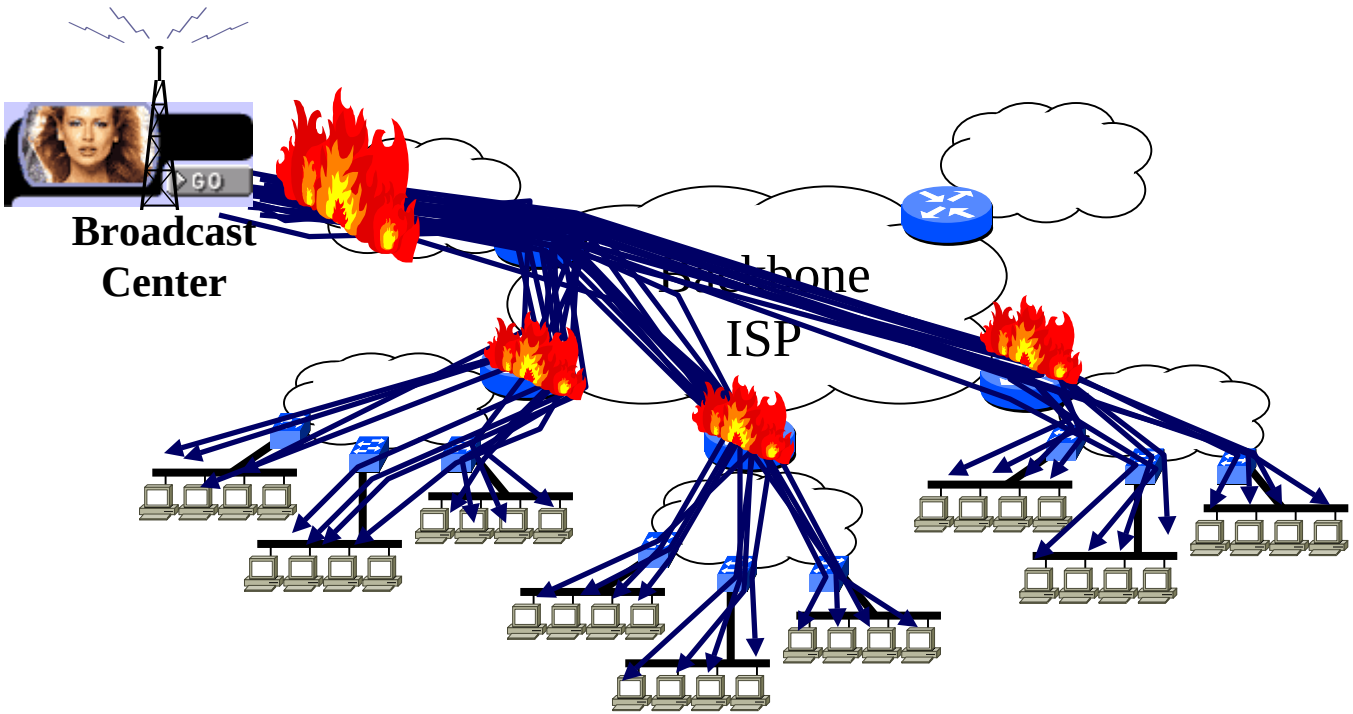
Contents

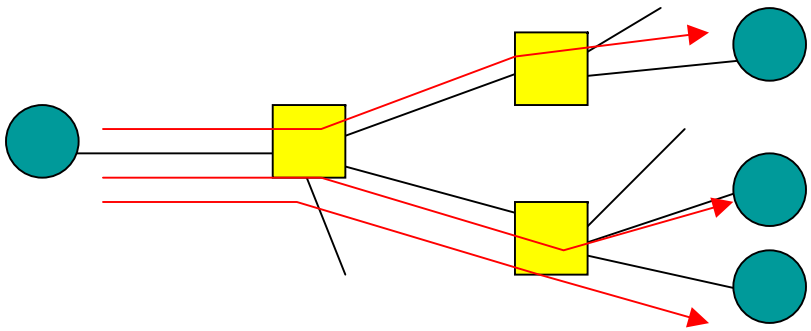
Motivation: ISPs charge by bandwidth



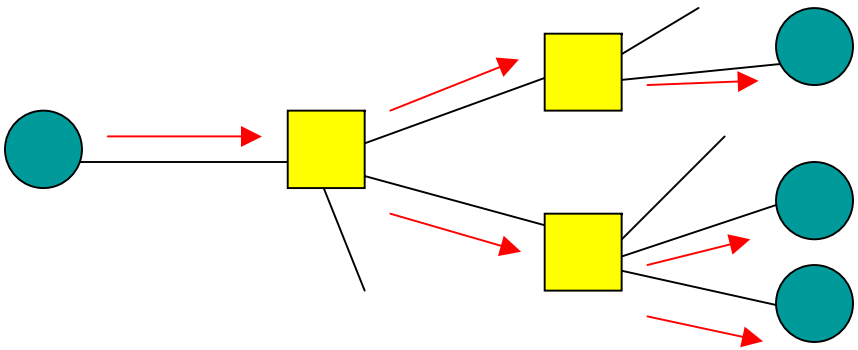
They make more money
without multicast

Overview

- Multicast = Send to a group of hosts
- Key ideas:
  - Replace multiple unicasts by a tree (good algorithms for building tree)



