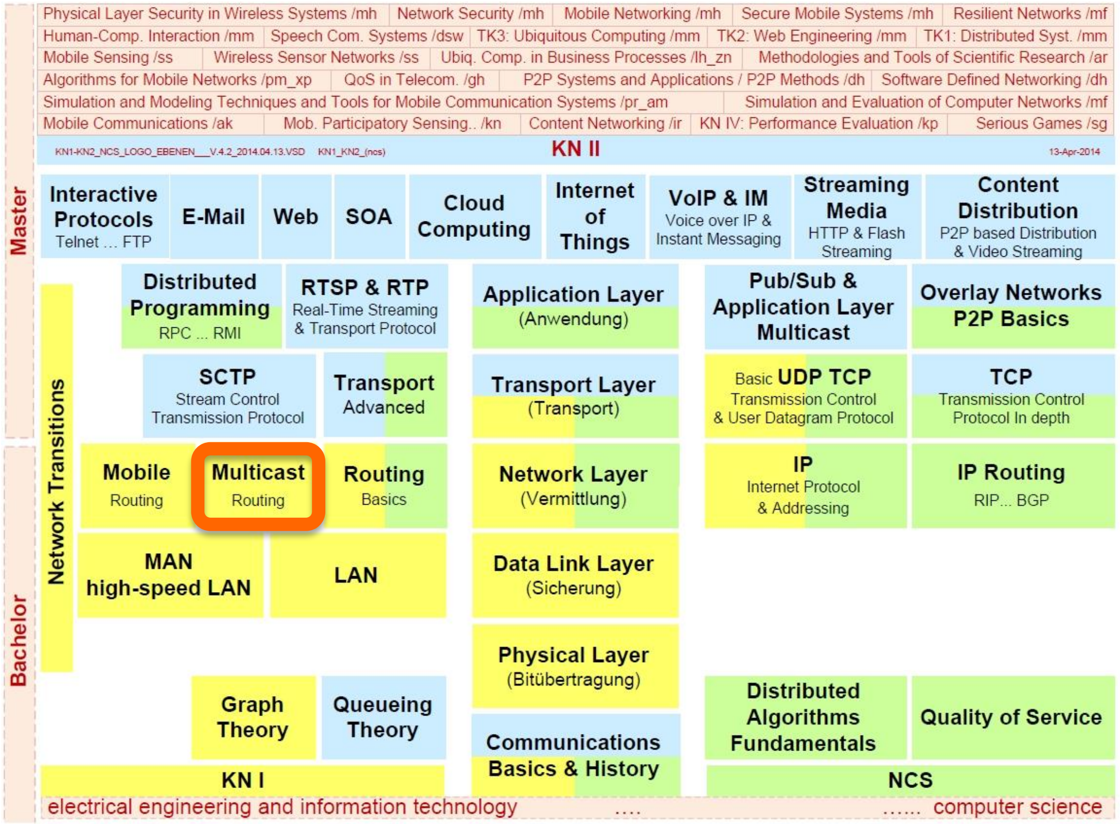
Communication Networks I

Multicast and Broadcast

TECHNISCHE UNIVERSITÄT DARMSTADT



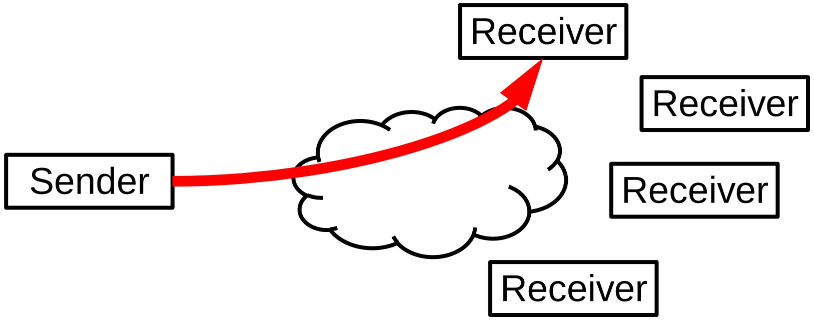
Prof. Dr.-Ing. Ralf Steinmetz
KOM - Multimedia Communications Lab

Overview

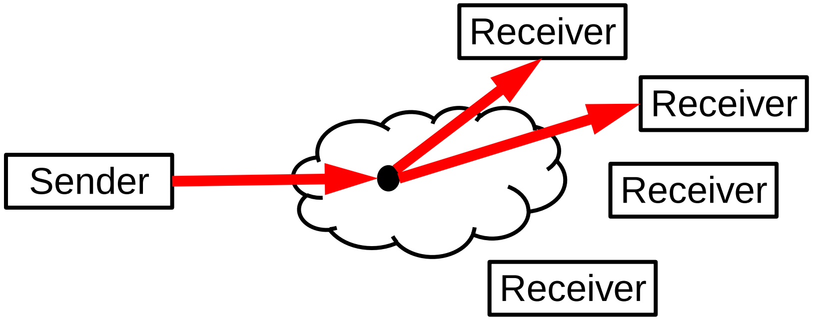
1 Unicast, Multicast, Broadcast

Terminology

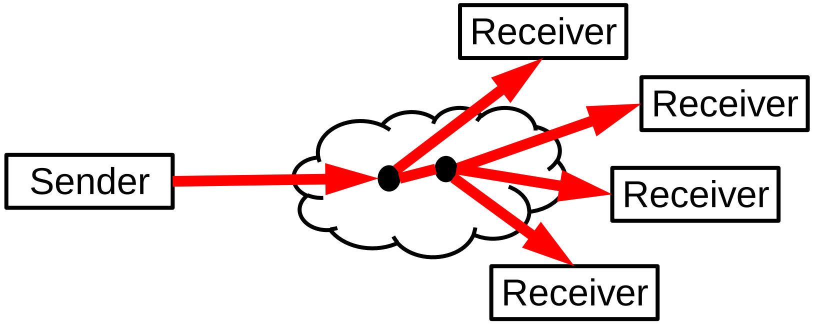
- Unicast: 1 : 1 communication



- Multicast: 1 : n communication



- Broadcast: 1 : all communication



Several methods have been proposed for broadcasting

Simple approaches:

1. **Individual sending to every destination (distinct packets)**
   - Requires no special feature from the network
   - Waste of bandwidth
   - Sender has to know all destinations

2. **Flooding**
   - Too many duplicates

3 … n

- → see in the following

TECHNISCHE UNIVERSITÄT DARMSTADT

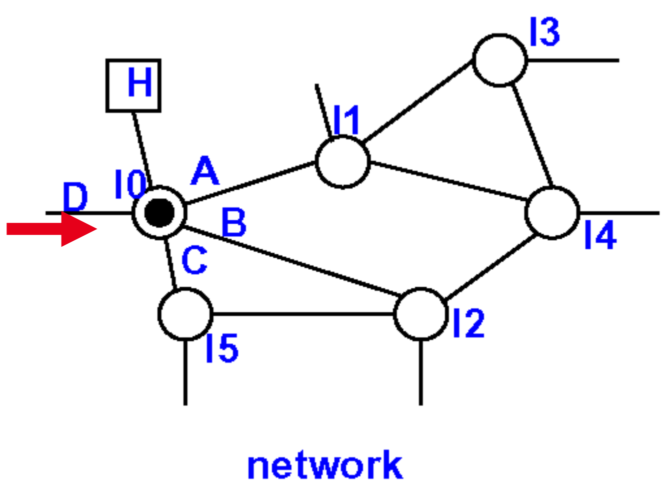
Each Packet CONTAINS A LIST OF DESTINATIONS

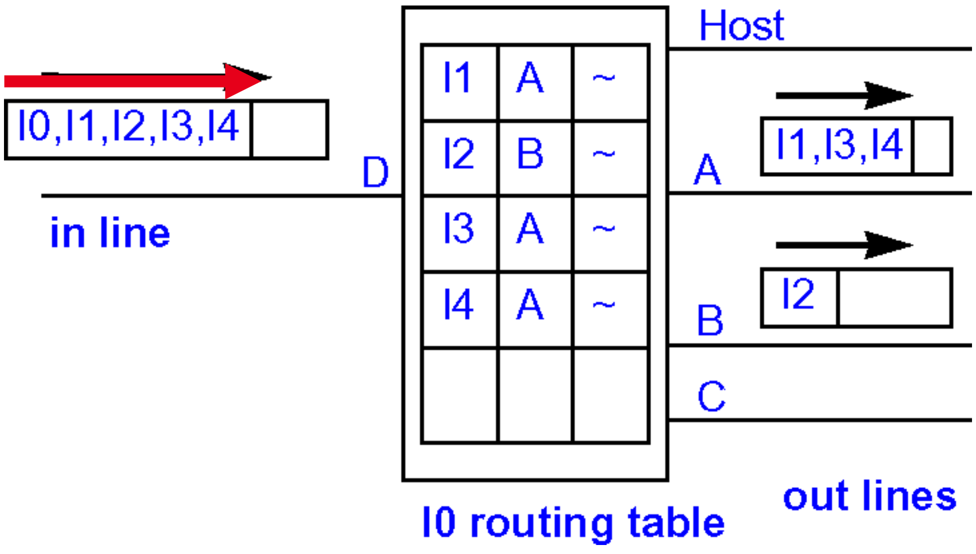
Steps performed at each IS

- Examine, which outgoing links are required
- Generate a copy of a packet for each REQUIRED outgoing link
- Packet copy contains ONLY destinations which can be reached via this line

