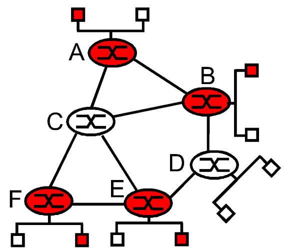
Advanced Networking

Multicast

Renato Lo Cigno – Alessandro Russo
LoCigno@disi.unitn.it - Russo@disi.unitn.it
Homepage:
disi.unitn.it/locigno/index.php/teaching-duties/advanced-networking

The Multicast Tree problem

- Problem: find the best (e.g., min cost) tree which interconnects all the members

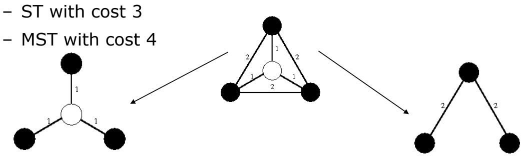


Steiner tree problem

- Given a graph $G = (V, E, w)$
  - G connected and undirected
  - Weight function $w: E \rightarrow R$
  - FIND $G_{ST} = (V_{ST}, E_{ST}, w)$
    - $T \subseteq V_{ST}$ multicast set
    - $w(G_{ST}) = \sum_{(i,j)\in E_{ST}} w(i,j)$ is minimum
  - $T - V_{ST}$ are called Steiner nodes
  - The Steiner Tree Problem is NP-Complete, i.e.,
    - It is an NP Problem (solution can be verified in polynomial time)
    - And it is NP-hard (any NP problem may be converted into it)
  - Special case: when T=V → Minimum Spanning Tree (MST), which can be solved in polynomial time

Steiner and Minimum Steiner tree

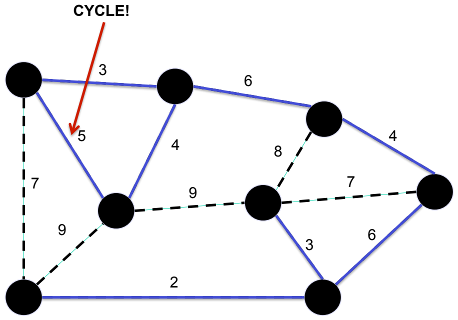
- ST has more solutions that MST
  - MST is “easier”
  - NOTE:MST is a problem: Dijkstra is an algorithm to find a MST!
- Given a Steiner Tree, we have two solutions
  - ST with cost 3
  - MST with cost 4



Minumum Spanning Tree (MST)

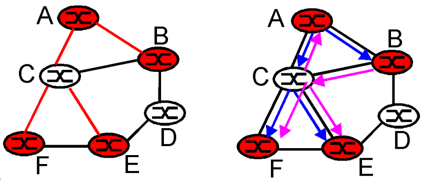
- Given a ST
  - Compute all possible STs, and
  - Pick the minimum one
- Solutions:
  - Dynamic programming might be used to solve the MST
  - Several Heuristics (greedy) to find a solution quickly:
    - Pruned Dijkstra Heuristic: Dijkstra over the source
    - Shortest Path Heuristic: Incrementally from the source
    - Kruskal based: Incrementally over links weights

Kruskal Example



Multicast Tree options

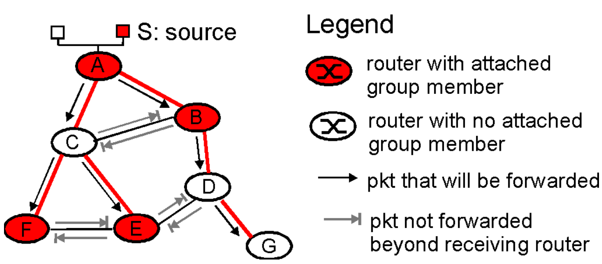
- **GROUP SHARED TREE**: just one spanning tree; the root is the “**CORE**” or the “**Rendez Vous**” point; all messages go through the **CORE**
- **SOURCE BASED TREE:** each source is the root of its own tree connecting to all the members; thus N separate trees



Group Shared Tree

- Predefined **CORE** for given m-cast group (e.g., posted on web page)
- New members “join” and “leave” the tree with explicit join and leave control messages
- Tree grows as new branches are “grafted” onto the tree
- CBT (Core Based Tree) and PIM Sparse-Mode are Internet m-cast protocols based on GSTree
- All packets go through the **CORE**

