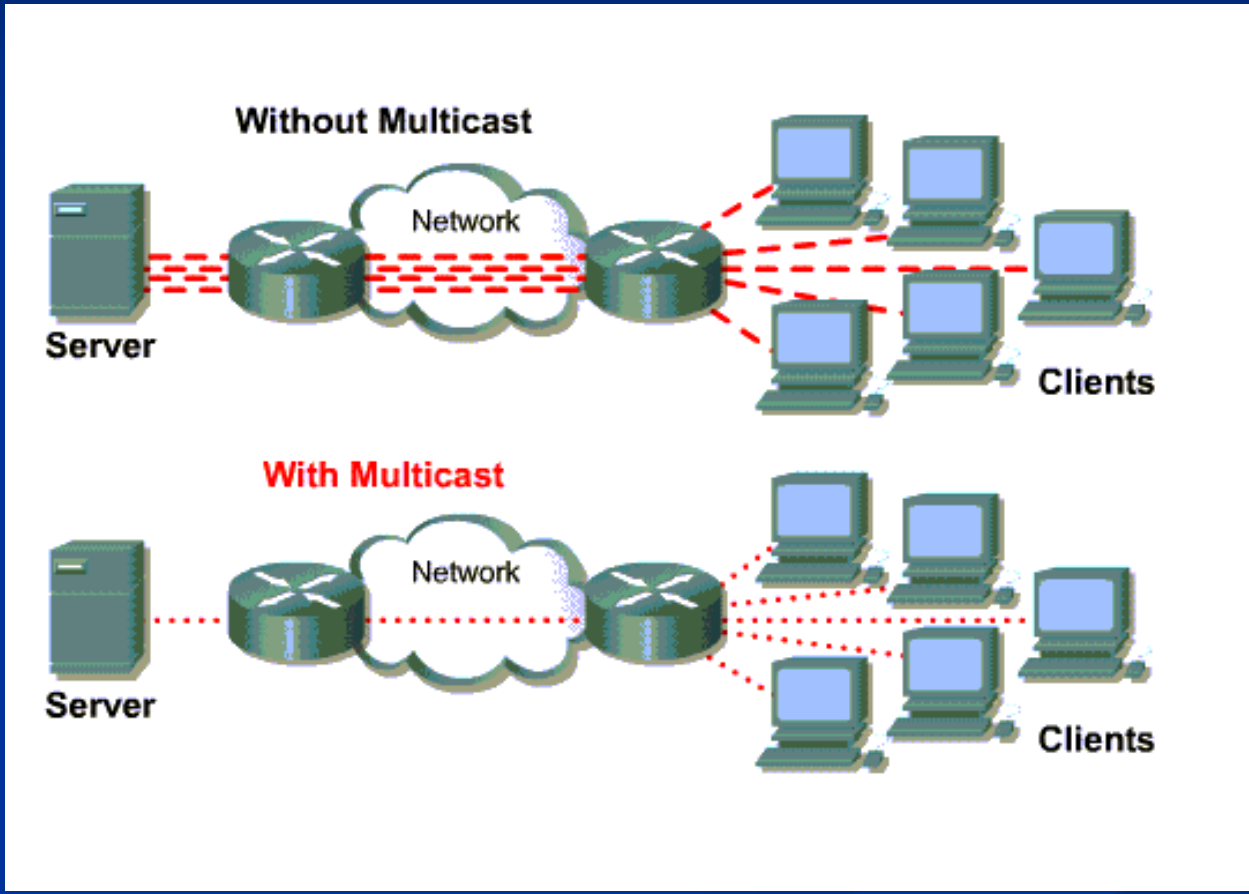
IP Multicast

Petr Grygárek

Multicast principle and usage



Multicast advantages

- Consumes less bandwidth on links which carry traffic for multiple receivers
- Packets duplicated only in routers where data flow has to be split into more branches with multicast receivers

Typical Multicast Applications

- Audio and video broadcasting
  - (unidirectional)
- Videoconferencing
- Service discovery
- Distributed simulations (including games)
- …

Paradigm of IP Multicast groups

- "Open" groups
  - Even station which is not part of a group can send packets into that group
  - Every station may became member of whatever group it wishes
- One station may be member of multiple groups simultaneously

