University of Duisburg-Essen
Networked Embedded Systems Group
Institute for Computer Science and
Business Information Systems (ICB)
Schützenbahn 70
D-45127 Essen, Germany

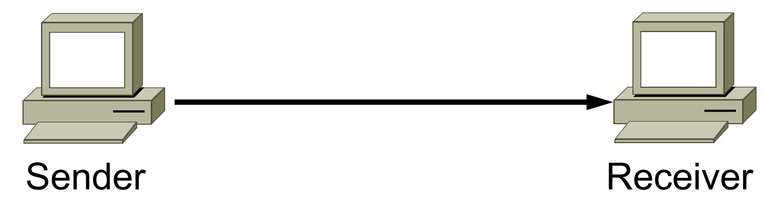
Kommunikationsnetze 2
6 – Multicast

Prof. Dr. Pedro José Marrón

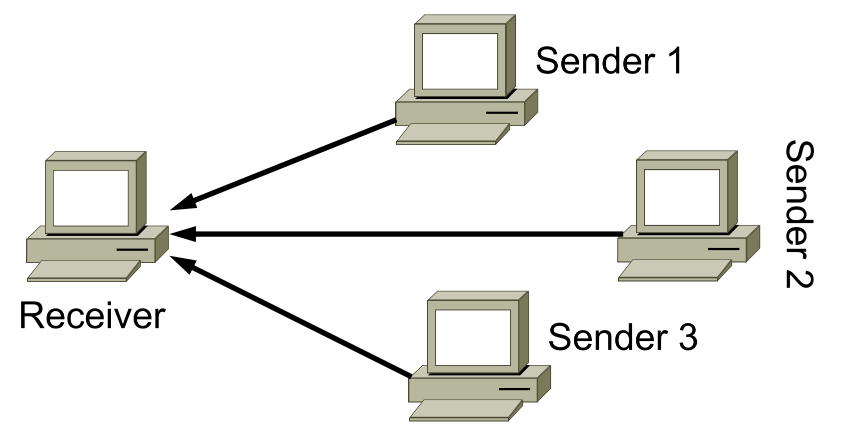
Original slides from Prof. E.P. Rathgeb and M. Becke

Communication Forms (1): Unicast and Concast

- Unicast – 1:1 (one to one)
  - 1 sender, 1 receiver
  - Most widespread communication form in the Internet
  - Application: WWW, Mail, …

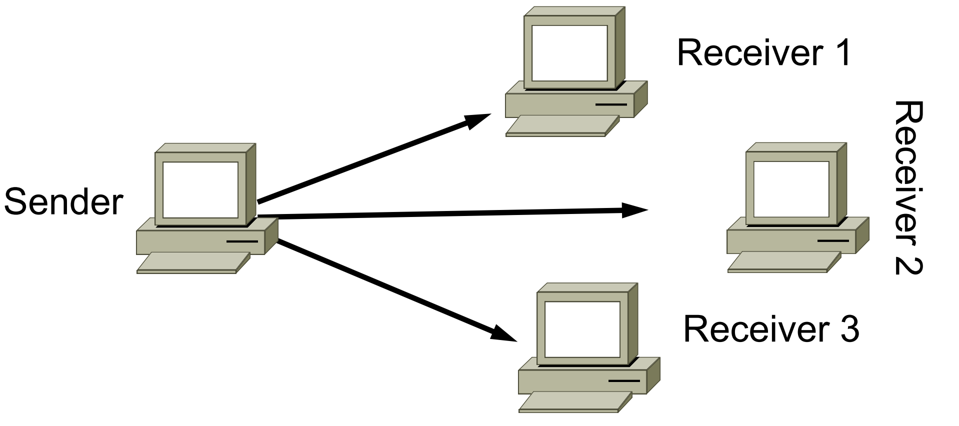


- Concast – M:1 (many to one)
  - m senders, 1 receiver
  - Application: status report, sensor data gathering

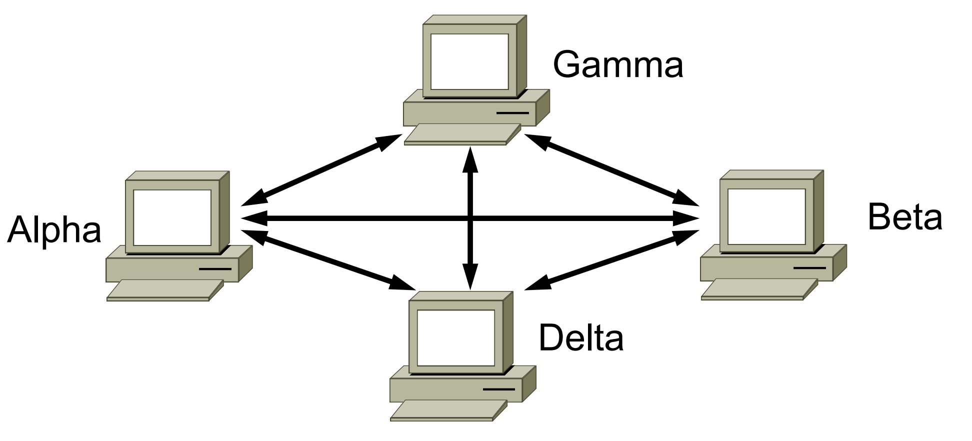


Communication Forms (2): Multicast and Multipeer

- Multicast – 1:N (one to many)
  - 1 sender, n receivers
  - Application: software distribution



- Multipeer – M:N (many to many)
  - m senders, n receivers
  - Application: file sharing



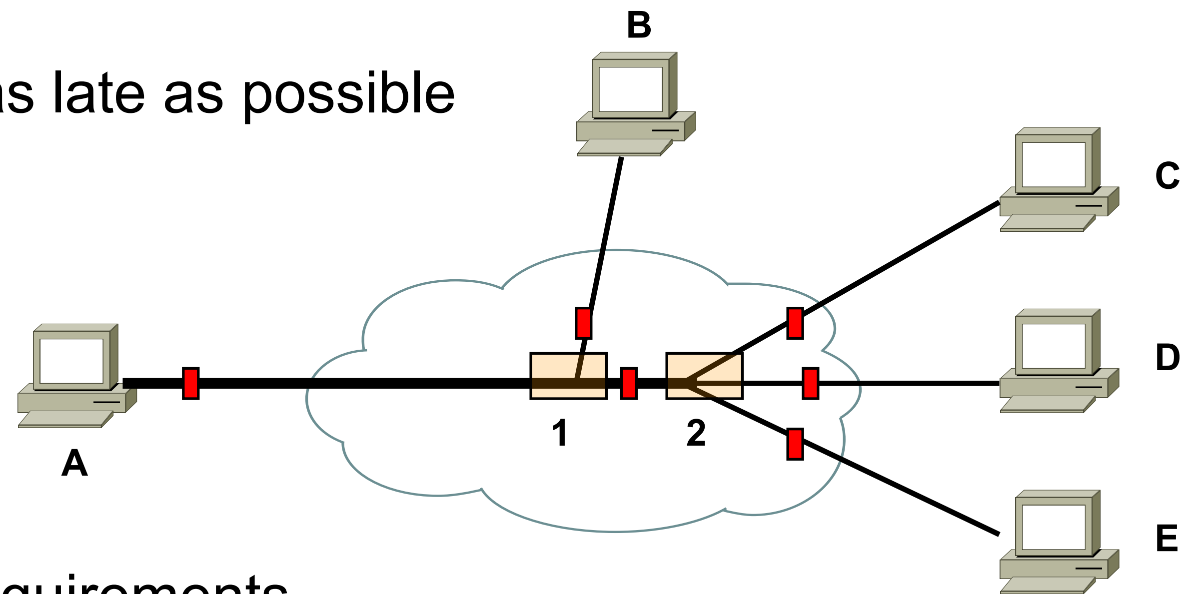
Communication Forms (3): Broadcast and Anycast

- Broadcast (one to all)
  - Unrestricted number of receivers
  - No management of groups necessary
  - Application: radio/TV broadcast

- Anycast (one to any)
  - One to any
  - Network routes to any (the nearest node)
  - Application: sending DNS requests to any DNS server in the network

