Internet Technology

10. Multicast Routing

Paul Krzyzanowski
Rutgers University
Spring 2016

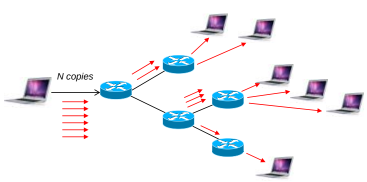
Broadcast & Multicast

- Broadcast routing
  - Deliver a packet to all nodes in the network
- Multicast routing
  - Deliver a packet to some subset of nodes in the network

N-Way Unicast

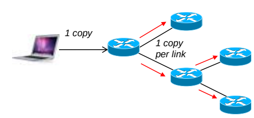
The initial sender (source node) makes *N* copies of a datagram, one for each destination node, and transmits them

- Potentially a lot of overhead: *N* copies go over the first link
- How does it know all the recipients?



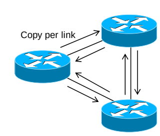
Uncontrolled Flooding

It would be more efficient for a router to make copies for each neighboring link ⇒ **In-network duplication**

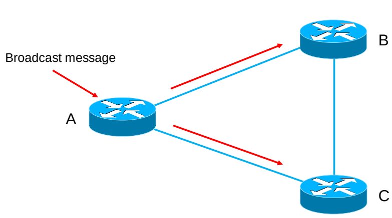


Cycles will result in a broadcast storm

- Endless replication of packets

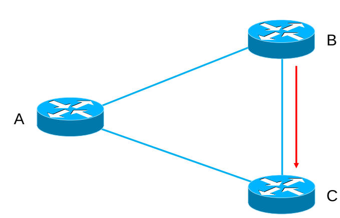


Uncontrolled Flooding: Broadcast Storm



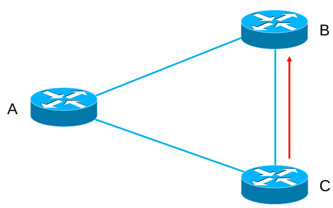
A receives broadcast & forwards to B & C

Uncontrolled Flooding: Broadcast Storm



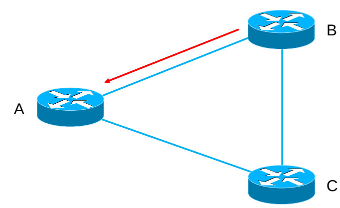
B forwards the message to C
(does not forward to A since that's where it came from)

Uncontrolled Flooding: Broadcast Storm



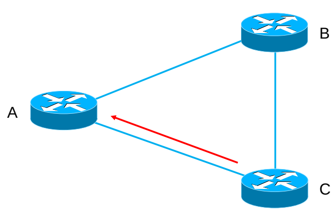
C forwards it message from A to B

Uncontrolled Flooding: Broadcast Storm



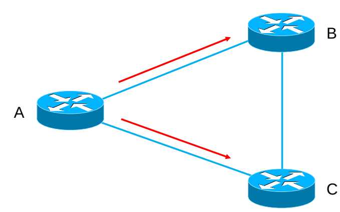
B forwards its message from C to A

Uncontrolled Flooding: Broadcast Storm



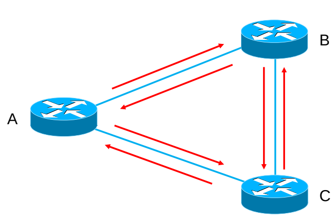
C forwards its message from B to A

Uncontrolled Flooding: Broadcast Storm



A receives a message from B: forwards it to C
A receives a message from C: forwards it to B

Uncontrolled Flooding: Broadcast Storm

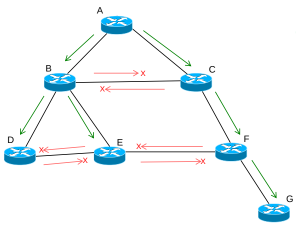


We're back where we started!

Controlled Flooding

- Sequence number controlled flooding
  - Sender places its source address and a broadcast sequence number into the packet
  - Each node keeps a list of *{source address, sequence number}* of each packet that was forwarded
  - Before copying & forwarding a packet, check the list
    - If we saw it, drop it
- Reverse path forwarding (RPF)
  - Packet is duplicated & forwarded ONLY IF it was received via the link that is the shortest path to the sender
  - Shortest path is found by checking the forwarding table to the source address

Reverse Path Forwarding (RPF)



Drop received packets that do not originate from the link for the shortest path back to the source

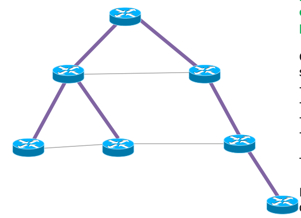
Example:
A message is initially sent to router A. It duplicates it and sends it to routers B & C.