Example

- Network with ‘I0’ as the considered IS





2.2 Broadcast Routing: Spanning Tree

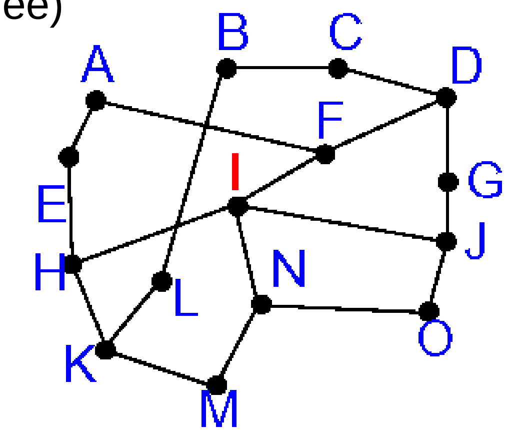
Idea

- Use sink tree (or other spanning tree) for router initiating broadcast

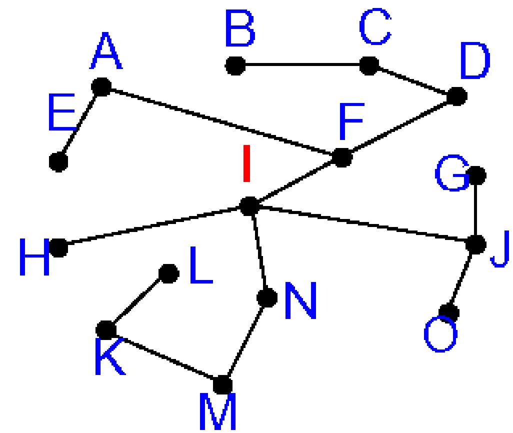
Spanning tree:
subset of subnets including all routers with no loops

Example

- Network, IS 'I' as the sender



Net



a Spanning Tree

Prerequisite

- Spanning Tree is known to the IS
- IS generates minimum number of packet copies
- IS generates a copy of a packet for each required outgoing line
  - all spanning tree lines except incoming one

Main issue

- How to determine a Spanning Tree?
  - sometimes available, e.g., from link state routing
  - sometimes not, e.g., with distance vector

2.3 Broadcast Routing: Reverse Path Forwarding (RPF)

Also called "Reverse Path Flooding" (RPF)

- Variation of the Spanning Tree

Principle

- Each sender has its own Spanning Tree
- But IS do not need to know the Spanning Trees

Considerations

- Each router has information which path it would use for (unicast)-packets
  - because of the unicast routing algorithms

Algorithm (for a packet arriving at an IS)

- Has this packet arrived at the IS entry port over which the packets for this station/source are usually also sent?

Yes:

- Packet used the BEST route until now
- Action: resend over all edges (not including the incoming one)

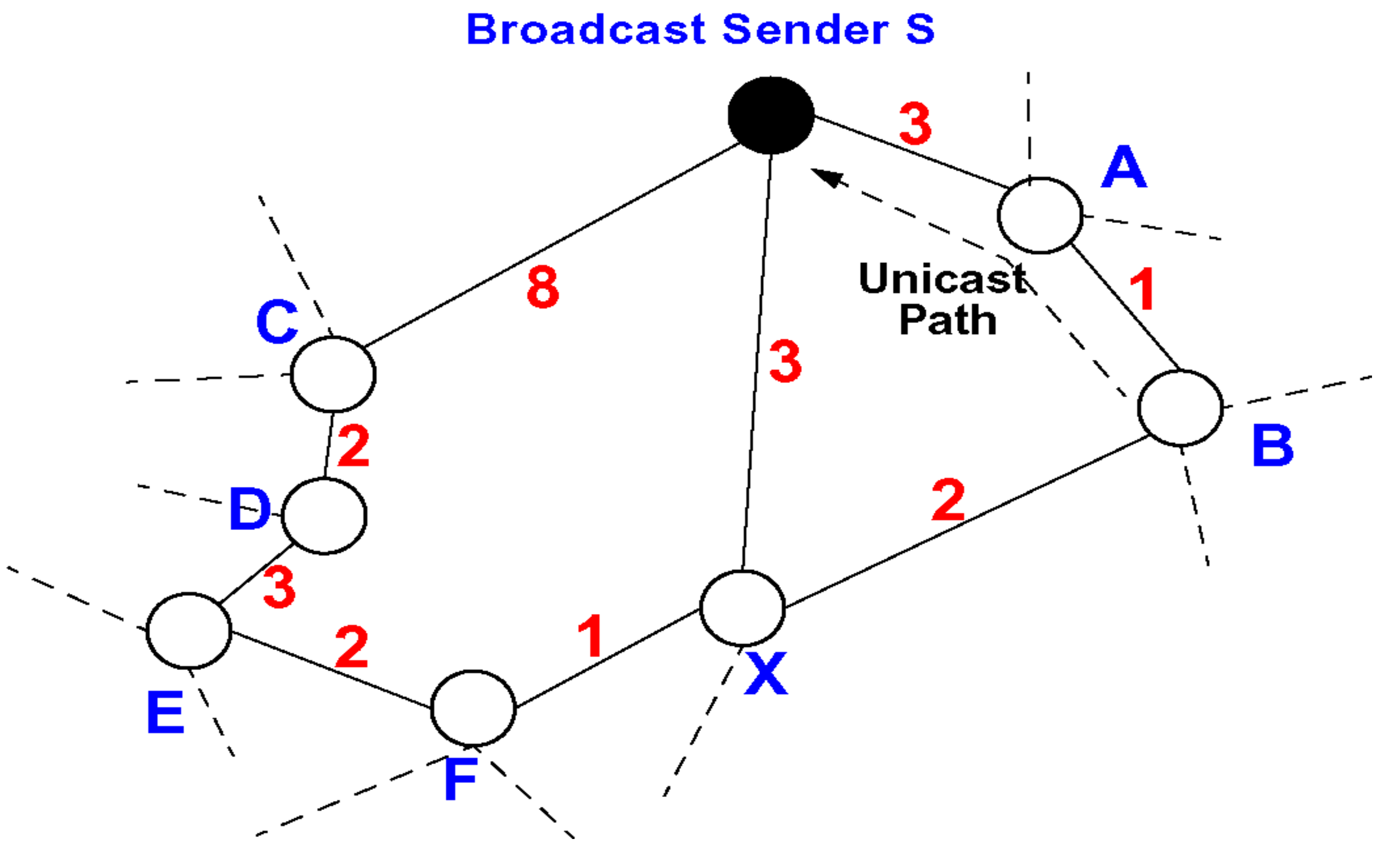
No:

- Action: discard packet (most likely duplicate)

Broadcast Routing: Reverse Path Forwarding (RPF)

Example

- In the example B will send its unicast packets to S via A (shortest route)

