

Multicast Routing

(/udd/bcousin/ITI-Caire/Cours/3.multicast-routing.fm- 24 April 2002 14:56)

PLAN

- Generalities on multicast
- Multicast addressing
- Multicast trees
- IGMP protocol
- Multicast Routing
- Conclusion

Bibliography

- S. Paul. Multicasting on the Internet and its applications. Kluwer academic publishers, 1998
- C. Huitema. Routing in the Internet, Prentice Hall, 2001
- W. Stallings. High Speed Networks. Prentice Hall, 1998
- C. Comer. TCP/IP: architectures, protocoles, applications. InterEditions, 1998
- R.Wittmann, M.Zitterbart. Multicast Communication : Protocols and Applications. Morgan Kaufmann publishers, 2001

1. Generalities on multicast

1.1. Presentation

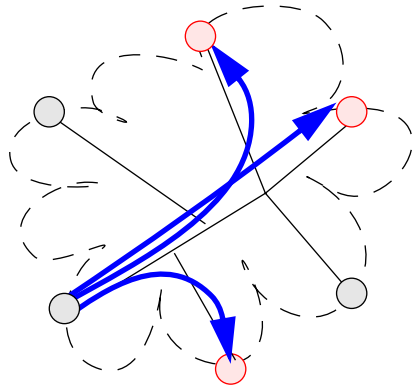
Many applications broadcast informations to **several** destinations:

- Audioconference, videoconference, web radio, etc.
 - ex: CU-SeeMe, IVS, Netmeeting, web phone
- Group oriented applications: whiteboard, group editing, etc.
- Data distribution: news-group, distributed computing, etc.

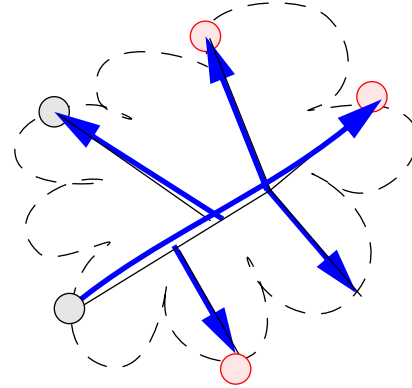
To build a broadcasting network:

- at **application** level:
 - many copies, sub-optimal, bipoint link, server based
 - . - e.g.: CU-SeeMe reflectors
- at **network** level:
 - multicast forwarding = multicasting
 - the network knows the best route!
 - . good efficiency

Some broadcasting techniques:

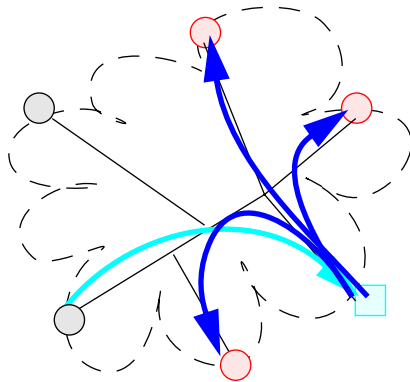


Multi- single transmission

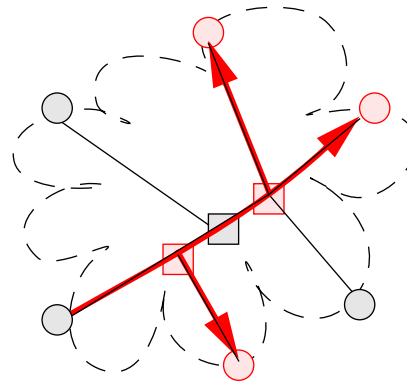


Total broadcasting

- host
- group member
- router



Broadcast server



Multicasting

1.2. Multicast protocols types

Multicast routing protocols

- Network layer
- Multicast routing data
- e.g.: (IGMP), DVMRP, MOSPF, CBT, PIM, etc.

Multicasting protocols

- Transport or above layer
- End-to-end multicast data transmission
 - congestion and error control
 - . e.g.: UDP!, ST2 (Stream Transport v2), XTP (eXpress Transport Protocol), MTP (Multicast Transport Protocol), RMP (Reliable Multicast Protocol), SRP (Scalable Reliable Protocol), RMTP (Reliable Multicast Transport Protocol), RAMP (Reliable Adaptive Multicast Protocol), etc.

Other protocols:

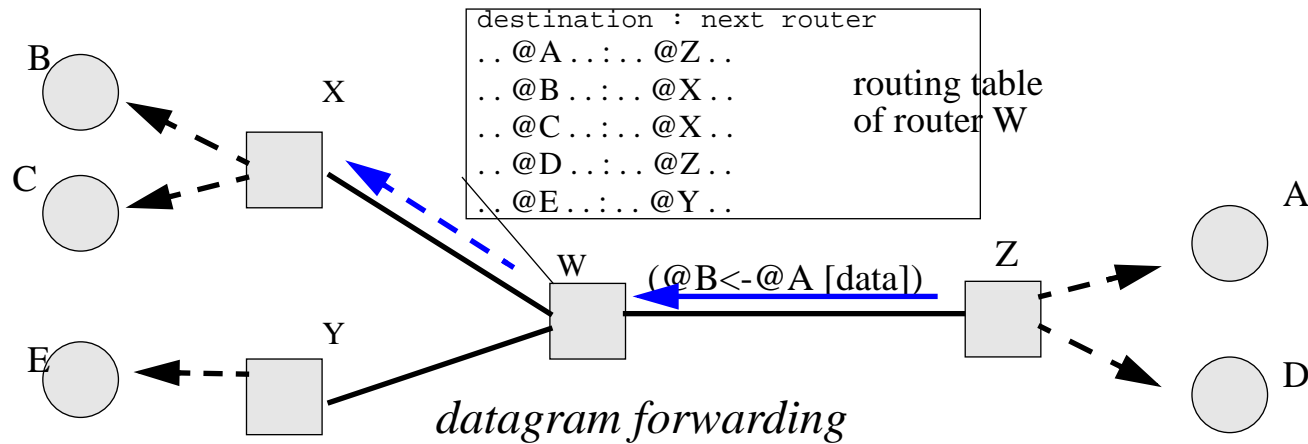
- group management protocols
 - e.g.: SAP (Session Announcement Protocol), SIP (Session Initiation Protocol) HTML oriented, SDP (Session Description Protocol), etc.
- QoS management protocols
 - e.g.: RSVP (ReSerVation Protocol), RTP & RTCP (Real Time &Control Protocol), QoSPP (QoS extension to OSPF), etc.
- application-oriented protocols
 - e.g.: MFTP (Multicast File Transfer Protocol): MCP + MDP (Control + Data)
- multicast address management
 - e.g.: MSDP (Multicast Source Distribution Protocol)

 [Multicast routing protocols](#)

1.3. Unicast forwarding

- datagram forwarding

⇒ next router selection (next-hop).