IP multicast group addressing

IP multicast group addresses

- Uses D-class addresses:
  224.0.0.0 – 239.255.255.255
- Only valid as Destination addresses
- Source address is always unicast
  - Many multicast distribution mechanisms based on that fact

Reserved local (multicast) addresses

- 224.0.0.0 - 224.0.0.255
- Limited to local segment, TTL always 1
- Used mainly for routing protocols and router discovery
  - Communication between neighboring routers

Some reserved local addresses

Link Local IP Address	Usage
224.0.0.1	All systems on this subnet
224.0.0.2	All routers on this subnet
224.0.0.5	OSPF routers
224.0.0.6	OSPF designated routers
224.0.0.9	RIP Version 2 routers
224.0.0.10	EIGRP routers
224.0.0.12	DHCP server/relay agent
224.0.0.13	All P1M routers
224.0.0.22	IGMP
224.0.0.25	Router-to-switch (such as RGMP)

Globally Scoped Addresses

- 224.0.1.0 - 238.255.255.255
- assigned by IANA for common applications
  - example: 224.0.1.1 - Network Time Protocol, NTP

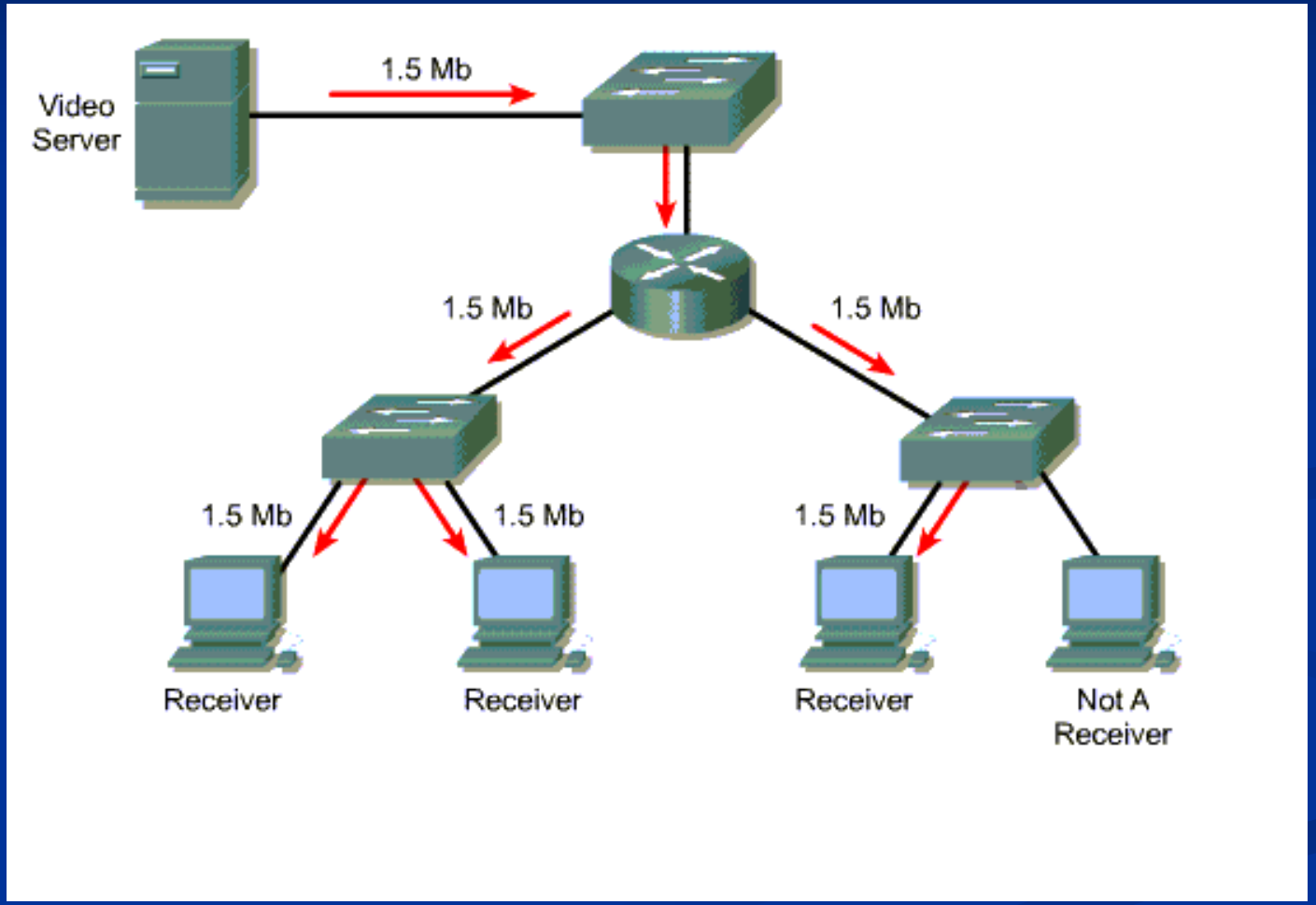
GLOP addresses

- 233.0.0.0 - 233.255.255.255
- Implicitly assigned to AS operators
  - Second and third byte encodes AS number
  - 255 AS-scope multicast groups

Limited Scope Addresses (RFC 2365)

- 239.0.0.0 - 239.255.255.255
- Similar to private IP addresses
- Used multiple times in independent networks
  - Scope defined by router configuration

Multicasting in LANs and WANs



Challenges of multicasting

- Mapping of IP multicast addresses into multicast MAC addresses
  - ARP works only for unicast addresses
- Distribution of multicast packets only into network branches with interested receivers
  - In routed and/or switched network

Multicasting in tree topology

- Multicast packets copied into branches where at least one receiver of the respective multicast group exists
  - Except incoming interface
- Layer 2 bridged/switched LAN is always tree

Multicasting in topology with loops

- Loops would cause cycling of multicast packets
- Need to construct distribution tree
- Distribution tree root choices
  - Router of network where multicast source resides
    - "First-hop" router
  - Some predefined router (Rendezvous Point).
    - Source sends multicast packets to Rendezvous Point using unicast tunnel or source is part of the distribution tree (upstream to RP)
- Every router has to know which interface leads to distribution tree root and which interfaces lead to branches with multicast group receivers
  - Stored in special kind of routing table for multicasts

