5.4 Routing for Multicast Networks

Definition of Multicast

The transmission of a data stream from one sender to multiple receivers is called **multicast**.

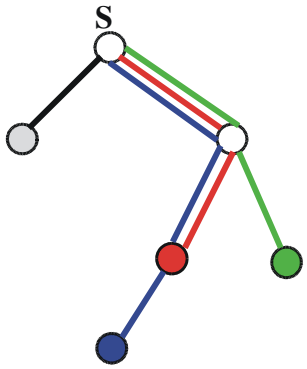
Multicast is especially important for multimedia data streams:

- Multimedia applications often require 1:n communication. Examples:
  - Video conferences
  - Tele-cooperation (CSCW) with a shared work space
  - Near-Video-on-Demand
  - Broadcast of radio and TV
- Digital video streams have very high data rates. A transmission over n point-to-point connections can eassily cause an overload of the network.

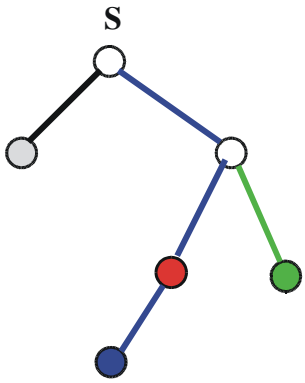
Motivation for Multicast

More „intelligence" in the inner nodes of the network reduces

- the load of the sender
- the load on the links



n end to the end connections



one multicast connection

Multicast in LANs

Ethernet, Token Ring, Wireless LAN, etc.

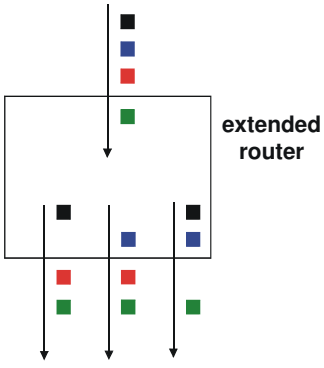
- The topology has broadcast characteristics.
- The layer-2 addresses according to IEEE 802 allow the use of group addresses for multicast. Thus, multicast can easily and efficiently be realized in a LAN segment.
- **But:** For a long time, in layer 3 and higher layers of the Internet protocol architecture, only peer-to-peer (unicast) addresses were supported!

And worldwide group communication requires multicast algorithms for wide-area traffic as well as for LANs.

Multicast in the Network Layer

Principle: Duplication of packets as "deep down" in the multicast tree as possible. Multicast in WANs requires

- a multicast address mechanism in layer 3 and more "intelligence" in layer 3 routers.
- extensions to the routing tables
- new routing algorithms