Router B duplicates it & sends it to its outgoing links: to C, D, and E.

Each router checks whether the link that the message arrived on is part of the shortest path to the destination.

When C receives the message from B, it checks that the shortest path for the message is through A, not C. Hence, the message is rejected.

Spanning Tree Broadcast



A source node would send a packet only onto the links that are part of the spanning tree

A node only has to keep track of which of its neighbors are part of the spanning tree

Ensure that every node receives only one copy of the broadcast packet.

Create an overlay network that is a subset of the connected network:
- contains all nodes in the network
- subset of edges
- connected graph
- no cycles

This is a spanning tree

If each link has a cost
cost(tree) = sum of link costs
tree whose cost is the minimum of all possible spanning trees is a minimum spanning tree

Building a spanning tree

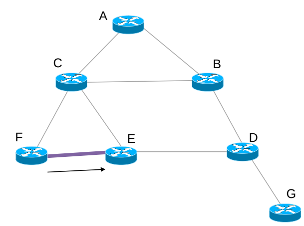
Center-based approach

- Define a center node (rendezvous point)
- Nodes send tree-join messages to this center node
  - The message is forwarded toward the center node until it
    - arrives at a node that already belongs to the spanning tree
    - or arrives at the center
  - The path that the tree-join message traverses defines a branch of the spanning tree

Example: building a spanning tree

Arbitrarily pick *E* as the center

F sends the first tree-join message
EF link becomes a branch

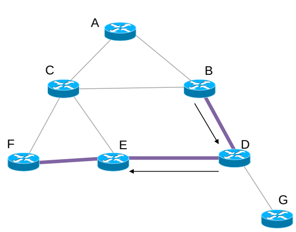


Example: building a spanning tree

Arbitrarily pick *E* as the center

F sends the first tree-join message
EF link becomes a branch

B sends the next tree-join message
BD, DE become branches
(could have been *BC, CE*)



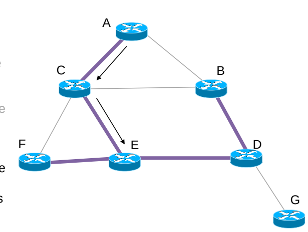
Example: building a spanning tree

Arbitrarily pick *E* as the center

F sends the first tree-join message
EF link becomes a branch

B sends the next tree-join message
BD, DE become branches
(could have been *BC, CE*)

A sends the next tree-join message
This one goes through C.
AC and *CE* become branches



Example: building a spanning tree

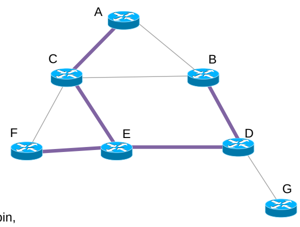
Arbitrarily pick *E* as the center

F sends the first tree-join message
EF link becomes a branch

B sends the next tree-join message
BD, DE become branches
(could have been *BC, CE*)

A sends the next tree-join message
This one goes through *C*.
AC and *CE* become branches

If a node connected to *C* needs to join, *C* is already a part of the tree.



Example: building a spanning tree

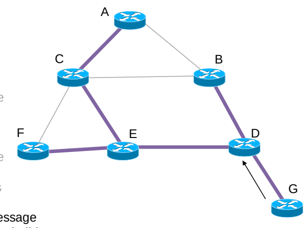
Arbitrarily pick *E* as the center

F sends the first tree-join message
EF link becomes a branch

B sends the next tree-join message
BD, DE become branches
(could have been *BC, CE*)

A sends the next tree-join message
This one goes through *C*.
AC and *CE* become branches

G sends the next (last) tree-join message
GD becomes a link (no need to build the tree to *E*, since *D* is already part of it)



IP multicast routing

Multicast routing

- Deliver messages to a subset of nodes
- How do we identify the recipients?
  - Enumerate them in the header?
    - What if we don't know?
    - What if we have thousands of recipients?