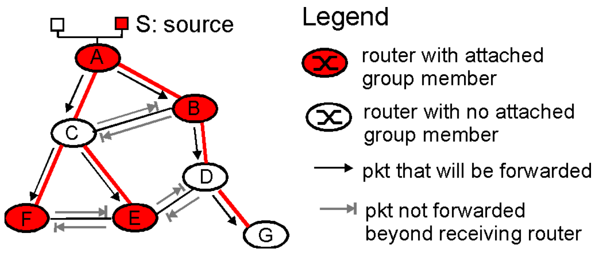
Group Shared Tree



Source Based Tree

- Each source is the root of its own tree: the tree of shortest paths
- Packets delivered on the tree using "reverse path forwarding" (RPF); i.e., a router accepts a packet originated by source S only if such packet is forwarded by the neighbor on the shortest path to S
- In other words, m-cast packets are "forwarded" on paths which are the "reverse" of "shortest paths" to S

Source Based Tree

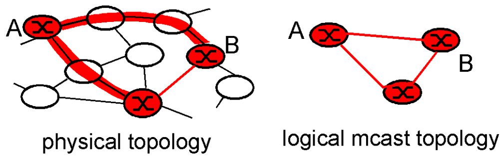


Source-Based tree: DVMRP

- DVMRP was the first m-cast protocol deployed on the Internet; used in Mbone (Multicast Backbone)
- Initially, the source broadcasts the packet to ALL routers (using RPF)
- Routers with no active Hosts (in this m-cast group) "prune" the tree; i.e., they disconnect themselves from the tree

Source-Based tree: DVMRP

- Recursively, interior routers with no active descendents self-prune. After timeout (2 hours in Internet) pruned branches "grow back"
- Problems: only few routers are mcast-able; solution: tunnels



PIM (Protocol Independent Multicast)

- Is becoming the de facto inter AS m-cast protocol standard
- "Protocol Independent" because it can operate on different routing infrastructures
  - Extract required routing information from any unicast routing protocol
  - Work across multiple AS with different unicast routing protocols
- PIM can operate in two modes:
  - Sparse Mode RFC 2362
  - Dense Mode RFC 3973

PIM Strategy

- Flooding is inefficient over large sparse internet
- Little opportunity for shared spanning trees
- Focus on providing multiple shortest path unicast routes
- Dense mode
  - For intra-AS
  - Alternative to MOSPF
- Sparse mode
  - Inter-AS multicast routing

PIM – Sparse Mode

- A sparse group:
  - Number of networks/domains with group members present significantly small than number of networks/domains in internet
  - Internet spanned by group not sufficiently resource rich to ignore overhead of current multicast schemes

PIM – Sparse Mode

- For a group, one router designated rendezvous point (RP)
- Group destination router sends join message towards RP requesting its members be added to group
  - Use unicast shortest path route to send
  - Reverse path becomes part of distribution tree for this RP to listeners in this group
- Node sending to group sends towards RP using shortest path unicast route
- Destination router may replace group-shared tree with shortest path tree to any source
  - By sending a join back to source router along unicast shortest path
- Selection of RP dynamic
  - Not critical

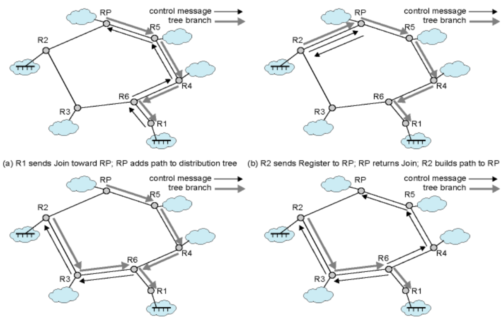
PIM – Sparse Mode

- Initially, members join the “Shared Tree” rooted on a Randez Vous Point
  - Join messages are sent to the RP, and routers along the path learn about the “multicast” session, creating a tree from the RP to the receivers
  - The source send the packet to the router which encapsulate such a packet in a unicast message towards the RP. The RP unpack the message and send the packet on the tree.
- Later, once the “connection” to the shared tree has been established, opportunities to connect **DIRECTLY** to the source are explored (thus establishing a partial Source Based tree)
  - e.g., load exceeds the forwarding threshold