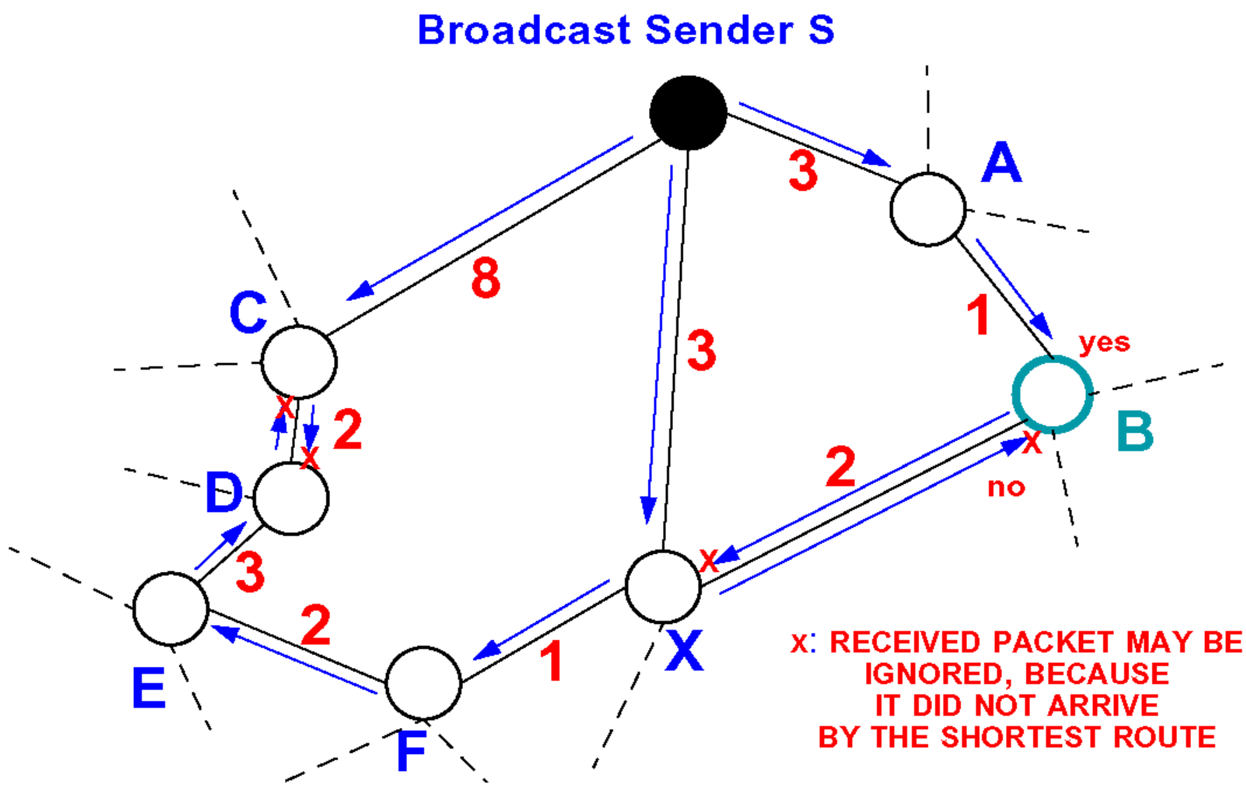


Broadcast Routing: Reverse Path Forwarding (RPF)

TECHNISCHE UNIVERSITÄT DARMSTADT

Example

- Within the RPF algorithm of the above example, e.g.
  - Router X forwards broadcast from Sender S to B, F, ...
  - Router B uses the unicast routing information to ignore all broadcast packets received from S, which did not arrive via node A



2.4 Broadcast Routing: Reverse Path Broadcast (RPB)

Motivation: disadvantages of Reverse Path Forwarding

- If packets are forwarded, then they are forwarded over ALL edges (not including the incoming one)
- It would be better if packets are forwarded over only one SUITABLE edge

Algorithm: packet from S(ource) to D(estination)

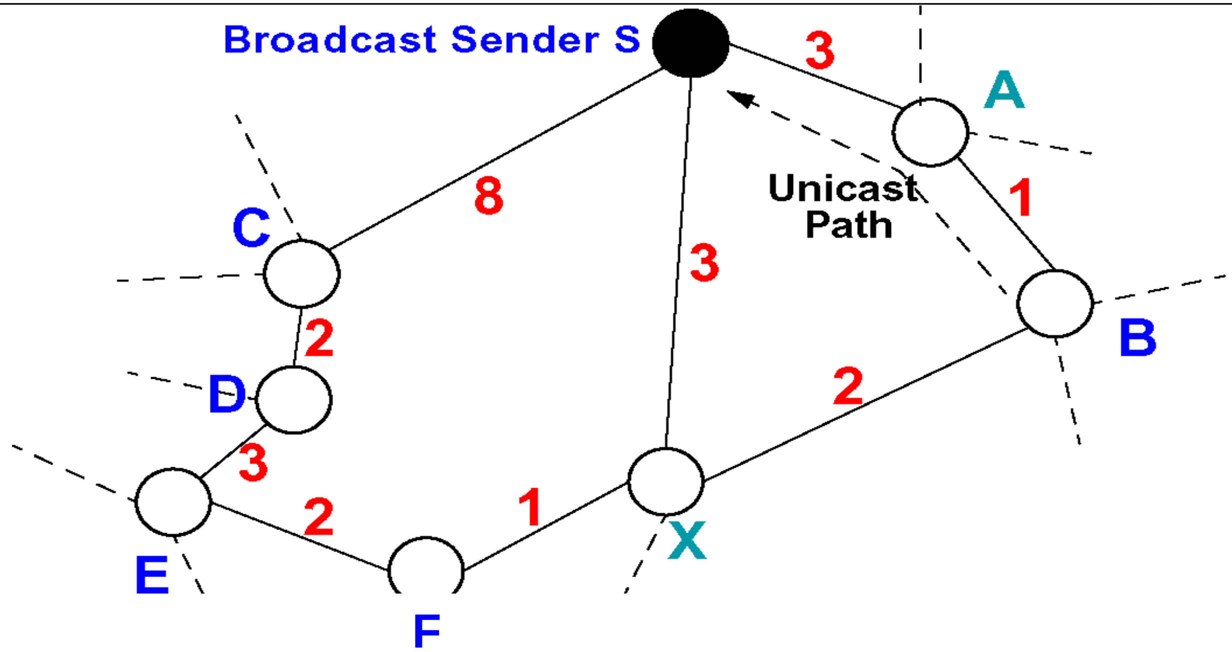
- Like REVERSE PATH FORWARDING with specific selection of the outgoing links

Has this packet arrived at THE IS entry over which the packets for this station/source S are usually also sent?

- Yes: Packet used the BEST route until now?
  - Yes: → select the edge at which the packets arrived and from which they are then rerouted to source S (in reversed direction)
  - No: → DO NOT send over all edges (without the incoming one), i.e., not as in Reverse Path Forwarding (RPF)

- No: → discard packet (is most likely a duplicate)

Broadcast Routing: Reverse Path Broadcast (RPB)



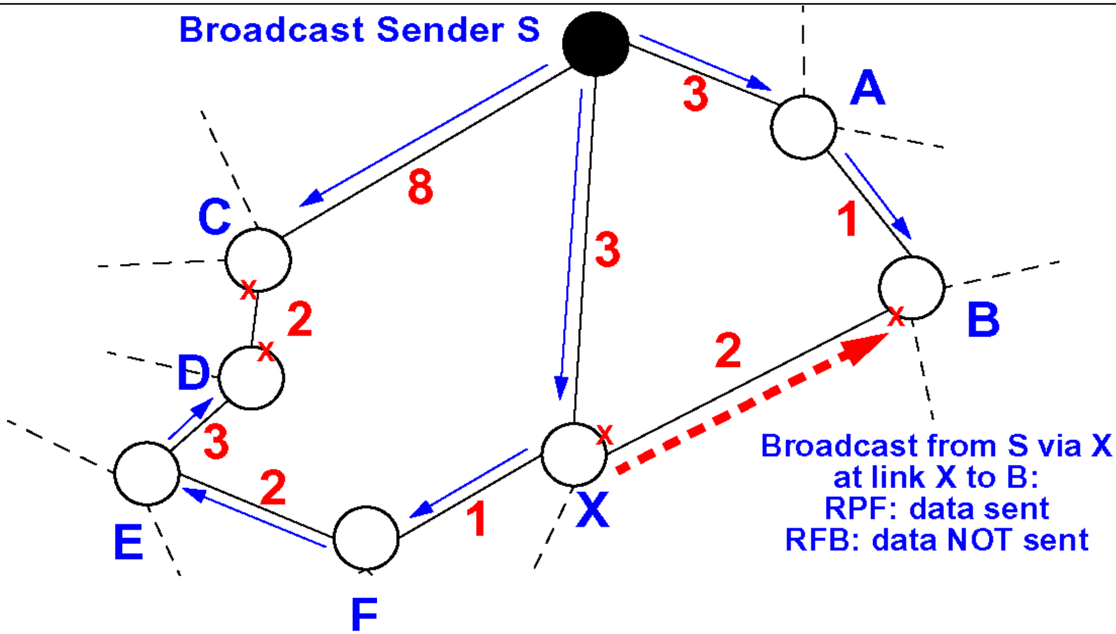
Example:

In the example

- A can learn by inspecting the unicast packets
  - that it is located on the unicast path from B to S
- X can learn by packets failing to appear
  - that it is not located on the unicast path from B to S