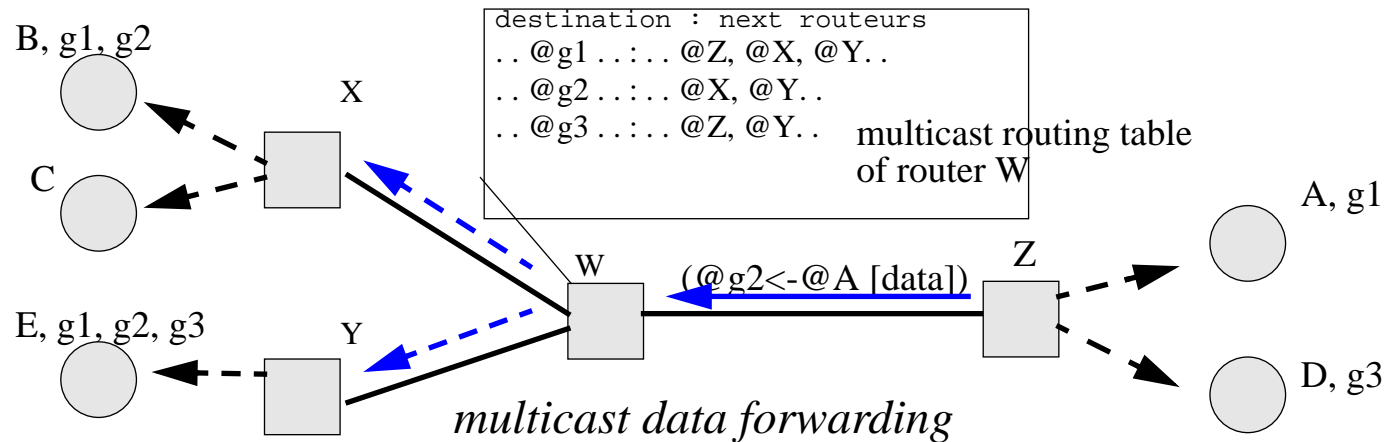
- routing table update

⇒ routing protocols

1.4. Multicast routing

- multicast forwarding

⇒ next routers selection (“next-hops”)



- multicast table update

⇒ multicast routing protocols

1.5. Scalability

Routers process thousands of packets per second

- forwarding process optimization
 - processing time is function of the routing table size
 - . search algorithm complexity: $n \log(n)$
 - . n : number of destination hosts \implies billions!

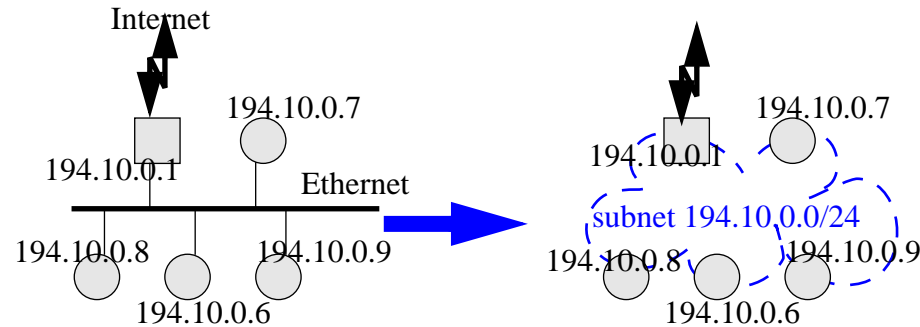
Solutions:

- Aggregation
- Hierarchisation

1.6. Aggregation

- Subnetworking

- hosts sharing an address prefix belong to the same subnet (“netid”)
- e.g.: hosts on an Ethernet LAN could belong to the same subnet



- CIDR (“Classless Inter Domain Routing”)

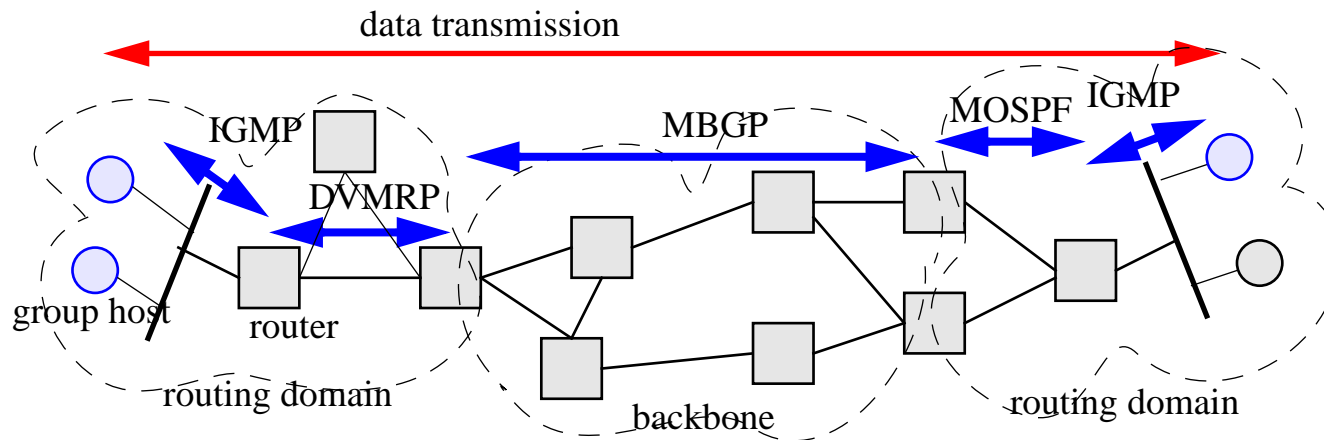
- extension of the previous concept
- one routing table entry for all destinations having consecutive IP addresses reached by the same next router
 - . european prefix: $194.0.0.0/7 = [194.0.0.0, 195.255.255.255]$

⇒ inapplicable to multicast forwarding: no locality of group members!

1.7. Hierarchical levels of multicast routing

3 multicast routing levels:

- IGMP protocol monitors group existence on its local subnets. IGMP messages are sent between **group hosts and multicast routers**.
- **Interior multicasting protocols** (e.g.: DVMRP or MOSPF) manage routing data sent between routers into a routing domain.
- **Exterior multicasting protocols** (e.g.: MBGP) manage routing data sent between routers into the backbone domain.



👉 Hierarchy assists scalability.

2. Multicast addressing

2.1. Introduction

Each IP multicast address **identifies a group** (of hosts)

- IP address of class D: address prefix = “1110₂”,
- Address range = [224.0.0.0, 239.255.255.255]

Reception:

- Hosts with a multicast address receive every multicast packets with the multicast address (as destination address).

Emission:

- To send a multicast packet to a group, no need to belong to the group.

Beware:

- never use multicast address as source address.
- No group management: outside IP scope.
- One multicast address can be use simultaneously by several applications. Every multicast packets are received by all hosts of all applications. Multicast packet filtering is required

Multicast packet: a packet with a multicast address

2.2. Reserved multicast address

2.2.1 Broadcast address

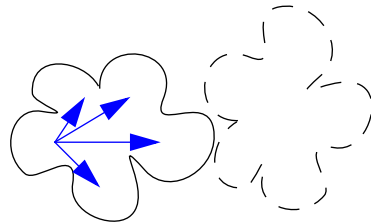
- Full “1”!

- **Local broadcasting:** 255.255.255.255

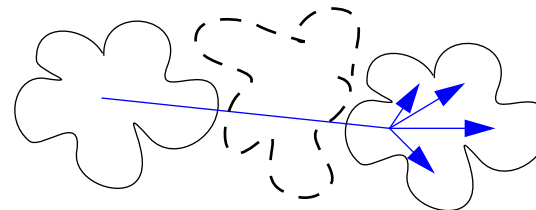
- . broadcasting to every hosts in the local subnet (source subnet)

- **Distant broadcasting:** netid-A.255.255.255, netid-B.255.255, netid-C.255

- . broadcasting to every hosts in the “netid” subnet (distant subnet)



local broadcasting



distant broadcasting

2.2.2 Specific multicast address

Specific services could be identified using multicast address.

