Broadcast and Multicast Routing

HIT

Shimritd Tzur-David
Distinction

• Broadcast Routing –
  – Delivering a packet sent from a source node to all other nodes in the network.

• Multicast Routing –
  – Delivering a packet sent from a source node to a subset of the other nodes in the network.
Broadcast Routing: N-way-unicast

• There are N destinations.
• The source sends a copy of the packet to each destination.
• The source node makes N copies of the packet addresses each copy to a different destination.
• The source transmits the N copies to the N destinations using unicast routing.
N-way-unicast Drawbacks

• Drawback:
  – Inefficiency: if the source node is connected to the rest of the network via single link, N separate copies of the same packet will traverse this single link.

• Solution:
  – Sending only a single copy of the packet over the first hop.
Source-duplication versus in-network duplication. (a) source duplication, (b) in-network duplication
N-way-unicast Drawbacks – Cont.

- Drawback:
  - An assumption of the N-way-unicast: The sender knows the addresses of the recipients.

- Solution:
  - destination-registration protocol.

- Solution’s drawback:
  - Overhead, complexity

- Drawback:
  - Routing algorithms use multicast to create and update unicast routes.

Any Ideas?
Broadcast Routing: Uncontrolled Flooding

• The source node sends a copy of the packet to all of its neighbors.
• When a node receives a broadcast packet, it duplicates the packet and forwards it to all of its neighbors (except the sender).
• If the graph is connected, at the end of the process every node in the graph will have a copy of the packet.

Any Problems?
Uncontrolled Flooding Drawbacks

• Drawback:
  – Graph with cycles: the packet will cycle indefinitely.
Drawbacks – Cont.

• Worse Drawback:
  – When a node is connected to more than two other nodes, it will create and forward multiple copies of the broadcast packet.
  – Each destination with more than two neighbors will create multiple copies of the packet and so on.
  – The result: endless multiplication of broadcast packet.
  – This is called broadcast storm.

Again, Ideas?
Broadcast Routing: Controlled Flooding

• **Sequence number controlled flooding:**
  – A source node puts its address and broadcast sequence number into a broadcast packet.
  – The source sends the packet to all of its neighbors.
  – Each node maintains a list of the source address and sequence number of each broadcast packet it has received.
  – When a node receives a broadcast packet, it checks whether the packet is in the list. If so, the packet is dropped, otherwise, the packet is duplicated and forwarded.
Broadcast Routing: Controlled Flooding

• Reverse path forwarding (RPF):
  – Router transmits the packet to all of its outgoing links only if the packet arrived on the link that is on its own shortest path back to the source.
  – Otherwise, the router discard the packet.
  – Each router has to know the next neighbor on its unicast shortest path to the sender.
Reverse Path Forwarding Example

Packet will be forwarded

Packet not forwarded beyond receiving router
Does Controlled Flooding Work?

• Both sequence number controlled flooding and RPF avoid broadcast storms.
• Are there any transmissions of redundant broadcast packets?

Nodes B, C, D, E and F receive either one or two redundant packets.

**Ideally, every node should receive only one copy of the broadcast packet.**
Broadcast Routing: Spanning-Tree

• Spanning tree definition:
  – A **spanning tree** of a graph $G=(N,E)$ is a graph $G’=(N,E’)$ such that $E’$ is a subset of $E$, $G’$ is connected, $G’$ contains no cycles, $G’$ contains all the original nodes in $G$.
  – If each link has an associated cost and the cost of a tree is the sum of the links costs, then a spanning tree whose cost is the minimum of all the graph’s spanning trees is called a **minimum spanning tree**.
Spanning Tree Example

• G’ is the tree consisting of the nodes connected by thick lines.
• If broadcast packets were forwarded only along links within G’, each and every network node would receive exactly one copy of the broadcast packet.
Spanning Tree Solution

- The network nodes first construct a spanning tree.
- A source node sends a broadcast packet out on all of the incident links that belong to the spanning tree.
- A node receiving a broadcast packet forwards the packet to all its neighbors in the spanning tree.
- In this way, we eliminate redundant broadcast packets.
- Note:
  - A node need not to be aware of the entire tree. It simply needs to know which of its neighbors in G are spanning-tree neighbors.
Broadcast along a spanning tree

Once the spanning tree is in place, it can be used by any node to begin the broadcast.
Creating the Spanning Tree

• We will consider a simple algorithm that uses the center-based approach.