→ This information is used by the RPB algorithm

Broadcast Routing: Reverse Path Broadcast (RPB)



Example

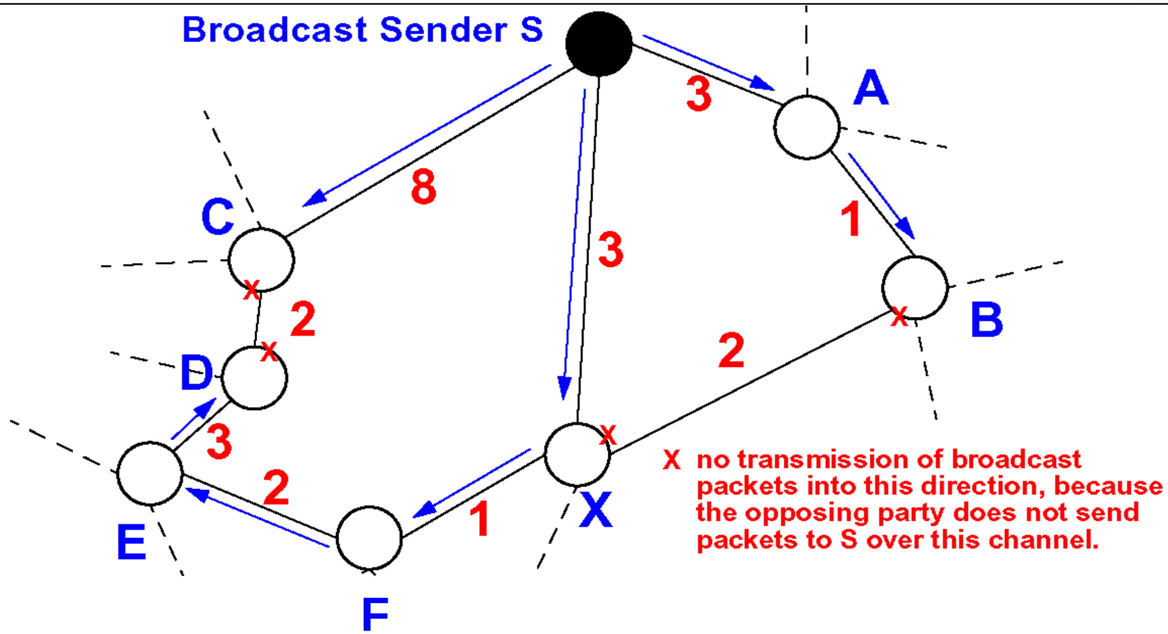
Within the (former, simpler) RPF algorithm of the above example

- Router X forwards broadcast from sender S to B, F, ...

In the example with the RPB ALGORITHM

- X DOES NOT FORWARD a broadcast packet from S to B

Broadcast Routing: Reverse Path Broadcast (RPB)



Example:

In the example with the RPB algorithm

- X does not forward a broadcast packet from S to B, because X knows that B does not receive unicast packets via X
- X sends them over a different node instead with this other node then receiving the broadcast packet

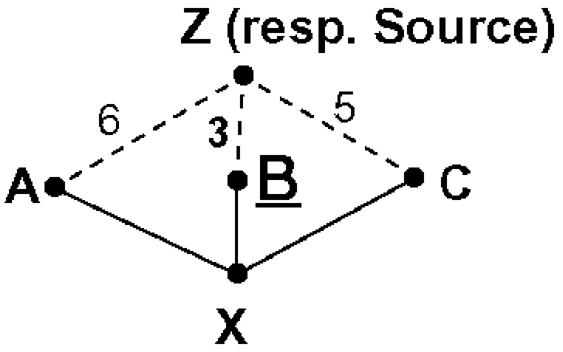
→ Connection X-B relieved in comparison to the RPF algorithm

Broadcast Routing: Reverse Path Broadcast (RPB)

Comment
- When distance is the same: IS with the shortest address is selected
- IS utilize the routing information, to exploit this parent-child relationship

(BELOW ANOTHER EXPLANATION)

Principle
- As in REVERSE PATH FORWARDING, i.e., only packets which arrived over the "best" path are forwarded, but...
- Collision avoidance (additional discarding of packets) by defining a PARENT-CHILD RELATIONSHIP
  - provided that knowledge of the Spanning Tree exists
- Or parent-child relationship:
  - IS B is the parent of the adjacent IS X, IF its distance to source Z is shorter than the distances of all other neighbours of X
  - (In the example: B is parent of X)



Broadcast Routing: Reverse Path Broadcast (RPB)

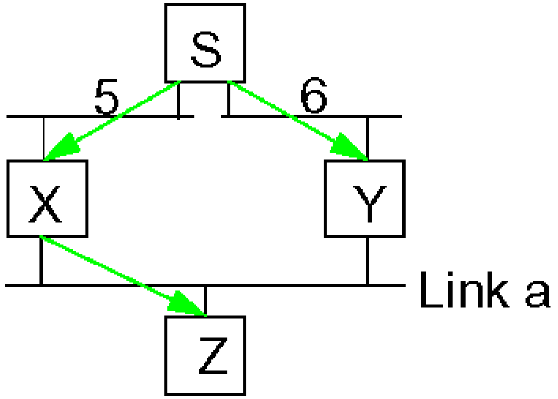
Algorithm for selecting the outgoing links

- X is the PARENT of a link, IF its distance to the source is shorter than Y's distance (or than all other ones)
- If distance is the same: decision is based on the shorter address
- Router exchange routing information with each other to determine parent-child relationship

Example

- Link 'a' is the child of X, not of Y
- Packets forwarded only over child links

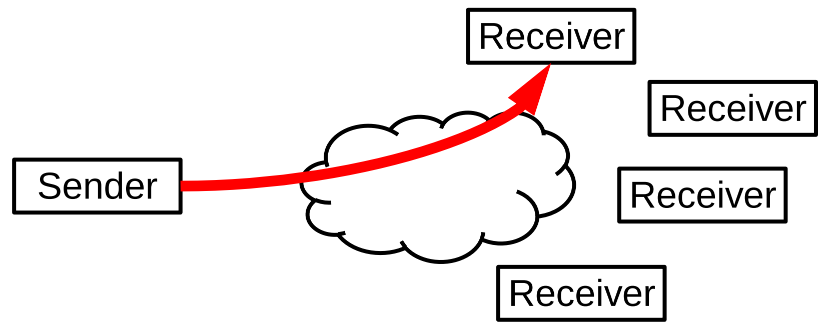
  → this results in the Spanning Tree



3 Multicast - Basics

Multicast Definition

- Unicast: 1:1 communication
- Multicast: 1:n communication

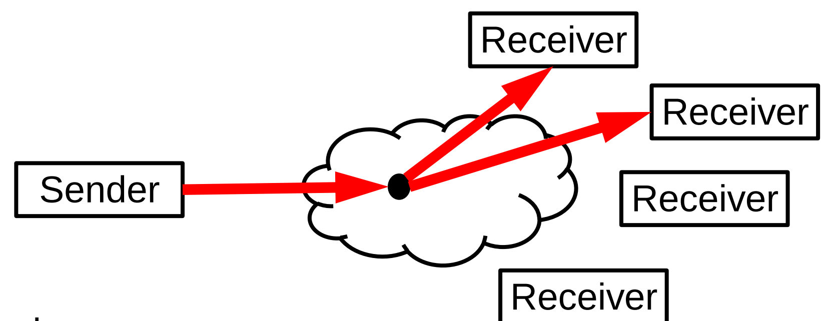


Tasks

- Send data to a group of end systems
- One-time sending instead of multiple sending
- Maintain the overall load at a low level

Results

- Lower load in the network
- Lower load at the sender



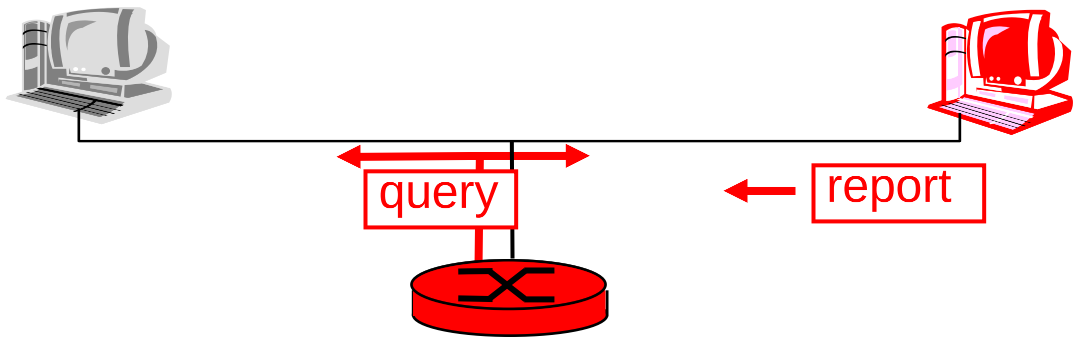
Precondition: group addressing

- Group membership may change, managed for example by:
  - Internet Group Management Protocol (IGMP)
    - Group management (create, destroy, join, leave)
  - Somehow related protocols for session maintenance
    - Session Description Protocol (SDP)
    - Session Announcement Protocol (SAP)
    - Session Initiation Protocol (SIP)

