receiver-based solution

Packet received by B on shortest path
Create state for router B:
- **(C,Broad) : ->A, ->E**

Packet is retransmitted on all links in outgoing list



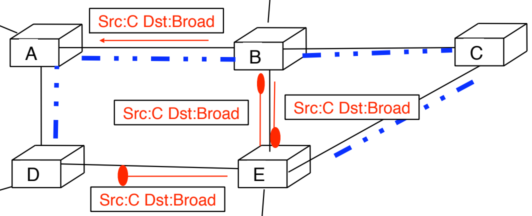
Packet received by E on shortest path
Create state for router E:
- **(C,Broad) : ->B, ->D**

Packet is retransmitted on all links in outgoing list

Improving RPF
Receiver-based solution (2)

Packet received by A on shortest path
Create state for router A
- **(C,Broad) : ->D**
Packet is transmitted on all other links

Packet received by B from non shortest path
Packet is discarded
Send Prune message (C,Broad) to router E

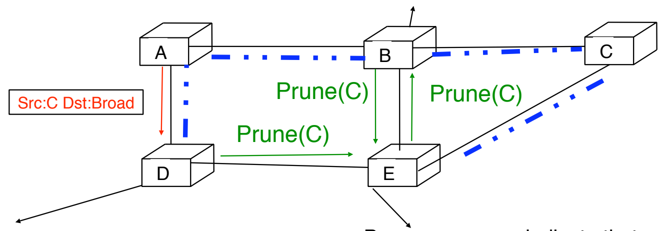


Packet received by D on non shortest path
No state created
Packet is discarded
Send Prune message (C,Broad) to router E

Packet received by E on non shortest path
State for router E unchanged:
- **(C,Broad) : ->B, ->D**
Packet is discarded
Send Prune message (C,Broad) to router B

Improving RPF
Receiver-based solution (3)

Prune message indicates that broadcast packets from C should not be retransmitted towards E
New state for router :
- **(C,Broad) : ->A**



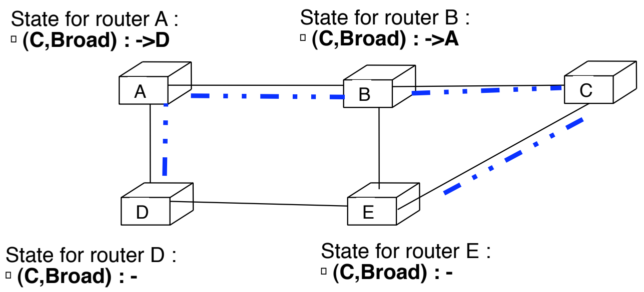
Packet received by D on shortest path
Create state for router D :
- **(C,Broad) : ->E**

Packet is transmitted on all other links

Prune messages indicate that broadcast packets from C should not be retransmitted towards B and D
New state for router E :
- **(C,Broad) : -**

Improving RPF
Receiver-based solution (4)

- Final router states



  - next broadcast packets from C will only flow on shortest path tree
  - state is created dynamically when multicast packets are received on shortest path from source
  - a timer is associated with each state
    - state is removed when timer expires

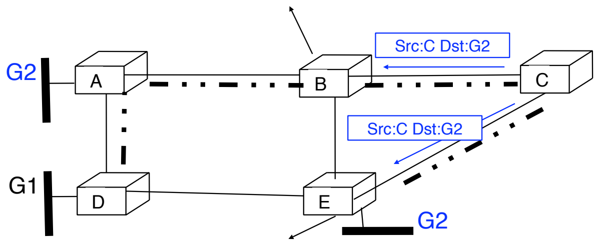
Using RPF for Multicast

- Principles
  - Multicast packets are initially transmitted by routers as if they were broadcast packets
  - Routers create (S,G) state upon reception of first multicast packets from S to G
  - Two types of Prune messages are sent
    - Source specific Prune (S,G) message if packet is received on non-shortest path
    - Group Prune message if there is no group member downstream of this router
  - (S,G) state is updated by reception of Prune msgs

RPF for Multicast : Example

Packet received by B on shortest path
(C,G2) state is set to all router interfaces except incoming interface
New state for router B:
- **(C,G2) : ->A, ->E**

Packet is retransmitted on all branches of source tree



Packet received by E on shortest path
(C,G2) state is set to all router interfaces except incoming interface
New state for router B:
- **(C,G2) : ->B, ->D, LAN**

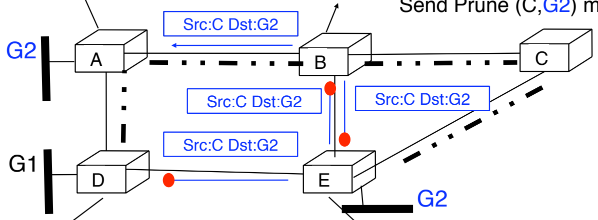
Packet is retransmitted on all branches of source tree