Group Destination Router
Group Source Router

- Group Destination Router
  - Has local group members
  - Router becomes destination router for given group when at least one host joins group
    - Using IGMP or similar
- Group source router
  - Attaches to network with at least one host transmitting on multicast address via that router

Example of PIM Operation



(a) R1 sends Join toward RP; RP adds path to distribution tree (b) R2 sends Register to RP; RP returns Join; R2 builds path to RP

(c) R1 sends Join to R2; R2 prunes path to RP (d) R6 sends Prune to RP; RP prunes path to R1

PIM – Dense Mode

- A dense group:
  - Designed for wired-static networks
  - Routers are close to each other
  - Dense clients within an area (e.g., an organization)
  - Flood and Prune Protocol
  - Heavily use of Timers

PIM – Dense Mode

- For a group there is one or many sources
- Each source floods data towards each interface
- Routers check whether there are nodes/ routers on interfaces interested in that multicast group
  - Forward packets towards such interfaces except the RPF
  - Otherwise send a prune to the RPF’ (i.e., next hop on the RPF), actually leaving the group

PIM – Dense Mode

- Routers get/keep in touch through Hello msgs
- Routers send
  - **Prune** to leave a group: no clients on that group;
  - **Join** to receive again data from that group after a Prune;
  - **Graft** when the RPF’ changes or a client joins a pruned group
  - **GraftAck** to ack a graft on downstream interfaces

PIM – Dense Mode

```
+-------------+                        +-------------+
|             |     olist == NULL      |             |
|   Forward   |----------------------->|   Pruned    |
|             |                        |             |
+-------------+                        +-------------+
     ^    |                                 ^    |
     |    |                                 |    |
     |    |RPF`(S) Changes     olist == NULL|    |
     |    |                                 |    |
     |    |          +-------------+        |    |
     |    +--------->|             |--------+    |
     |               | AckPending  |             |
     +---------------|             |<------------+
  Rcv GraftAck OR    +-------------+ olist != NULL
Rcv State Refresh
   With (P==0) OR
S Directly Connect

      Figure 1: Upstream Interface State Machine
```

Sparse vs. Dense Mode

- RP must be configured
- Explicit join
- Traffic flows to where it's needed
- Just routers along paths keep the state
- Scales better than DM

- Flood and Prune -> congestion?
- Routers must keep (S,G) state information
- Routers negotiate traffic forwarding: assert msgs
- More reliable on dynamic network:
  - Routers knows all (S,G)
  - No RP constraints

IPv4 Multicast Space

- Host Extension for IP Multicasting
  - Groups may be permanent or transient:
    - Permanent groups have well-known addresses
    - Transient groups receive address dynamically
  - The multicast addresses are in the range 224.0.0.0 through 239.255.255.255.
  - Which authority coordinates addresses assignments??
    - *http://www.iana.org*
    - *The Internet Assigned Numbers Authority (IANA) is the body responsible for coordinating some of the key elements that keep the Internet running smoothly;*

IANA Activities

- From IANA websites:
- IANA's various activities can be broadly grouped in to three categories:
  - Domain Names: IANA manages the DNS root, the .int and .arpa domains, and an IDN practices resource.
  - Number Resources: IANA coordinates the global pool of IP and AS numbers, providing them to Regional Internet Registries
  - Protocol Assignments: Internet protocols' numbering systems are managed by IANA in conjunctionwith standards bodies

Assigning Multicast Addresses

- How does IANA assign IPv4 multicast addresses:
  - Local Network Control Block: range 224.0.0.0 - 224.0.0.255 reserved for routing protocols and low-level topology discovery or maintenance protocols

224.0.0.0	Never assigned
224.0.0.1	All Hosts on this Subnet
224.0.0.2	All Routers on this Subnet
224.0.0.4	DVMRP Routers
24.0.0.13	All PIM Routers (hello messages…)

Assigning Multicast Addresses

- There are several addresses blocks already assigned
  - Internetwork Control Block (224.0.1.0 - 224.0.1.255 (224.0.1/24))
  - AD-HOC Block I (224.0.2.0 - 224.0.255.255)
  - RESERVED (224.1.0.0-224.1.255.255 (224.1/16))
  - SDP/SAP Block (224.2.0.0-224.2.255.255 (224.2/16))