Group membership may change, managed for example by

- Internet Group Management Protocol (IGMP)
  - group management (create, destroy, join, leave)

- Somehow related protocols for session maintenance
  - Session Description Protocol (SDP)
  - Session Announcement Protocol (SAP)
  - Session Initiation Protocol (SIP)



IP v.4

- IGMP: Internet Group Management Protocol

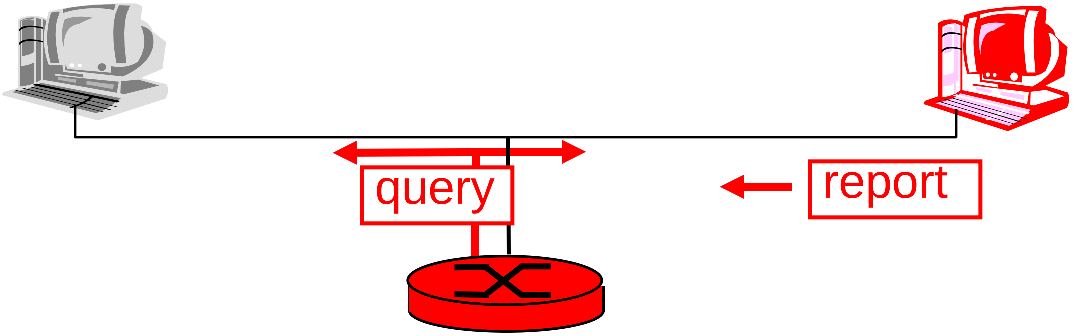
IP v.6

- Multicast Listener Discovery (MLD)
- Similar

```
Network Working Group                                          R. Vida, Ed.
Request for Comments: 3810                                    L. Costa, Ed.
Updates: 2710                                                          LIP6
Category: Standards Track                                         June 2004


        Multicast Listener Discovery Version 2 (MLDv2) for IPv6
```

IGMP: Internet Group Management Protocol



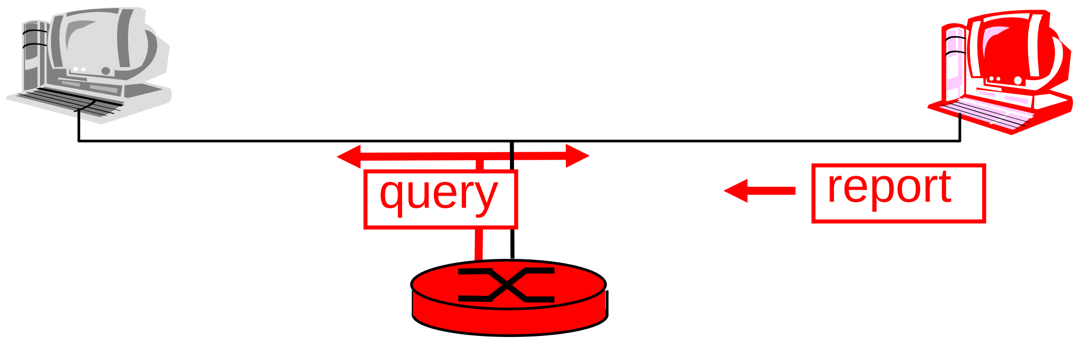
Host: sends IGMP report when application joins multicast group

- IP_ADD_MEMBERSHIP socket option
- Soft state
  - Host need not explicitly “unjoin” group when leaving

Router: sends IGMP query at regular intervals

- Host belonging to a multicast group must reply to query
  - at least one per subnet

IGMP: Internet Group Management Protocol



Router:

Host Membership Query msg

- broadcast on LAN to all hosts

Host:

Host Membership Report msg

- to indicate group membership
  - Randomized delay before responding
  - Implicit leave
    - by not replying to query

Router query

- Query may be group-specific

Host Leave Group msg

- Last host replying to Query can send explicit Leave Group msg
- Router performs group-specific query to see if any hosts left in group
- Introduced in RFC 2236

TECHNISCHE UNIVERSITÄT DARMSTADT

Some Details

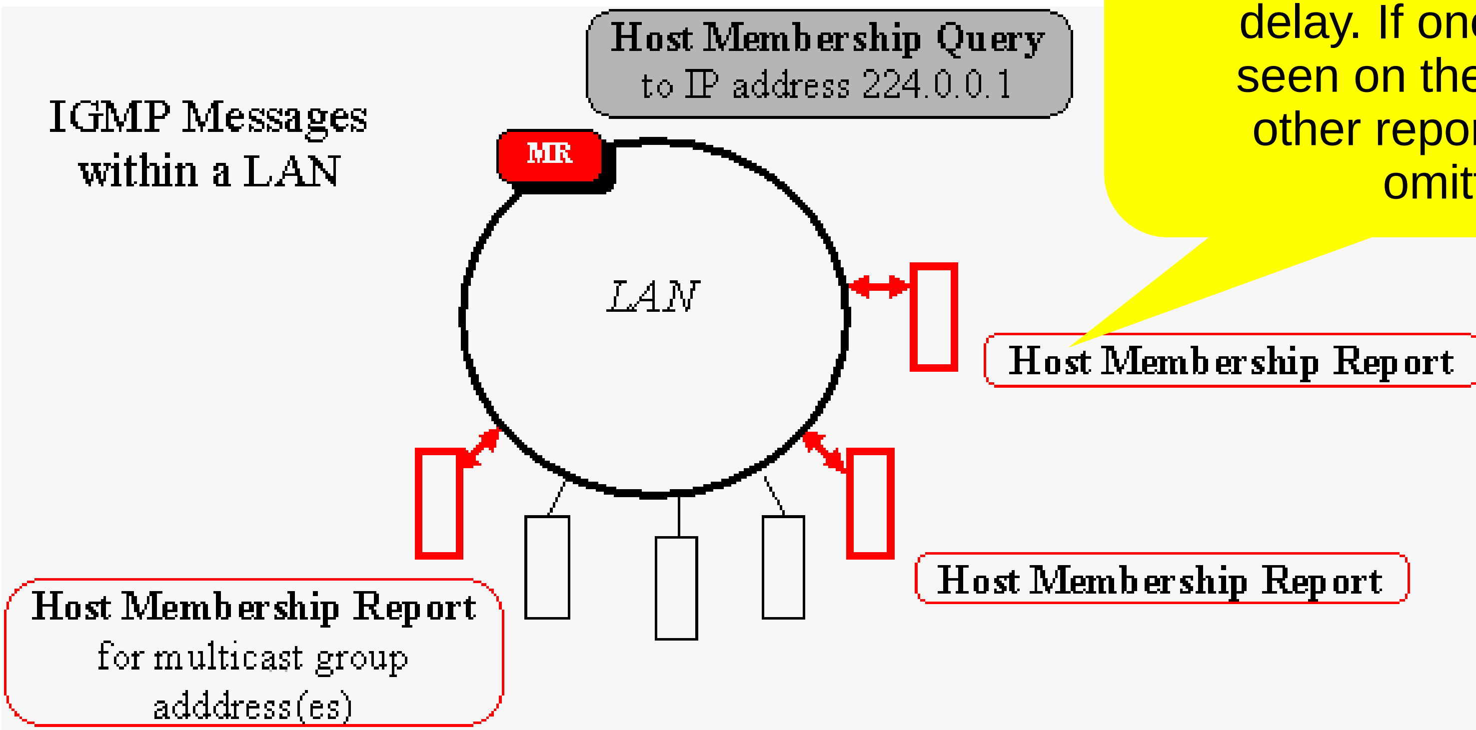
One host joins a group

- Newly joined host in a group
  - sends IGMP message to group multicast address declaring membership
- Local multicast router
  - receives the message
  - establishes necessary routing path

Group membership report

- Router sends Host Membership Query to 224.0.0.1
  - (all multicast hosts on a subnet)
- Host responds with Host Membership report for each group to which it belongs
  - sent to group address
- Other hosts in the same group “suppress” reports
- Router periodically broadcasts query to detect if groups have gone away