- Optimization:
 - less inopportune processor interruptions due to broadcasting of messages

Examples:

- 224.0.0.1: every hosts on the local subnet
- 224.0.0.2: every routers on the local subnet
- 224.0.0.9: every RIP-2 routers on the local subnet
- etc.

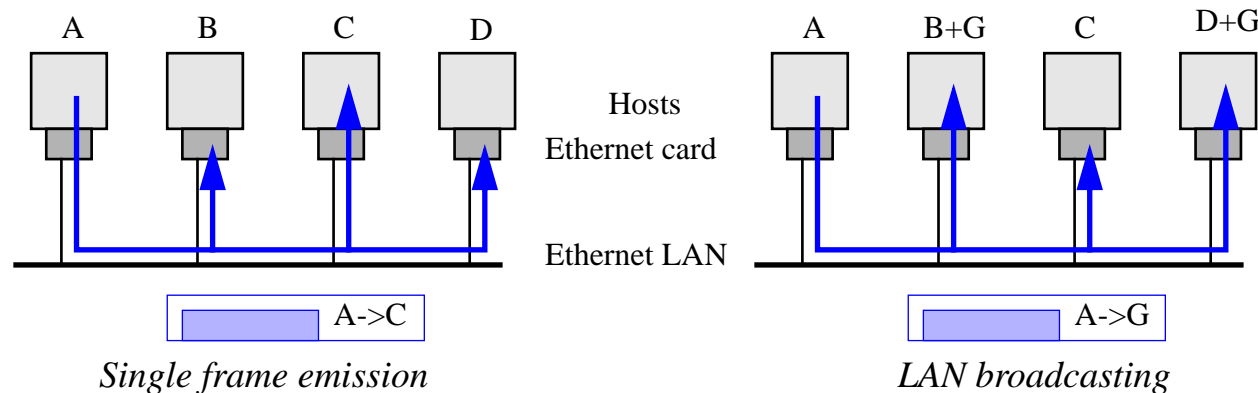
2.3. Broadcasting in Local Area Network

Broadcasting is natural on Lan

☞ LAN == shared medium.

- All LAN hosts received a copy of any LAN frames.
- frame emission cost to a single destination is strictly equal to LAN broadcasting cost
- Network access card receiving an unwanted frame, discarded it!

. note : Network access cards are dedicated to network, so processing of unwanted frames does not hinder the CPU



. Bit G of LAN address identifies group address type.

2.4. Multicast address resolution

- multicast IP address  MAC group address

- **multicast** address:

1110 xxxx xabc defg hijk lmno pqrstu

23 least significant bits

- IEEE 802 **groupe** address:

0000 0001 0000 0000 0101 1110 0abc defg hijk lmno pqrstu

Note:

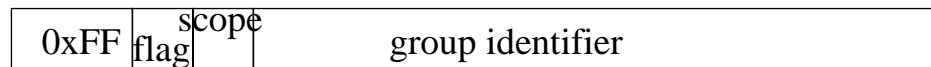
- several (16) multicast addresses are associated with one IEEE 802 address
 . e.g.: 224.1.2.3 ou 225.129.2.3 \Rightarrow 01 00 5E 01 02 03

2.5. Adressage multicast IPv6

2.5.1 Presentation

IPv6 address proposes a multicast address type (specific **prefix**: FF₁₆):

address IPv6 : 16 bytes
 8 bits 4 bits 4 bits 112 bits



- *Flag* field [4 bits]:
 - permanent address or not
 - T=0: permanent
 - T=1: non-permanent
- *Scope* field [4 bits]
 - 1 : node-local scope
 - 2 : link-local scope
 - 5 : site-local scope
 - 8 : organization-local scope
 - E : global scope
 - . 0, F: reserved; 3: subnet scope; 4, 6, 7, 9, A, B, C, D: unused
 - . substituted for IPv4 TTL field.
- *Group identifier* field [14 octets]

2.5.2 Reserved multicast addresses

- Discovery process optimization
- Examples:
 - No host group:
 - . FF0s::0
 - All nodes group:
 - . FF0s::2 avec s={1, 2, 5, 8, E}
 - . e.g.: all organization nodes: FF08::1
 - All hosts group:
 - . FF0s::2 avec s={1, 2, 5, 8, E}
 - . all node interfaces = FF01::2
 - All routers group:
 - . FF0s::3 avec s={1, 2, 5, 8, E}
 - . e.g.: all site-local routers = FF05::3
 - All NTP servers group
 - . FF0s::43 avec s={1, 2, 5, 8, E}

3. Multicasts trees

3.1. Introduction

Multicast tree:

- The Source host is the root of the tree
- The Destination hosts are the tree leaves
- Network routers are the tree nodes

Given a network topology, given a source and destinations, many multicasts trees could be build.

Multicast tree properties:

- Shortest path tree
- Optimum tree

3.2. Optimum tree

Optimum Spanning tree:

- to minimize the resource cost: link utilization

- all tree nodes belong to the group:

⇒ Total spanning tree

- not all tree nodes belong to the group:

⇒ Steiner tree

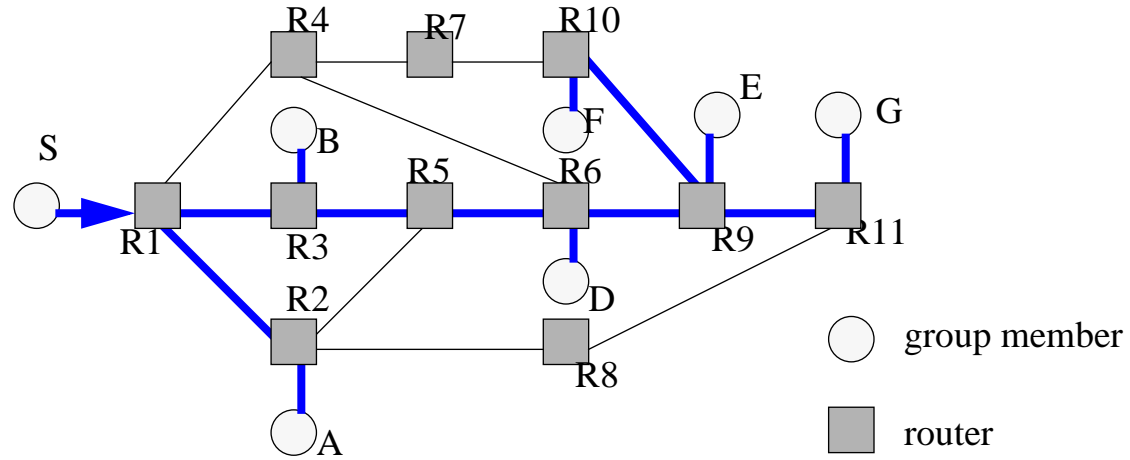
Global optimization of network resources.