RPF for Multicast : Example (2)

Packet received by A on shortest path
(C,G2) is set to all interfaces except incoming interface
New state for router A:
- **(C,G2) : ->D, LAN**

Packet is retransmitted on all branches of source tree

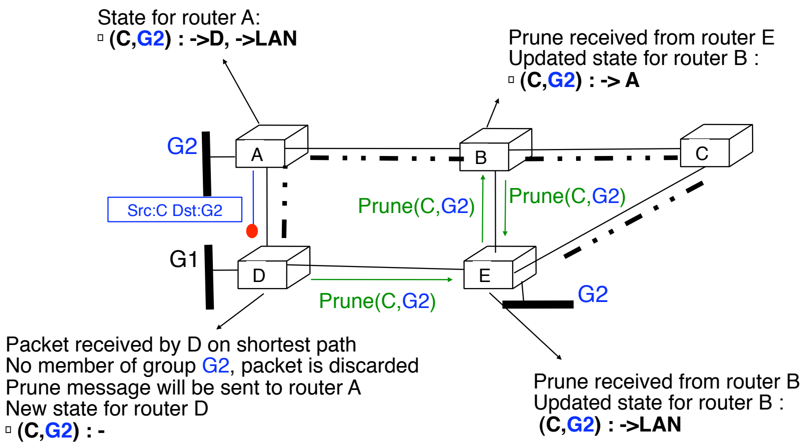
Packet received by B on non shortest path
Packet is discarded
No update to state
Send Prune (C,G2) msg to E



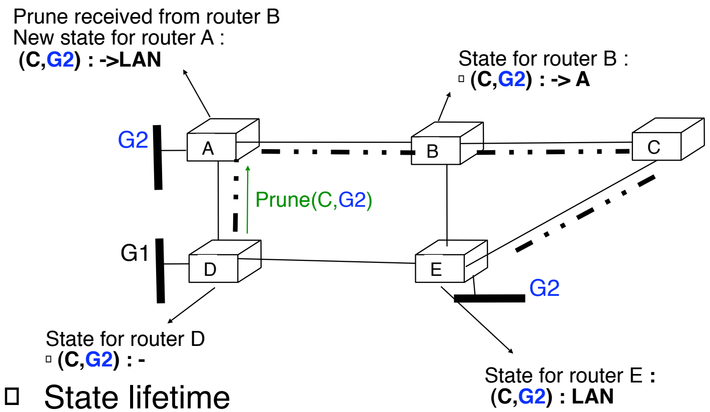
Packet received by D on non shortest path
Packet is discarded
No state is created in router D
Send Prune (C,G2) msg to E

Packet received by E on non shortest path
Packet is discarded
No update to state
Send Prune (C,G2) to B

RPF for Multicast : Example (3)



RPF for Multicast : Example (4)



- State lifetime
  - periodically state info is removed and flooding restarts
  - flood and prune operation
- Graft msg can be used by D to join the tree later

Example protocols

- DVMRP
  - first multicast routing protocol deployed
  - special unicast distance vector routing protocol
    - sender-based optimization for RPF
    - special poison reverse message to inform upstream router that it is the best path for a given source
  - timer associated with router state
  - periodic flood and prune
- PIM-Dense Mode
  - periodic flood and prune
  - does not rely on a special unicast routing protocol
    - RPF is based on existing unicast routing table
    - receiver-based optimization for RPF

DVMRP is defined in :

RFC1075 Distance Vector Multicast Routing Protocol. D. Waitzman, C. Partridge, S.E. Deering. Nov-01-1988.

DVMRP was used to build the MBONE. The MBONE was an overlay network built by establishing tunnels between routers to distribute multicast packets through networks that do not support multicast.

See e.g.
Michael R. Macedonia and Donald P. Brutzman, MBone Provides Audio and Video Across the Internet , http://www-itg.lbl.gov/mbone/Macedonia.html

PIM Dense mode is defined in :

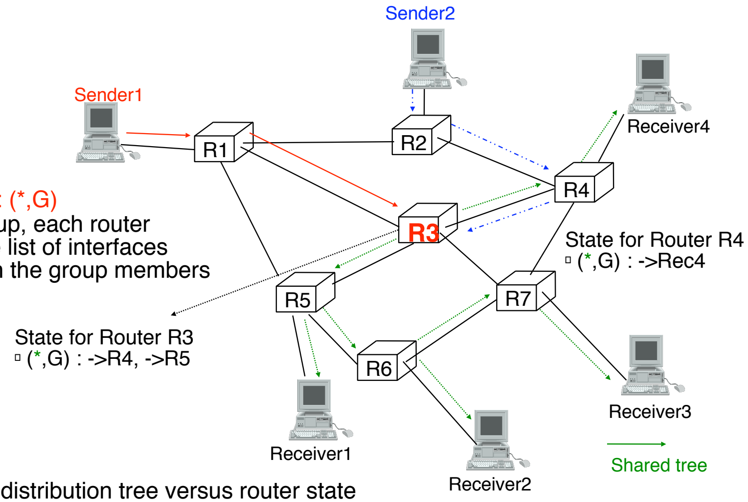
A. Adams, J. Nicholas, W. Siadak, Protocol Independent Multicast - Dense Mode (PIM-DM): Protocol Specification (Revised), Internet draft, draft-ietf-pim-dm-new-v2-01.txt, work in progress, Feb. 2002