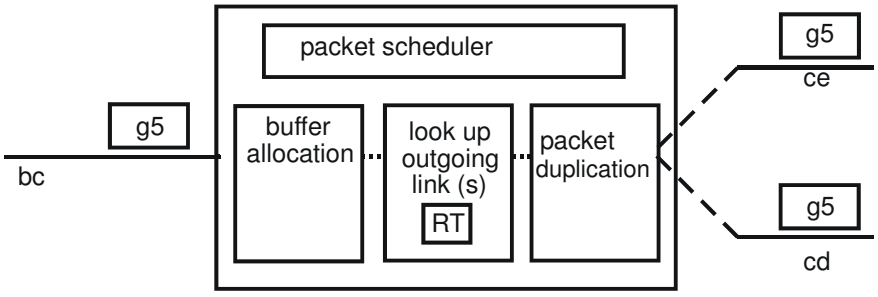


Router with Multicast Extension

RT **Routing Table**

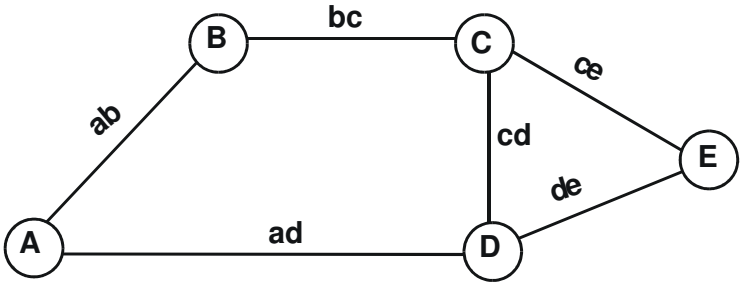
From C to	link	cost
g5	{ce, cd}	

Router at node C

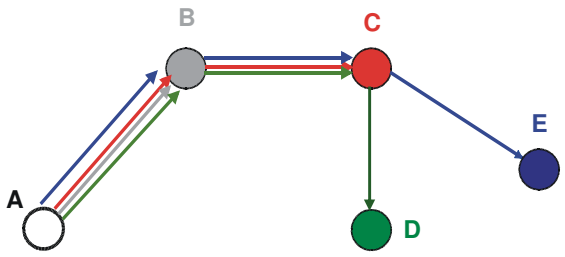


g5 = IP address of group 5

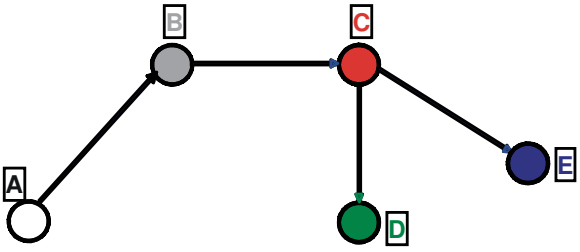
Example Topology



The Advantage of Multicast in the Example Topology



(a) four unicast connections



(b) one multicast connection

Routing Algorithms for Multicast

Multicast routing has been realized for the Internet in layer 3 (**multicast IP**).

The employed routing algorithms are extensions of the unicast routing algorithms; they are compatible with these.

Multicast in the Internet is **receiver-oriented**. For a multicast session all participants (sender and receivers) agree on a multicast address. The sender begins to send to this address. Each node in the Internet can then decide whether it would like to be included into an existing multicast session and receive the data traffic.

Principles of Multicast IP

- IP packets are transmitted to a group address (IP address of type D)
- connectionless service (datagram service)
- best-effort principle (no quality of service guarantees):
  - no error control
  - no flow control
  - no guarantee that the packet order is maintained
- receiver-oriented:
  - The sender sends multicast packets to the group.
  - The sender does not know the receivers, has no control of them.
  - Each host on the Internet can join a group.
- A restriction of the transmission range is only possible by the Time-To-Live parameter (TTL = hop counter in the header of the IP packet).

Multicast Addresses in IP

The IP group address was standardized as IP **address of class D**. It has no netid/hostid structure, is just a flat number space.

Group addresses are assigned dynamically. There is no mechanism for the unique assignment of a group address in IP! Higher layers have to take care of address assignment. For this purpose the software tool *sdr* (*session directory*) was developed.

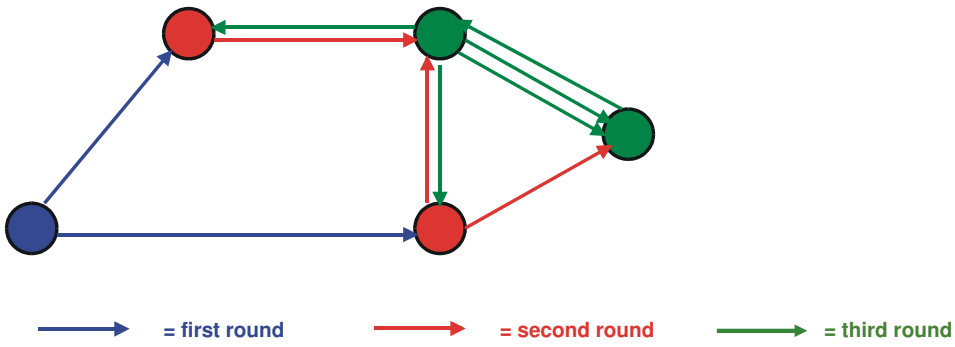
Routing Algorithms for Multicast

Flooding

The simplest possibility for reaching all receivers of a group would be Flooding (Broadcasting).