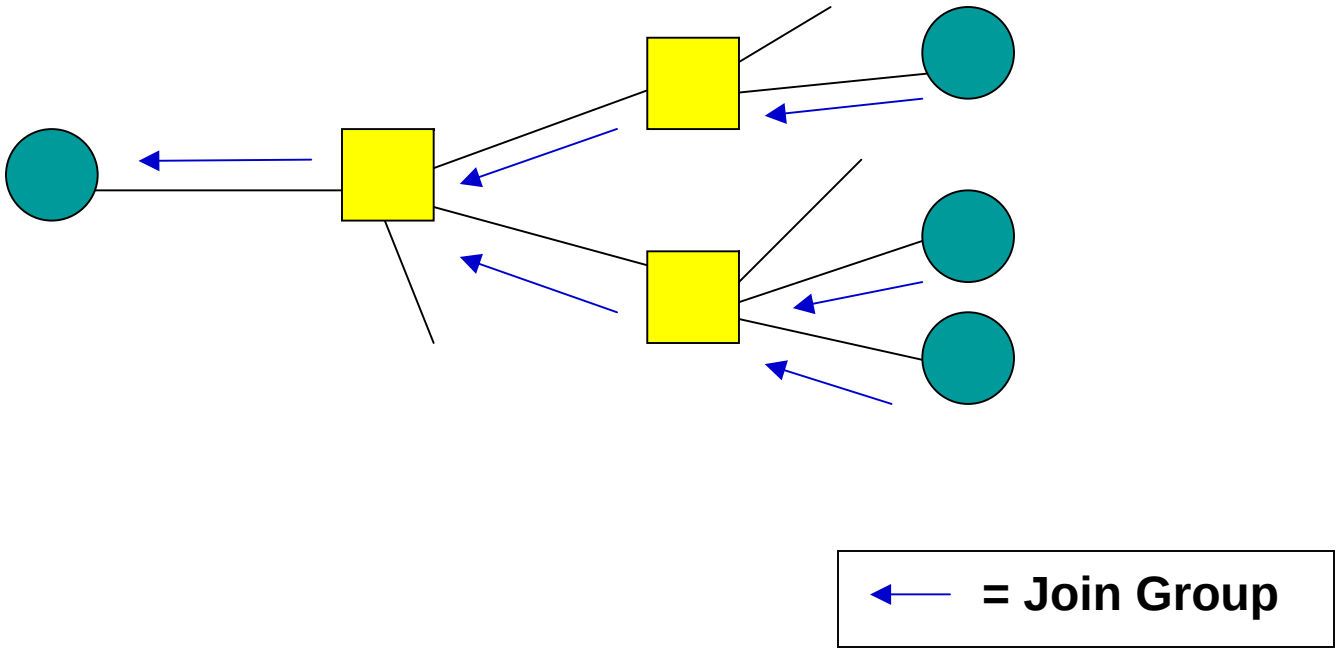
Multiple Unicasts



Multicast Tree

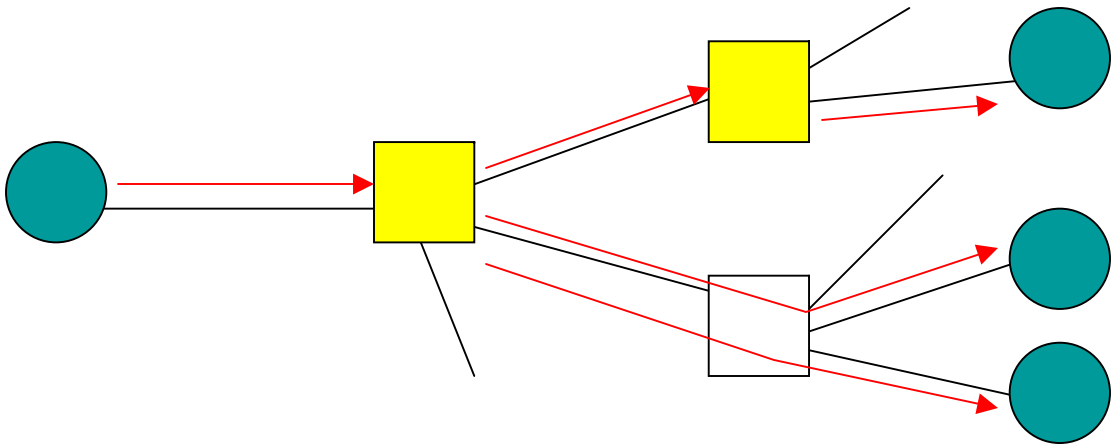
Overview

- Key ideas …:
  - Subscribe to a group; soft state



Overview

- Key Ideas…:
  - MBONE: When only a subset of routers multicast



= multicast router

= non multicast router

Implementation Issues

- Multicast can be implemented at different layers
  - data link layer
    - e.g. Ethernet multicast
  - network layer
    - e.g. IP multicast
  - application layer
    - e.g. as an overlay network like Kazaa
- Which layer is best?

Implementation Issues …

- How are multicast packets addressed?
- How is join implemented?
- How is send implemented?
  - How does multicast traffic get routed?
    - This is easy at the link layer and hardest at the network layer
- How much state is kept and who keeps it?

Ethernet Multicast

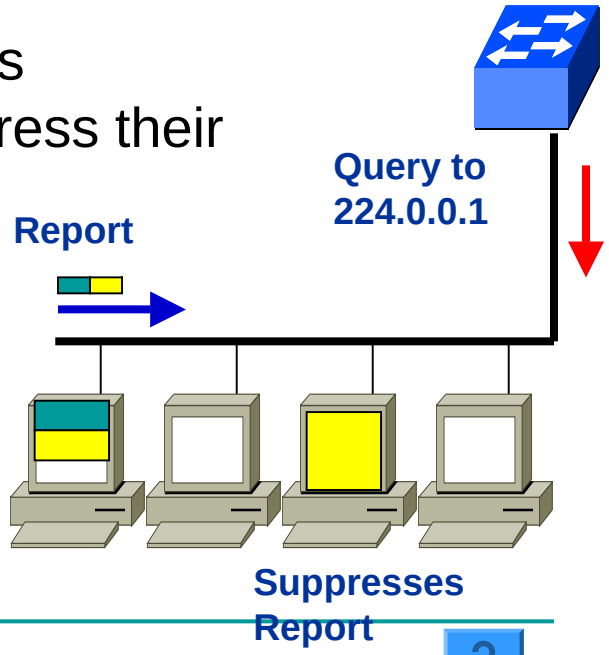
- Reserve some Ethernet MAC addresses for multicast
- To join group G
  - network interface card (NIC) normally only listens for packets sent to unicast address A and broadcast address B
  - to join group G, NIC also listens for packets sent to multicast address G (NIC limits number of groups joined)
  - implemented in hardware, so efficient
- To send to group G
  - packet is flooded on all LAN segments, like broadcast
  - can waste bandwidth, but LANs should not be very large
- Only host NICs keep state about who has joined → scalable to large number of receivers, groups
- Limited to single LAN
- More generally: Mapping of IP Group address to LAN group address

Internet Group Management Protocol IGMP

- Operates between Router and local Hosts, typically attached via a LAN (e.g., Ethernet)
  - Query response architecture

1. Router periodically queries the local Hosts for group membership information
   - Can be specific or general
2. Hosts receiving query set a random timer before responding
3. First host to respond sends membership reports
4. All the other hosts observe the query and suppress their own repots.