Issues for network-level multicast

1. How to localise all the members of each multicast group ?
   - Internet Group Management Protocol
2. How to efficiently distribute multicast packets from one or more sender(s) to a group of receivers ?
   - A loop-free graph must be built to connect the sender(s) to all the interested receivers
     - source tree
     - unidirectional shared tree
     - bi-directional shared tree

Multicast shared distribution tree

Router state : (*,G)
For each group, each router maintains the list of interfaces used to reach the group members



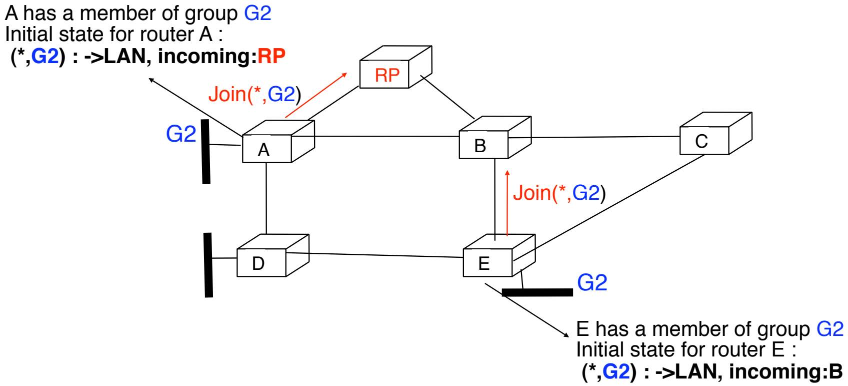
Tradeoff
▫ optimum distribution tree versus router state

In this example, R3 is a special router acting as the root for the multicast shared distribution tree

Using unidirectional shared trees

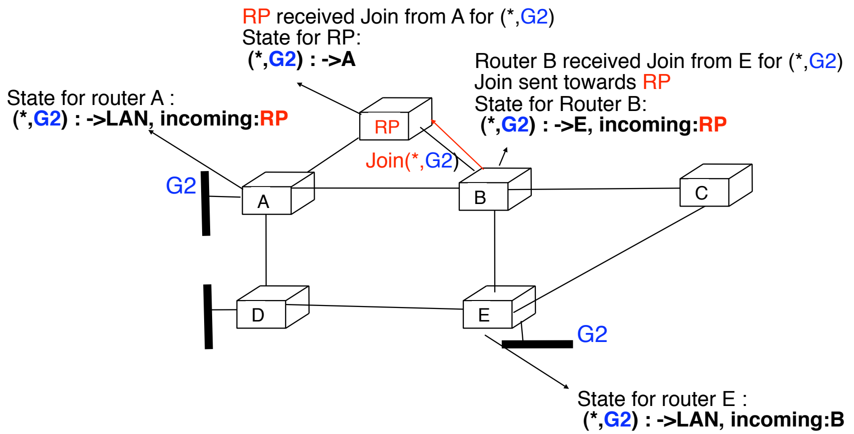
- Principles
  - One router is configured as Rendez-vous Point
  - All other routers know the address of the RP
  - RP is the root of the shared (*,G) tree
  - Multicast packets sent by hosts are intercepted by multicast routers and sent to the RP
    - RP redistributes these packets over the (*,G) tree
  - Receivers join dynamically the shared (*,G) tree

Construction of the shared tree

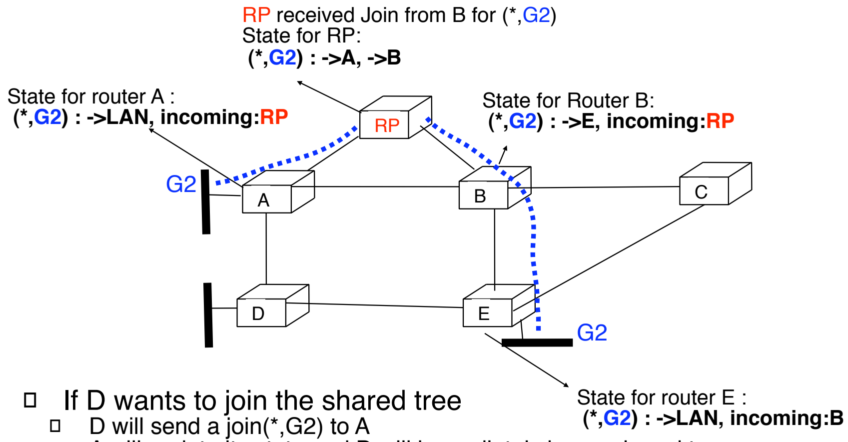


- Routers attached to group members send Join (*,G) msgs towards RP to build shared tree

