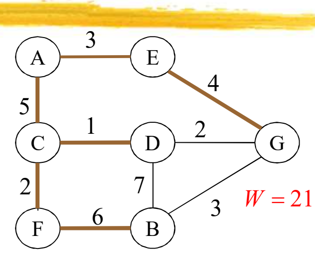
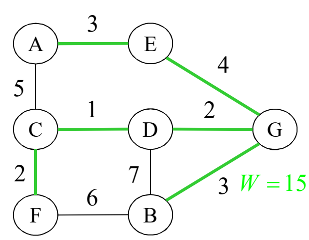
Trees and Minimum Weight Spanning Trees

- A *tree* is a connected graph that contains no loops (cycles)
- A *spanning tree* of a graph G is a subgraph of G that is a tree and contains all nodes of G
- Cost of a tree is the sum of the weights of all its links (packets traverse all tree links once)
- A *spanning tree* of a graph of N nodes contains N-1 links (arcs)
- *A Minimum Weight Spanning Tree (MST)* is a spanning tree with minimum sum of its links' weights

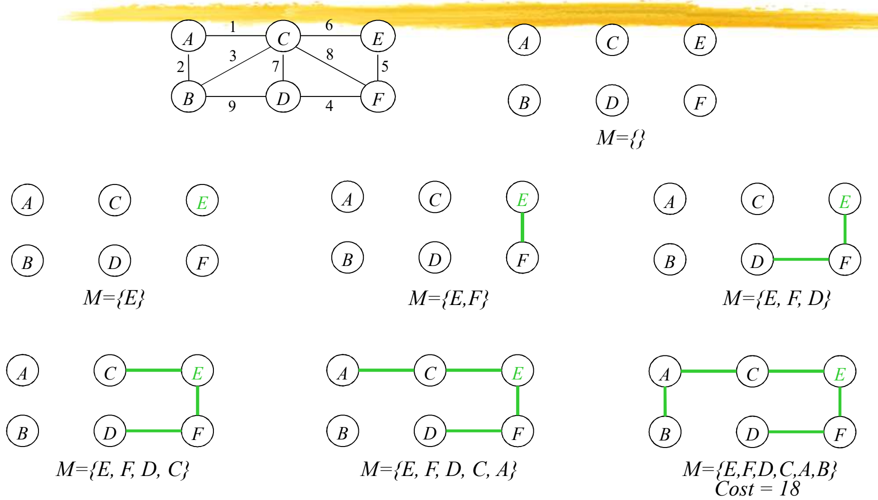




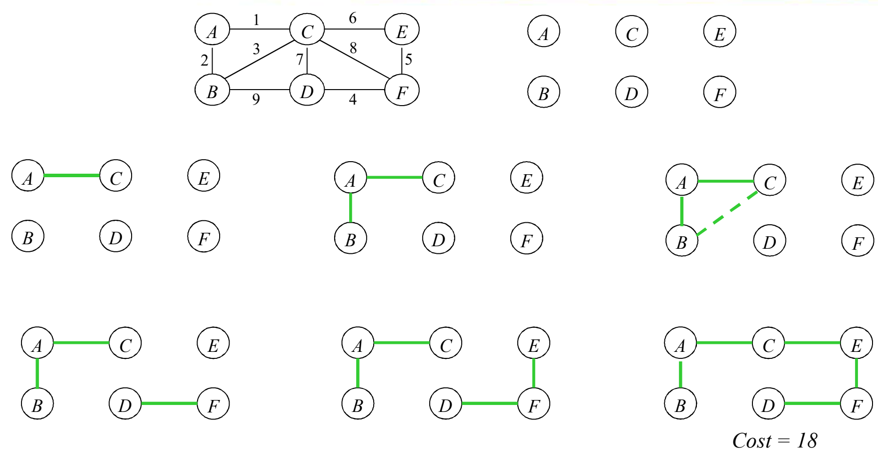
Minimum Weight Spanning Tree Algorithms

- *Prim's Algorithm*:
  - Add an arbitrary node to set *M*
  - Successively add nodes whose distance to set *M* is minimal
  - Stop when the graph is a *spanning tree*
- *Kruskal's Algorithm*:
  - Organize the links in increasing order of their weights
  - Successively add links from the list, ensuring that no cycles are created.
  - Stop when the graph is a *spanning tree*