IP multicasting

- Can span multiple physical networks
- Dynamic membership
  - Machine can join or leave at any time
- No restriction on number of hosts in a group
- Machine does not need to be a member to send messages
- Efficient: Packets are replicated only when necessary

IP multicast addressing

- Use a special address to identify a group of receivers
  - A copy of the packet is delivered to all receivers associated with that group
  - Class D multicast IP address
    - 32-bit address that starts with 1110 (224.0.0.0/4 = 224.0.0.0 – 239.255.255.255)
  - Host group = set of machines listening to a particular multicast address

IP multicast addresses

- Addresses chosen arbitrarily for an application
- Well-known addresses assigned by IANA
  - Internet Assigned Numbers Authority
  - See http://www.iana.org/assignments/multicast-addresses/multicast-addresses.xml
  - Similar to ports – service-based allocation
    - For ports, we have:
      - FTP: port 21, SMTP: port 25, HTTP: port 80
    - For multicast, we have:

      224.0.0.1: all systems on this subnet
      224.0.0.2: all multicast routers on subnet
      224.0.23.173: Philips Health
      224.0.23.52: Amex Market Data
      224.0.12.0-63: Microsoft & MSNBC

IGMP

- Internet Group Management Protocol (IGMP)
  - Operates between a host and its attached router
  - Goal: allow a router to determine to which of its networks to forward IP multicast traffic
  - IP protocol (protocol number 2)
- Three message types
  - Membership_query
    - Sent by a router to all hosts on an interface to determine the set of all multicast groups that have been joined by the hosts on that interface
  - Membership_report
    - Host response to a query or an initial join or a group
  - Leave_group
    - Host indicates that it is no longer interested
    - Optional: router infers this if the host does not respond to a query

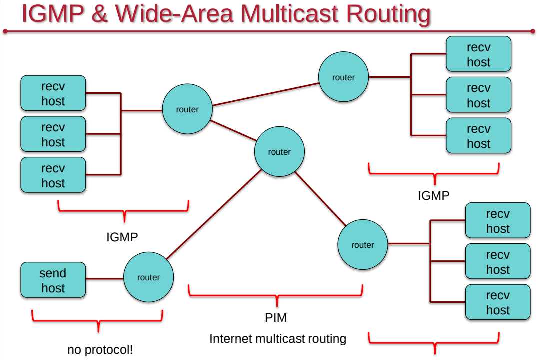
Multicast Forwarding

- IGMP allows a host to *subscribe* to a multicast stream
- What about the source?
  - There is no protocol for the source!
  - It just sends to a class D address
  - Routers have to do the work

Multicast Forwarding

- IGMP: Internet Group Management Protocol
  - Designed for routers to talk with hosts on directly connected networks
- PIM: Protocol Independent Multicast
  - Multicast Routing Protocol for delivering packets across routers
  - Topology discovery is handled by other protocols

IGMP & Wide-Area Multicast Routing



Flooding: Dense Mode Multicast

Source-based tree

- Relay multicast packet to all connected routers
  - Use reverse path forwarding (RPF) to avoid loops
  - Cutoff if there are no multicast receivers on a link
    - A router sends a *prune* message
    - Periodically, routers send messages to refresh the prune state
  - Flooding is initiated by the sender's router
- Advantage:
  - Simple
  - Good if the packet is desired in most locations
- Disadvantage:
  - wasteful on the network, wasteful extra state & packet duplication on routers

Sparse Mode Multicast

- Initiated by the routers at each receiver
  - Only network segments with receivers that joined a group will be forwarded multicast traffic
- Each router needs to ask for a multicast feed with a PIM *Join* message
  - Initiated by a router at the destination that gets an IGMP *join*
  - Spanning tree constructed
    - *Join* messages propagate to a pre-defined *rendezvous point*
    - Sender transmits only to the rendezvous point
  - A *Prune* message stops a feed
- Advantage
  - Packets go only where needed
  - Creates extra state in routers only where needed

IP Multicast in use

- Initially exciting:
  - Internet radio, NASA shuttle missions, collaborative gaming
- But:
  - Few ISPs enabled it
  - For the user, required tapping into existing streams (not good for on-demand content)
  - Industry embraced unicast instead

IP Multicast in use: IPTV

- IPTV is emerging as the biggest user of IP multicast
- Cable TV systems: aggregate bandwidth ~ 4.5 Gbps
  - Video streams: MPEG-2 or MPEG-4 (H.264)
  - MPEG-2 HD: ~30 Mbps
  - MPEG-4 HD: ~6-9 Mbps; DVD quality: ~2 Mbps

IP Multicast in use: IPTV

- Traffic is within the provider's network
  - QoS: typically mix of ATM and/or IP
    - 2.5 Mbps VBR video
    - 256 kbps CBR voice
    - Remainder: ABR for IP traffic
  - Unicast for video on demand
  - Multicast for live content
    - When you select a channel, you join a multicast group via IGMPv2
      - Local office checks if you are authorized.
      - If yes, routers add the user to the group
    - Burst of unicast data to get the I-frame to ensure 150 msec channel switching times.
  - Multicast for
    - STB system integration, music on hold, conferencing

The end