Multicast

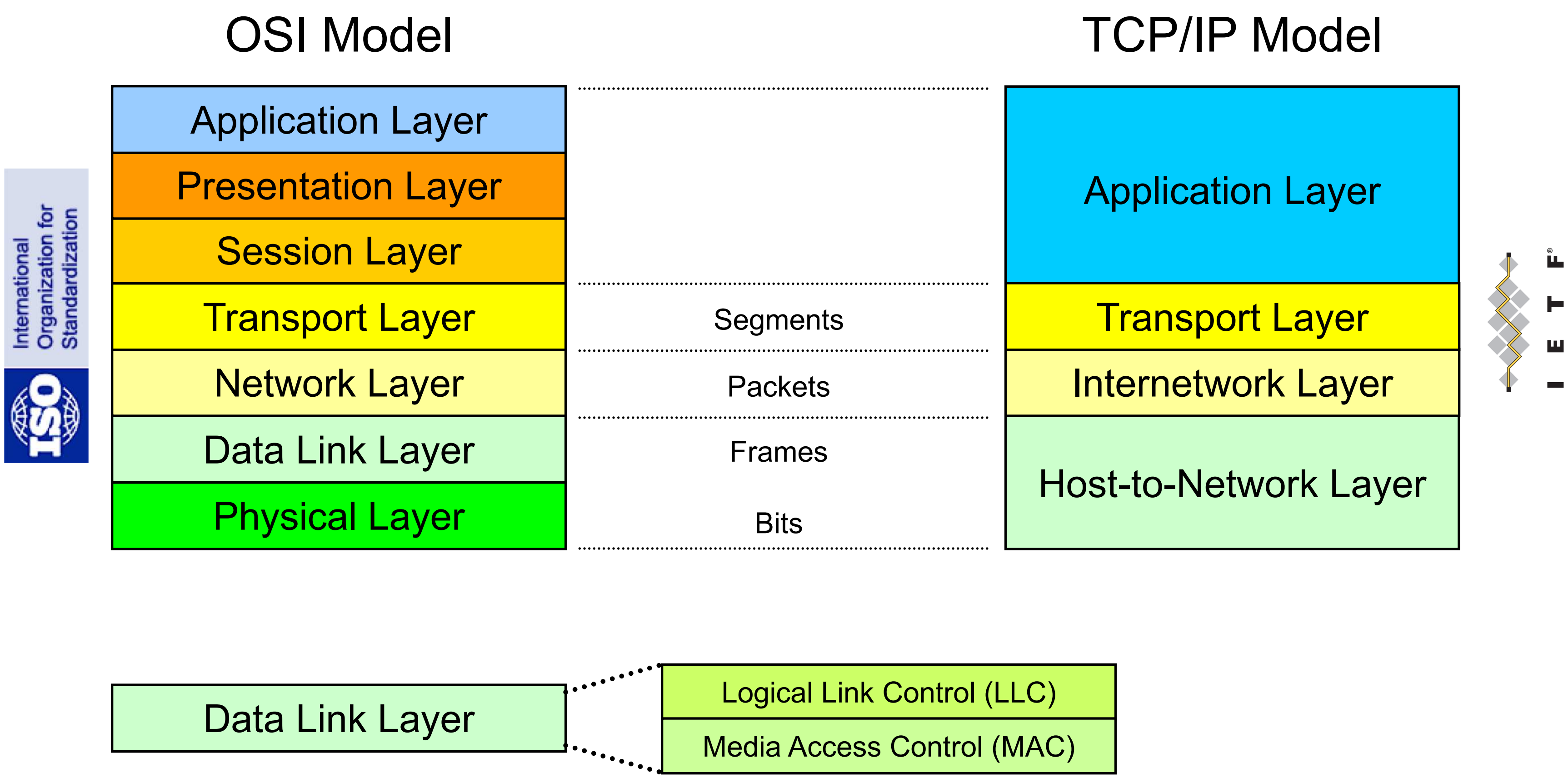
- Basic idea
  - A packet is duplicated as late as possible



- Advantage
  - Minimizes bandwidth requirements
- Disadvantages
  - State information in routers, for each group
  - Support required by network (i.e., routers, etc.)

Multicast on the OSI Layers

- Multicast support necessary on different OSI layers

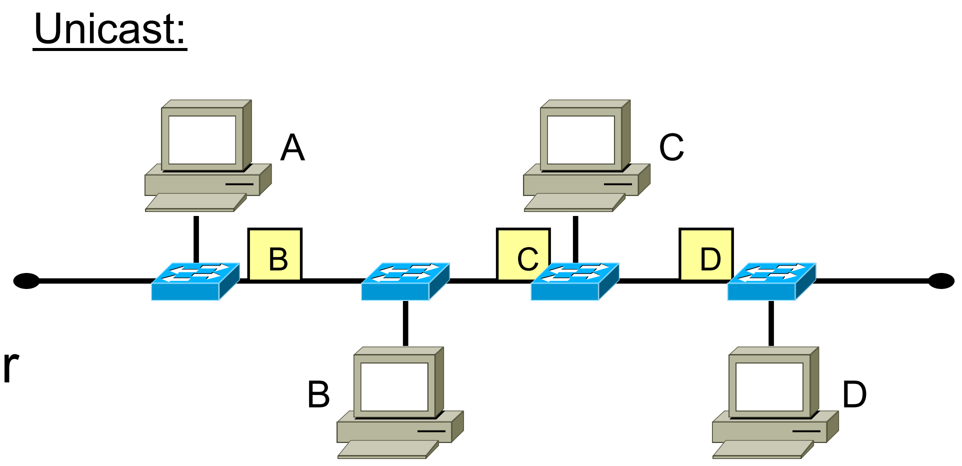


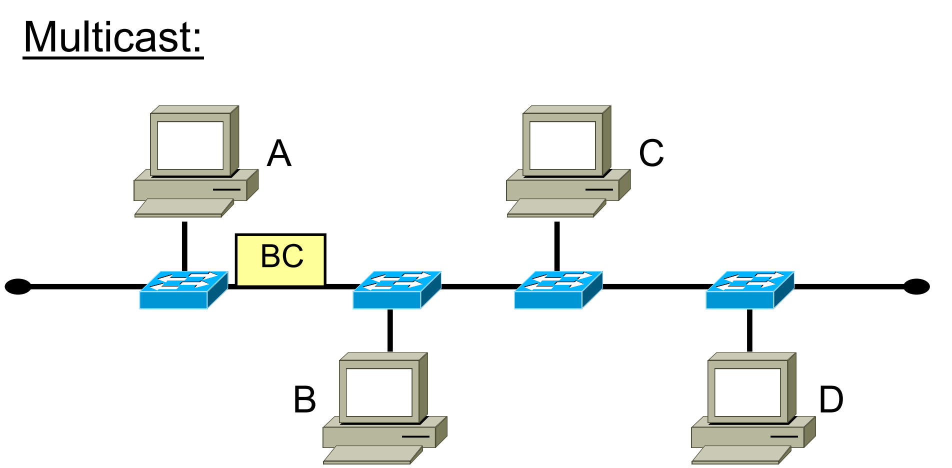
Multicast on the Data Link Layer

- Implementation of multicast depends on medium used
  - Broadcast (like Ethernet)
  - Without broadcast (like ATM)

Multicast for Ethernet (1)

- Shared medium
- Only one multicast packet for several receivers
  - Filtering by NIC, using
    - Hash table (usually 512 entries) or
    - Address table (usually 32 entries) or
    - Both
  - Switch may learn group membership of ports
    - Newer switches
    - Older ones: handle like broadcast





Multicast for Ethernet (2)