- criterion: global cost = total number of links used by the tree

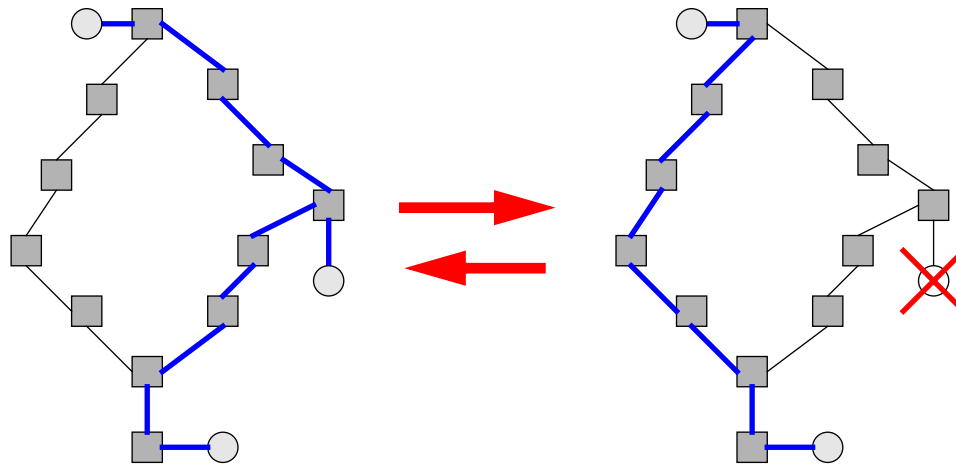
Algorithm complexity of optimum tree building:

- **NP-complete** problem!

Example: unity cost function, global cost = 7+7, maximum path length = 5+2



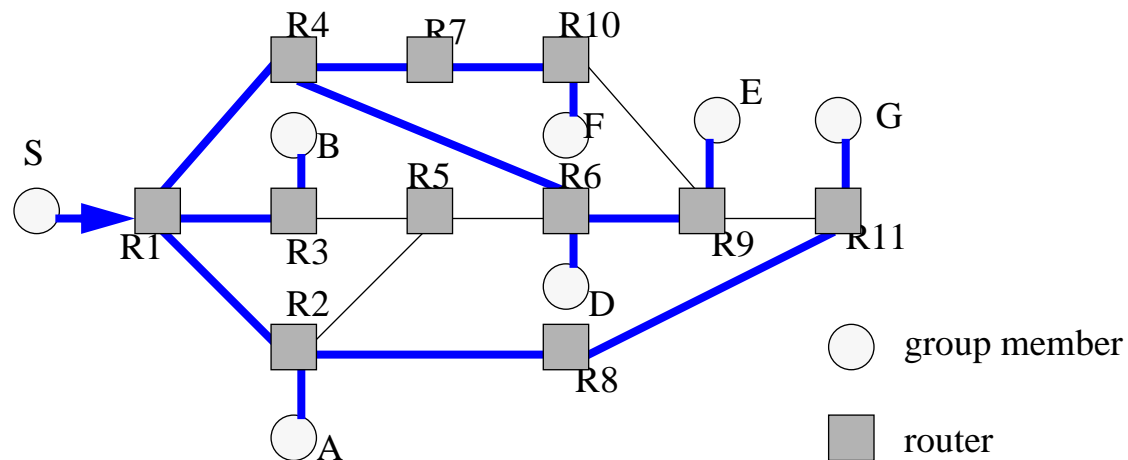
Optimum tree instability:



3.3. Shortest path tree

SPF (“shortest path tree”):

- multicast packets follow the shortest path from the source to each destination
- **minimum delay** (usual application requirement)
 - criterion: maximum path length
- **Example:** unity cost function, global cost = 9+7, maximum path length = 3+2



⇒ used by current multicast routing protocols

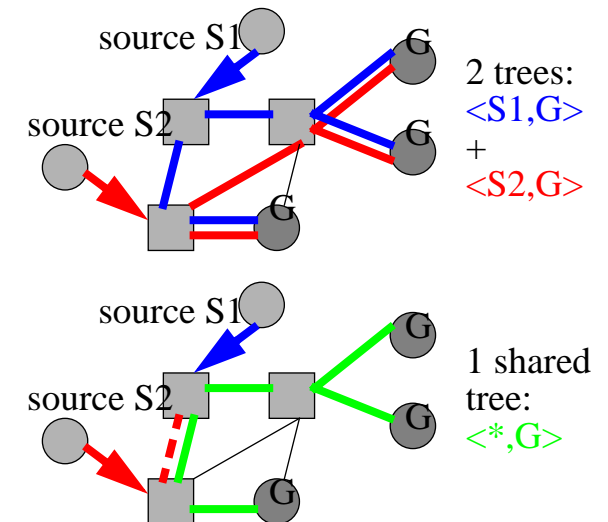
3.4. Shared tree

Sharing of trees:

- one tree for all groups
 - total network spanning tree
 - e.g.: Spanning Tree protocol for LAN bridging
 - drawbacks: non optimal forwarding, traffic concentration on root links
- trees for each group
 - e.g.: Internet multicasting tree
 - drawbacks: numerous trees, difficult multicast route management

Sharing of source-based trees belonging to the same group:

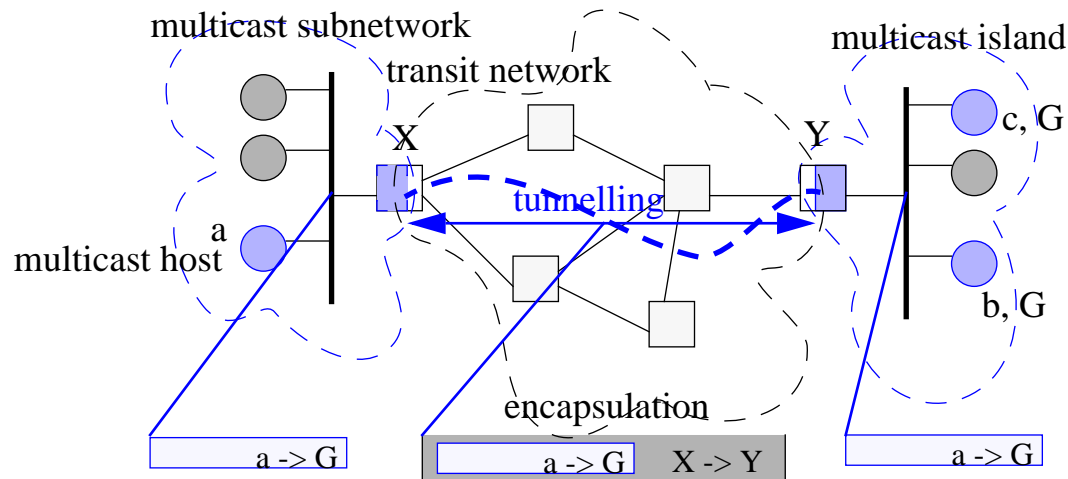
- for a group, different sources produce different trees
 - **source based** tree: $\langle S, G \rangle$
 - . shorter path but resource waste
 - **shared** tree: $\langle *, G \rangle$
 - . resource is more efficiently used but longer path



3.5. Mbone infrastructure

Multicast Backbone: multicast network interconnection

- every multicast network is an multicast island
- interconnected by “tunnels”
 - virtual link between multicast networks
- the transit network isn't required to forward multicast packets: current Internet



Multicast datagram encapsulation in unicast datagrams (IP in IP):

- encapsulation in ingress router of transit network
- dis-encapsulation in egress router

4. IGMP Protocol

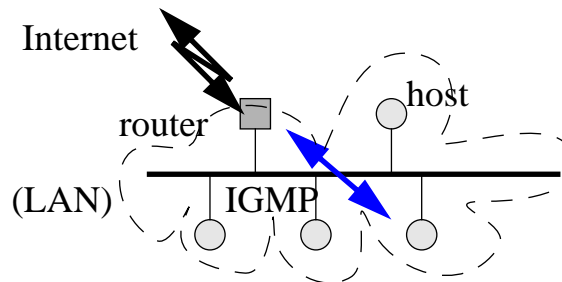
4.1. Introduction

Routers use IGMP protocol to monitor the activity of groups on their LAN interfaces.