• Steps:
  – Define a center node.
  – Nodes unicast <tree-join> messages to the center node.
  – A <tree-join> messages is forwarded using unicast towards the center until it either arrives at a node that already belongs to the spanning tree or arrives at the center.
  – The path that the <tree-join> message has followed defines the branch of the spanning tree between the edge node that initiate the tree-join message and the center.
Center-Based Construction

E - center node

The order of joining nodes: F, B, A, C, G

(a) Stepwise construction of spanning tree

(b) Constructed spanning tree
Broadcast in Practice

• Application layer - Gnutella:
  – Broadcasts *Gnutella Query* messages.
  – Uses sequence-number-controlled flooding: 16 bits id and 16 bits payload descriptor (identify message type).

• Network layer – OSPF
  – Broadcasts link-state advertisements
  – Uses sequence-number-controlled flooding: 32-bits sequence number and 16 bits age.
Multicast Routing

• Delivering a packet sent from a source node to a **subset** of the other nodes in the network.

• There are two problems:
  – How to identify the receivers of a multicast packet?
  – How to address the packet sent to these receivers?
  – Possible solution: each multicast packet will carry the IP addresses of all of the multiple recipients.

Do we have these problems at unicast or broadcast?
Multicast Routing Solution

• A multicast packet is addressed using address indirection.
• The group of receivers get a single identifier.
• A copy of the packet that is addressed to the group using the single identifier is delivered to all of the multicast receivers associated with that group.
• Class D represents a multicast address.
• Every node in the multicast group has two IP addresses: its unique address and the multicast address.
Questions…

• How does a group get started and how does it terminate?
• How is the group address chosen?
• How are new hosts added to the group?
• Do group members know the identities of the other group members?
• How do the network nodes interoperate with each other to deliver a multicast datagram?
• …
IGMP – Internet Group Management Protocol

• Operates a host and its directly attached router.
• Provides the means for a host to inform the attached router that an application running on the host wants to join a specific multicast group.
• Another protocol is required to coordinate the multicast routers.
### IGMP Messages

<table>
<thead>
<tr>
<th>Type</th>
<th>Max resp time</th>
<th>Checksum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multicast group address</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Encapsulated within an IP datagram
- There are 3 messages types:
  - membership_query – sent by the router to all hosts on an attached interface to determine the set of multicast groups at that interface (can ask about specific group).
  - membership_report – sent by the hosts to report a multicast group. Can be sent as a result of membership_query message or as a result of application joining a group.
  - leave_group – sent by hosts when leaving a group.
IGMP Messages – Cont.

- The router periodically send membership_query messages.
- If it does not get any membership_report for a specific multicast address, it knows that this address is no longer valid for the attached interface.
- The leave_group message is optional.
IGMP – Joining a Group

• When a router gets a membership_report message from a host, it will start delivering multicast datagram to that host.

• Joining multicast group is receiver driven.

• The sender does not have to know the recipients IPs but neither it can control who’s going to get its datagrams.
Multicast Routing: Problem Statement

- **Goal:** find a tree which connects routers that have local multicast group members
  - tree: not all paths between routers used
  - source-based: different tree from each sender to receivers
  - shared-tree: same tree used by all group members
Building Multicast Trees

Approaches:

• **source-based tree**: one tree per source
  – shortest path trees: reverse path forwarding

• **group-shared tree**: group uses one tree
  – minimal spanning (Steiner)
  – center-based trees

...we first look at basic approaches, then specific protocols adopting these approaches
Source-Based Tree: Shortest Path Tree

- Reverse path forwarding relies on router’s knowledge of unicast shortest path from it to sender.
- How a node knows its shortest path to each other node?
- The nodes perform a link state routing algorithm.
  - Dijkstra’s algorithm
Shortest Path Tree