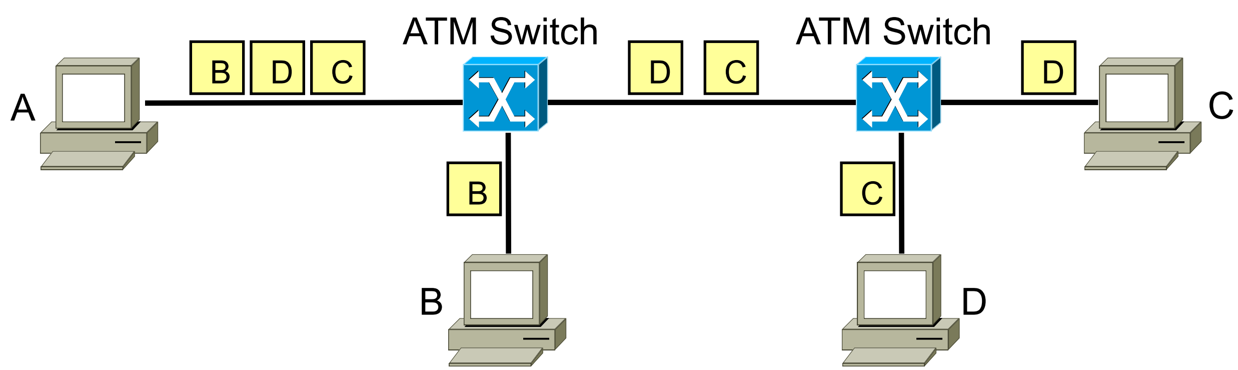
- Ethernet frame

7 Bytes	1 Byte	6 Bytes	6 Bytes	2 Bytes	46 - 1500 Bytes	4 Bytes
Pre-amble		Dest addr	Src addr	Type	Data	FCS

- Destination address
  - Destination address is first field (early processing of frame)
  - First byte
    - Bit 0
      - 0 = individual a single host
      - 1 = group a multicast group
    - Bit 1
      - 0 = global global scope (i.e., unique MAC address)
      - 1 = local local scope (only valid in same domain)

Multicast for ATM

- ATM has no shared medium
- Emulation of multicast by unicast transmissions
  - Multiple transmissions of the same content
  - High bandwidth usage on links to multiple destinations



Multicast on the Network Layer: Address Mapping

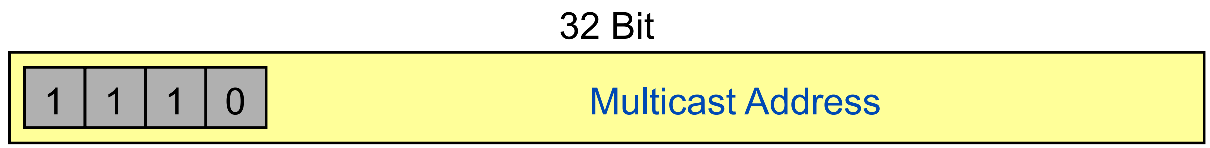
- Mapping between Network and Data Link Layer multicast addresses
- Challenge
  - Ethernet multicast address space is smaller than IPv4/IPv6 multicast address space
    - IPv4, class D 28 bits for multicast group
    - IPv6, FF00::/8 120 bits for multicast group
- Mappings
  - IPv4
    - **01:00:5E:00:00:00**
    - with the last 23 bits of the IPv4 multicast address (RFC 1112, sec 6.4)
  - IPv6
    - **33:33:00:00:00:00**
    - with the last 32 bits of the IPv6 multicast address (RFC 2464, sec. 7)

Multicast on the Network Layer: Address Mapping Examples

- IPv4
  - 239.0.0.1 → 01:00:5E:00:00:01
  - 239.0.0.50 → 01:00:5E:00:00:32
  - 224.1.2.3 → 01:00:5E:01:02:03
  - 224.129.2.3 → 01:00:5E:01:02:03 (!)
  - 224.127.99.10 → 01:00:5E:7F:63:0A
- IPv6
  - FF02::1 → 33:33:00:00:00:01
  - FF08::50 → 33:33:00:00:00:50
  - FF05::AB:0:0:50 → 33:33:00:00:00:50 (!)
  - FF0E::50:50 → 33:33:00:50:00:50 (!) FF0E::0050:0050
  - FF08::1:2:3:4 → 33:33:00:03:00:04

Multicast on the Network Layer: IPv4 Multicast

- Address space
  - Class D: 224.0.0.0/4 - 224.0.0.0 – 239.255.255.255



- Scope (defined in RFC 2365)
  - Link-local: 224.0.0.0/24
  - Site-local: 239.255.0.0/16 239.253.0.0/16
  - Organisation-local: 239.192.0.0/14
  - Global: 224.0.1.0 to 238.255.255.255
- Well-known addresses (allocated by IANA)

  http://www.iana.org/assignments/multicast-addresses
  - 224.0.0.1 all systems
  - 224.0.0.2 all routers
  - 224.0.0.5 all OSPF routers (RFC 2328)
  - 224.0.0.6 all OSPF designated routers (RFC 2328)
  - 224.0.0.9 all RIPv2 routers (RFC 1723)

Multicast on the Network Layer: IPv6 Multicast

- Address space
  - FF0x::/8 x = Scope
- Scopes
  - 1 = node-local (the node itself)
  - 2 = link-local (the Link Layer domain)
  - 5 = site-local (defined by administrator)
  - 8 = organization-local (defined by administrator)
  - E = global
- Well-known addresses (allocated by IANA)

  http://www.iana.org/assignments/ipv6-multicast-addresses
  - FF02::1 all nodes, link-local
  - FF02::2 all routers, link-local
  - FF02::5 all OSPF routers, link-local (RFC 2328)
  - FF02::6 all OSPF designated routers, link-local (RFC 2328)
  - FF02::9 all RIP routers, link-local (RFC 2080)

Multicast on the Transport Layer

- Management tasks (see also RFC 1458)
  - Connection establishment to multiple receivers
  - Adding/removing participants
  - Reliable transport?
    - Acknowledgements (e.g., negative acks only)
    - Difficult (joining/leaving nodes, all other nodes should wait until one node got its retransmissions, …)
  - Flow control
  - Congestion control
  - Quality of service
- Approaches
  - TCP extensions
  - Multicast Transport Protocol (MTP, RFC 1301)
  - Mostly academic, no widely deployed implementations

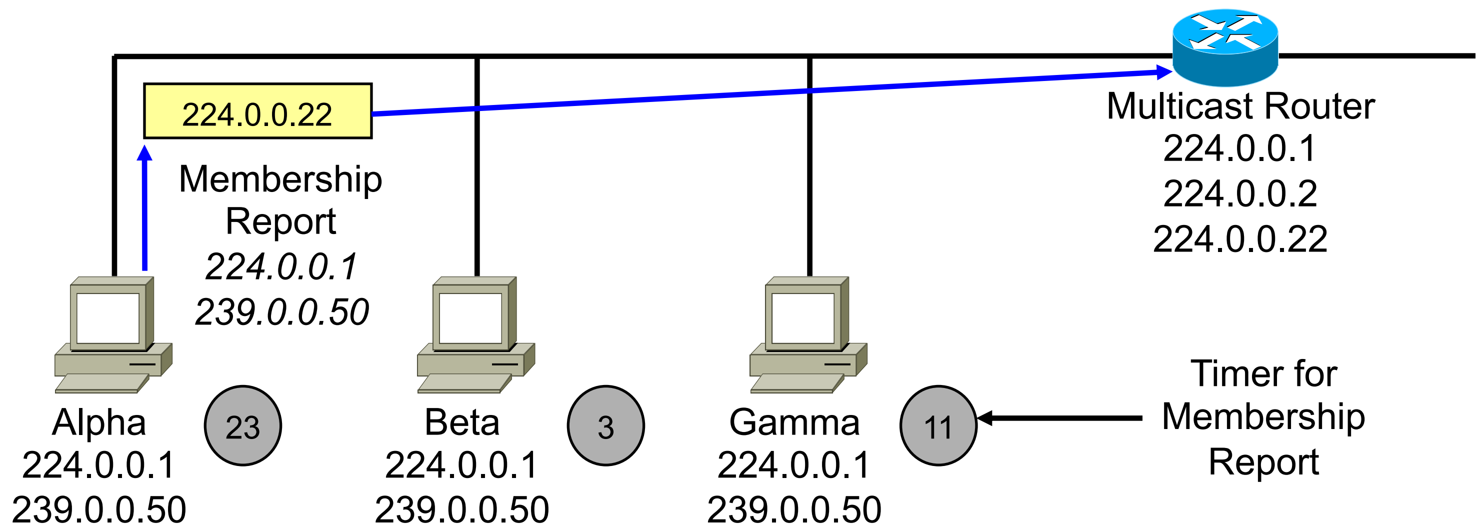
Multicast on the Application Layer

- Problem in today's Internet
  - No support for IPv4/IPv6 multicast
  - Economic reasons (business models of ISPs)
  - Political reasons
- Simple solution
  - Multicast realised in Application Layer
    - No need for Data Link/Network/Transport Layer support
    - No new underlying protocol necessary
  - Quite inefficient
    - Every packet is copied for each recipient
    - One transport connection to each recipient