- IGMP: “Internet Group Management Protocol”
- rfc 1112: “Host extensions for IP multicasting”

Definition:

- an active group is detect on a LAN iff a host is a member of this group.



IGMP messages are encapsulated into IP datagrams:

- datagram Protocol field = 2
- note: ICMP has been merged into IGMP for IPv6

4.2. Principle

4.2.1 Membership

When a host want to join a group:

- the host sends an “IGMP report” message
 - its “Group Address” field contains the group address.
- the “IGMP report” message is encapsulated in a datagram with “Destination address” is the group address
 - 👉 multicast routers listen to any multicast packet
- without response, IGMP report message transmission is repeated after a random delay
 - 👉 loss recovery

4.2.2 Active group lists

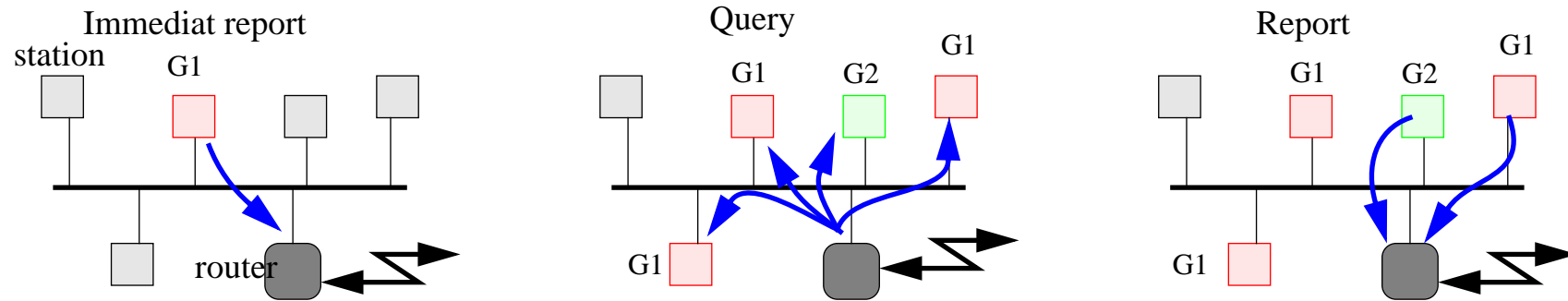
Multicast router monitors group activity.

- routeurs send “IGMP query” message
 - periodically (but not too often to limit overhead): > 1 mn
 - entry which is not refresh in time is discarded
 - ☞ host failure
- “IGMP query” messages are encapsulated into datagrams with All Nodes Destination address (= 224.0.0.1) and TTL field = 1.

Every node is a member of the group 224.0.0.1. This group is always active. No monitoring is required for that particular group.

Any group member, for each group, sends back:

- an “IGMP report” message after a **random delay** [0 - 10 s]!
- “IGMP report” messages are encapsulated into datagrams with “Destination address” = the group address
- when an IGMP report message on the same group is received by the group member, the delayed IGMP report messages is discarded
 - ☞ one IGMP report, each period and for every active group (traffic minimization)



4.2.3 Group leave

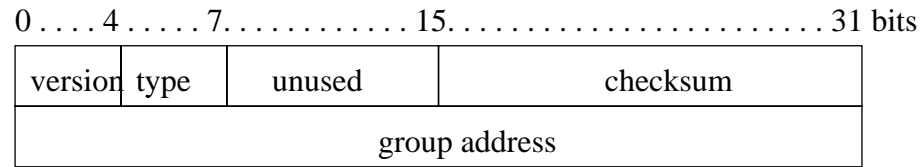
When a host **leaves a group**:

- just do nothing: cease all message group activity
- if the host is the last group member, when the next query, the router will received no report for that group.

☞ multicast routing will be sub-optimal during this transient phase

4.3. IGMP message format

Message length: 8 octets



Version:

- Current version = 1: rfc 1112, 1989 (old version = 0: rfc 988)

Two IGMP message types:

- “Host membership **query**” = 1
- “Host membership **report**” = 2

Checksum processing:

- 1-complement addition of the 16 bit words
- same processing than TCP, UDP or IP checksum.

“Group address”:

- multicast IP address identifying the group

4.4. Error Management

Corruption

- Corrupted datagrams are silently discarded.

Loss

- message retransmission on timeout:

 datagram forwarding may be inconstant

Failure

- Designated routeur failures can be detected and replaced by redundant routers

4.5. IGMP versions

Version 2:

- hosts can send explicit IGMP leave message.
- IGMPv2: rfc 2236, 1997

Version 3:

- any host can send source-specific IGMP report message.

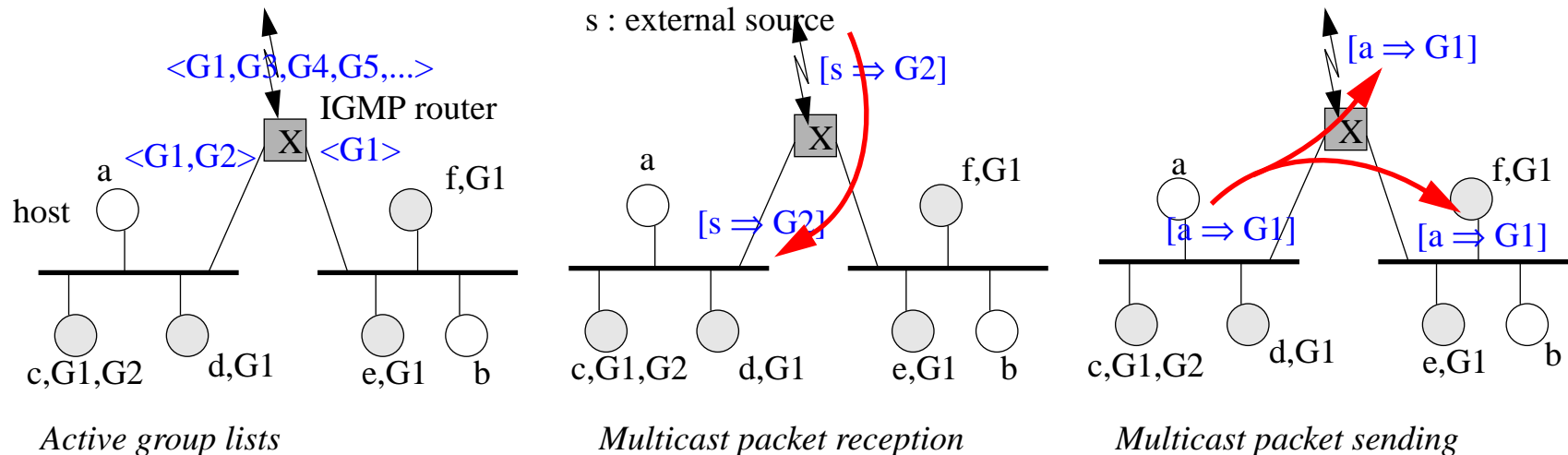
4.6. Multicasting on IP subnet

When a router receives a multicast packet for a group G

- it broadcasts the packets on every subnet where the group G is active (except the subnet where the packet has come from).

Multicast routers monitor group activity.

☞ every IGMP router keeps a list of active groups for all of its interface.