IGMP: Internet Group Management Protocol



Each host responds to the query with a random delay. If one report is seen on the LAN, the other reports will be omitted

IGMP: Internet Group Management Protocol

IGMP version 1:

- Basic message formats & procedures

IGMP version2:

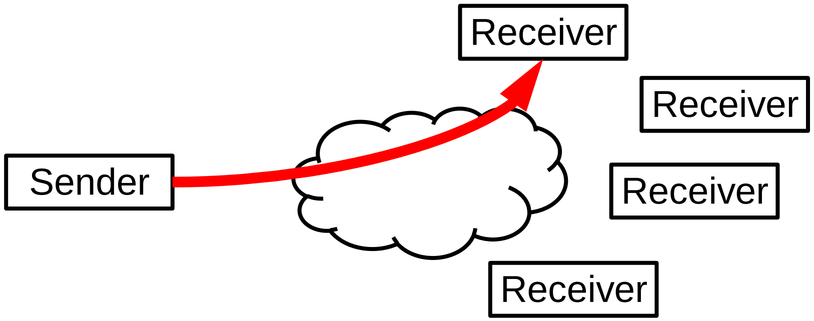
- Procedure for the election of multicast querier for each LAN
- New type of Query message-the Group-Specific Query message
- “Leave Group” message

IGMP version 3:

- Support for Group-Source Report messages
  - so that a host can elect to receive traffic from specific sources of a multicast group
- Support for Leave Group messages
  - first introduced in IGMP Version 2
  - enhanced to support Group-Source Leave messages.
- Feature allows a host to leave an entire group or to specify the specific IP address(es)
  - of the (source, group) pair(s) that it wishes to leave

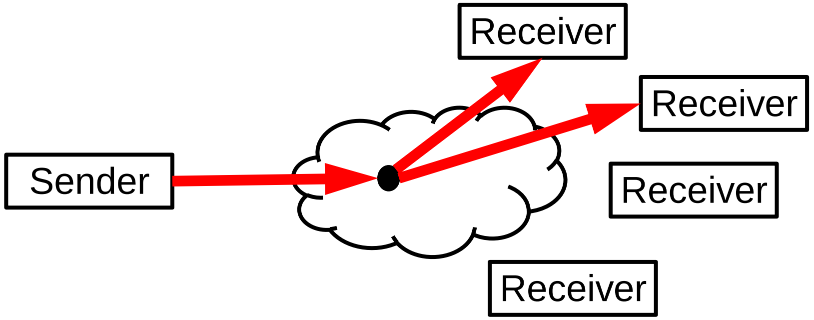
Multicast Definition

- Unicast: 1:1 communication
- Multicast: 1:n communication



Tasks

- Send data to a group of end systems
- One-time sending instead of multiple sending
- Maintain the overall load at a low level



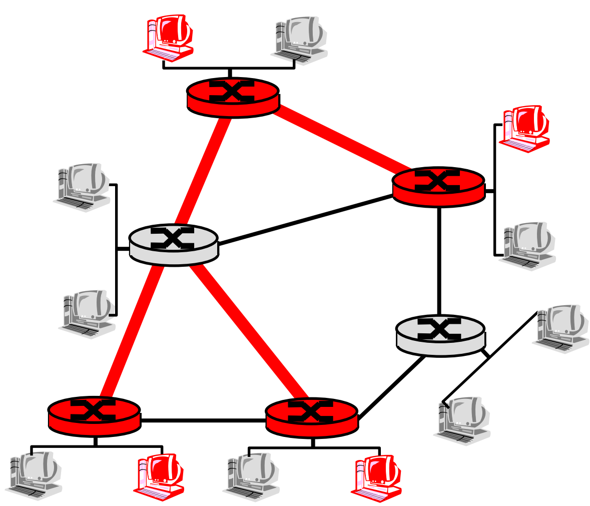
Results

- Lower load in the network
- Lower load at the sender

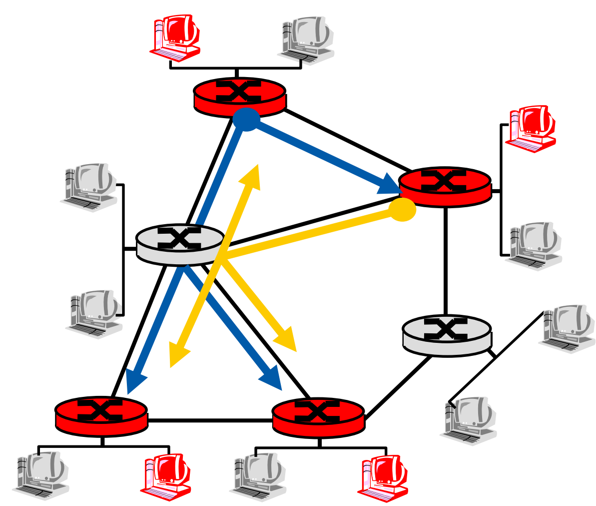
And on the backbone..

Goal: To find a tree (or trees) connecting routers having local multicast group members

- Tree: not all paths between routers used
  - Shared-tree: same tree used by all group members
  - Source-based: different tree from each sender to receivers



Shared tree

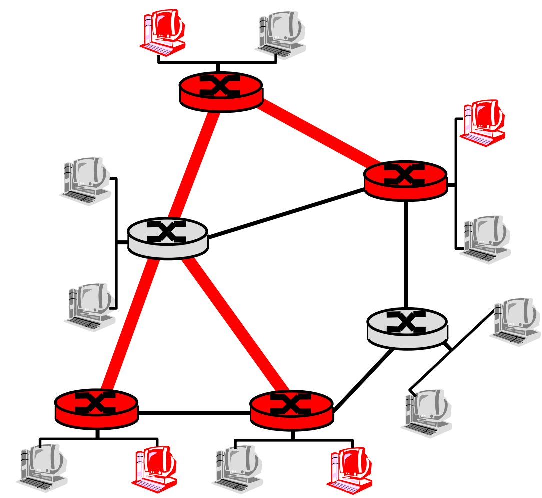


Source-based trees

Approaches for Multicast Traffic Delivery

Potential routing and delivery approaches

- Flooding: no explicit forwarding topology

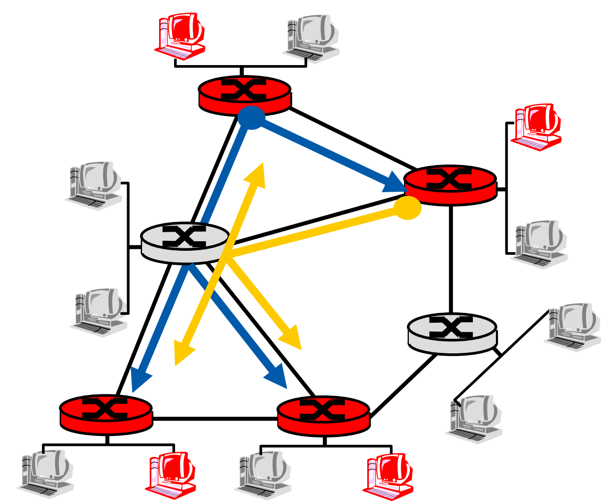


Group-shared tree: group uses one tree

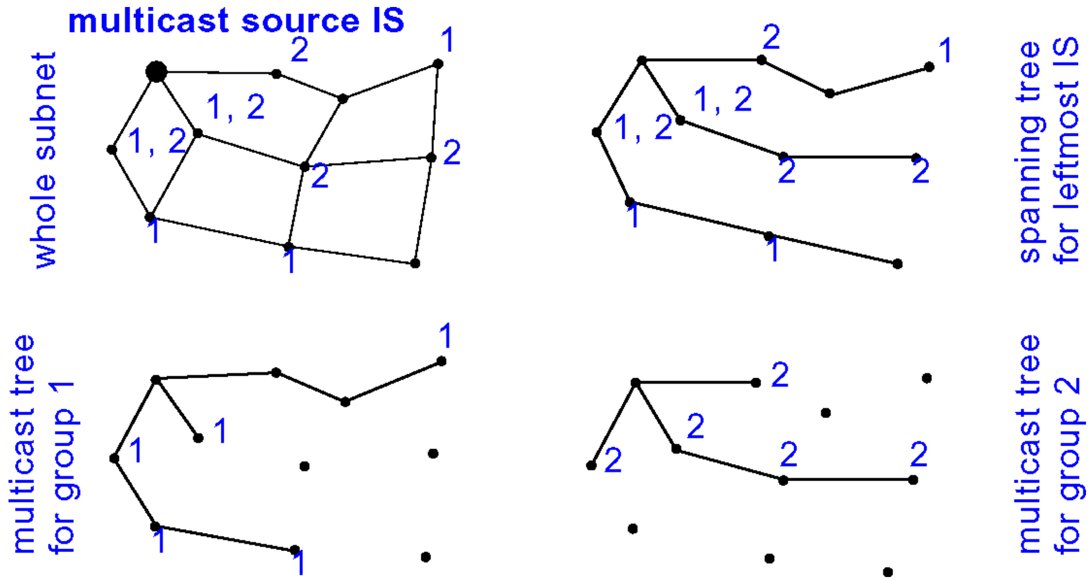
- Minimal spanning (Steiner)
- Core based trees