Group Management

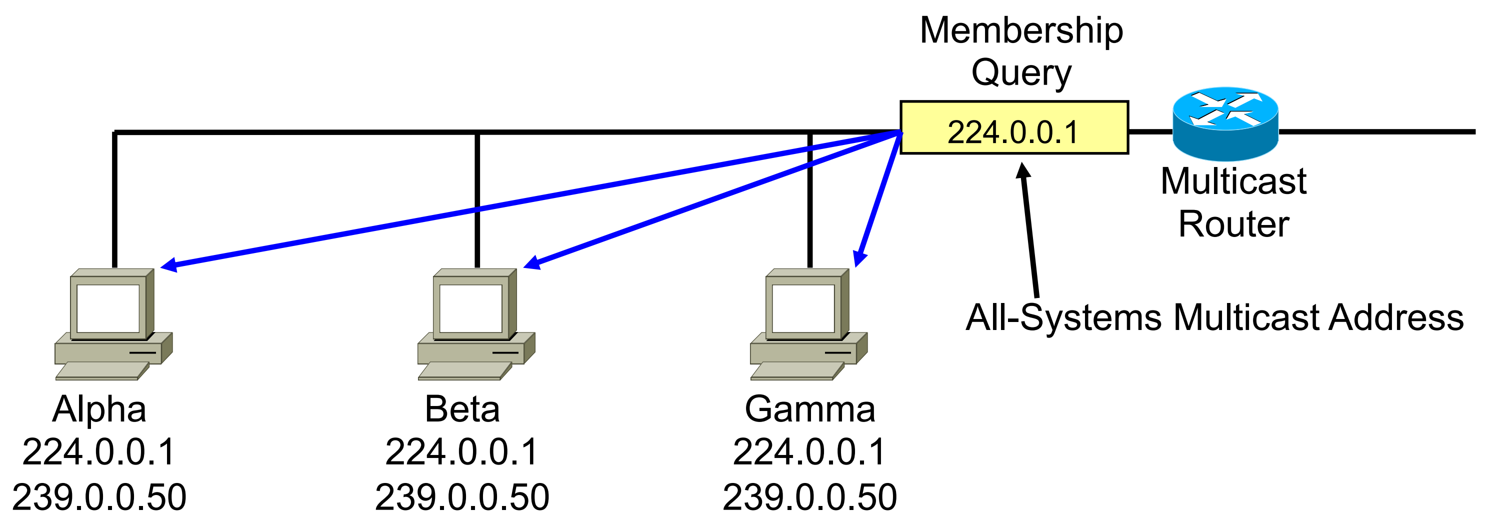
- Requirement
  - Router needs knowledge of multicast memberships
  - In all of its directly connected subnets
- Group ⇔ multicast address
  - Join
  - Leave
- Protocols
  - IPv4: Internet Group Management Protocol (IGMP) [RFC 3376]
  - IPv6: Multicast Listener Discovery (MLD) [RFC 3810]
    - MLD is part of ICMPv6

IGMPv3 Group Management (1): At a Host



- Joining or leaving group
  - Transmission of a membership report to 224.0.0.22, TTL=1
- Regular transmission of membership report
  - Timer per network interface

IGMPv3 Group Management (2): At a Multicast Router



- Multicast router
  - Member of 224.0.0.22
  - Periodic IGMP General Query to 224.0.0.1 (all systems), TTL=1
  - Group-specific queries: directly to group address

MLDv2 Group Management

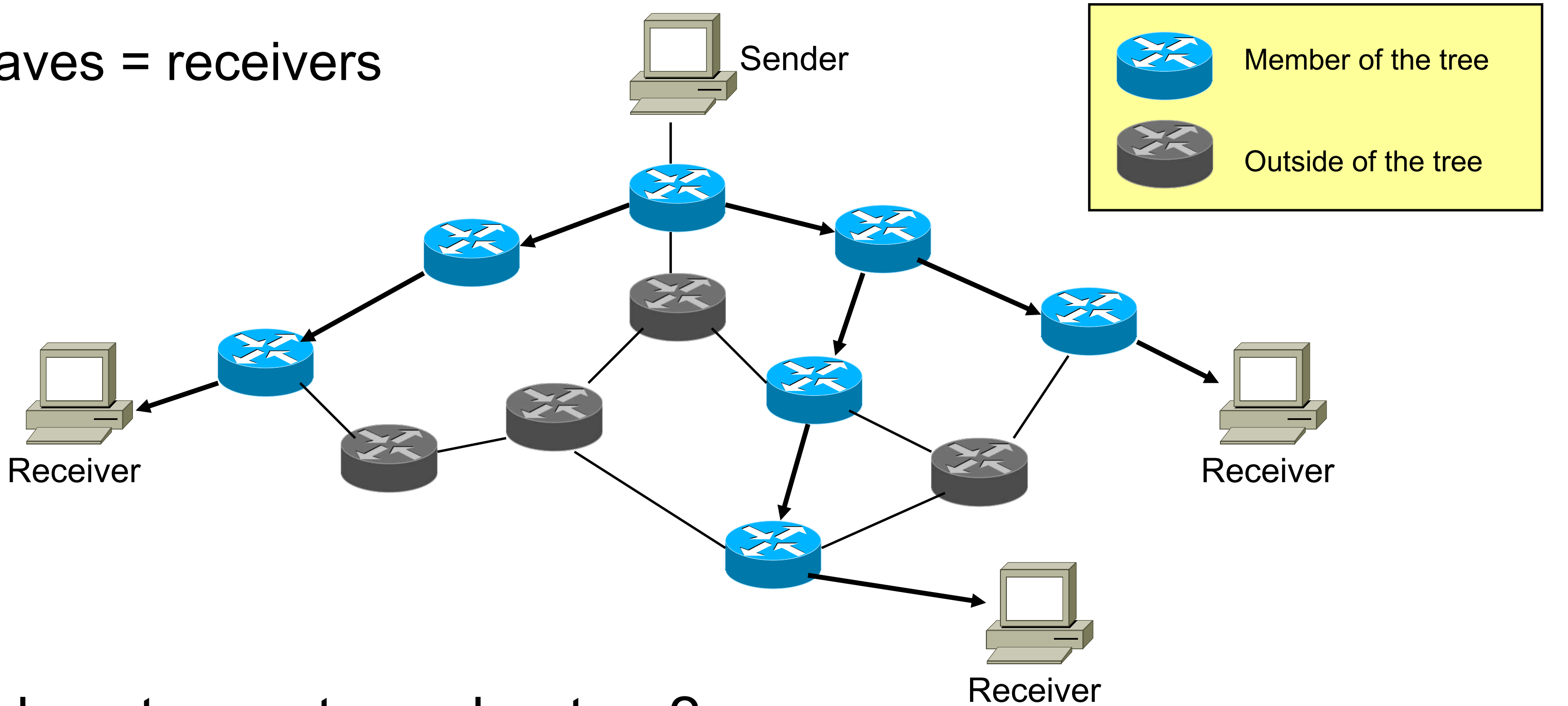
- Joining or leaving group
  - MLDv2 Membership Report
- Router querying membership
  - MLDv2 Membership Query to FF02::1 (“all nodes”)
- Answer to Membership Query
  - After random waiting time
    - To avoid congestion
  - To FF02::16 (“all MLDv2-capable routers”)

Multicast Routing: Flooding

- Flooding
  - Forwarding packet over all interfaces…
  - ..except the interface on which it has been received
  - Advantages
    - All receivers are reached, similar to broadcast
    - Very simple
  - Disadvantages
    - Loops possible (TTL/Hop Count helps, but inefficient)
    - Waste of bandwidth (in network parts without receivers)
- Improvements
  - No forwarding of duplicates => no loops
  - Requires identification of duplicates (e.g., copy or hash value)
- Any better approach?