5. Multicast routing

Building of a broadcast tree is costly and complex.

Some solutions:

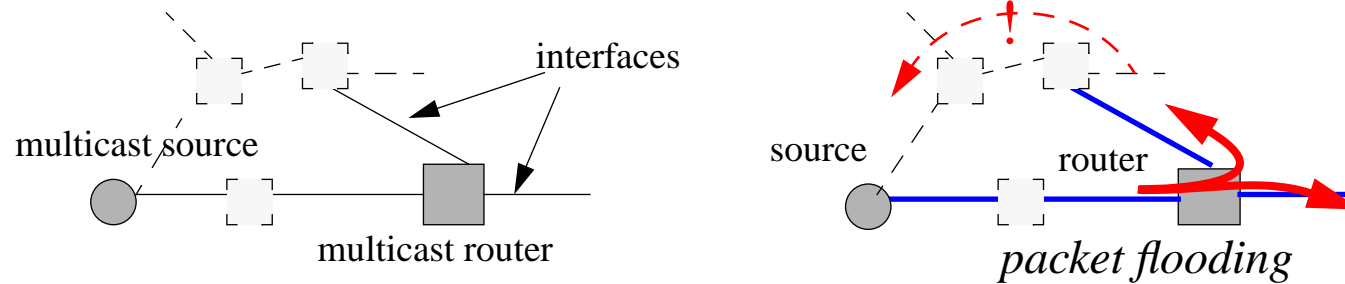
- **network topology** knowledge
 - information coming from “link state” unicast routing protocols
 - . e.g.: OSPF => MOSPF
- routing information:
 - information coming from **unicast routing table**
 - 2 main techniques:
 - . “**Reverse Path Forwarding**”
 - . “Core Based Tree”

 RPM => DVMRP

5.1. Flooding

- Flooding broadcast

- received packets are forwarded to all output interface, except the input interface
- simple, reliable, optimal (!)
- destination location independent: imprecise
- many problems: congestion, cycles and multiple packet copies!



5.2. Reverse Path Forwarding

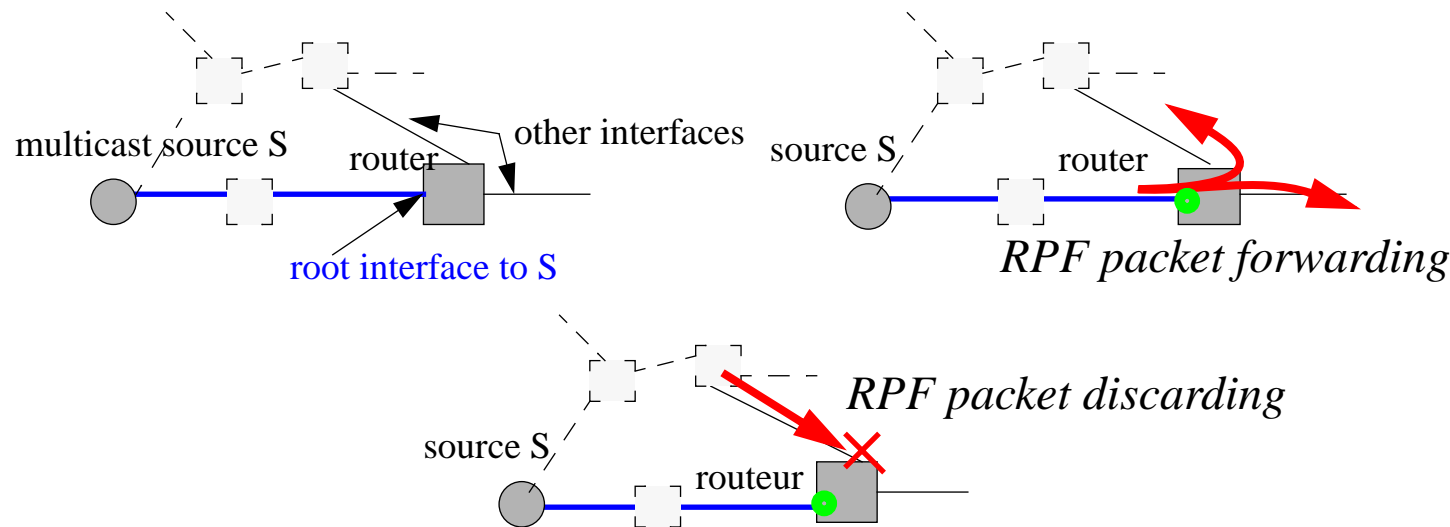
Cycle suppression added to flooding.

Principle:

- every packet received **on root interface is broadcast** to all other interfaces.
- every packet received on **other interfaces are discarded**.

For each source, the root interface of a router is:

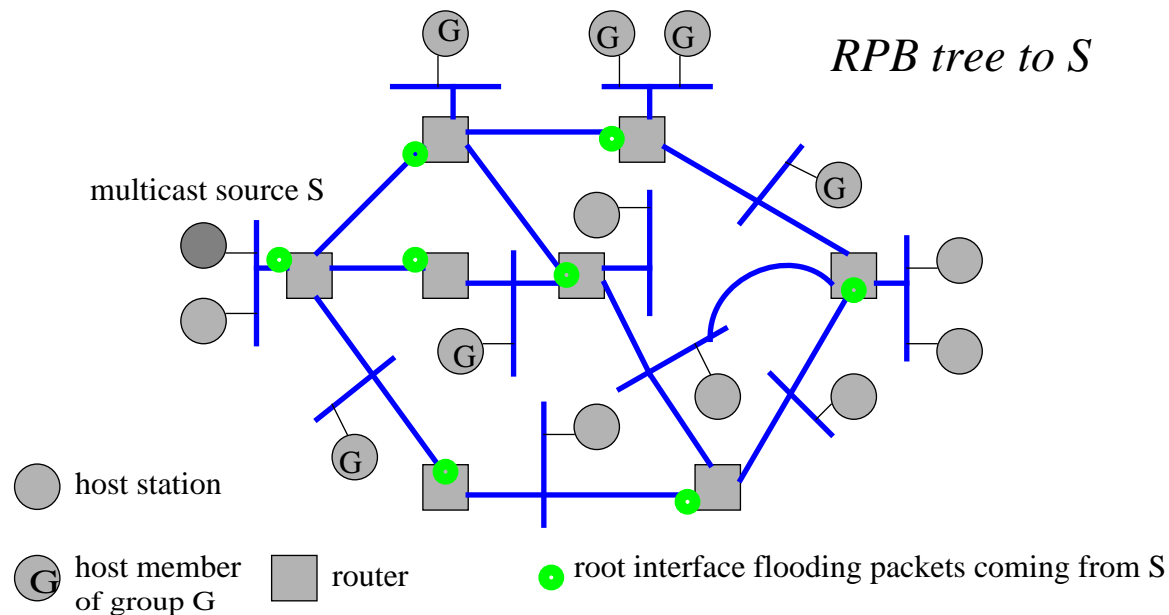
- the interface designated by the unicast routing protocol as the (shortest) path from the router to the source.