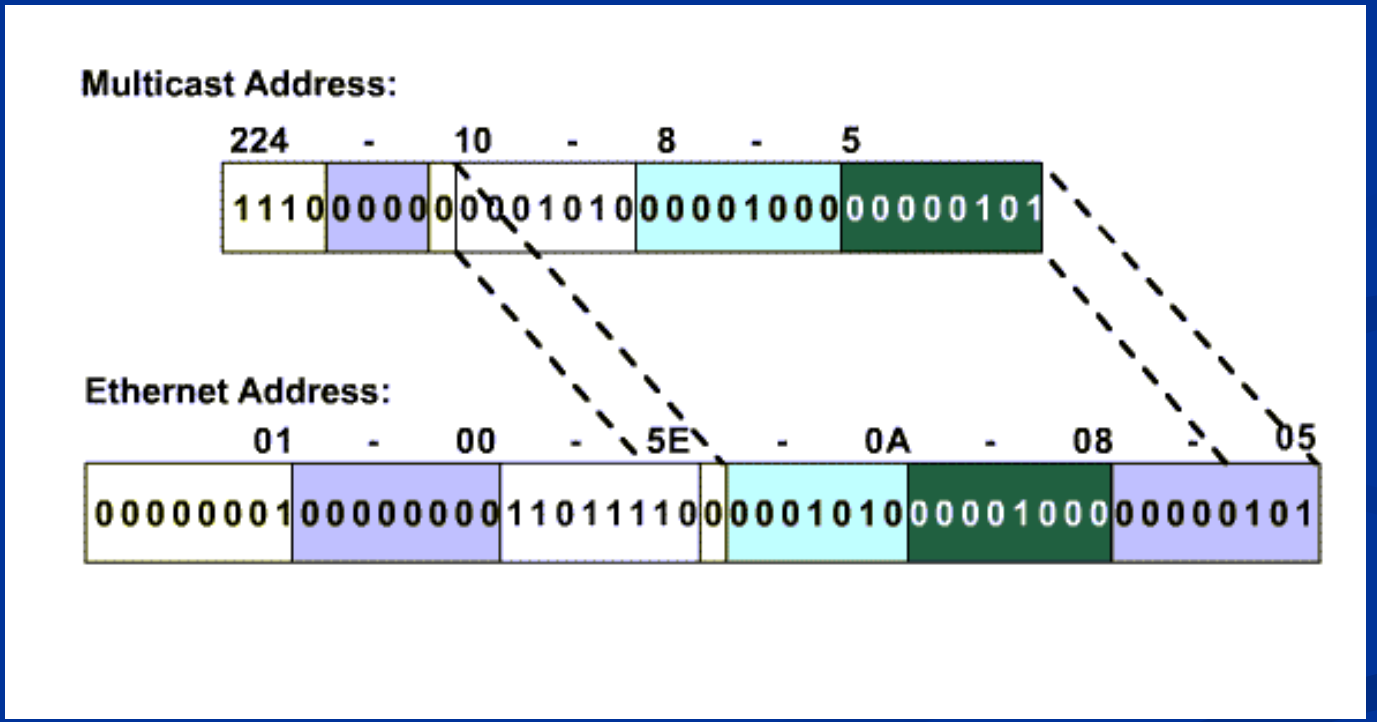
IP to MAC multicast address mapping

IP to MAC multicast address mapping process

- Multicast MAC addresses start with 01-00-5E
  - MSB of fourth byte always 0
  - Only 23 bits of MAC available for mapping
- Multicast IP address always start with 1110 (class D)
  - remaining 28 bits have to be mapped
- Last 23 bits of multicast IP address copied into last 23 bits of MAC address
- 5 bits of multicast IP address not mapped
  - 2^5 (32) IP multicast groups mapped to the same MAC address
    - Additional filtering apply at layer 3 (device drivers)