Multicast Tree

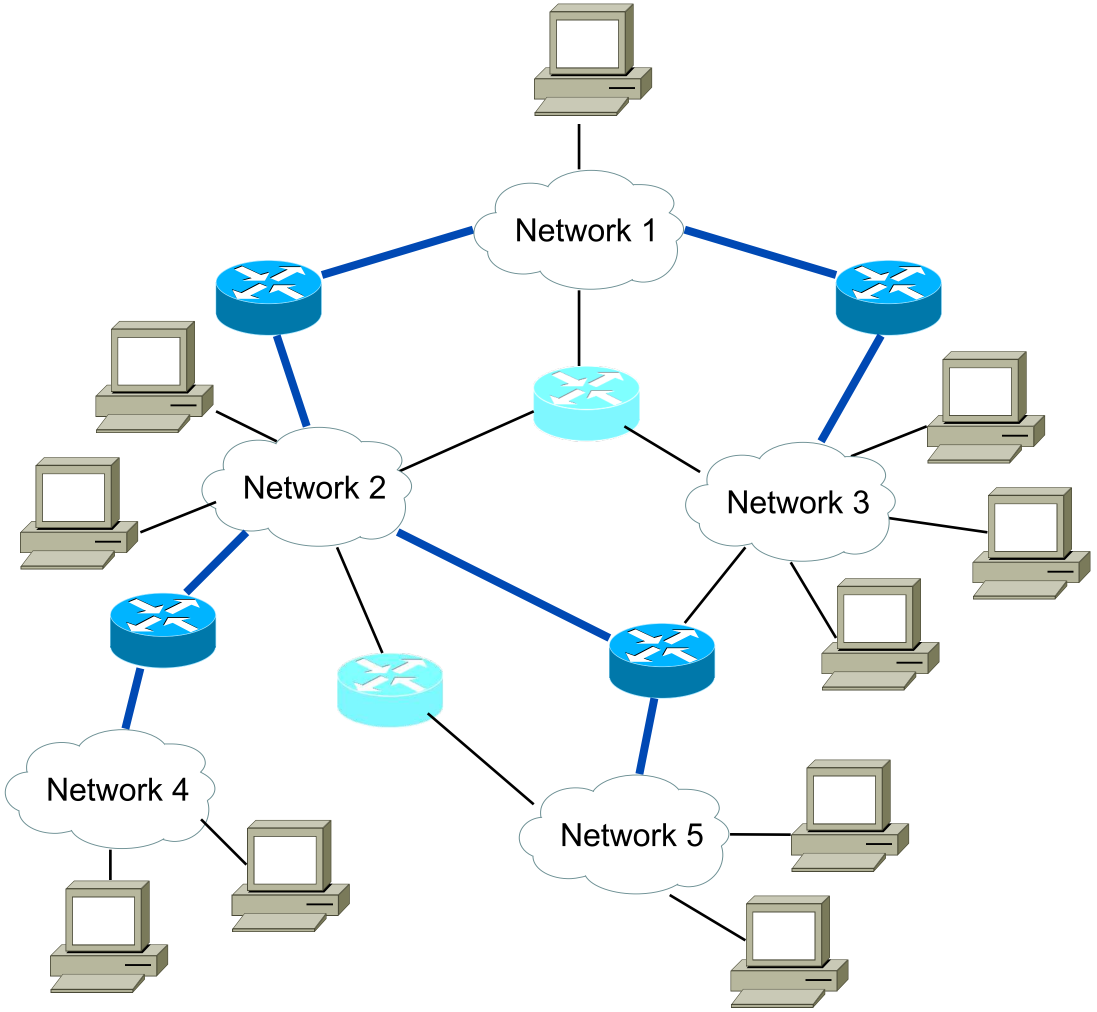
- Appropriate structure: a tree
  - Root = sender
  - Leaves = receivers



- … but how to create such a tree?

Spanning Tree

- Minimum-cost spanning tree
  - Given link costs
  - Easy to compute (see Kruskal’s algorithm)
- Remaining problem
  - No use of group membership
  - Forwarding into each network…
  - …even if there are no receivers
- Goal
  - Take group membership into account

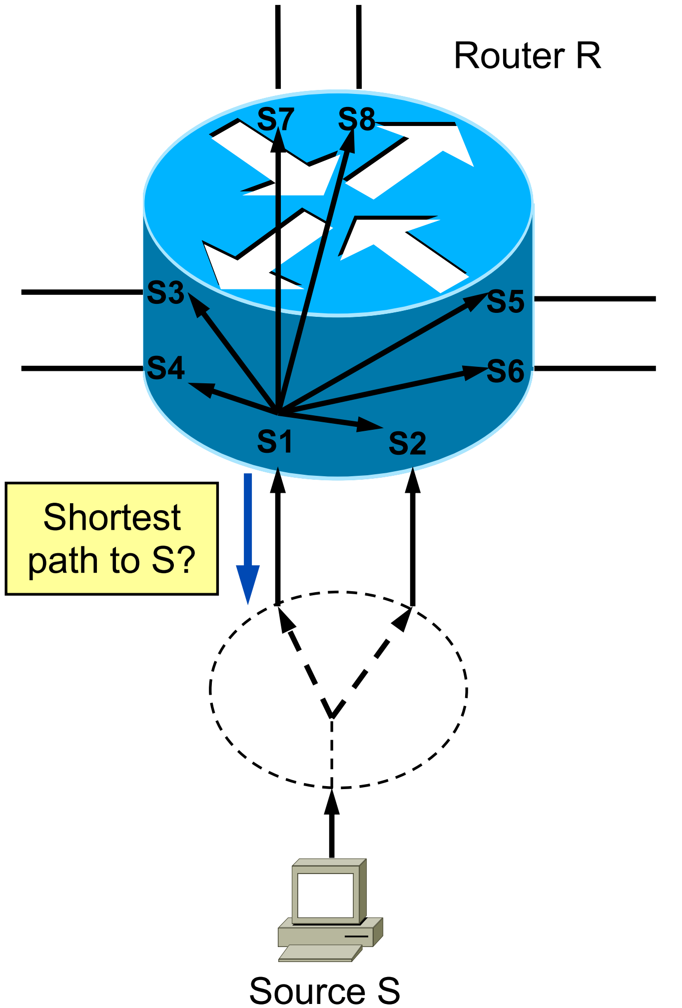


Challenges and Approaches

- Challenges
  - Packets must find route to receivers
    - Creation of an efficient multicast tree
    - Short/inexpensive paths
    - No unnecessary paths
  - Receivers need to find path to sender or to multicast tree
- Approaches
  - Flood and prune
    - First, flooding parts where might be receivers
    - Then, pruning parts without receivers
  - Rendezvous points
    - Suitable points in tree to add nodes

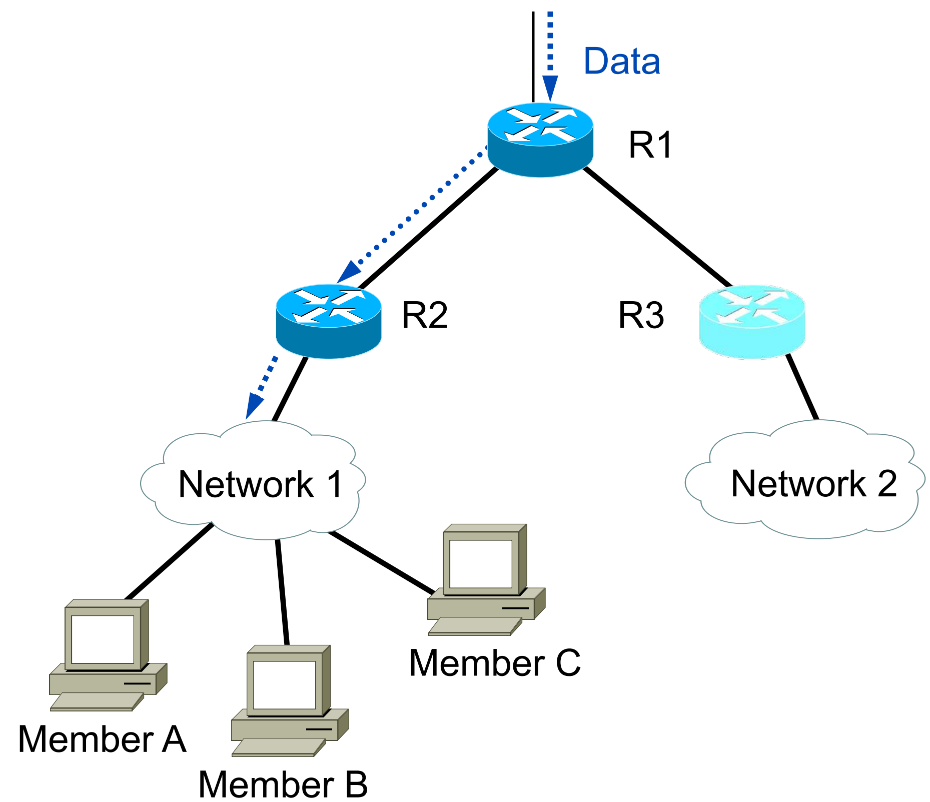
Reverse Path Forwarding

- Principle of Reverse Path Forwarding (RPF)
  - Router R receives packet from source S via interface S1
  - Would packet in reverse direction (from R to S) go out via S1?
    - Yes => flood the packet
    - No => drop it
- Advantage
  - Efficient and easy to implement
- Disadvantage
  - Still no use of membership
  - Nodes may receive duplicates