Construction of the shared tree (2)

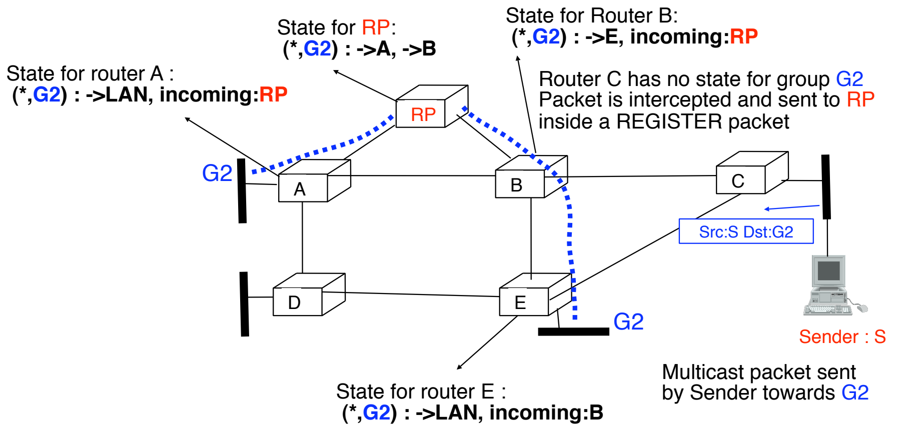


Construction of the shared tree (3)



- If D wants to join the shared tree
  - D will send a join(*,G2) to A
  - A will update its state and D will immediately be on shared tree

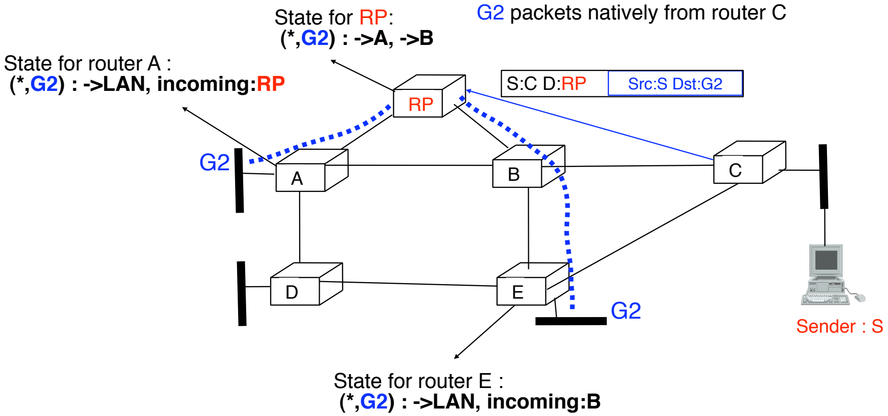
Transmission of multicast packets



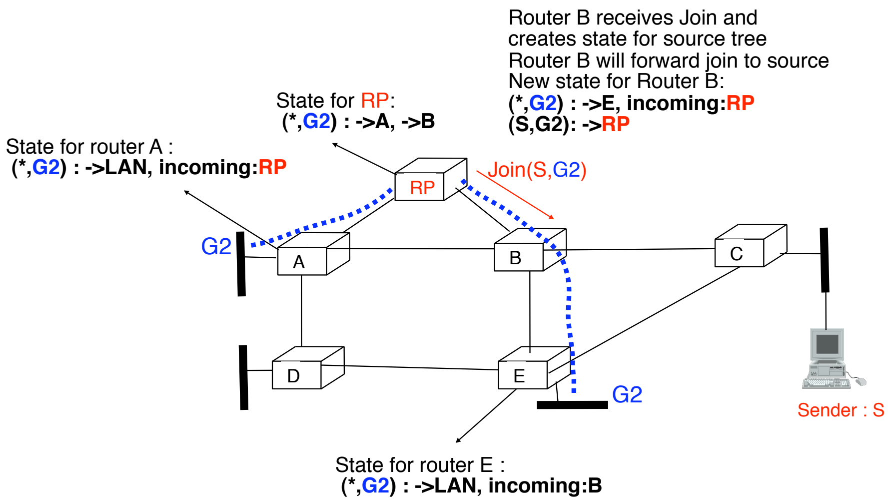
The Register packet used by router C is a kind of tunnel. Router C places the packet received from the sender inside an IP packet whose source is C and whose destination is the RP.