IP to MAC multicast address mapping example

224.10.8.5 -> 01.00.5e.0a.08.05

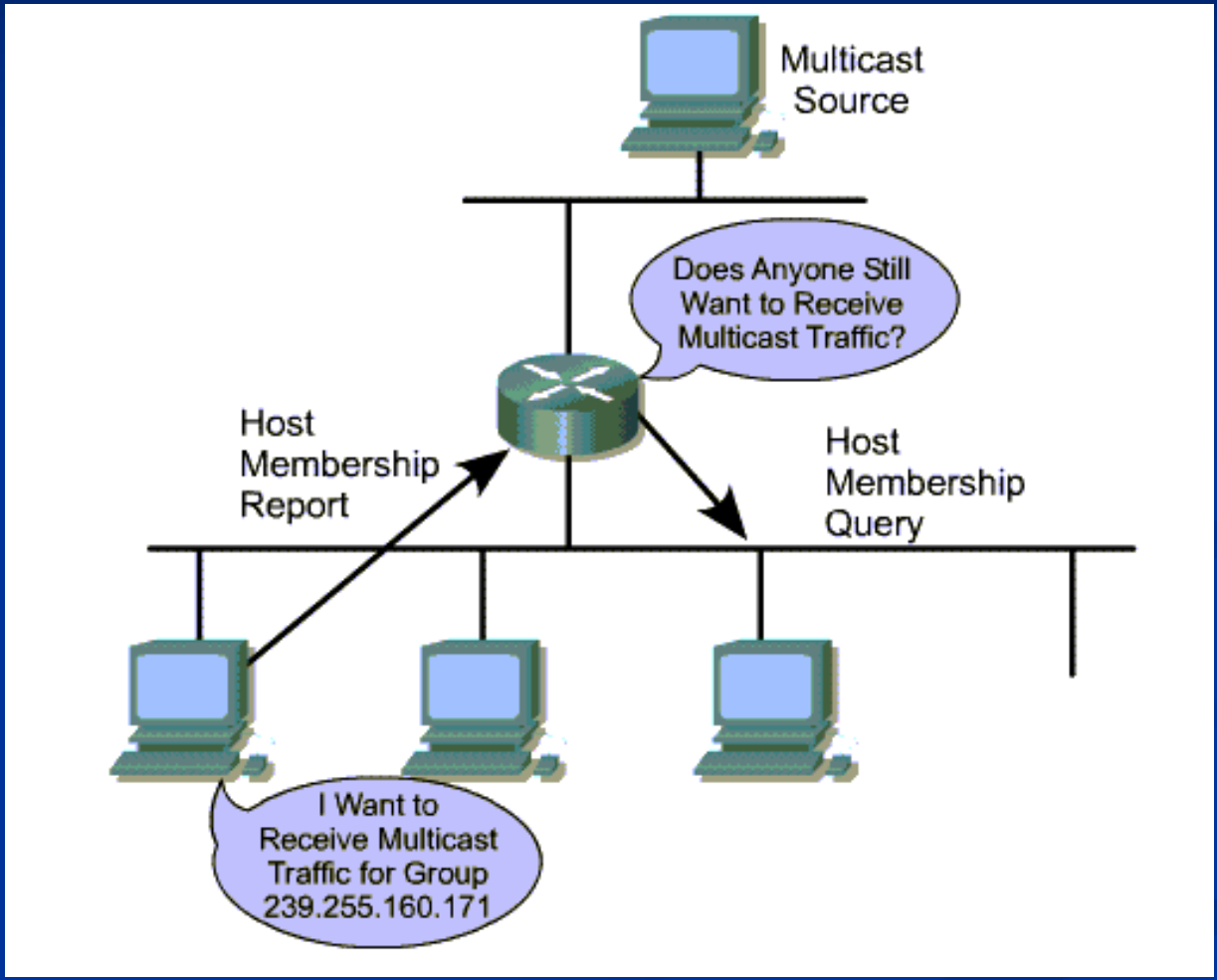


Internet Group Membership Protocol (IGMP)

Purpose of IGMP

- Used by multicast receivers to join a multicast group
  - Join message processed by router designated to transmit multicasts to the local segment
  - IGMP version 2 also allows receiver to inform about desire of group leave
- Used by router to query whether there is still some receiver of a multicast group on it's particular interfaces
- Also used by routers running multicast routing protocol PIM v.1 to join to distribution tree

IGMP operation



IGMP version 1 messages

- MEMBERSHIP REPORT
  - Unsolicited join to multicast group
  - Positive response to router's membership query
    - delayed a random interval do avoid response bursts

  Sent to address of the respective multicast group
- MEMBERSHIP QUERY
  - Periodically sent from interfaces of multicast router to query whether there still exist (any) multicast group receiver on each interface

IGMP version 2

Additional messages:

- LEAVE GROUP message
  - explicit deregistration from multicast group
- Group-specific MEMBERSHIP QUERY
  - used by router to verify presence of group receivers after reception of LEAVE GROUP

IGMP version 3

- allows to filter required multicast traffic based on source address
  - defense against abusers of multicast group
- allows to list stations from which receiver wants or doesn't want multicast traffic

Multicast processing on L2 switches

L2 multicast processing options

- Simple switches may simply handle multicasts the same way as broadcasts