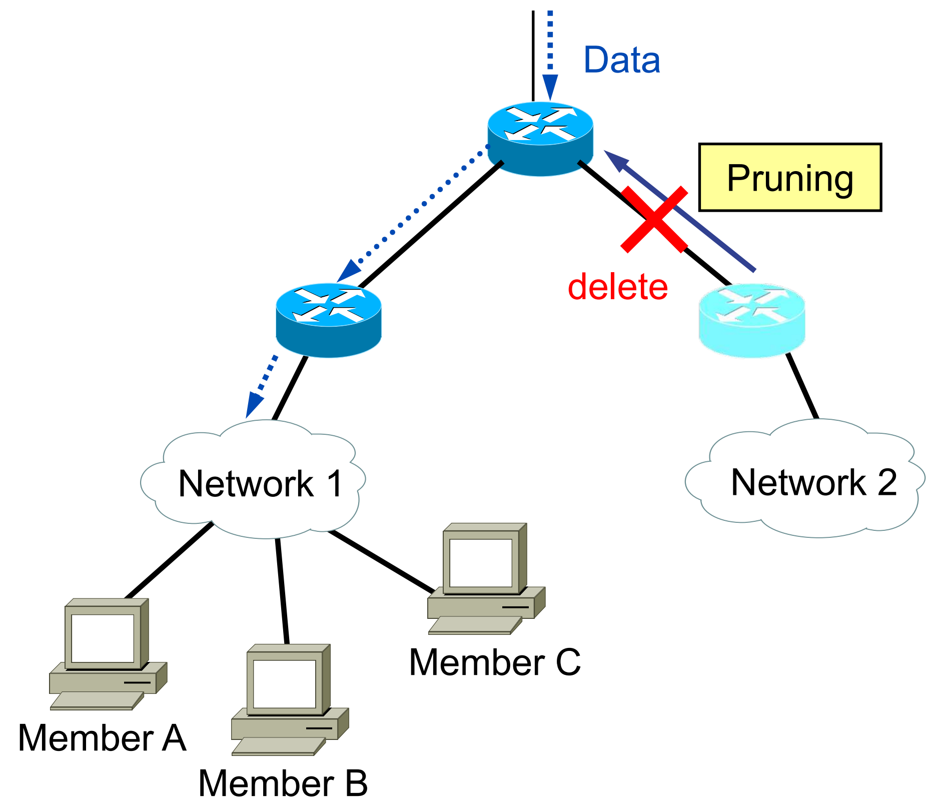
Adding Membership Information

- Goal
  - No transmission into networks without group members
- Requirement
  - Membership information in routers
- Example
  - Network 2 has no group members
  - Router R3 does not forward into network 2



Pruning

- Goal
  - No transmission to subnets without group members
- Procedure: flood and prune
  - Periodic flooding of data
  - Routers send Pruning message if there are no group members
- Further optimisation
  - Explicit Join message
  - Requires information about sender

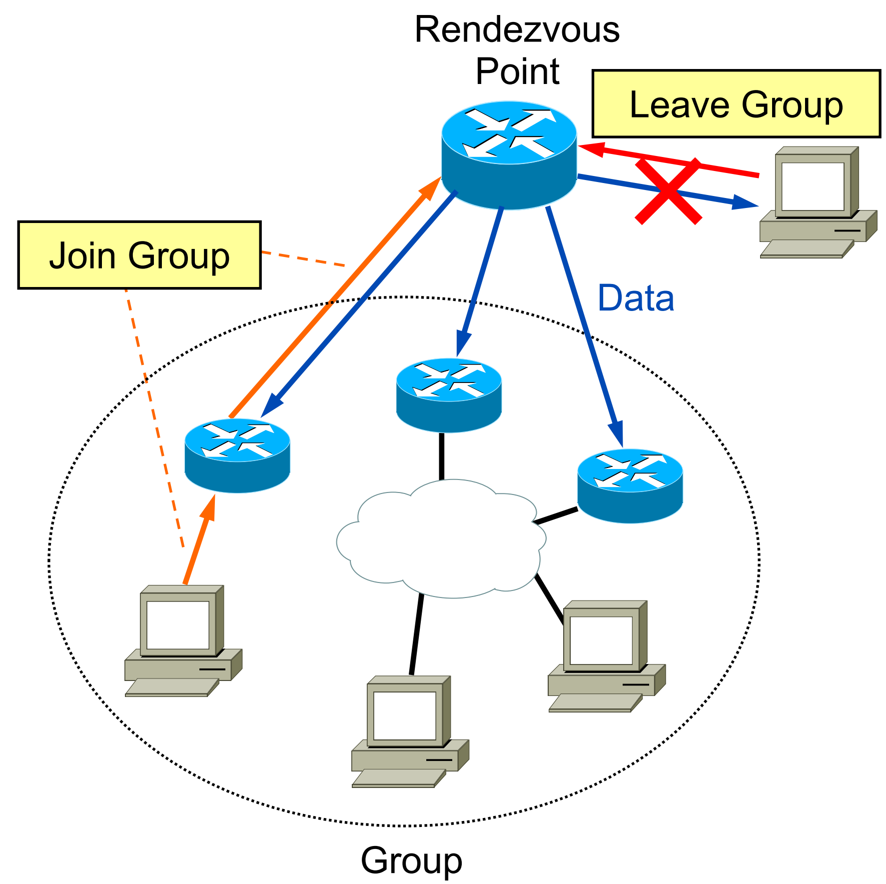


Flooding-Based Routing Mechanisms: A Comparison

Approach	Broadcast or Multicast?	Forwarding	Data Units to Forward	Forwarding Network Area
Flooding	Broadcast	All interfaces	All data units	Complete network
Reverse Path Forwarding (RPF)	Broadcast	All interfaces	Data units from shortest path	Complete network
RPF + Membership Information	Multicast	Interface on shortest path	Data units from shortest path	Not into subnets without members
RPF + Pruning	Multicast	Interface on shortest path	Data units from shortest path	Network areas with members only

Trees with Rendezvous Points

- Goal
  - Avoid flooding upon startup
- Procedure
  - One or more designated rendezvous points (RP) per group
  - Members join or leave at a rendezvous point
  - One spanning tree per group
- Disadvantage
  - Traffic concentration at rendezvous points



Comparison of Routing Approaches

Property	Flood and Prune	Rendezvous Points
Flooding	yes, but limited	no
Group Density	high	low
Traffic Concentration	no	yes
Complexity	low	moderate

Routing Protocol Approaches (1): Any-Source Multicast (ASM)

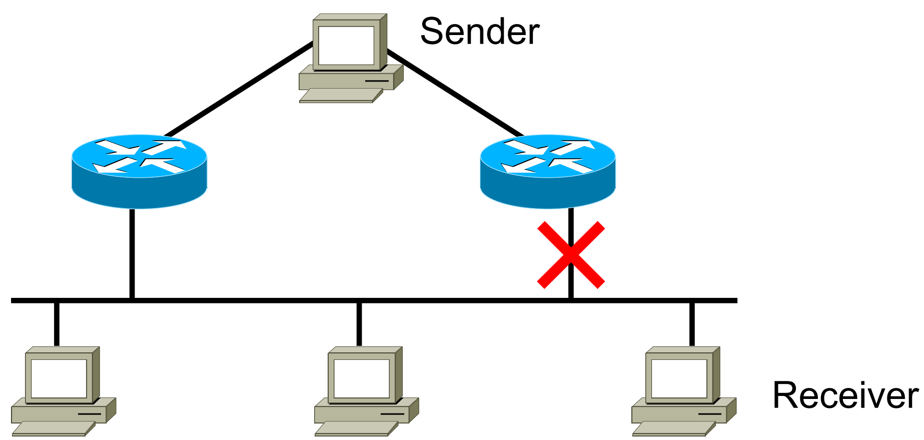
- Any-Source Multicast (ASM)
  - Arbitrary nodes can send to a multicast group…
  - …inclusive non members
  - Provides multipeer communication (M:N)
    - But denoted as “multicast” in RFC 1112 (first IGMP version)
- Challenges
  - No access control
    - Source cannot prevent other sources from transmitting
    - Reception cannot be restricted to certain sources
  - Inefficient
    - If there is only one source
    - Requires address allocation for group