- Multicast forwarding tree: tree of shortest path routes from source to all receivers

LEGEND:
- Router with attached group member
- Router with no attached group member
- Link used for forwarding
- \( i \) indicates order link added by algorithm
Reverse Path Forwarding

- Rely on router’s knowledge of unicast shortest path from it to sender
- Each router has simple forwarding behavior:

  \[
  \text{if} \ (\text{multicast datagram received on incoming link on shortest path back to sender})
  \]
  \[
  \text{then} \ \text{flood datagram onto all outgoing links}
  \]
  \[
  \text{else} \ \text{ignore datagram}
  \]
Reverse Path Forwarding: example

• What is the problem here?
• There is no need to forward the packet to R5 and R7.

LEGEND
- router with attached group member
- router with no attached group member
- datagram will be forwarded
- datagram will not be forwarded
Reverse Path Forwarding: Pruning

- Forwarding tree contains sub-trees with no multicast group members
  - no need to forward datagrams down subtree
  - “prune” msgs sent upstream by router with no downstream group members

**LEGEND**
- router with attached group member
- router with no attached group member
- prune message
- links with multicast forwarding
Group-Shared-Tree: Steiner Tree

• **Steiner Tree**: minimum cost tree connecting all routers with attached group members
• Problem is NP-complete
• Excellent heuristics exists
• Not used in practice:
  – computational complexity
  – information about entire network needed
  – monolithic: rerun whenever a router needs to join/leave
Center-Based Trees

- Single delivery tree shared by all
- One router identified as “center” of tree
- To join:
  - edge router sends unicast $join-msg$ addressed to center router
  - $join-msg$ “processed” by intermediate routers and forwarded towards center
  - $join-msg$ either hits existing tree branch for this center, or arrives at center
  - path taken by $join-msg$ becomes new branch of tree for this router
Center-Based Trees: Example

Suppose R6 chosen as center:

LEGEND
- Router with attached group member
- Router with no attached group member
- Path order in which join messages generated
Multicast Routing in the Internet

• DVMRP: Distance Vector Multicast Routing Protocol
• PIM: Protocol Independent Multicast
Internet Multicasting Routing: DVMRP

- **flood and prune**: implements source-based tree with reverse path forwarding and pruning.
  - Each router uses distance vector algorithm to compute the next hop on its shortest path back to each possible source.
  - RPF uses this information.
  - DVMRP computes a list of dependent downstream routers for pruning.
DVMRP – Cont.

• A DVMRP *prune* message contains a prune lifetime that indicates how long a pruned branch will remain pruned before it automatically restored.

• DVMRP *graft* messages are sent by a router to its upstream neighbor to force previously pruned branch to be added back to the multicast tree.
Protocol Independent Multicast: PIM

• Not dependent on any specific underlying unicast routing algorithm (works with all)

• Two different multicast distribution scenarios:
  • Dense:
    – Group members densely packed, in “close” proximity.
    – Bandwidth more plentiful
  • Sparse:
    – Number of networks with group members small with respect to number interconnected networks
    – Group members “widely dispersed”
    – Bandwidth not plentiful
## PIM – Dense Mode vs. Spare Mode

<table>
<thead>
<tr>
<th>Dense Mode</th>
<th>Sparse Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group membership by routers <em>assumed</em> until routers explicitly prune</td>
<td>No membership until routers explicitly join</td>
</tr>
<tr>
<td><em>data-driven</em> construction on multicast tree (e.g., RPF)</td>
<td><em>receiver-driven</em> construction of multicast tree (e.g., center-based)</td>
</tr>
<tr>
<td>Bandwidth and processing: <em>profligate</em></td>
<td>Bandwidth and processing: <em>conservative</em></td>
</tr>
</tbody>
</table>
PIM - Dense Mode

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
PIM - Sparse Mode

- Center-based approach
- Router sends join msg to rendezvous point (RP)
  - intermediate routers update state and forward join
- After joining via RP, router can switch to source-specific tree
  - increased performance
sender(s):
• Unicast data to RP, which distributes down RP-rooted tree
• RP can send *stop* msg if no attached receivers
  – “no one is listening!”