5.3. Reverse Path Broadcasting

If every network router in network uses Reverse Path Broadcasting, then:

- RPB tree
- algorithm with no additional states
- any routers, any LAN host (**regardless of its membership**) receives a copy (**at less**) of every packet

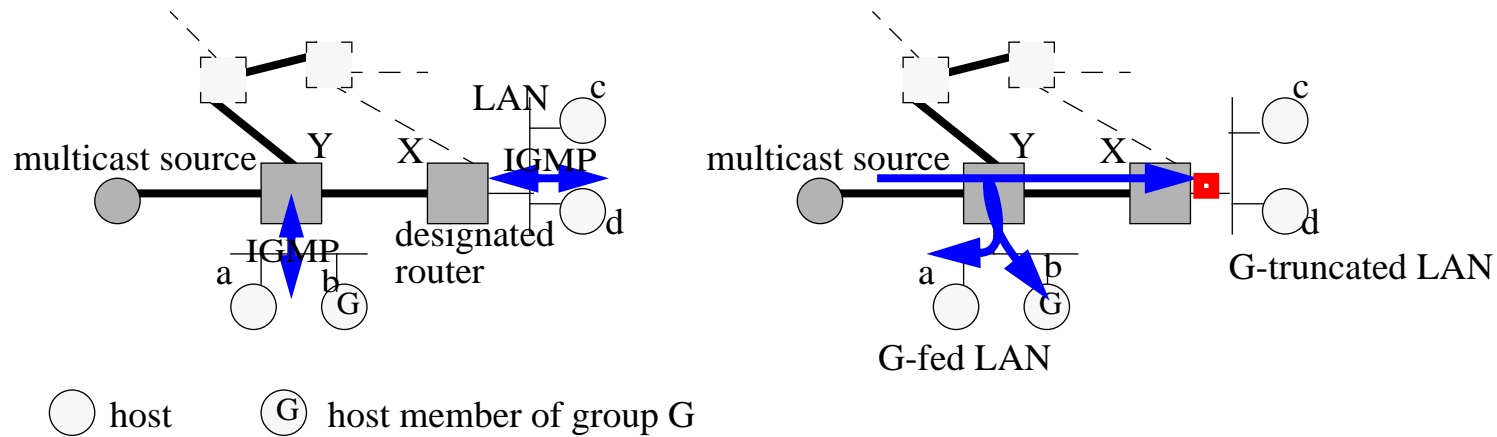


5.4. Truncated Reverse Path Broadcasting

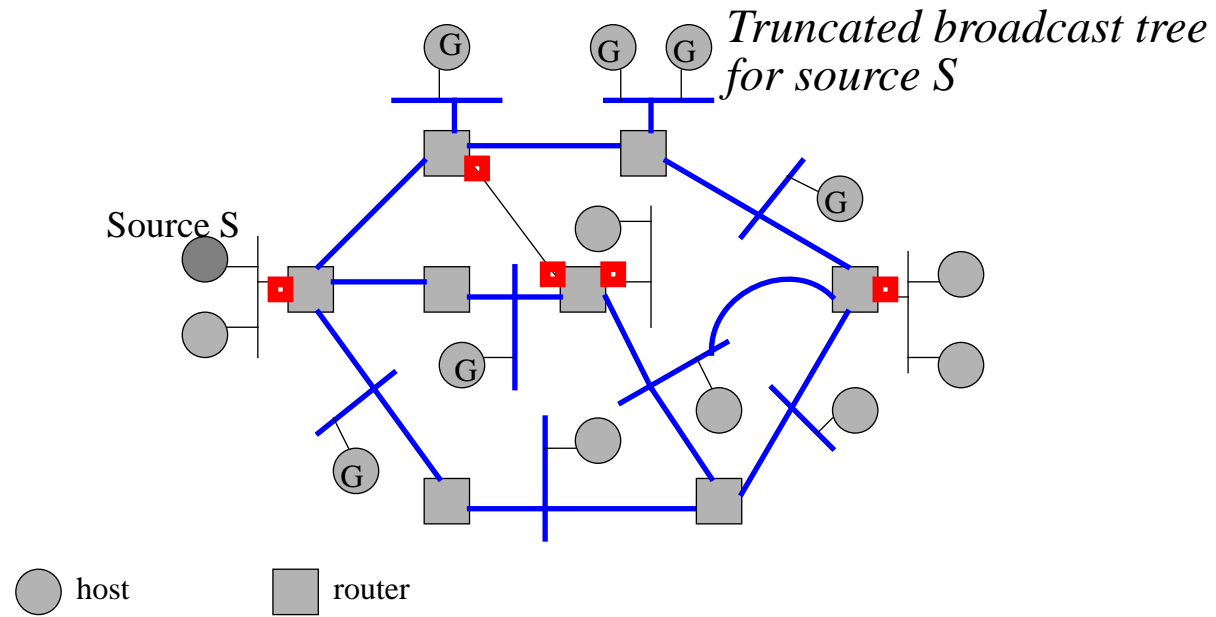
To truncate inactive LANs from the broadcast tree.

For a group, a LAN is inactive iff there is no LAN host members of the group.

- IGMP protocol is designated to monitor Group activity on every router interface.



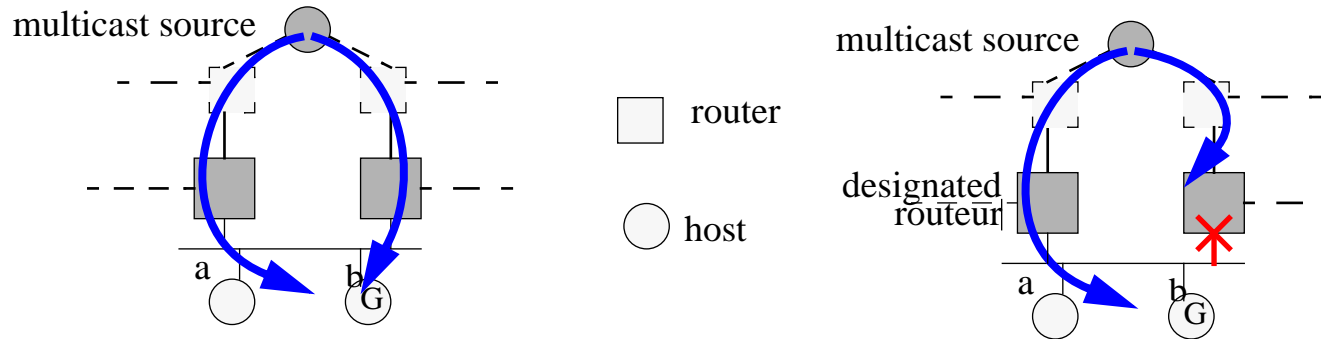
5.5. Improved broadcast tree



5.6. Multi router LAN

LAN may have redundant connections through several routers:

- Hosts on multi-connected LAN receive multiple copies of every packets.