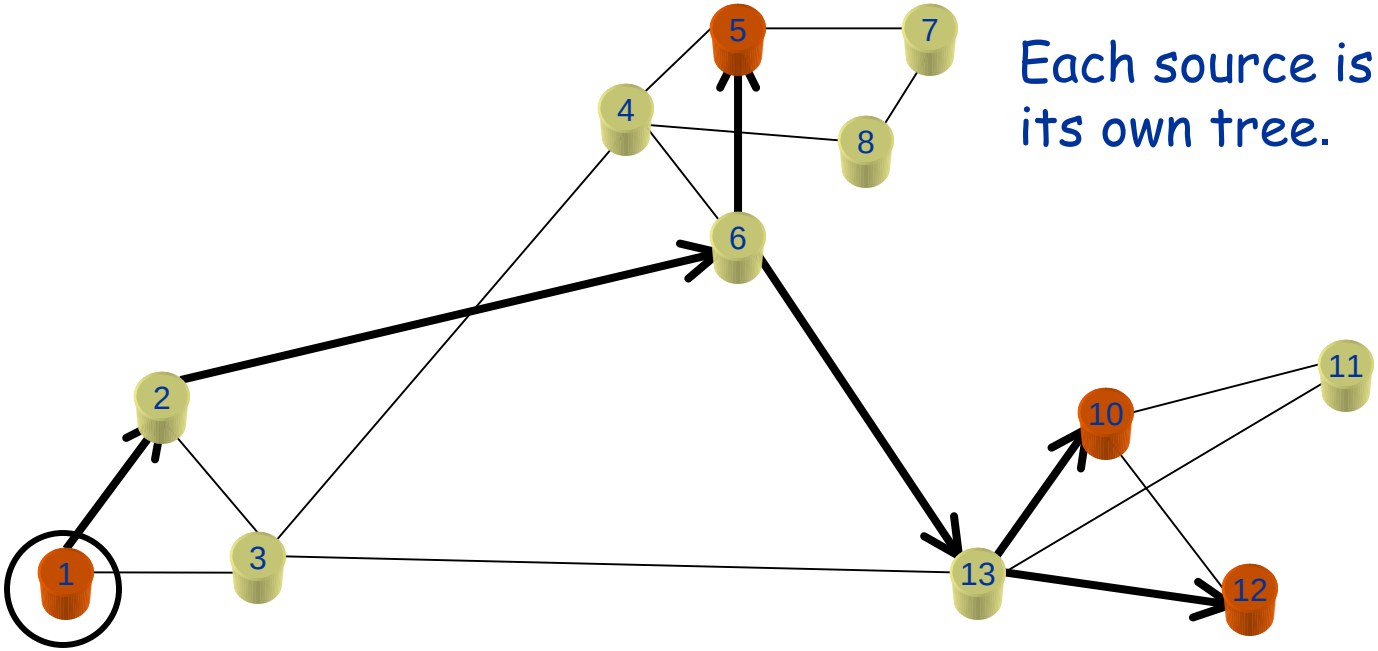
- To Join: send a group send an unsolicited Join
  - Start a group by joining it
- To leave: don't do anything
  - Soft state



Routing: Approaches

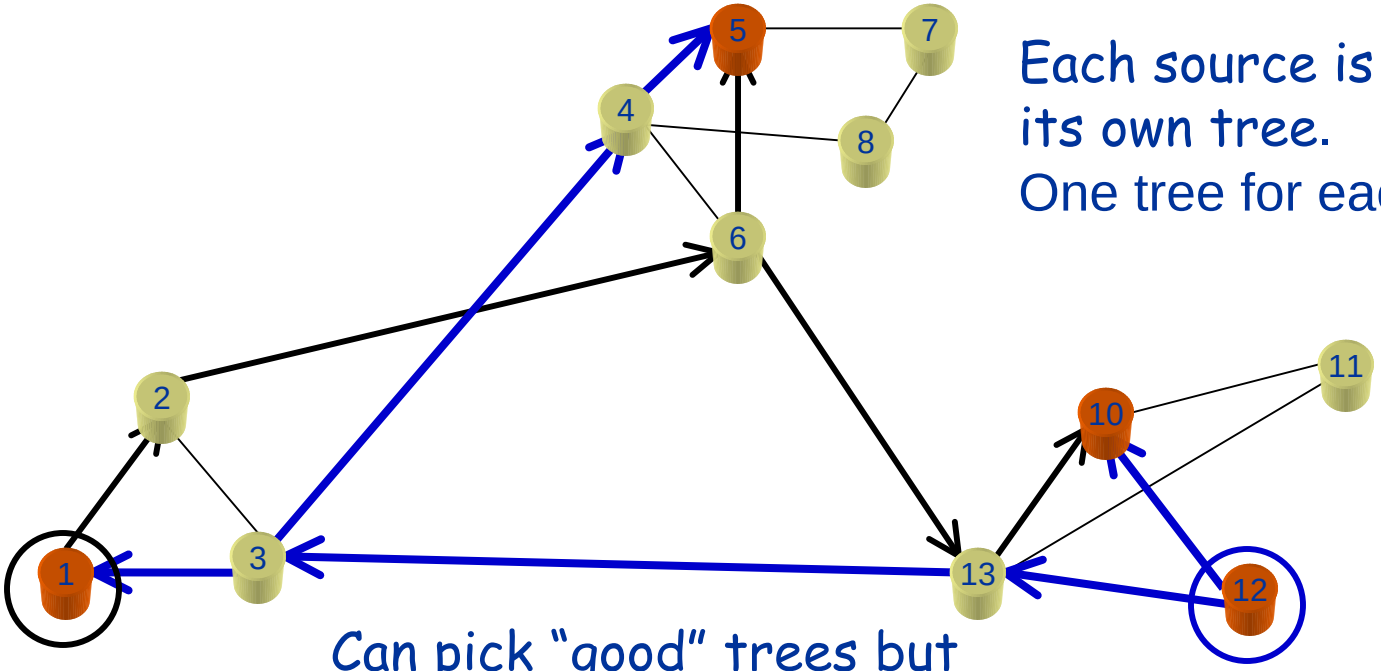
- Kinds of Trees
  - Source Specific Trees
  - Shared Tree
- Tree Computation Methods
- Intradomain
- Interdomain: BGMP
  - This is still evolving…

Source Specific Trees



Each source is the route of its own tree.