Algorithm Flooding

When a packet arrives a copy is sent to each outgoing link except the one on which it came.



Reverse Path Broadcasting (RPB)

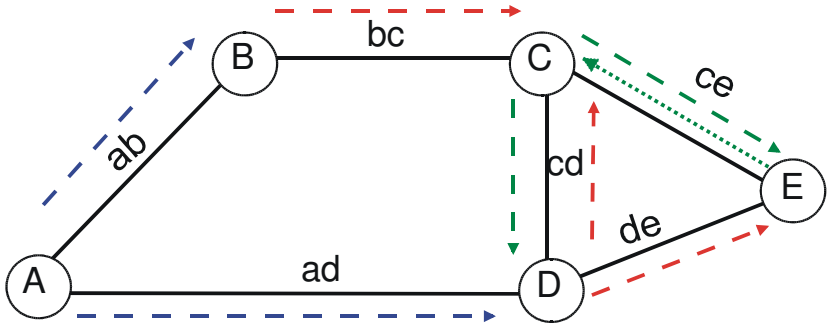
More efficient than flooding is the **Reverse Path Broadcasting** algorithm (RPB). It uses the fact that each node knows its shortest path to the sender from the classical (point-to-point) routing table! This path is called the **Reverse Path**.

The first idea is now that a node forwards only those packets to his neighbours that have arrived on the shortest path from the sender.

This algorithm generates substantially fewer packets than flooding.

Example of Reverse Path Broadcasting (incomplete algorithm)

For our example topology the (so far still incomplete) RPB algorithm works as follows:

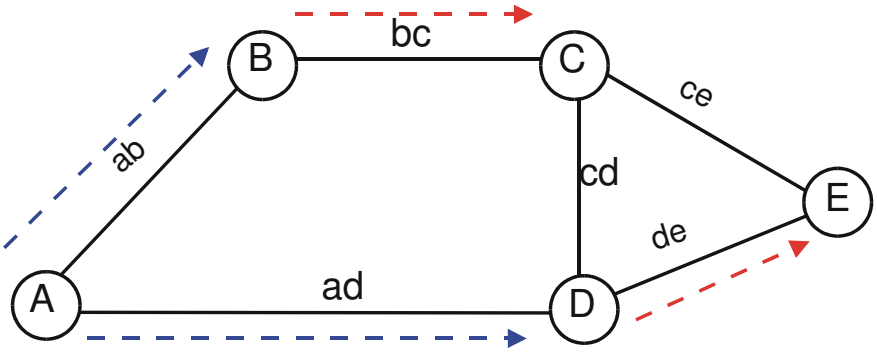


As we see, there are still redundant packets: nodes D and E receive every packet twice, node C even three times.

Reverse Path Broadcasting (complete algorithm)

If each node communicates additional information to his neighbors, RPB can prevent all redundant packets. The additional information consists of the fact whether the neighbor is on his shortest path to the sender.

In our example, E informs his neighbors C and D of the fact that D lies on his shortest path to A. A node will then forward packets only to those "sons" from whom he knows that he lies on their shortest path to the sender. The diagram below shows the packet flow for the full RPB algorithm.



Truncated Reverse Path Broadcasting (TRPB)

TRPB limits the distribution of the data to those subnetworks that contain multicast group members. Only LANs that are leaves of the routing tree are considered.

A simple group management protocol was defined for this purpose: the router asks the hosts in his LAN whether they are interested in receiving the packets of a certain group. They reply with *yes* or *no*. If a router has no interested host in his LAN, he will no longer forward packets with this group address into his LAN (**IGMP: Internet Group Management Protocol**).