Prim's Algorithm



Kruskall's Algorithm



Minimum Weight Spanning Tree Algorithms

- *Prim's Algorithm*:
  - Add an arbitrary node to set *M*
  - Successively add nodes whose distance to set *M* is minimal
  - Stop when the graph is a *spanning tree*
- *Kruskal's Algorithm*:
  - Organize the links in increasing order of their weights
  - Successively add links from the list, ensuring that no cycles are created.
  - Stop when the graph is a *spanning tree*
- Minimum Spanning Tree vs. Least-Cost Routing
  - Can MST be built from LCR paths?
  - Can a LCR paths be used to form a MST?
  - MST paths are NOT LCR paths!!!

Motivation for Multicast

- Rationale similar to link-layer multicast
- Two types: *one-to-many* and *many-to-many*
- *IP multicast* provides a **scalable** solution to routing
- Original IP service model: one-to-one data delivery
  - One sender sending its data to one receiver at a time
- Introduction of new apps with multi-receiver semantics
  - Audio/video conferencing, news dissemination, Internet TV, etc.
- Unicast not designed to efficiently support multi-receiver apps
- Solution
  - *Multicast support for IP*
  - Two versions of IP Multicast
    - *Any Source Multicast (ASM)* – original service model
    - *Single Source Multicast (SSM)* – proposed later on
      - Also called as Source Specific Multicast (SSM)

Multicast semantics

- Open group semantic
  - Zero or more receivers form a multicast group – *host group*
    - Hosts can join/leave at will – no registration/synchronization
  - A group is represented by a class D IP address (more to follow)
  - Anyone knowing group address can send to it (open group)
    - This has been modified in SSM (more later)
  - IP based best effort delivery semantics
    - Multicast supports UDP only – *no TCP !*
- Advantages:
  - Sources do not need to know individual receivers
  - Receivers simply join the group and receive data
- Disadvantages:
  - Difficult to protect from unauthorized senders/receivers

Fundamentals of Multicasting

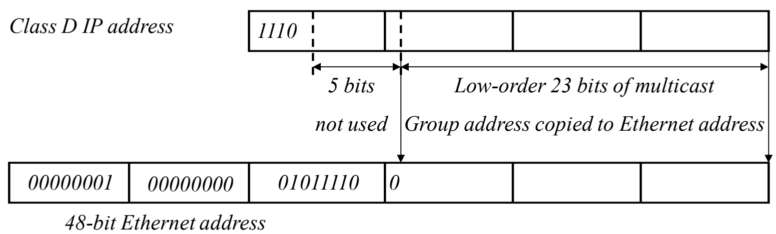
- Basic host model
  - Hosts
    - Send and receive multicast data with no/minimum extra effort
  - When sending data, normal IP-Send operation
  - When receiving data, tell your router what group you are interested in
    - join/leave a multicast group (start/stop receiving data from source(s) sending to the group) (IGMP – see RFC 3376 and RFC 4605)
- Basic router model
  - Routers
    - Have the task of connecting multicast sources to multicast receivers
    - Prepared to receive data from all multicast group addresses
    - Know when to forward or drop packets
    - Keep track of interfaces leading to receivers

Fundamentals of Multicasting (con't)

- Multicast group addresses
  - Class D IP addresses (224.0.0.0 – 239.255.255.255)
  - Implicit scoping
    - 224.0.0.0 – 224.0.0.255: link scoped
    - 224.0.1.0 – 238.255.255.255: global scoped
    - 239.0.0.0 – 239.255.255.255: admin scoped
  - Explicit scoping
    - Use TTL value for scoping
    - Use TTL thresholds at routers for scoping
- How to do actual delivery to a receiver host?
  - Map IP multicast address to an Ethernet multicast address
  - Receiver NIC is configured to receive packets destined to this Ethernet address

Fundamentals of Multicasting (con't)

- Mapping an IP address to an Ethernet address



Fundamentals of Multicasting (con't)