Routing Protocol Approaches (2): Sender-Specific Multicast (SSM)

- Source-Specific Multicast (SSM) [RFC 4607]
  - 1:N Multicast
  - Channel concept
    - Channel (S,G) consists of source S and group address G
    - Unique identification
  - Solves many problems
    - No address allocation necessary (channel is already unique)
    - Access control
    - Routing efficiency
      - Sender is known => join into direction of sender
      - First router in multicast tree can satisfy join
      - => No flooding necessary
  - Support
    - IPv4 with IGMPv3: 232.0.0.0/8
    - IPv6 with MLDv2: FF3x::/32

Who is Responsible? Designated Router

- Problem
  - Multiple routers in the same subnet
  - => Incoming data can reach network over multiple paths



- Solution
  - Selection of designated router
    - E.g., selection of router with lowest IPv4/IPv6 address
  - Designated router becomes responsible for this network

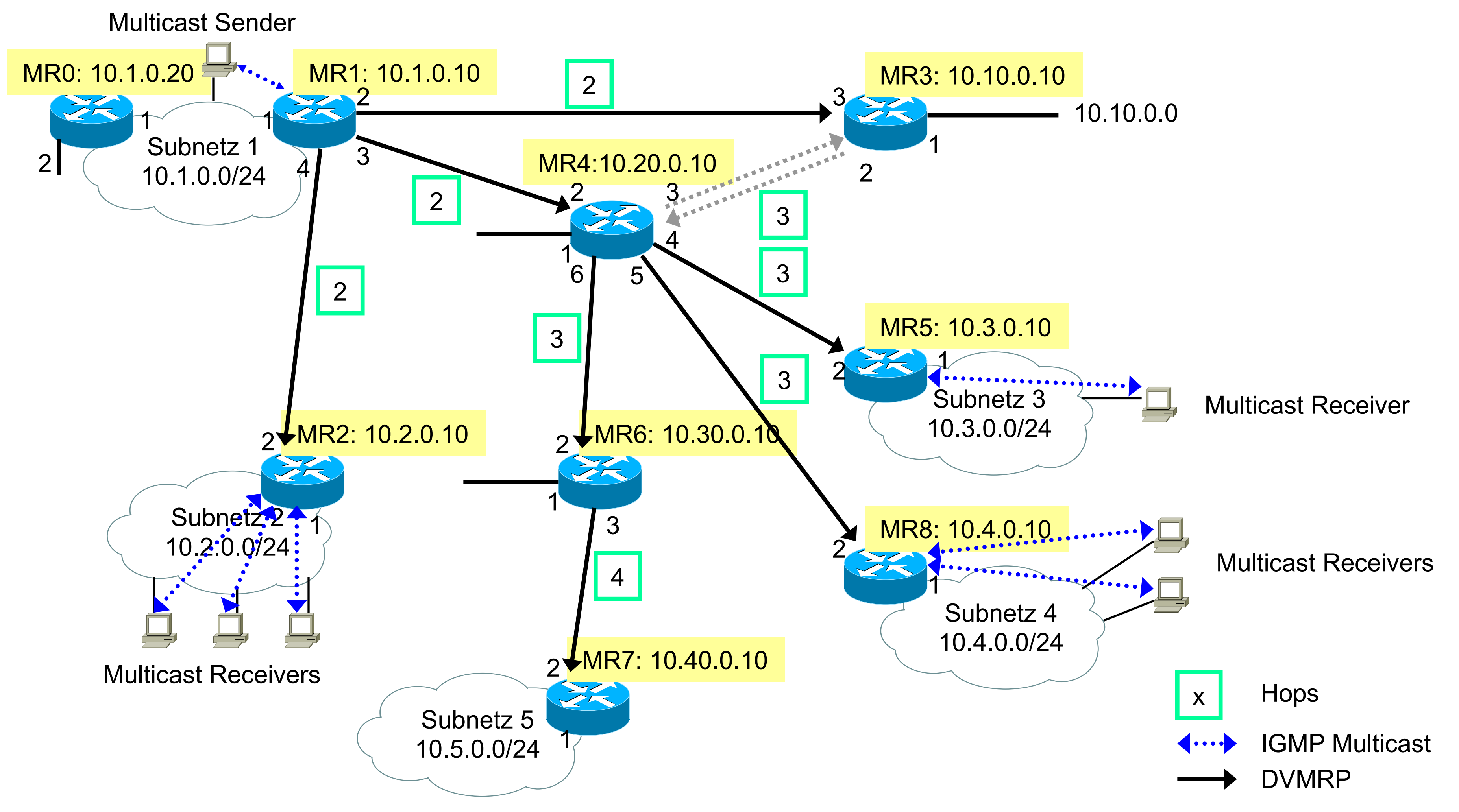
Overview of Multicast Routing Protocols

- Intra-Domain
  - Distance Vector Multicast Routing Protocol (DVMRP)
  - Multicast Open Shortest Path First (MOSPF)
  - <u>Protocol Independent Multicast (PIM)</u>
    - Sparse Mode (PIM-SM)
    - Dense Mode (PIM-DM)
    - Bidirectional Mode (BIDIR-PIM)
- Inter-Domain
  - Multicast Source Discovery Protocol (MSDP)
  - Border Gateway Multicast Protocol (BGMP)

Distance Vector Multicast Routing Protocol (DVMRP) [RFC 1075]

- Assumption
  - High group density
- Features
  - DVMRP is a distance vector protocol (derived from RIP)
  - For each sender: mapping to designated router (DR)
    - DR election in subnet: router with lowest IP address
  - For each group: generation of a multicast tree
  - Uses reverse path forwarding
  - Router can send pruning message if it has no receivers
    - Perform IGMP Membership Query to find out
  - Router can send grafting message if it gets receivers later
    - Grafting messages must be acknowledged by upstream router

DVMRP Example (1): Multicast Tree



DVMRP Example (2): Routing Table

Multicast Router	Source Network	From Router	Metric	Input Port	Output Port
MR0	10.1.0.0/24		1	1	2
MR1	10.1.0.0/24		1	1	2,3,4
MR2	10.1.0.0/24	10.1.0.10	2	2	1
MR3	10.1.0.0/24	10.1.0.10	2	3	1,2
MR4	10.1.0.0/24	10.1.0.10	2	2	1,~~3~~,4,5,6
MR5	10.1.0.0/24	10.20.0.10	3	2	1
MR6	10.1.0.0/24	10.20.0.10	3	2	1,3
MR7	10.1.0.0/24	10.30.0.10	4	2	1
MR8	10.1.0.0/24	10.20.0.10	3	2	1

Multicast Open Shortest Path First (MOSPF) [RFC 1584]