Assigning Protocol Number

- IANA assigns IP protocol numbers
  - TCP has IP protocol number 6
  - UDP is 17
  - PIM messages have IP protocol number 103
- IANA port numbers
  - ssh is 22
  - echo is 7
  - telnet 23
  - PIM over Reliable Transport: is a “congestion-control” modification for JP messages
    - pim-port 8471 tcp
    - pim-port 8471 sctp

Useful Links

- IETF Datatracker
  - The IETF Datatracker is a web-based system for managing information about Internet-Drafts (I-Ds), RFCs and several other important aspects of the IETF process
  - http://datatracker.ietf.org/
- IETF TOOLS team
  - The purpose of the TOOLS team is to provide IETF feedback and guidance during the development of software tools to support various parts of IETF activities
  - http://tools.ietf.org/

Wired and Wireless Multicasting

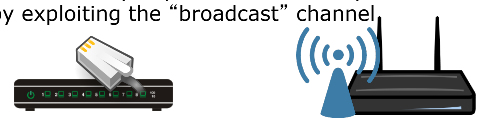
- Traditional multicast algorithms have been designed for wired networks
  - Static networks
  - Several configurations are done manually (e.g., RP)
- Is the world completely wired?
  - Smartphones
  - Tablets
  - Netbooks
  - Next generation cars

Multicast over MANETs

- MANET: Mobile Ad hoc Networks
  - More about MANETs on “Nomadic Networks” course…
- Why Multicast over Wireless network?
  - Mobile inventory: tracks good and services location, determining the delivery time
  - Group-oriented mobile commerce
    - Mobile auction and reverse auction: offering, selling, and bidding
  - Military command and control systems
  - Intelligent distributed transportation systems
  - Emergency rescue operation
    - Locate and lead people to safe areas

Multicast over MANETs

- Multicast protocol designed over wired networks cannot be used over MANET because of:
  - Limited bandwidth availability
  - Error-prone shared broadcast channel (Pro & cons)
  - Node mobility and dynamic topology
  - Lack of infrastructure
  - The hidden terminal problem
  - Energy constrains
- Multicast may improve the efficiency of wireless links by exploiting the “broadcast” channel



Wireless Multicast Routing Protocol

- The goal is always the same: High Packet Delivery Ratio
- Robustness
  - Data packets may be lost/dropped due to node mobility and shared broadcast channel (collisions) → low packet delivery ratio
- Control overhead
  - Control messages are required to build and maintain the multicast tree and to identify multicast clients, as well
- Efficiency
  - The ratio between the total number of data packets received by the receivers and the total number of packets transmitted in the network

Wireless Multicast Routing Protocol

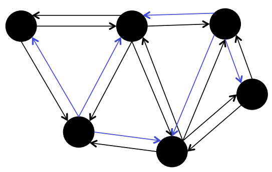
- Quality of Service
  - QoS in multicast routing protocol is a fundamental issue that involves the following parameters: throughput, delay, and jitter
- Dependency on the unicast routing protocol
  - (**+**) Designed on a given unicast routing protocol to perform better,
    (**-**) however they cannot operate in heterogeneous networks
  - (**+**) Independent ones can operate in heterogeneous networks,
    (**-**) perhaps cannot use “features” provided by unicast routing protocols
- Resource management
  - Nodes should reduce the number of packets transmitted to limit the energy consumption, as well as limit the amount of information tracked/stored to reduce the memory usage

Multicast over MANETs

- Does PIM work over unstructured wireless networks?
  - Wireless Routing protocol
    - OLSR, DSDV, AODV, and more
  - PIM was designed for wired networks, it is not supposed to work on MANETs, however PIM
    - is protocol independent
    - needs several changes to work on MANETs

MANET: Multicast Routing Protocols

- Flooding
  - Reliable
  - Considerable overhead
  - Inefficient



MANET: Multicast Routing Protocols

- Tree-Based approaches: source or shared tree based (efficient against data forwarding)
  - Ad hoc Multicast Routing protocol utilizing Increasing id-numberS (AMRIS): assigns ID to nodes building a tree rooted at a Sid Node (e.g., source)
    - Tree initialization: create and advertise a multicast session
    - Tree maintenance: data driven tree rejoin through *Branch Reconstruction* (direct parent join/broadcast parent join within K hops)
  - Multicast Ad hoc On-Demand Distance Vector (MAODV)
    - Similar to AODV, it discovers routes on demand broadcasting RREQ and RREP messages
    - High control overhead