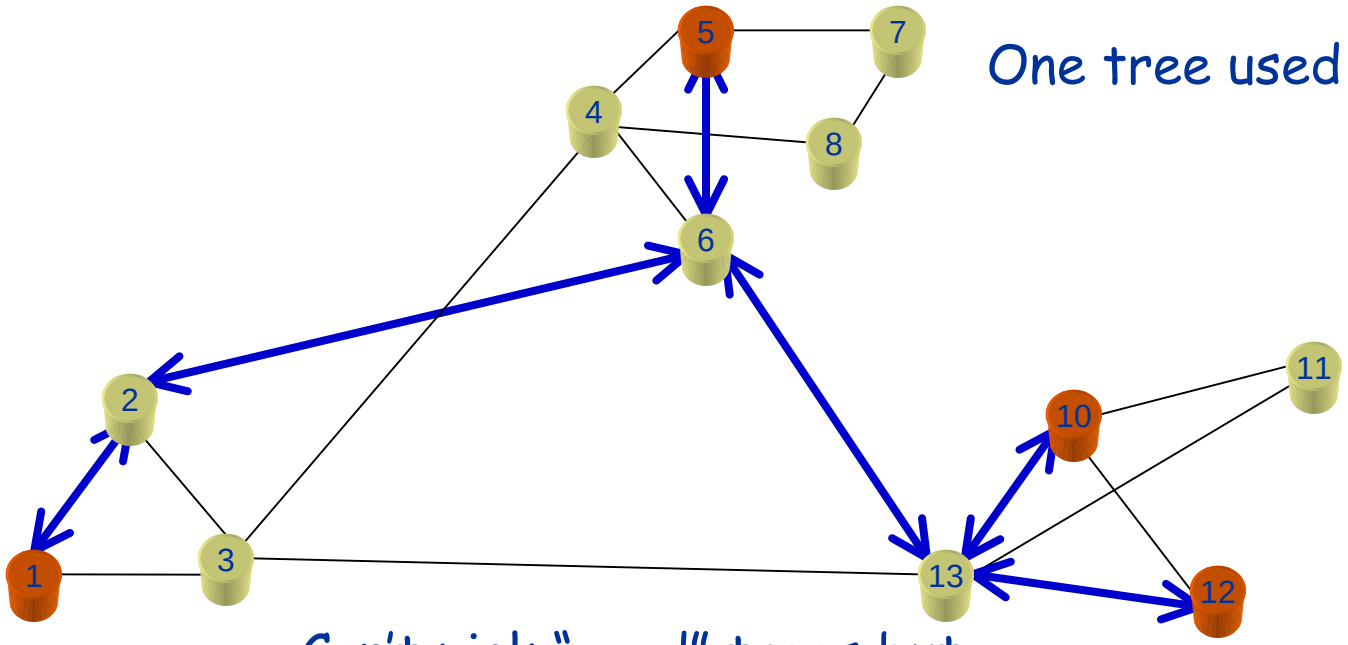
Source Specific Trees



Each source is the route of its own tree.
One tree for each source

Can pick "good" trees but lots of state at the routers!