- Multicast packet format
  - Source IP is unicast IP address corresponding to source of the packet
  - Destination IP is multicast group address
  - In general, we show this (Source IP, Dest IP) couple as
    (S,G) where S is source IP, G is destination IP (multicast group address)
- How data is forwarded to multiple receivers?
  - Routers in the network build *forwarding trees* connecting sources and receivers
    - On-tree routers keep *multicast forwarding states* for each group
  - Source data propagates on this tree toward the receivers

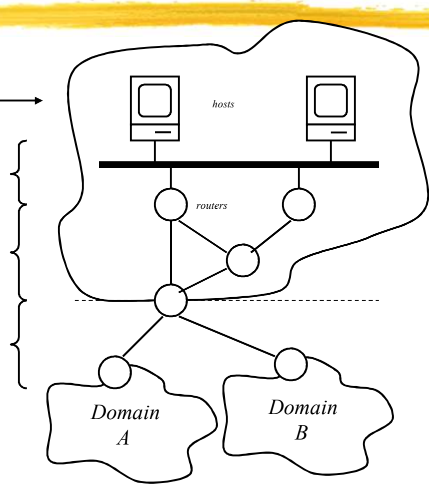
IP Multicast Architecture

application API

host-to-router (IGMP)

intra-domain routing

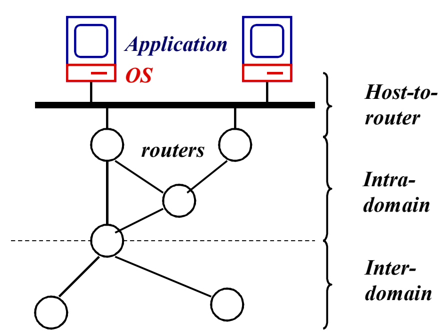
inter-domain routing



Application API

At multicast receiver (Application API socket):
`IP_ADD_MEMBERSHIP`
`IP_DROP_MEMBERSHIP`

At multicast source: … it depends

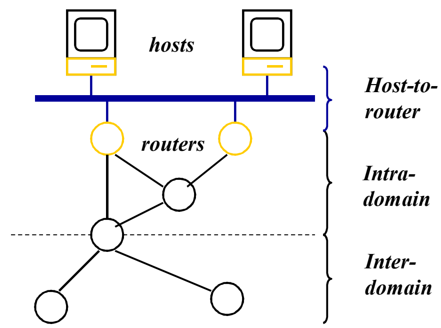


Host-to-Router

Internet Group Management Protocol

Kernel informs router that an application wants to join a specific group G

See RFC 2236 and RFC 3376



IGMP (Internet Group Management Protocol)

- Used by hosts to indicate their interest in receiving packets addressed to a particular multicast group *G*.
  - IGMPv1 (RFC 1112)
    - Routers
      - *General Membership Query*
    - Hosts
      - *Membership Reports*
      - *Unsolicited Group Membership Reports*
  - IGMPv2 (RFC 2236)
    - Added explicit *Leave Group* and *Group Specific Membership Query* messages
  - IGMPv3 (RFC 3376)
    - Added source filtering capabilities
  - IGMP messages aren't forwarded by routers
    - This is for the ones defined in above RFCs

In IPv6: *Multicast Listener Discovery (MLD)* protocol.

Multicast Routing

- Building forwarding trees between sources and receivers
- Hard due to the open service model
  - Sources do not know who/where the receivers
  - Receivers *may* not know (in advance) who/where the sources

Routing approaches

- Flood and prune
  - Begin by flooding traffic to the entire network
  - Prune branches with no receivers
  - Examples: DVMRP, PIM-DM
  - Disadvantage: unwanted state where there are no receivers
- Core based protocols
  - Specify a *meeting place* or *core*
  - Sources send their packets to core
  - Receivers join group at core
  - Requires mapping between group addresses and cores
  - Examples: CBT, PIM-SM
- Link-state multicast protocols
  - Routers advertise groups for-which-they-have-receivers to entire network
  - Compute trees on demand
  - Example: MOSPF
  - Disadvantage: unwanted state where there are no senders