MANET: Multicast Routing Protocols

- Mesh-Based: multiple paths between source-receiver pair (robust against mobility)
  - On-Demand Multicast Routing Protocol (ODMRP)
    - On-demand route construction
    - Source sends periodic Join-Query
    - Intermediate nodes forward the packet, learning the path back to source
    - Multicast group members send Join-Reply in response to Join-Query
  - Core-Assisted Mesh Protocol
    - shared mesh among groups
    - guarantee reverse shortest path between source and receivers

MANET: Multicast Routing Protocols

- Stateless Multicast: no multicast info stored on routers, source lists receivers in the packet header: small groups, good for mobility!
  - Differential Destination Multicast (DDM)
    - multicast information are provided in each packet
    - source controls multicast membership
    - source encodes all group members within the data packet using the DDM header format
    - packets are sent to the next hop group member via single hop broadcast
    - REF: http://tools.ietf.org/id/draft-ietf-manet-ddm-00.txt

MANET: Multicast Routing Protocols

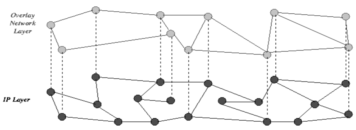
- Hybrid Protocols: provides either robustness or efficiency from mesh and tree approaches
  - Ad hoc Multicast Routing Protocol (AMRoute)
    - Tree based protocol
    - Each group has a core node which is responsible for member and tree maintenance
    - Based on Join-Req/Join-Ack, Tree-Create/ Tree-Create-Nak
    - Use virtual mesh links to establish the multicast tree
  - Multicast Core Extraction Distributed Ad hoc Routing (MCEDAR)
    - Multicast extension of the CEDAR unicast routing protocol

MANET: Multicast Routing Protocols

- Which one?
  - Reactive or Proactive
  - Stateful or stateless
  - Tree, Mesh or Hybrid
  - Control messages overhead: flood or not
  - Dependent on unicast protocol or not
  - Stable or harsh environment
- Answer: Actually, it depends on
  - the combination environment-application
  - the properties we are interested in (e.g., robustness)

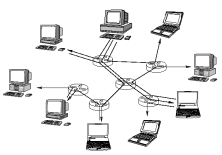
Application-Level Multicast

- Multicast protocols are not supported by all routers
  - ISPs won't replace such routers
- Reproduce multicast at application level
  - Use Peer-to-peer paradigm among nodes interested in the multicast session
  - Build a logical multicast-overlay over the physical network
  - Nodes have a small view of the network: they know few nodes called neighbors



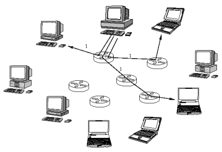
Application-Level Multicast

- The source sends packets towards a few neighbors
- Nodes forward data received to their neighbors
  - Exchanging messages about data received, avoiding duplicates
  - Two main strategies: Push and Pull
    - PUSH: nodes send data towards neighbors → sender oriented
    - PULL: nodes request data from neighbors → receiver oriented



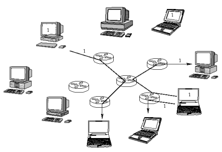
Example: Following One Chunk

- Source encodes the first piece of data
- It issues three copies of chunk number 1 to three neighbors
- Three nodes have the chunk 1 at the end of cycle 1



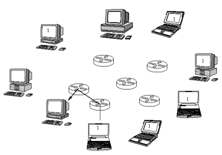
Example: Following One Chunk

- Three node has the chunk number 1
- Each node tries to PUSH the chunk to some neighbor
- In cycle 2 there are 3 + 3 = 6 nodes with chunk 1



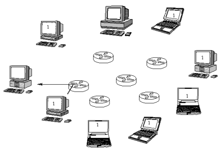
Example: Following One Chunk

- The chunk number 1 is still forwarded towards some neighbors
- At the end of cycle three, there are 7 nodes have that chunk



Example: Following One Chunk

- The forwarding process for chunk 1 ends
- The last node receives chunk number 1
- After 4 cycles all the nodes got the chunk
- Remember: it's an example! It's not a bound!!



Example: Following One Chunk

- Some questions:
  - Are chunks path always the same?
  - Is the chunks receiving order correct?
  - Are links stable?
  - It video streaming delay sensitive?
  - But that's another story…