Transmission of multicast packets (2)

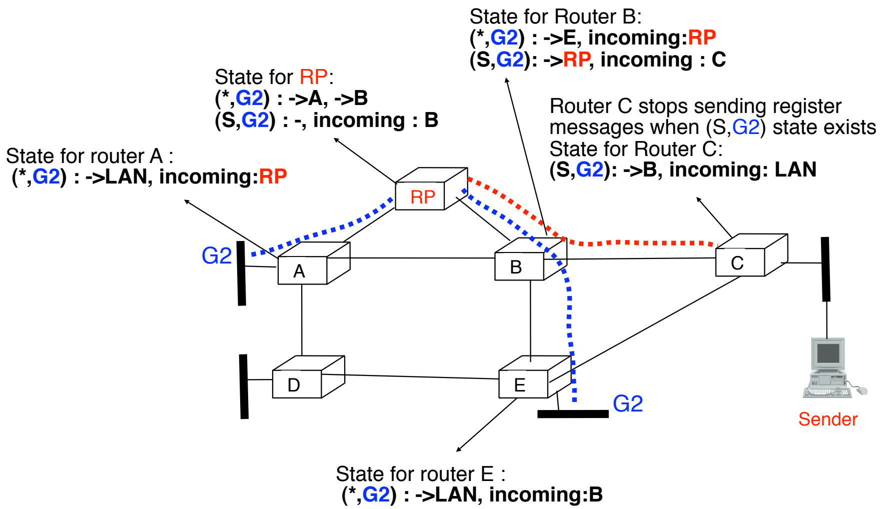
RP checks that there is state for G2
Packet is de-encapsulated and forwarded on (*,G2) shared tree
RP will create a source tree to receive G2 packets natively from router C



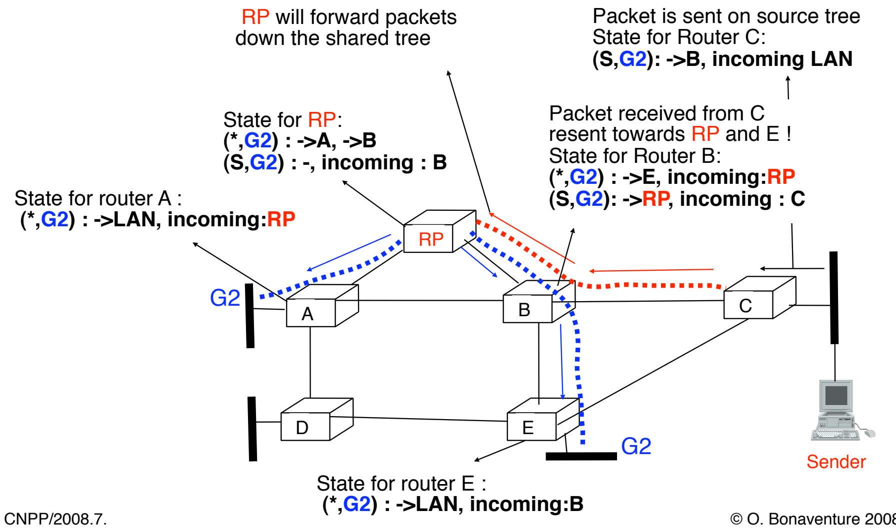
Construction of the source tree



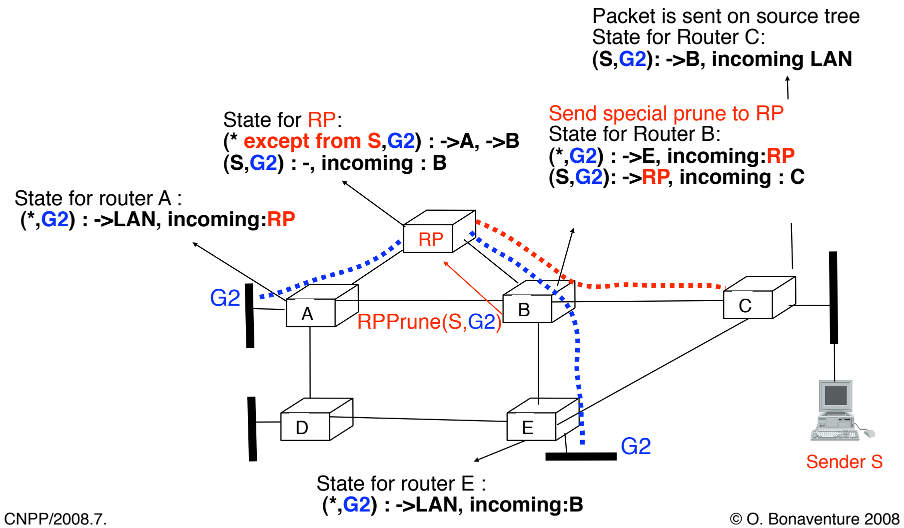
Construction of the source tree (2)