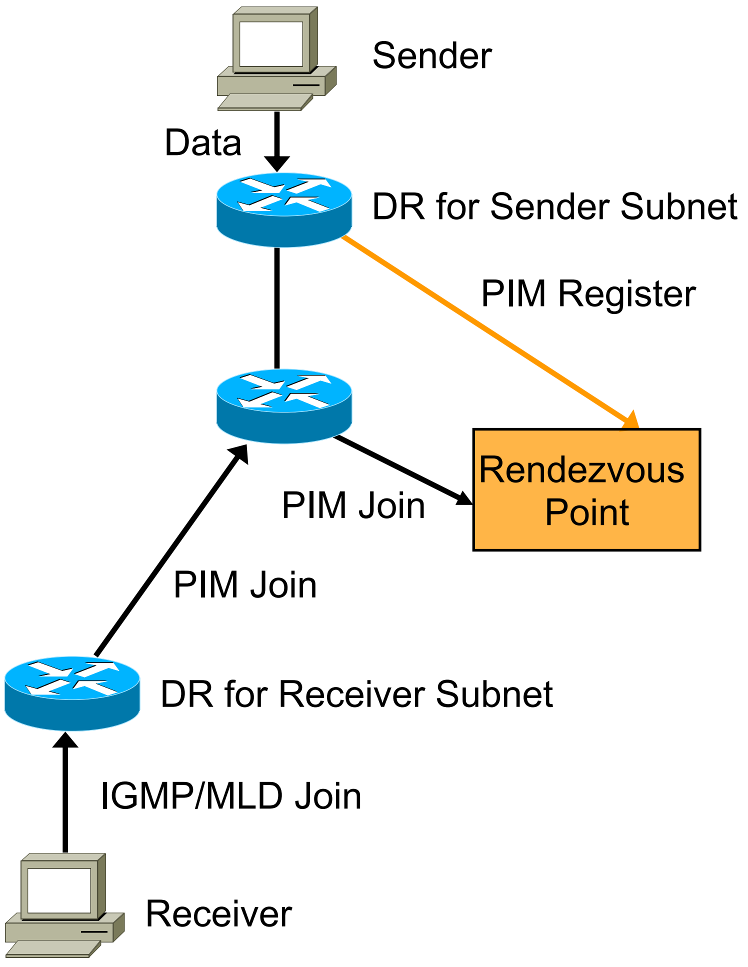
- Assumption
  - High group density
- Features
  - Extension of unicast OSPF with LSAs for multicast groups
    - Interoperability with (non-multicast) OSPF routers
  - MOSPF is link-state protocol => full topology knowledge
  - On-demand computation of multicast tree…
    - for (sender, group)-pair,
    - when source sends first packet to multicast group
  - Storage of computation in forwarding cache
- Advantages
  - Easy computation of multicast trees
- Disadvantages
  - Scalability: receiver-based state storage, even without senders

Protocol-Independent Multicast (PIM)

- Basic idea
  - Usage of existing underlying unicast routing protocol
- Variants
  - PIM-Dense Mode (PIM-DM) [RFC 3973]
    - For high group density
    - Based on flood and prune with reverse path forwarding
  - PRIM-Sparse Mode (PIM-SM) [RFC 4601]
    - For low group density
    - Based on rendezvous points
  - Bidirectional PIM-SM (BIDIR-PIM) [RFC 5015]
    - For multipeer communications (receivers are also senders)
    - Bidirectional tree
    - Only one multicast tree for all senders of a group

PIM-SM Routing (1)

- Phase #1: “Register”
- Receiver joining a group
  - Receiver joins a multicast group G
    - Using IGMP/MLD Join
  - Designated router (DR) in local subnet
  - DR sends PIM Join (*,G) in direction of rendezvous point (RP)
  - RP is root of the multicast tree
    - Consists of all routers receiving (*,G)
- Sender starting transmission
  - Sender S sends multicast packets
  - DR for sender network encapsulates them into PIM Register unicast messages to RP (PIM Register Tunnel)
  - Decapsulation/forwarding by RP

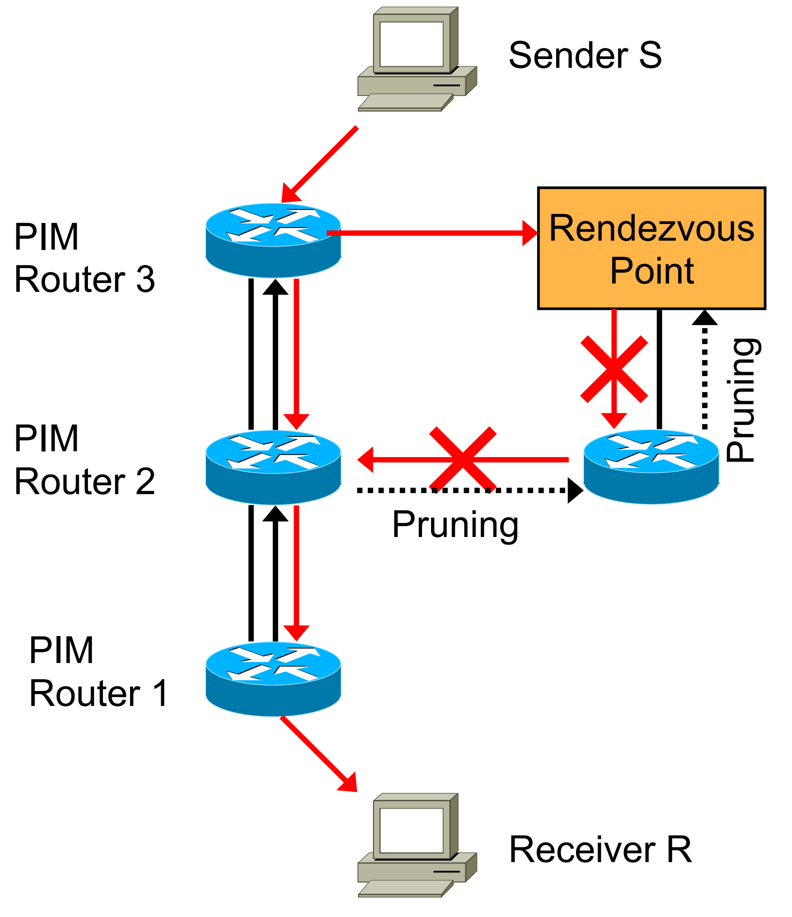


PIM-SM Routing (2)

- In first phase
  - Multicast between sender DR and RP as encapsulated unicast
  - => inefficient
- Phase #2: “Register Stop”
  - Upon reception of first PIM Register at RP
    - PIM Join (S,G) to sender S
  - Upon reception of first multicast message
    - PIM Register Stop to sender
    - Multicast transmission, no more unicast

PIM-SM Routing (3)

- Phase #3
  - "Shortest Path Tree" (SPT)
    - Goal: shorter path between sender S and receiver R
    - Creation of receiver-specific multicast tree (S,R)
    - Initiated by receiver
      - E.g., upon reaching a data rate threshold
    - Now: multicast reception over SPT and RP trees
    - Dropping of RP trees' packets
    - Pruning of RP tree by Pruning message (S,R,rpt) in direction of RP



PIM-SM Bootstrap Router

- Requirement for PIM-SM
  - Information about RP for given group address
  - Provided by Bootstrap Router (BSR)
- Procedure
  - Routers can be configured as:
    - Candidate for BSR (C-BSR)
    - Candidate for RP (C-RP)
  - Election of BSR
  - C-RP register at BSR, BSR selects set of active RPs (RP set)
  - BSR announces RP set within multicast domain
  - Routers use this RP set to find a RP for a group

Inter-Domain Multicast Routing Protocols

- Multicast Source Discovery Protocol (MSDP) [RFC 3618]
  - Interconnection of PIM-SM domains
  - For each group: RP per domain
    - Independence of remote domains, but…
    - … when sender gets active, notification to remote RPs
    - => scalability problems
  - Supports IPv4 only
  - IETF consensus on need for MSDP replacement
- Border Gateway Multicast Protocol (BGMP) [RFC 3913]
  - Supports IPv4 and IPv6
  - Designed to scale to the global Internet (like BGP)

Literature

- [RFC 1075]: “Distance Vector Multicast Routing Protocol”
- [RFC 1112]: “Host Extensions for IP Multicasting”
- [RFC 1301]: “Building a Network Information Services Infrastructure”
- [RFC 1458]: “Requirements for Multicast Protocols”
- [RFC 1584]: “Multicast Extensions to OSPF”
- [RFC 2080]: “RIPng for IPv6”
- [RFC 2328]: “OSPF Version 2”
- [RFC 2365]: “Administratively Scoped IP Multicast”
- [RFC 2464]: “Transmission of IPv6 Packets over Ethernet Networks”
- [RFC 3376]: “Internet Group Management Protocol, Version 3”
- [RFC 3550]: “RTP: A Transport Protocol for Real-Time Applications”
- [RFC 3618]: “Multicast Source Discovery Protocol (MSDP)”
- [RFC 3810]: “Multicast Listener Discovery Version 2 (MLDv2) for IPv6”
- [RFC 3913]: “Border Gateway Multicast Protocol (BGMP): Protocol Specification.”
- [RFC 3973]: “Protocol Independent Multicast - Dense Mode (PIM-DM): Protocol Specification (Revised)”
- [RFC 4601]: “Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)”
- [RFC 5015]: “Bidirectional Protocol Independent Multicast (BIDIR-PIM)”