Advantage

Avoids unneccessary traffic in the leaf LANs.

Disadvantage

Can eliminate only leaf subnetworks, does not reduce the data traffic at higher levels of the tree.

Reverse Path Multicasting (RPM)

It obviously makes sense to cut back the routing tree in the data phase of a multicast session so that packets are only forwarded to subtrees where there are interested receivers.

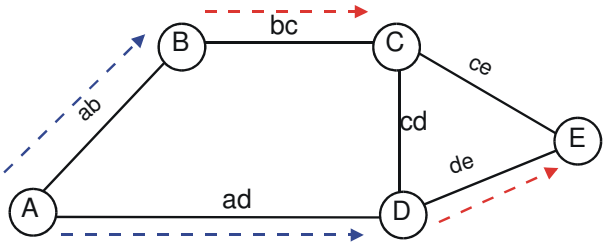
This is done with so-called "**prune messages**". They propagate from the leaves towards the root of the tree and communicate to the upstream nodes that there are no interested receivers further down in the tree. In this way a broadcast tree becomes a multicast tree. The algorithm is called **Reverse Path Multicasting (RPM)**.

In the Internet the protocol that implements the RPM algorithm is called **DVMRP (Distance Vector Multicast Routing Protocol).**

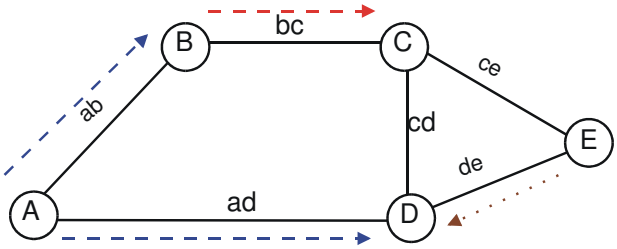
Algorithm Pruning

- A router whose sons are not interested in the multicast session sends a **Non-Membership Report** (NMR) to the upstream router, i.e., to the higher-level router in the multicast tree.
- A router who received NMRs from *all* his downstream routers send a NMR to his upstream neighbor.
- NMRs come with a timeout after which the pruning is canceled. This allows new, joining hosts at lower levels of the tree to find out about all ongoing multicast sessions.
- NMRs can also be canceled by an explicit message (the *craft* message) if a host below a link becomes interested in the session again.

Example for Reverse Path Multicasting (1)

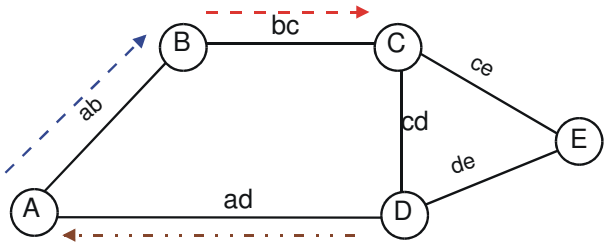


(a) tree in the initial RPB phase



(b) E has sent "prune message"

Example for Reverse Path Multicasting (2)



(c) D has sent a "prune message"

Advantages and Disadvantages of RPM

Advantage

- Reduction of data traffic compared to TRPB

Disadvantages

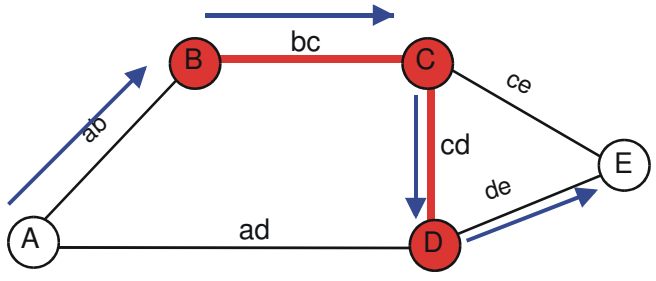
- Periodic flooding of the data to all routers is necessary so that they can reconsider their decision.
- Status information for each group and for each sender must be maintained in each node.
- For each pair (sender, group address) a separate routing tree must be created.