Source-based tree: one tree per source

- Shortest path trees
- Reverse path forwarding



5.2 Spanning Tree



Principle

- Global knowledge of the multicast group’s spanning tree (Multicast Tree)
- Initially only local knowledge

Distribution of Information

- First IS adapts spanning tree to the specific group, i.e., aligning (propagating) the spanning tree by
  - distance vector routing or
  - link state routing

5.3 Spanning Tree with Link State Routing

Basic Principle:

all IS have to know the multicast tree

I.e., each IS

- Knows to which group it belongs to
- But does not know (initially) which other IS belong to the group as well

Distribution of this information

- Depends on the underlying routing protocol
- Here: Link State Routing

Link State Packets:

A	B	C	D	E	F
Seq.	Seq.	Seq.	Seq.	Seq.	Seq.
Age	Age	Age	Age	Age	Age
B 4	A 4	B 2	C 3	A 5	B 6
E 5	C 2	D 3	F 7	C 1	D 7
	F 6	E 1		F 8	E 8

Link State Routing

All IS send link state packets periodically

- Containing information
  - distance to neighbors
  - expanded by information on multicast groups
- By broadcast to all the others

Each IS calculates a multicast tree

- From the now locally available and complete state information

Based on the information about the multicast tree

- IS determines the outgoing lines
- On which packets have to be transmitted

5.4 Spanning Tree with Distance Vector Routing

TECHNISCHE UNIVERSITÄT DARMSTADT

Basic Principle:

all IS have to know the multicast tree

I.e., each IS

- Knows to which group it belongs to
- But does not know (initially) which other IS belong to the group as well

Distribution of this information

- Depends on the underlying routing protocol
- Here: Distance Vector Multicast Routing Protocol DVMRP

Distance Vector

Distance Vector Multicast Routing

REVERSE PATH FORWARDING

- Algorithm (for a packet arriving at an IS)
  - Has this packet arrived at THE IS entry over which the packets for this station/source are usually also sent?
  - Yes:
    - Packet used the BEST route until now,
    - Action: resend over all edges (not including the incoming one)
  - No:
    - Action: discard packet (most likely duplicate)

REVERSE PATH FORWARDING WITH PRUNING

- Defined at Request For Comments 1075
- Pruning: Feedback in order to stop data transfer

Spanning Tree with Distance Vector Routing

Principle

Sender sends first multicast packet to everybody

- using the broadcast method Reverse Path Forwarding RPF

Then applies adaptation (PRUNING)

- because broadcasting too resource consuming

I.e., to adapt

- From broadcast communication
- To multicast structure

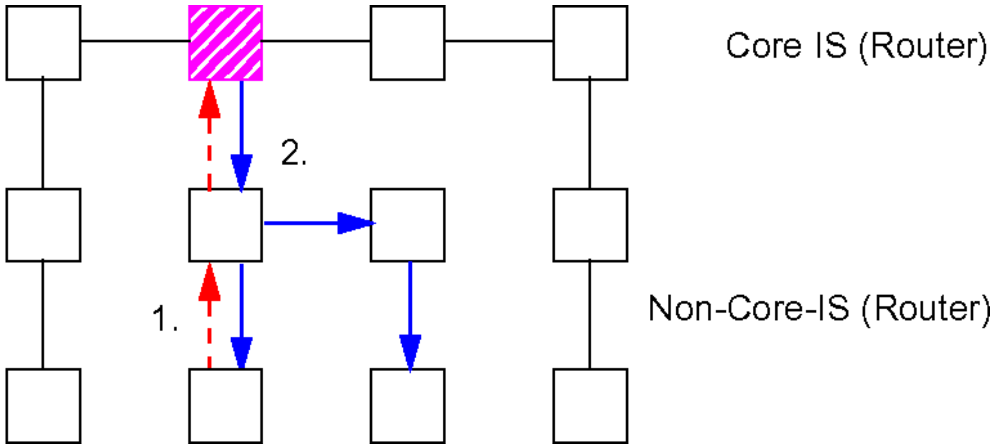
Benefit

- Pruning only on trees that are actually used
  - unused trees are cut coarsely
- Optimized for many receivers

Originating from leaves of the spanning tree:

- IF multicast packet arrives from IS leaf NOT belonging to multicast group:
  - Send a NON-MEMBERSHIP-REPORT (NMR) to the immediate predecessor
  - Propagate a Non-Membership-Report NMR, if
    - IS receives Non-Membership-Reports NMRs from all descendants
    - but does not belong to the group itself
- IF multicast packet arrives from IS leaf which DOES BELONG TO multicast group:
  - Nothing happens on the IS side
  - I.e., the following multicast packets are also send there again

5.5 Core Based Trees



Single delivery tree shared by all

One router identified as "center" of tree

Also known as "Trees with Rendezvous Points"

Principle

- The CORE is selected (an IS which is central to the group)
- The group's spanning tree from this node/IS is determined
- The sender transmits a packet to this central IS
- The core transmits this packet via the spanning tree

Properties

+ simple central calculation

+ one common tree for all n senders (instead of n trees)

- route to the central IS may not be optimized

5.6 Truncated Reverse Path Forwarding (TRPB)

TECHNISCHE UNIVERSITÄT DARMSTADT

Principle

- Enhancement of broadcast procedures "REVERSE-PATH-BROADCAST"
- Here packets are sent only on edge/leaf links which
  - contain group members
  - contain additional routers in their path (known from the message exchange between the routers)

Algorithm (when packet arriving at IS)

- Has this packet arrived from the same connection over which packets are sent to this station? (RPB)
- Yes: packet used the most favorable route up to now

  → Action:
    - send over all subnetwork edges/leaf links (not incl. the incoming one)
    - containing group members or
    - containing additional routes within their path
- No: packet has not used the most favorable route up to now

  → Action:
    - discard packet (is probably duplicate)

Comment on selecting the outgoing paths

- Recognizing leaf links by sending router messages
- Exchange membership information via IGMP
- Uncoupling of subnetworks only (not pruning procedure)

5.7 Truncated Reverse Path Forwarding (TRPB)

A) Principle

- Enhancement of broadcast procedures: "REVERSE-PATH-BROADCAST"
- Here packets are sent only on edge/leaf links which
  - contain group members
  - contain additional routers in their path
    (known from the message exchange between the routers)

C) Comment on selecting the outgoing paths

- Recognizing leaf links by sending router messages
- Exchange membership information via IGMP
- Uncoupling of subnetworks only (not pruning procedure)

B) Algorithm (when packet arriving at IS)

- Has this packet arrived from the same connection over which packets are sent to this station? (RPB)
- Yes: packet used the most favorable route up to now
  - →Action:
    - send over all subnetwork edges/leaf links (not incl. the incoming one)
      - containing group members or
      - containing additional routes within their path
- No: packet has not used the most favorable route up to now
  - →Action:
    - discard packet (is probably duplicate)

5.8 Additional Procedures & Topics

Variations (some additional ones)

- Steiner Trees (optimizing network resources)
- Reverse Path Multicast (RPM)
- Distance Vector Multicast Routing Protocol (DVMRP)
  - first version of DVMRP (RFC 1075) based on RPM
- Hierarchic DVMRP
  - two-tiered, non-overlapping domains/subnetworks
- Multicast Open Shortest Path First (MOSPF)
  - based on link state routing OSPF
- Protocol Independent Multicast (PIM)
  - for groups with small spatial density

Objectives: optimizations - constraints

- Edge optimization:
  - e.g. path with largest bandwidth
- Edge limited:
  - e.g. find a path that adheres to the constraints at every edge
- Path optimization:
  - e.g. path with the lowest overall costs
- Path limited:
  - e.g. path which does not exceed a certain overall delay

Reserving resources

- Resource Reservation Protocol (RSVP)
- Stream Protocol Version 2 (ST-2)

Quality of Service

- Negotiation
- With heterogeneous receivers (filtering)
- Adaptation (scaling)