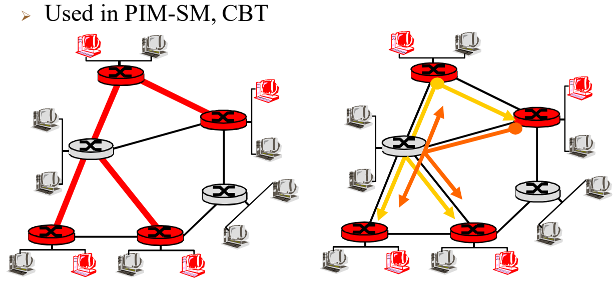
Types of forwarding trees

- Shared trees
  - Single tree shared by all members (sources)
  - Data flows on the same tree regardless of the sender
  - Example: CBT, PIM-SM
  - ☺: less states at routers
  - ☹: higher delay, traffic concentration at core
- Source-based trees
  - Separate shortest path tree for each sender
  - Example: DVMRP, MOSPF, PIM-DM, PIM-SM
  - ☺: Low delay, better load distribution
  - ☹: more state at routers (per source state)

Multicast Routing: Problem Statement

Goal: find a tree (or trees) connecting routers having local mcast group members

- **source-based:** different tree from each sender to receivers
  - Used in DVMRP, PIM-DM
- **shared-tree:** same tree used by all group members
  - Used in PIM-SM, CBT



Shared tree　　　　Source-based trees

DVMRP -Distance Vector Multicast Routing Protocol

- Defined by Deering and Cheriton in "*Multicast Routing in Datagram Internetworks and Extended LANs*" in ACM Transactions on Computer Systems in 1990. (Also, see RFC1075.)
- Flood and prune based tree construction algorithm
- Create source-based shortest path trees
  - Tree is rooted at the source site
  - It corresponds to shortest path between the source and each receiver
    - Main assumption is path symmetry *(links have the same costs in both directions)*
    - *Routers use Reverse Path Forwarding (RPF) rule*
- Observation
  - Every shortest-path multicast tree rooted at the sender is a subtree of a single shortest-path broadcast tree rooted at this sender
  - First build a shortest-path broadcast tree by **flooding**
  - Then **prune** that tree to get the multicast tree

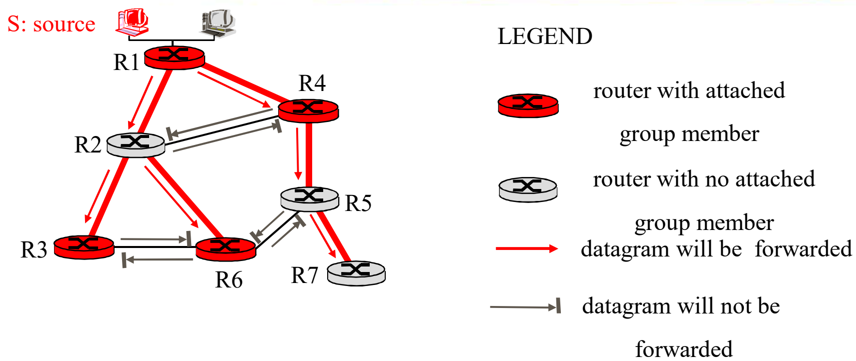
DVMRP -Distance Vector Multicast Routing Protocol

- Reverse Path Forwarding (RPF)
  - rely on router's knowledge of shortest path from it to sender
  - each router has simple forwarding behavior:

```
if (pkt is received on incoming link on shortest
    path back to source)
then accept datagram for forwarding
else ignore datagram
```

- "flood and prune": reverse path forwarding, source-based tree
  - RPF tree based on DVMRP's own routing tables constructed by communicating DVMRP routes
  - no assumptions about underlying unicast
  - initial datagram to mcast group flooded everywhere via RPF
  - routers not wanting group data: send upstream prune messages

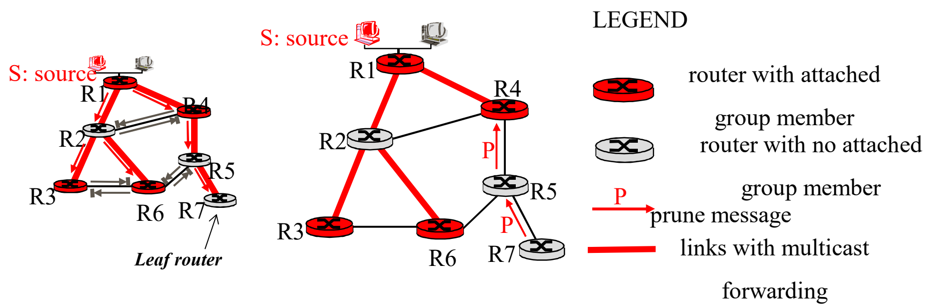
Reverse Path Forwarding: Flooding



- result is a source-specific reverse SPT
- may be a bad choice with asymmetric links