Core-Based Trees

All algorithms represented so far have the disadvantage that for every (sender, group) pair a separate multicast tree must be created and maintained. **Core based trees** avoid this disadvantage. Only one tree per group is built. Every sender sends to the same tree.

Today's most advanced multicast routing protocol in the Internet is called **PIM-SM** (Protocol Independent Multicast - Sparse Mode). It is based on the idea of core-based trees.

Example of a Core-Based Tree

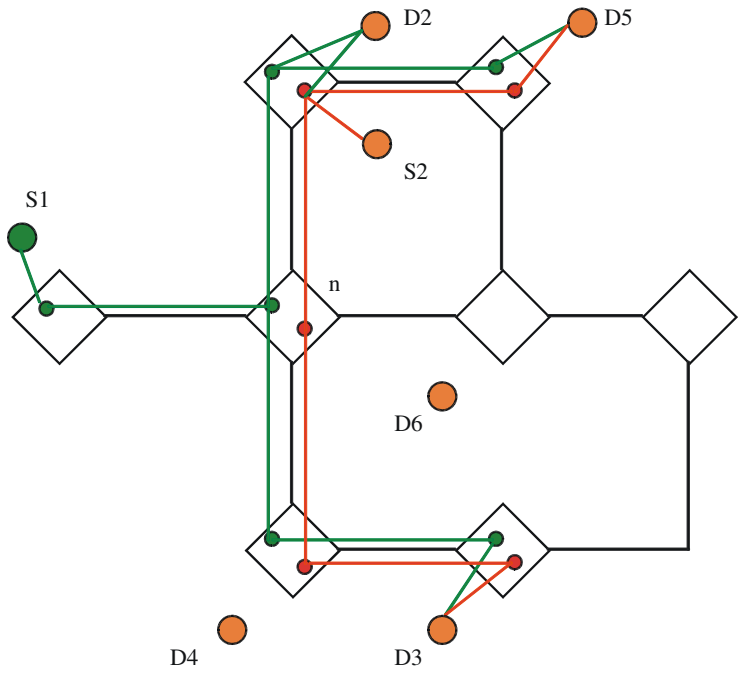


red = core tree

QoS-Based Routing

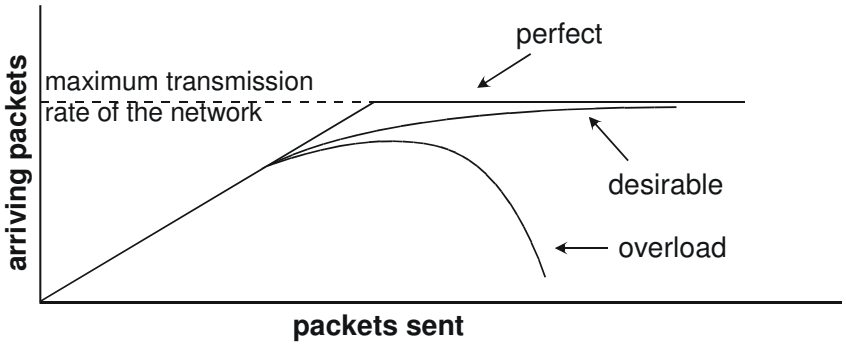
Multicast routing is still a current research topic. The problem of routing under quality-of-service constraints ("**QoS-based routing**") is still unresolved.

Dynamic Join and Leave with QoS Guarantees



5.5 Congestion Control

Problem



Reason for congestion: inner nodes too slow to route all traffic in real-time

Method 1: Reservation of Buffers (1)

Assumption: layer 3 uses virtual circuits.

- Reservation of all necessary buffers in the inner nodes when the connection is established

Example 1:

If the stop-and-wait protocol is used for flow control on the connection, one buffer at each node for each direction is sufficient.