Forwarding of multicast packets



Forwarding of multicast packets (2)

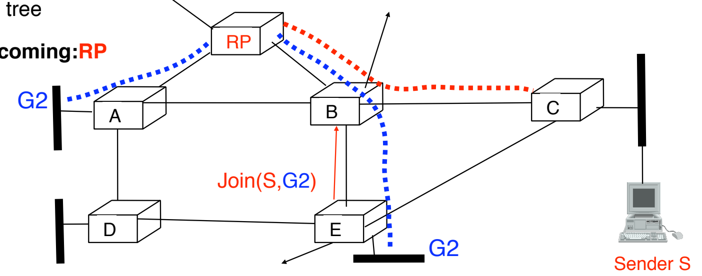


Switching to the source tree

State for RP:
(* except S,G2) : ->A, ->B
(S,G2) : -, incoming : B

When receiving Join(S,G2), B will add E to the (S,G2) source tree
New state for router B
(*,G2) : ->E, incoming:RP
(S,G2): ->RP, ->E, incoming : C

Asks to join source tree
State for router A :
(*,G2) : ->LAN, incoming:RP



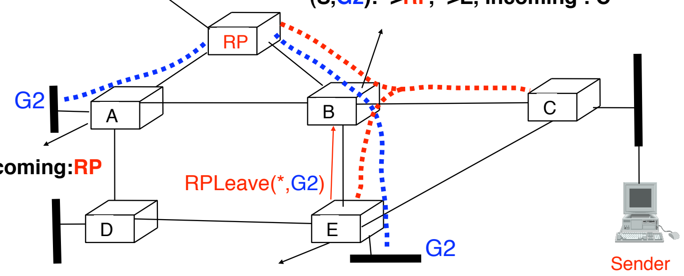
Asks to join the source tree
State for router E :
(*,G2) : ->LAN, incoming:B
(S,G2) : ->LAN, incoming:B

Switching to the source tree (2)

State for RP:
(* except S,G2) : ->A, ->B
(S,G2) : -, incoming : B

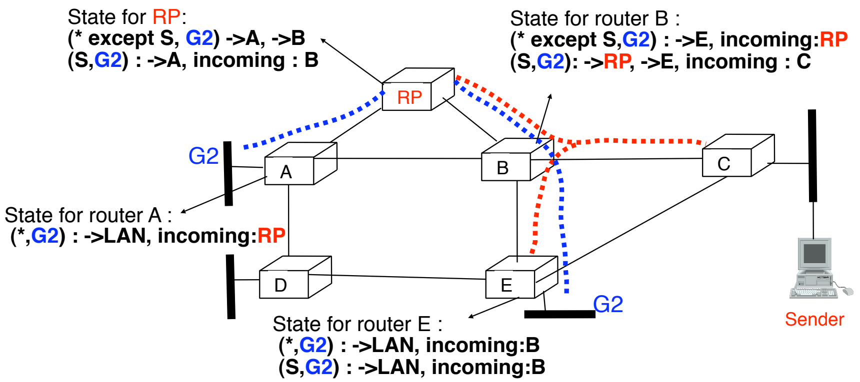
RPLeave will force B to discard packets from source S on (S,G2) tree
B will send RPLeave(S,G2) to RP
State for router B :
(* except S,G2) : ->E, incoming:RP
(S,G2): ->RP, ->E, incoming : C

State for router A :
(*,G2) : ->LAN, incoming:RP



When packets arrive on source tree, router E will ask its upstream to stop sending (S,G2) on shared tree
State for router E :
(*,G2) : ->LAN, incoming:B
(S,G2) : ->LAN, incoming:B

Switching to the source tree (3)



- Examples
  - Multicast packet from Sender S towards G2
  - Multicast packets from sender attached to D towards G2

PIM Sparse Mode

- Widely deployed multicast routing protocol
  - Possible configurations of Rendez-vous Points
    - RP addresses manually configured
    - using anycast to discover Rendez-vous Point
    - RP broadcast announces in dense-mode
  - Possible utilization of unidirectional shared tree
    - For all multicast packets
      - But RP could become overloaded with multicast
    - For first multicast packets sent to each group
      - Those packets are first sent to RP to let receivers hear from sources
      - first-hop routers can join the source-tree if there is a lot of traffic from a given source
      - popular PIM-SM implementation switches to source-tree as soon as traffic is >0

PIM Sparse-Mode is defined in :

RFC2362 Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification. D. Estrin, D. Farinacci, A. Helmy, D. Thaler, S. Deering, M. Handley, V. Jacobson, C. Liu, P. Sharma, L. Wei. June 1998.