Reverse Path Forwarding: Pruning

- forwarding tree contains subtrees with no mcast group members
  - no need to forward datagrams down subtree
  - "prune" messages sent upstream by router with no downstream group members



DVMRP: continued…

- Soft State: DVMRP router periodically (every 1 min.) “forgets” branches are pruned:
  - mcast data again flows down unpruned branch
  - downstream router: re-prune or else continue to receive data
- Routers can quickly re-graft to tree
  - following IGMP join at leaf
- The first multicast routing protocol implemented and deployed on the Internet
- Problem: Not scalable

PIM-SM (Protocol Independent Multicast-Sparse Mode)

- Motivating observations
  - Compared to the scale of the global Internet, receivers and senders of nearly any multicast content are sparsely populated over a very wide area.
    - 100,000 or more 2- and 3-member groups.
  - Flood and prune protocols discover receivers by sending packets everywhere, and pruning back when there are no receivers → inefficient in the wide area
- Founding principles
  - Avoid unnecessary flooding of multicast packets to network segments with no receivers.
  - Support good-quality distribution trees for heterogeneous applications.
  - Minimize bandwidth consumed by the protocol overhead.

PIM-SM

- The PIM-SM architecture is characterized by these features:
  - Shared and Shortest-Path distribution trees.
    - All participants for a group *G* can use a shared distribution tree. Permits very large scale multicast routing.
    - Last hop routers can initiate a switch to shortest-path trees for certain sources when needed (*e.g.,* as data rates or delay requirements warrant, and scale permits).
  - Explicit join/prune tree management. Routers with local or downstream receivers join (or leave) a distribution tree by sending explicit join (or prune) messages toward sources. We say that PIM-SM is thus *receiver driven.*
  - Routing protocol independence. Makes use of existing unicast routing functionality to adapt to topology changes, but in a manner that is independent of the particular unicast routing protocol used to create the unicast routing table.

PIM-SM

- Rendezvous Point (RP)
  - In PIM-SM, for every multicast group *G* there exists a router known as the *rendezvous point (RP)* in a domain
    - The RP is the root of the *shared tree* for group *G in this domain*. We also call this tree the *RP Tree* (*RPT).*
    - Sources for group *G* are discovered when their (*S,G*) packets are forwarded down the RPT.
    - Every router in a PIM domain must know the (domain-local) RP for every group *G*. In many cases, the same router is used as the RP for every group.
  - In a PIM domain, there is only one RP for any group *G* at any particular time.
  - We will use the notation *RP(G)* to denote the RP mapping for a group *G.*
    - Example: $RP(G) = A$ means that router *A* is the RP for group *G*. Note that *RP*(*G*) is a unicast IP address.

PIM-SM

- Distributed Tree Maintenance
  - The PIM *Join/Prune* message is used to construct and dismantle group distribution trees.
  - An edge router upon receiving IGMP-based join request for G, sends a PIM-Join(*,*G*) message toward RP.
  - Routers on the path establish a multicast forwarding entry for (*,*G*) and forward the PIM-Join message toward RP (using RPF checks).
  - Once the PIM-Join message reaches RP or another router who already has a multicast forwarding state for (*,*G*), the join process ends.
  - When an edge router has no more receivers for (*,*G*) group, it sends a PIM-Prune(*,*G*) message to prune itself from the tree.

PIM-SM

- How does a source send its data?
  - The *Register* message is one means by which packets from a source *S* reach the root of the shared tree for group *G*.
    - Register messages are sent as unicast PIM messages addressed to the RP.
    - The payload of a Register message is a complete multicast packet.
    - The RP decapsulates the Register message, and forwards (or drops) the packet according to prevailing state. (No RPF check occurs.)
- How to discover sources?
  - Source *S* for group *G* is *discovered* when the router receives a packet from *S* addressed to group *G*.
    - A directly connected source is discovered when the router receives a packet addressed to group *G* from the source on the connecting interface.
    - A distant (non-connected) source is discovered when the router receives a packet from the source on the shared tree for group *G*.