A router is designated to feed the shared LAN. This router:

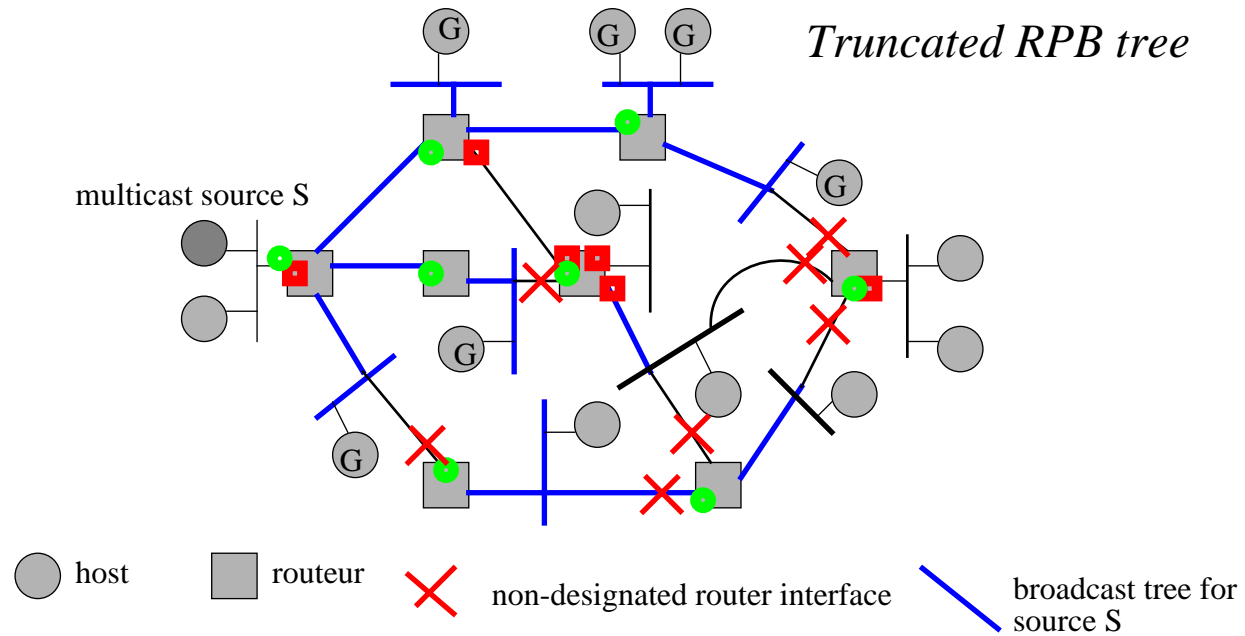
- is on the **shortest path** to the source
- has the **lowest address** (in case of length equality)
 - same technique used by “Spanning Tree” protocol

Routers could deduce other routers presence on a LAN through router specific message exchanges:

- routing protocol
- IGMP protocol

5.7. Truncated-Reverse Path Broadcasting

⇒ Truncated tree without message duplication based on Reverse Path Broadcasting



5.8. Pruning

Some broadcast tree branches **do not feed any group member!**

- T-RPB tree is a Total network spanning tree.

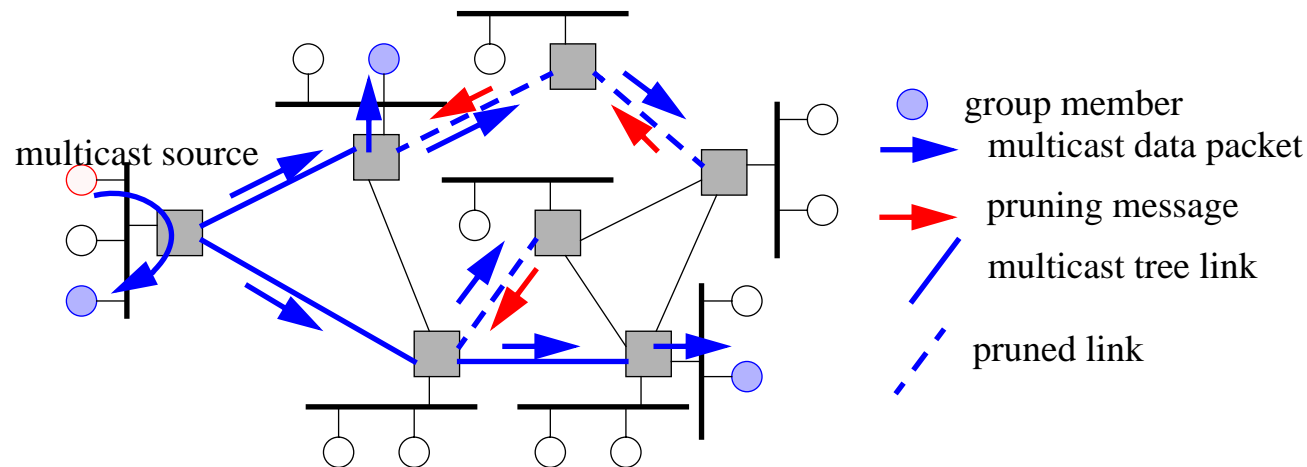
Specific messages will prune useless branches.

When a router receives a multicast data packet sent from a source to a group G and, if **all router interfaces are G-inactive**,

- the router sends back a G-prune message on the root interface toward S.

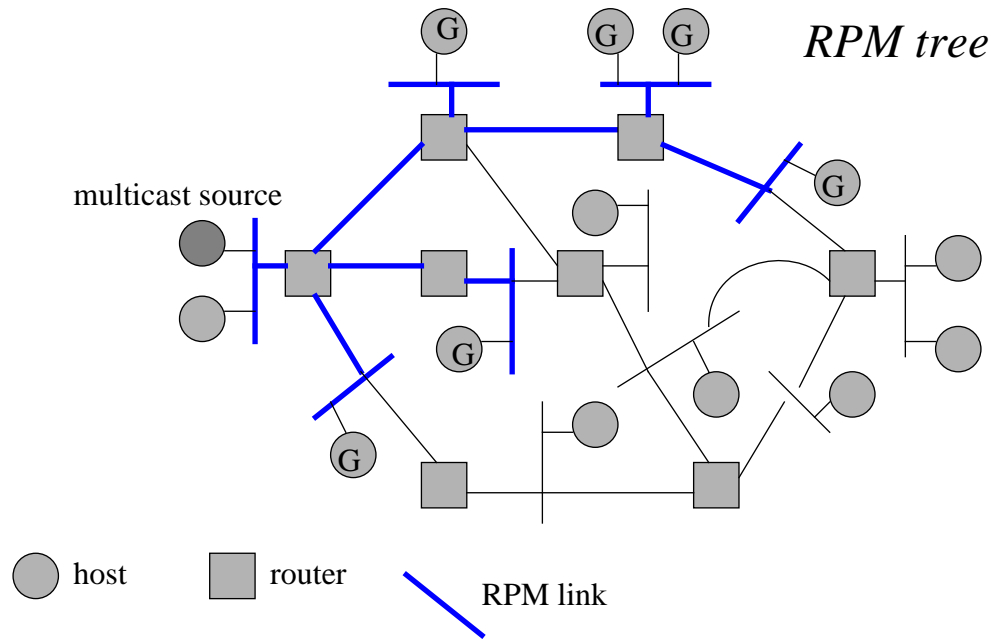
When a router receives a G-prune message on all of its interfaces:

- it forwards the G-prune message on the root interface.



5.9. RPM tree

⇒ Reverse Path Multicasting
 - T-RPB + pruning



5.10. RPM evaluation

Scalability

- on **every router interface**, for **every network group**, the activity of groups have to be monitoring
 - many and long active group lists

Tree monitoring is required due to

- topology modification:
 - link or router failures or insertions
- group membership join or leave

Trial and error

- a **timer** is associated with each list entry
 - . when the timer expires the entry is canceled
- group list entries are managed as soft state
- an entry will be recreated when multicast data packet will be received
- entry cancellation **leads to multicast data flooding**

 Core Based Tree protocols

6. Conclusion

Many operating systems integrate multicast IP driver (fact unknown of users)

Most routers integrate multicast functions required by multicasting (fact deliberately ignored by some network administrators whose don't want to know).

Many applications could benefit from multicasting.

Slowly the Mbone is spreading:

- through tunneling between multicast islands into a unicast sea.

Reliable multicast data transmission is required.