Shared Tree

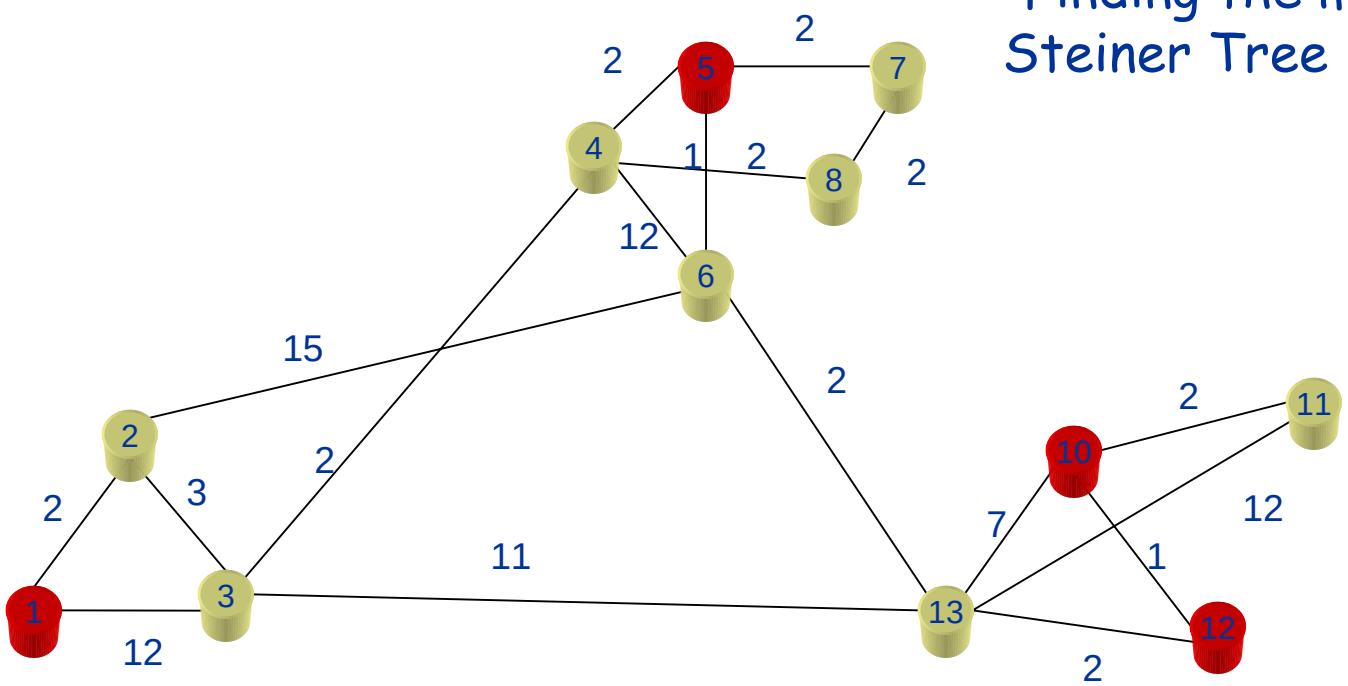


One tree used by all

Can't pick "good" trees but minimal state at the routers

Tree Computation

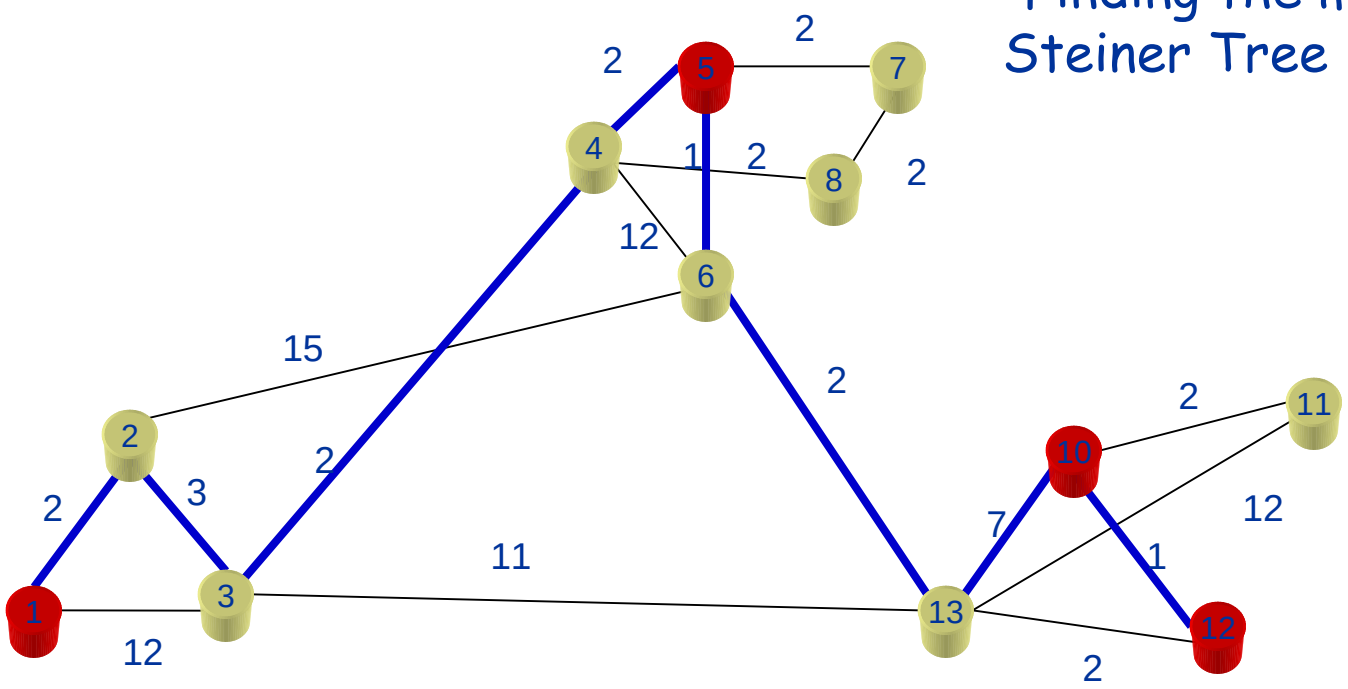
- A tree which connects all the group nodes is a Steiner Tree
- Finding the min cost Steiner Tree is NP hard



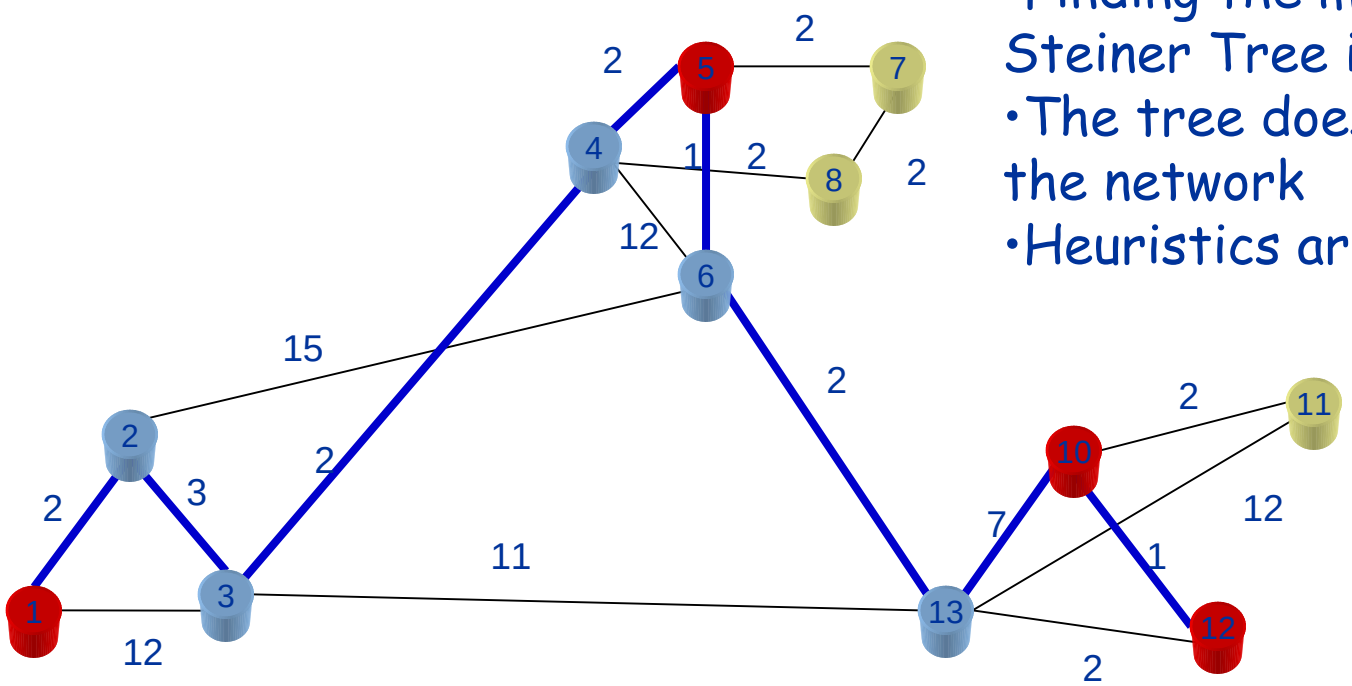
= group member

Tree Computation

- A tree which connects all the group nodes is a Steiner Tree
- Finding the min cost Steiner Tree is NP hard