PIM-SM

- How to switch to shortest path tree?
  - Once an edge router at the receiver site receives data on (*S,G*), it can optionally switch to (*S,G*) shortest path tree
    - Sends PIM-Join(*S,G*) toward S
- What is needed to implement PIM?
  - Need to build forwarding trees connecting sources and receivers
    - PIM-SM can be used for its suitability for sparse groups
  - How to learn about sources when using PIM-SM
    - Who is the RP?
    - Is one RP enough or do we have different RPs (one for each domain)
- PIM-SM is typically used within a single domain.
- The PIM-SM/MSDP model introduced to extend routing across domains
  - MSDP - Multicast Source Discovery Protocol
    - MSDP peer RPs are connected using TCP
    - MSDP messages are broadcasted using Reverse Path Broadcast.
    - MSDP communicates Source Advertisements (SAs) among domains, facilitating receivers joining the source-specific tree.

Issues with Any Source Multicast (ASM)

- The initial IP multicast model renamed as ASM
  - Standard multicast model discussed so far (the model defined in RFC 1112)
  - Attempts to provide solution to both one-to-many and many-to-many multicast apps
- Issues to consider:
  - Address scarcity/allocation/management
  - Complexity of routing architecture (PIM-SM/MSDP)
    - Source discovery
  - Denial of service attacks to groups
    - Anyone can send to a multicast group
  - Denial of service attacks to multicast infrastructure
    - MSDP Source Announcement (SA) message flooding
- The main problems:
    - Anyone is allowed to send to a group
    - The identities of senders are not known in advance

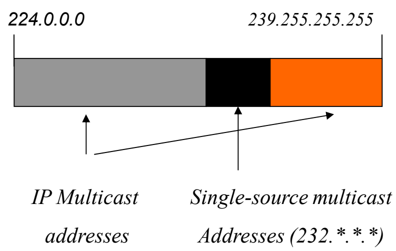
Issues with Any Source Multicast (ASM)

- The main problems:
  - Anyone is allowed to send to a group
  - The identities of senders are not known in advance
- Most popular applications are one-to-many: IP TV, VoD (Voice On Demand), IP Radio
- Simplify solution for well-known sources, particularly in cases where there is a single source sending to a given group.
- Once we know the IP address of the source, a receiver can join the shortest path tree of the source directly
  - No need for RP
  - No need for MSDP
  - Address allocation is much easier
- Optimality vs. scalability tradeoff: SSM vs. ASM

Source Specific Multicast (SSM)

- IP Multicast Channels
  - A *multicast channel* is a datagram delivery service identified by a pair (*S*, *G*) where *S* is the sender's source address and *G* is a channel destination address.
  - Only the source host *S* may send to the channel (*S*,*G*).
  - Example: (*S1*,*G1*) and (*S2*,*G1*) are two different IP Multicast channels
    - Routers will have different entries for (*S1*,*G1*) and (*S2*,*G1*).
    - On receiving multicast data from *S1* that is destined to group address *G1*, the router will forward it based on the state entry (*S1*,*G1*).
- Allows first-hop router to respond to receiver initiated join requests for specific sources within a group.
- Allows first-hop router to send (S,G) join directly to source without creation of shared tree.
- Support elimination of shared tree state in 232/8, simplifying address allocation.
- RFC 4608: Source-Specific Protocol Independent Multicast in 232/8

Single-source IP Multicast Addresses



- 2^{24} class D addresses allocated by IANA
- Routers identify a channel multicast datagram by its destination address (an address in 232/8 (232.0.0.0 – 232.255.255.255) range)

Advantages of SSM

- Source
  - 2^{24} channels per source (whole class D address range)
  - Address management is simplified
- Subscriber
  - Receives traffic only from the source it designates
- ISP
  - Provides basis for charging
  - SSM is relatively simple to implement and manage
- Packet Forwarding
  - Forwarding state entries at each router
  - Forwarding procedure is nearly identical to IP Multicast
- Advantages
  - Simple integrated protocol
    - Supports subscription, multicast channel maintenance
  - Minimal changes in host OS if it supports IP Multicast
  - Multicast traffic travel only along paths from source to subscribers