Example 2:

If the Sliding Window protocol is used for flow control on the connection, w buffers at each node for each direction are sufficient (w = window size).

Method 1: Reservation of Buffers (2)

Characteristics

- congestion is impossible

but

- buffers remain reserved all the time exclusively for specific connections even if traffic is bursty.

Therefore, buffer reservation is only used for specific applications where short delays and a high bandwidth are necessary.

Method 2: Packet Dropping (1)

Method

- no reservation of resources
- An arriving packet is dropped if it cannot be buffered by the router.

Datagram service: No further measures necessary since layer 3 gives no delivery guarantees anyway.

Connection-oriented service: dropped packets will be retransmitted by the sender after the timeout.

Method 2: Packet Dropping (2)

Characteristics

- very easy to implement

but

- packet retransmissions waste bandwidth

Optimization 1:

A router drops those packets first that have not traveled far; he evaluates the hop counter to find out.

Optimization 2:

Random Early Discard (RED): A router begins to drop packets before the buffer (input queue) is full. The fuller the buffer gets, the higher the dropping rate will be. Dropped packets implicitly notify the sender of the congestion problem, and he will reduce the sending rate.

Method 3: Isarithmic Congestion Control

Method

Restrict the number of packets in the network by a “permit“ system:

- A number of permits circulate in the network
- A permit is required for sending:
  - When a packet is sent, a permit is destroyed.
  - When a packet arrives, a new permit is created.

Problems

- Parts of the network can be overloaded while other parts are not much utilized.
- An even distribution of the permits over the network is difficult.
- Bandwidth is needed to transfer the permits.
- Inappropriate for the transmission of bulk data (e.g., file transfer, multimedia data stream)
- Detecting and repairing the loss of permits due to errors in the network is difficult.

Method 4: Abusing Flow Control

Method

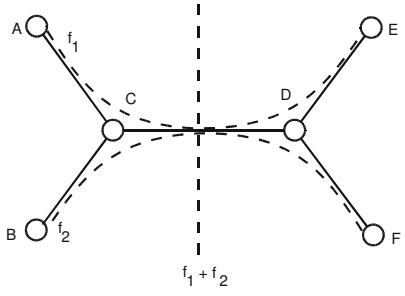
Abuse the end-to-end flow control algorithm to control congestion:

- Flow control is defined between pairs of end systems. Its purpose is to prevent overloading of the receiver by a fast sender.
- Abuse: The inner nodes of the network change the window size of the Sliding Window Flow Control Protocol. The packet flow on the connection will slow down.
- Implemented for example in layer 3 of IBM‘s Systems Network Architecture (SNA)

In the Internet a very strange variant of this idea is implemented. TCP (not UDP!) tries to guess when the network is congested. This is done on the basis of observed packet losses (missing ACKs). If the TCP sender notices congestion he voluntarily reduces his transmission rate. We will discuss the protocol details in the TCP chapter of this course.

Disadvantages of Congestion Control by Flow Control

- It is against the spirit of a layered architecture if congestion control takes place in layer 4: Layer 3 changes the flow control window parameter in the packet header of layer 4 although it should not know the protocol of layer 4.
- CC by flow control works only with connection-oriented communication, not with datagram communication! Problem: How would one implement a "TCP-friendly” flow control for UDP?
- Often several connections share a common link in the network. Will flow reduction take place in a fair way?



Method 5: Choke Packets

Method

Network management packets notify the senders of the congestion condition.

Example of an algorithm

- Each outgoing line of a router is provided with a variable u ($0 \leq u \leq 1$) indicating the current utilization.
- $u >$ threshold: the line switches to condition „warning"
- As long as the condition “warning“ is on, the router sends a „*choke*" packet for each arriving packet to the sender.
- If the sender receives a *choke* packet, it reduces the data traffic.

Variant

There are several thresholds for u leading to different warning levels. The rate reduction of the sender depends on the warning level it receives.