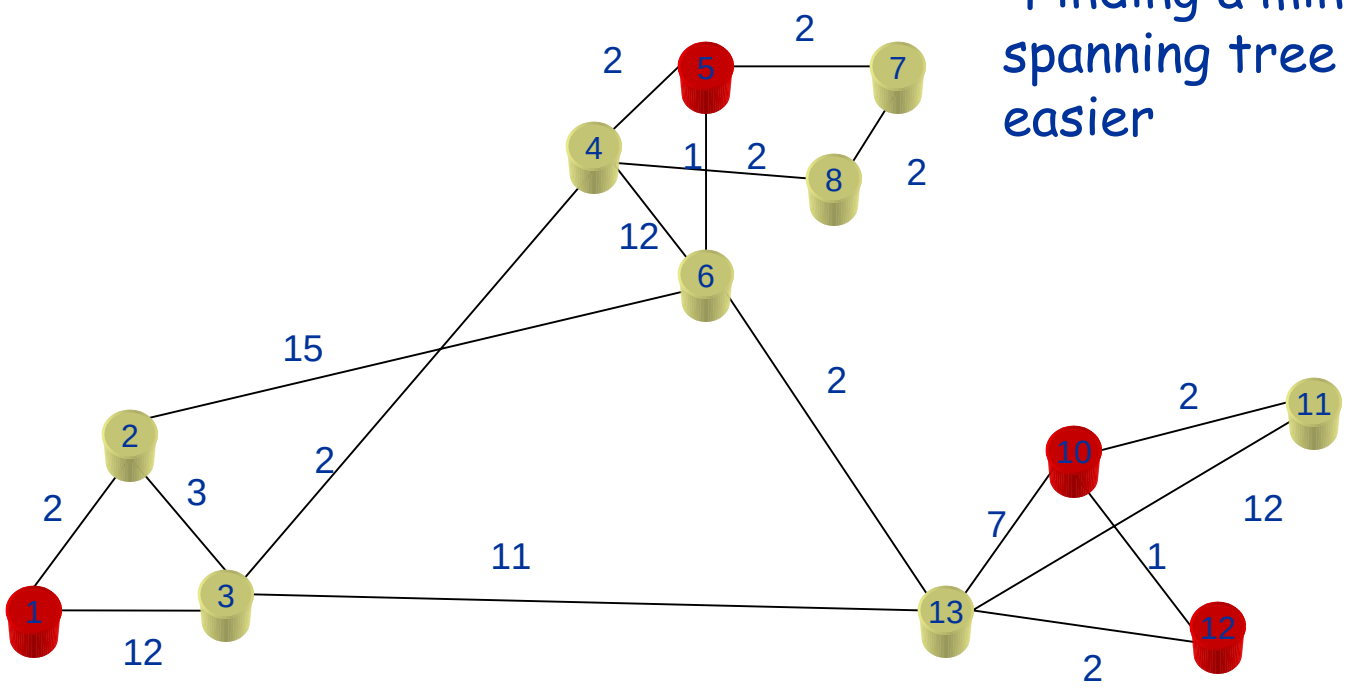
Tree Computation



- A tree which connects all the group nodes is a Steiner Tree
- Finding the min cost Steiner Tree is NP hard
- The tree does not span the network
- Heuristics are known

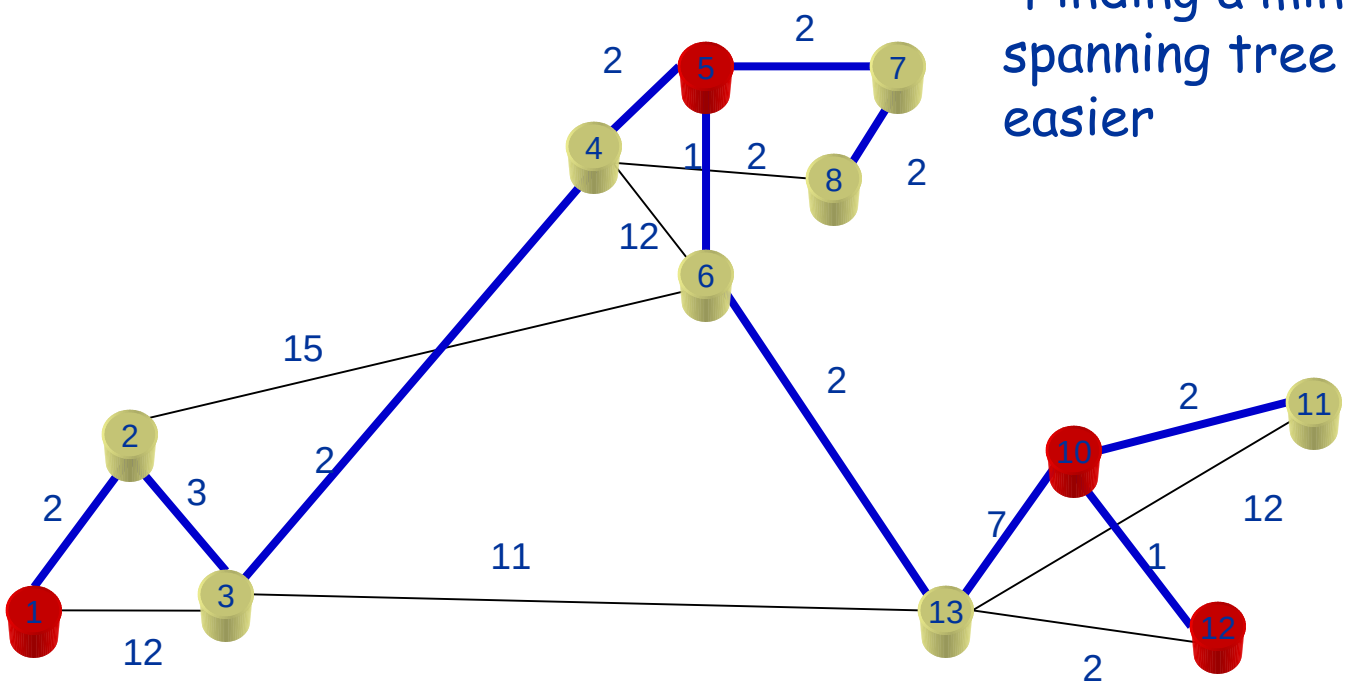
Tree Computation

- A tree that connects all of the nodes in the graph is a spanning tree
- Finding a minimum spanning tree is much easier