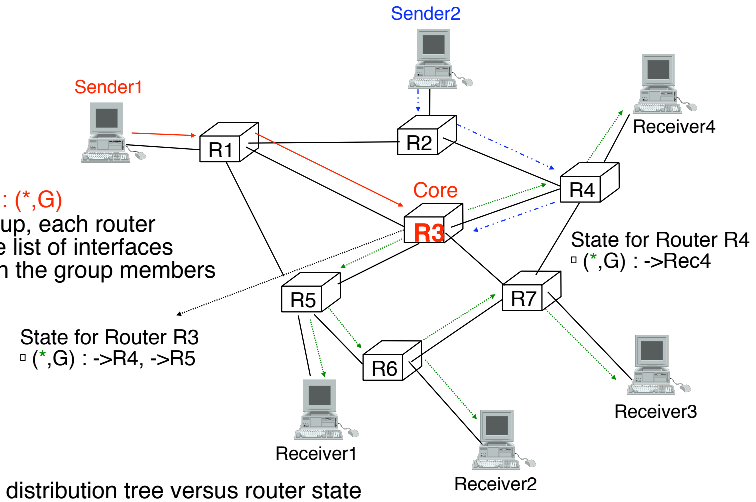
Issues for network-level multicast

1. How to localize all the members of each multicast group ?
   - Internet Group Management Protocol
2. How to efficiently distribute multicast packets from one or more sender(s) to a group of receivers ?
   - A loop-free graph must be built to connect the sender(s) to all the interested receivers
     - source tree
     - unidirectional shared tree
     - bi-directional shared tree

Using bi-directional shared trees

- Principles
  - One router is configured as core for each group
  - All other routers know the address of the core
  - Core is root of bi-directional shared tree
  - Bi-directional shared tree is used to forward multicast packets sent by any source
  - Routers join dynamically the shared tree

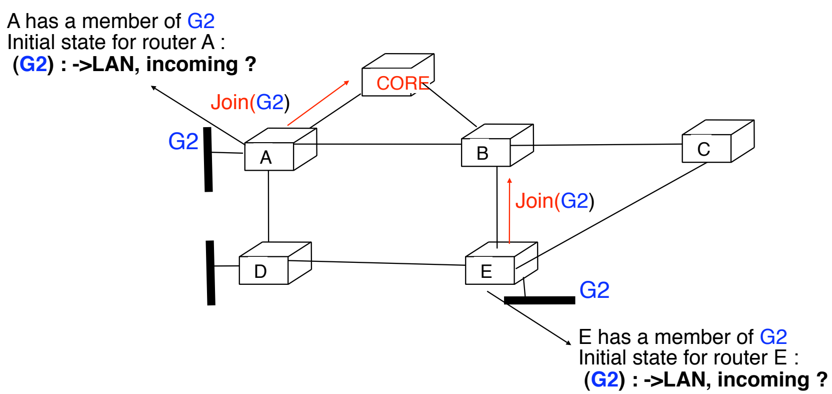
Multicast shared distribution tree



Router state : (*,G)
For each group, each router maintains the list of interfaces used to reach the group members

Tradeoff
- optimum distribution tree versus router state

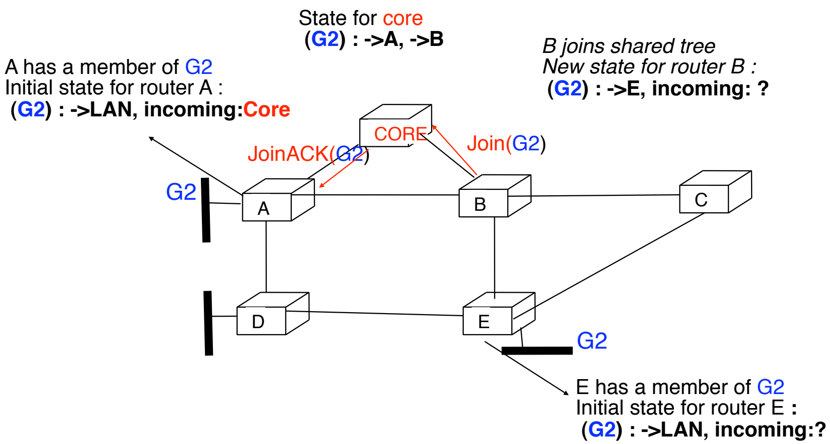
Construction of the shared tree



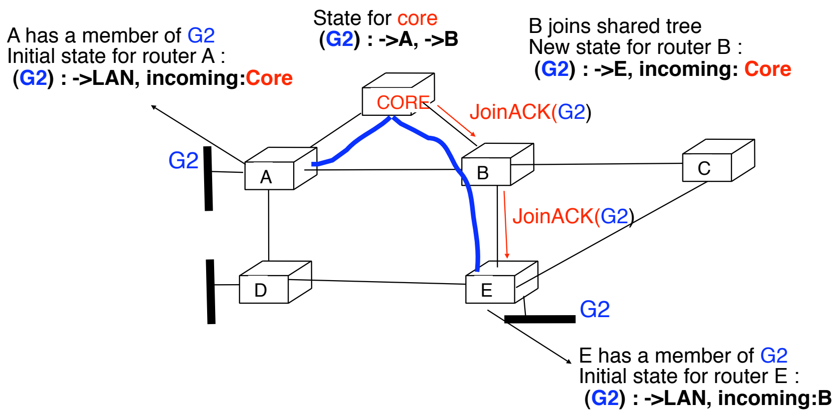
CNPP/2008.7.