DVMRP: Distance vector multicast routing protocol,

- RFC1075

Protocol Independent Multicast PIM

- E.g., RFC 4601
- Protocol Independent Multicast - Sparse Mode (PIM-SM) (Revised)

DVMRP: Distance vector multicast routing protocol, RFC1075

Derived from Routing Information Protocol (RIP)

- RIP forwards the unicast packets based on the next-hop towards a destination
- DVMRP constructs delivery trees based on shortest previous-hop back to the source

Flood and prune: reverse path forwarding, source-based tree

- RPF tree based on DVMRP's own routing tables
  - constructed by communicating DVMRP routers
- No assumptions about underlying unicast
- Initial datagram to multicast group flooded everywhere via RPF
- Routers not wanting group:
  - send upstream prune messages
- Since version 3, RPM is used (prior versions used TRPB)

Internet Multicasting Routing: DVMRP

Soft state: DVMRP router periodically "forgets" that branches have been pruned

- Multicast data again flows down unpruned branch
- Downstream router:
  - reprune or else continue to receive data

Routers can quickly rejoin the tree

- Following IGMP join at leaf
- Using "grafting" messages

Characteristics:

- Works well in small autonomous domains
- Commonly implemented in commercial routers
- Mbone routing done using DVMRP

Limitations:

- Distance-vector → slow to adapt to topology changes;
- Must store source-specific state even when not on tree → scaling problems

If router C can receive datagrams from both A and B, then:

- It will receive from A if A’s metric to the source is smaller than B’s, or
- If they are equal:
  - if A has the smaller IP address on its downstream interface

Initial TTL	Scope
0	Restricted to the same host
1	Restricted to the same subnetwork
32	Restricted to the same site
64	Restricted to the same region
128	Restricted to the same continent
255	Unrestricted in scope

TECHNISCHE UNIVERSITÄT DARMSTADT

Deployment of IGMP/DVMRP complicated

- (and not very successful)

→ PIM - Protocol Independent Multicast proposed as follow-up

- Not dependent on specific underlying unicast routing algorithm

Two different multicast distribution scenarios

Dense:

- Group members densely packed, in “close” proximity

Sparse:

- Group members “widely dispersed”

Sparse-Dense Dichotomy and the Consequences

Dense:

- Group members densely packed, in “close” proximity
- Bandwidth more plentiful

Sparse:

- Number of networks with group members small wrt
  - number of interconnected networks
- Group members “widely dispersed”
- Bandwidth not plentiful

I.e.

- Group membership by routers assumed until routers explicitly prune
- Data-driven construction of multicast tree (e.g., RPF)
- Bandwidth and non-group-router processing wasteful

I.e.

- No membership until routers explicitly join
- Receiver- driven construction of multicast tree (e.g., core based)
- Bandwidth and non-group-router processing conservative

TECHNISCHE UNIVERSITÄT DARMSTADT

Flood-and-prune RPF, similar to DVMRP, but

- Underlying unicast protocol provides RPF info for incoming datagram
- Less complicated (less efficient) downstream flood than DVMRP
  - reduces reliance on underlying routing algorithm
- Has protocol mechanism for router to detect it is a leaf-node router

6.4 PIM - Sparse Mode

Sparse Mode
- For case of small (or wide spread) groups
- Developed due to scaling issues
  - (flooding is generally a really bad idea…)

Based on creating routing tree for a group with Rendezvous Point (RP) as a root for the tree
- RP (core) is a focus for both senders and receivers

Explicit join model
- Receivers send Join towards the RP
- Senders send Register towards the RP

Supports both shared trees (default) AND source trees

RPF check depends on tree type
- For shared tree (between RP and receivers), uses RP address
- For source tree (between RP and source), uses source address

PIM - Sparse Mode

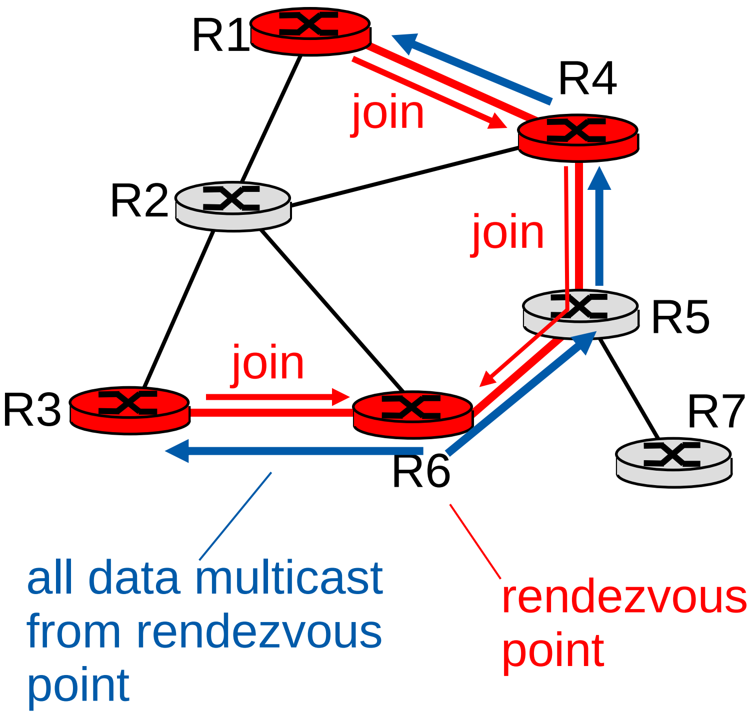
Core based approach

Router sends join msg to rendezvous point (RP)

- Intermediate routers
  - update state and
  - forward join

After joining via rendezvous point RP, router can switch to source-specific tree

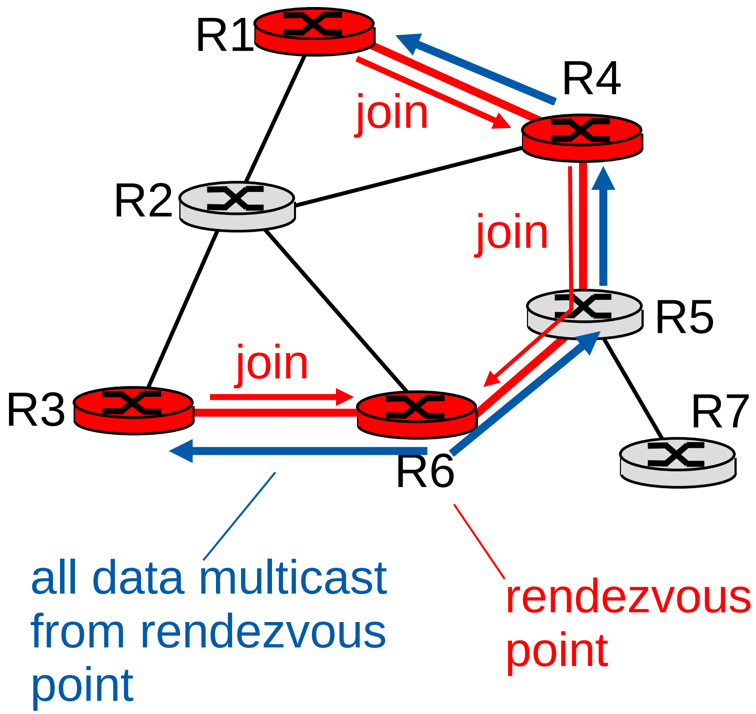
- Increased performance:
  - less concentration
  - shorter paths



TECHNISCHE UNIVERSITÄT DARMSTADT

Sender(s)

- Unicast data to rendezvous point RP which distributes down RP-rooted tree
- Rendezvous point RP
  - can extend multicast tree upstream to source
  - can send stop msg if no attached receivers (“No one is listening!”)



TECHNISCHE UNIVERSITÄT DARMSTADT

What if source is located in remote domains?

- PIM-Sparse Mode requires
  - group-RP (rendezvous point) mappings to be advertised to all PIM-Sparse Mode domains

- Use Multicast Source Discovery Protocol,
  - functionality is similar to BGP

Inter-domain multicast to be managed by Border Gateway Multicast Protocol (BGMP)

TECHNISCHE UNIVERSITÄT DARMSTADT

Multicast technology already widely deployed, but multicast service not yet turned on – why?

- Multicast is hard to manage
  →decentralized assignment of addresses!?
  →Terribly hard to debug
- There is not yet a real “killer” application
  - It only decreases cost, but does not really facilitate anything new…
- IP multicast service model not inline with requirements for commercial deployment
  - No group membership control
  - Open to denial of service attacks
  - Domain dependency
- Lack of a business models for multicast pricing