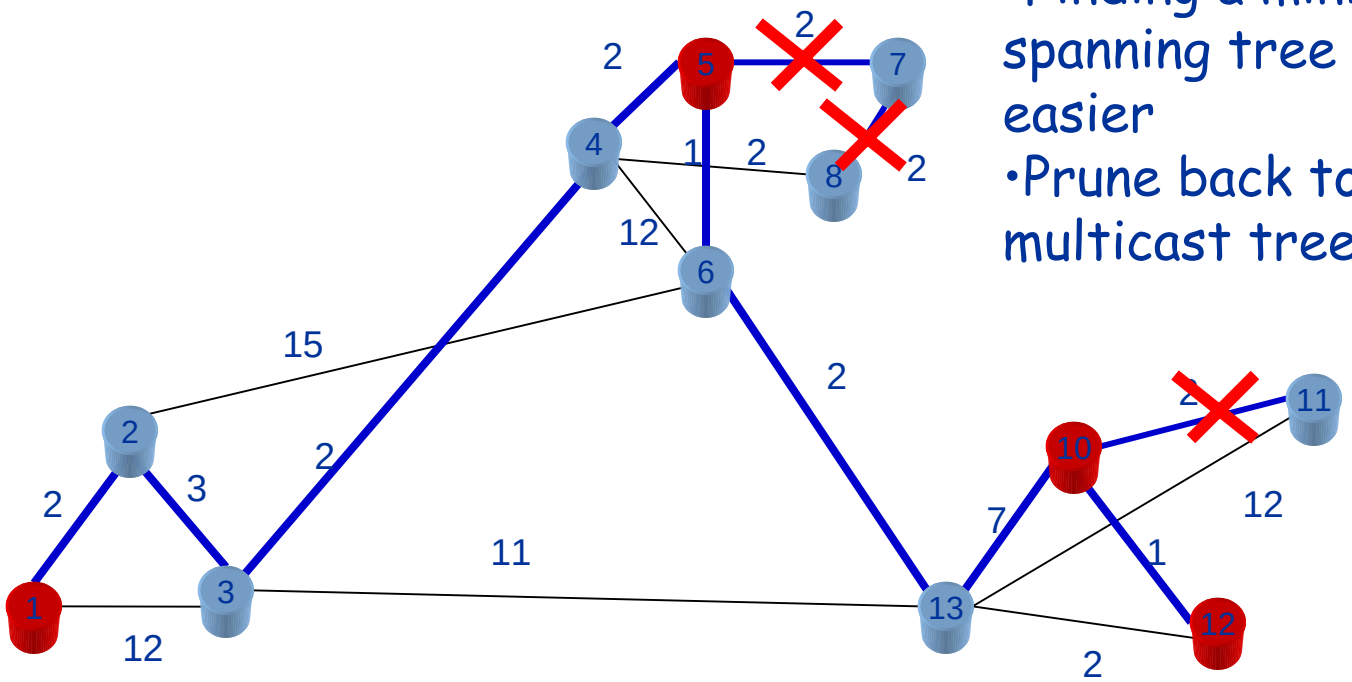


Tree Computation

- A tree that connects all of the nodes in the graph is a spanning tree
- Finding a minimum spanning tree is much easier

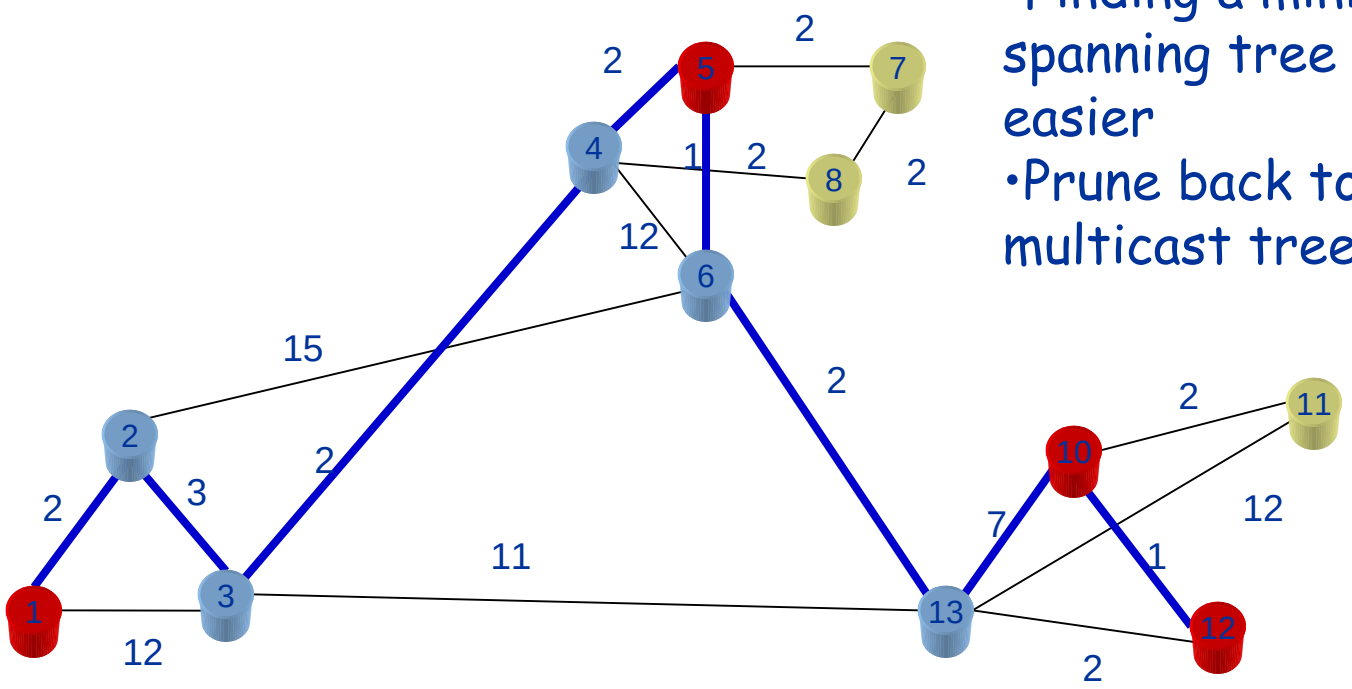


Tree Computation



- A tree that connects all of the nodes in the graph is a spanning tree
- Finding a minimum spanning tree is much easier
- Prune back to get a multicast tree

Tree Computation



- A tree that connects all of the nodes in the graph is a spanning tree
- Finding a minimum spanning tree is much easier
- Prune back to get a multicast tree

Intradomain Multicast Routing

- Build on unicast Link State: MOSPF
- Build on unicast Distance Vector: DVMRP
- Protocol Independent: PIM
- Problems

Link State: MOSPF

- Use in conjunction with a link state protocol for unicast
- Enhance the LSP updates with group membership
- Compute best tree from source
- Flood Membership in link state advertisements
- Dynamics are a problem

Distance Vector: DVMRP

- An elegant extension to DV routing
- Use shortest path DV routes to determine if link is on the source-rooted spanning tree (tree of shortest paths to source, break ties with lowest router ID)
- Prune back if no group members down the subtree
- Improve scalability: Multicast to a “shared tree” (core) and from there to rest of hosts