- Sophisticated switches try to determine what unicast (MAC) addresses are interested in receiving individual multicast group's traffic and send frames destined to each multicast group only to ports where these address reside
  (based on normal switching table)
  - mechanism of mapping of IP multicast groups to multicast MAC addresses is known

How switch learns about multicast group members ?

1. Special protocol between LAN segment multicast router and switches on that segment
2. IGMP Snooping

1. Special router-switch protocol

- Commonly implemented on low-performance switches without support for IGMP Snooping
  - Proprietary protocol (Cisco: CGMP)
- Multicast router of LAN segment informs all switches of that segment about unicast MAC addresses interested in each multicast group traffic
  - Router gets mapping information from IGMP Joins
  - Router sends mapping information to reserved multicast MAC address to all switches
  - Switches flood mapping to other switches

2. IGMP Snooping

- Switch “snoops” into IGMP (L3+) information carried by passing frames
  - MEMBERSHIP REPORT, LEAVE GROUP
- Requires intensive processing, hardware support needed
  - Snooping is limited only to multicast frames
  - Only frames carrying IGMP are inspected

Multicast Distribution Trees

Distribution tree

- Subset of network topology
- Covers all networks where receivers interested in traffic of multicast group are located
- Needed to avoid loops and assure delivery of particular group's multicast traffic to every interested receiver
- As receivers deregister or new ones appear, multicast tree has to be updated ("pruning" and "grafting" of tree branches)