This example is based on the utilization of CBT. Other protocols may behave differently

Construction of the shared tree (2)

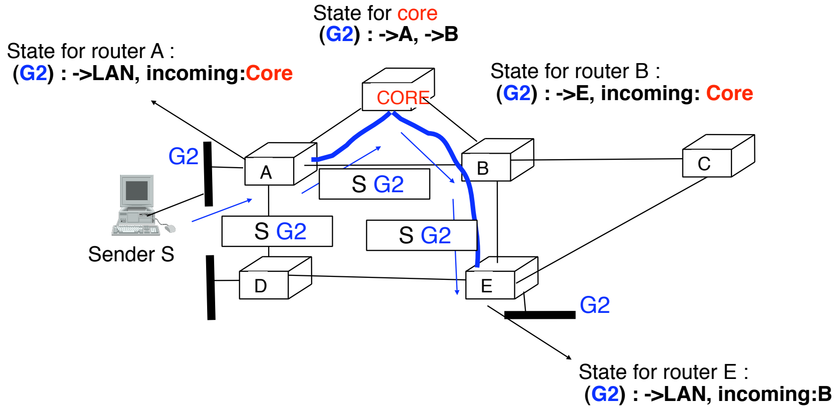


Construction of the shared tree (3)



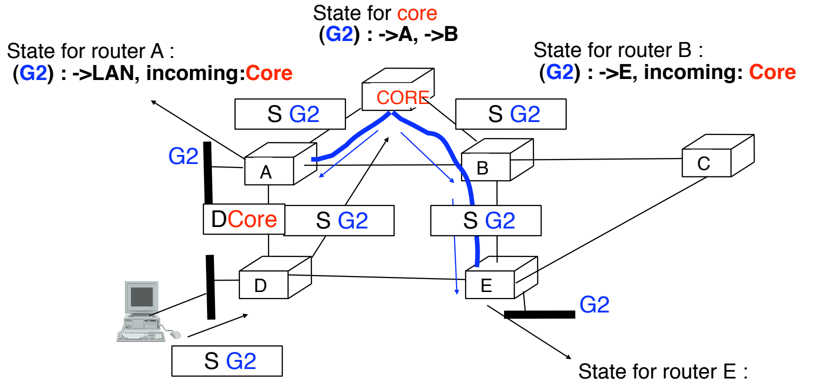
Transmission of multicast packets

- Member transmission



Transmission of multicast packets (2)

- Non-member transmission



Router D will send the non-member transmission to the Core inside a tunnel

CBT : Core Based Tree

- A proposed multicast routing protocol
  - still only a proposal
  - based on bi-directional shared trees
    - requires less state than PIM-SM
    - does not switch to source trees to reduce delay
  - Core placement and reliability are key issues
    - Core is more involved in the forwarding of packets than the RP in PIM-SM
      - all multicast packets for a given group are retransmitted by core
    - Core should correctly be placed for large groups
      - otherwise packets main the send unnecessary on expensive links
    - Core locations can be configured on a per group basis
    - Core candidates send regular broadcast messages and the best core for each group is selected

CBT is defined in :
RFC2189 Core Based Trees (CBT version 2) Multicast Routing. A. Ballardie. September 1997

Besides CBT, an extension to PIM that supports bidirectionnal shared trees is being developed, see :

Mark Handley, Isidor Kouvelas, Tony Speakman, Lorenzo Vicisano, Bi-directional Protocol Independent Multicast (BIDIR-PIM), Internet draft, draft-ietf-pim-bidir-04.txt , work in progress

Multicast without trees

- Existing multicast solutions force each router on the tree to maintain per-group state
  - acceptable with a small number of large groups
  - not acceptable with many small groups

- Alternative solution
  - Connectionless/explicit multicast for small groups
  - Principle
    - Modify the IP header to encode the IP addresses of all the members of the group inside the IP header
    - Routers will forward and duplicate packets on this basis
      - routers do not need to maintain per-group state
      - routers must instead perform per-packet processing
      - solution could be useful in specific cases (e.g. ADSL)

See e.g. R. Boivie et al. Explicit Multicast (Xcast) Concepts and Options, RFC5058, Nov. 2007