PIM

- Popular intradomain method
  - UUNET streaming using this
- Recognizes that most groups are very sparse
  - Why have all of the routers participate in keeping state?
- Two modes
  - Dense mode: flood and prune
  - Sparse mode: Core-based shared tree approach

Problems with Network Layer Multicast

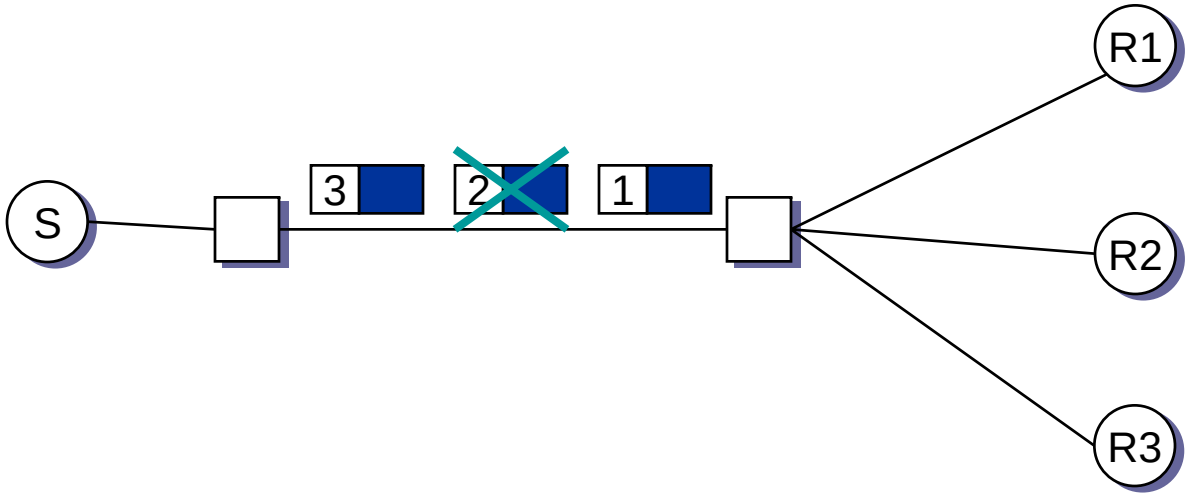
- Scales poorly with number of groups
  - A router must maintain state for every group that traverses it
  - many groups traverse core routers
- Supporting higher level functionality is difficult
  - NLM: best-effort multi-point delivery service
  - Reliability and congestion control for NLM complicated
- Deployment is difficult and slow
  - Difficult to debug problems given the service model

Reliability

- Assume reliability through retransmission
- Sender can not keep state about each receiver
  - e.g., what receivers have received
  - number of receivers unknown and possibly very large
- Sender can not retransmit every lost packet
  - even if only one receiver misses packet, sender must retransmit, lowering throughput
- N(ACK) implosion
  - described next

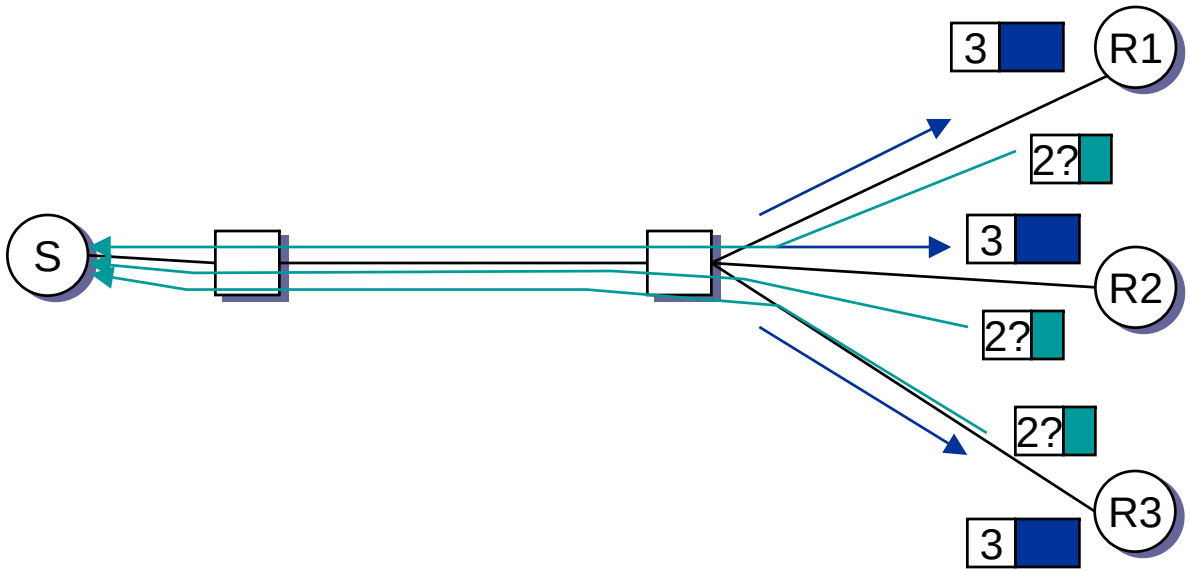
(N)ACK Implosion

- (Positive) acknowledgements
  - ack every n received packets
  - what happens for multicast?
- Negative acknowledgements
  - only ack when data is lost
  - assume packet 2 is lost



NACK Implosion

- When a packet is lost all receivers in the sub-tree originated at the link where the packet is lost send NACKs



Scalable Reliable Multicast (SRM)

- Randomize NACKs (request repairs)
- All traffic including request repairs and repairs are multicast
- A repair can be sent by any node that heard the request
- A node suppresses its request repair if another node has just sent a request repair for the same data item
- A node suppresses a repair if another node has just sent the repair

Application Layer Multicast

- Provide multicast functionality above the IP unicast
- Gateway nodes could be the hosts or multicast gateways in the network
- Advantages
  - No multicast dial-tone needed
  - Performance can be optimized to application
    - Loss, priorities etc.
  - More control over the topology of the tree
  - Easier to monitor and control groups
- Disadvantages
  - Scale
  - Performance if just implemented on the hosts (not gateways)

Summary

- Large amount of work on multicast routing
- Major problems
  - preventing flooding
  - minimizing state in routers
  - deployment
- Multicast can be implemented at different layers
  - lower layers optimize performance
  - higher layers provide more functionality
- IP Multicast still not widely deployed
  - Ethernet multicast is deployed
  - application layer multicast systems are promising