Distribution tree types

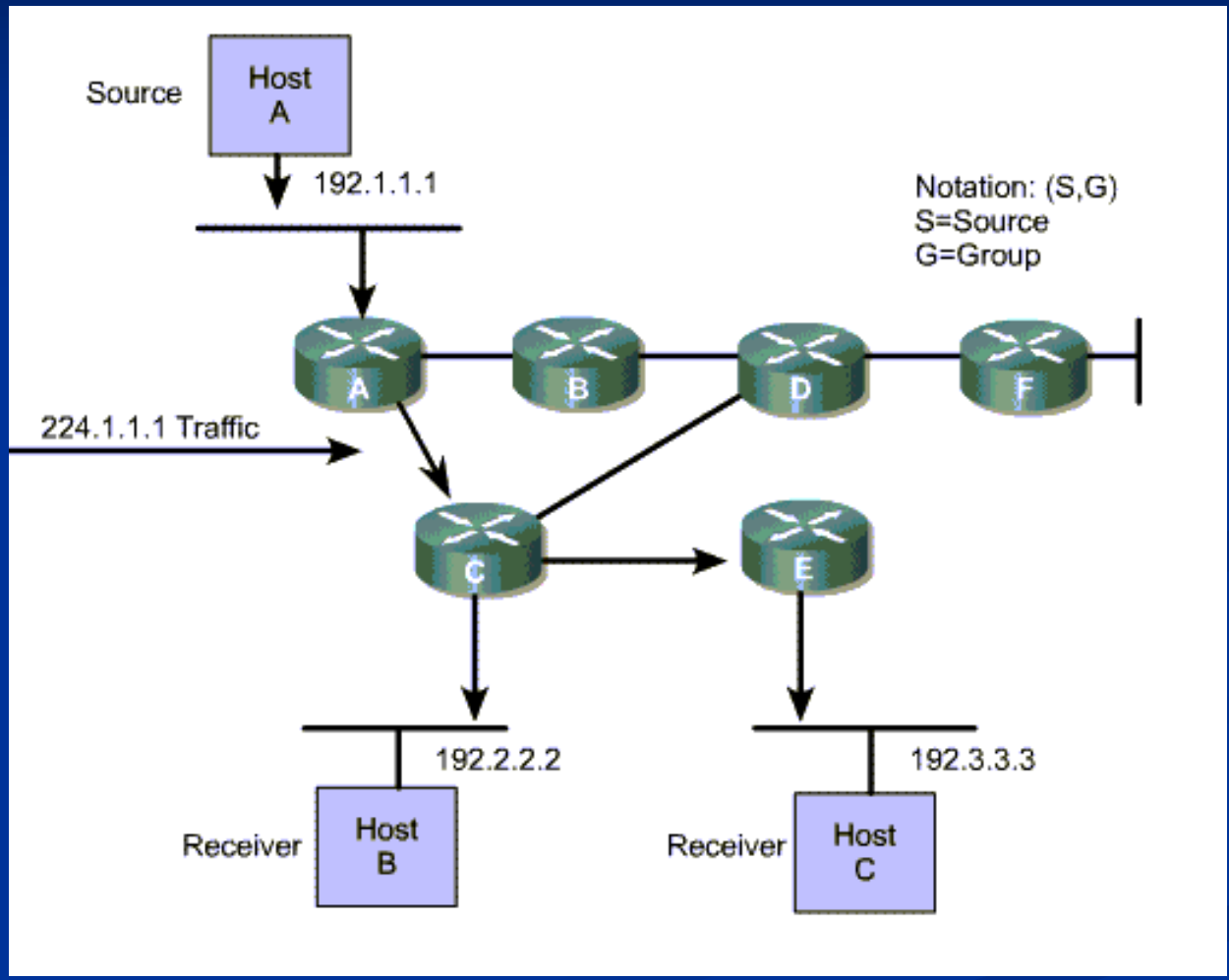
- Shortest-Path Tree (Source Tree)
- Shared Tree

- Unidirectional Tree
  - from root to leaves
- Bidirectional Tree
  - from source up to the root and down to branches simultaneously

Shortest-Path Tree (SPT) (Source Tree)

- Shortest paths tree rooted in particular multicast source, cover all networks where receivers of particular group are located
- Denoted as < S,G >
  - S = source address
  - G = multicast group address
- Separate tree for every source sending to every multicast group has to be maintained
  - Optimal, but poorly scalable

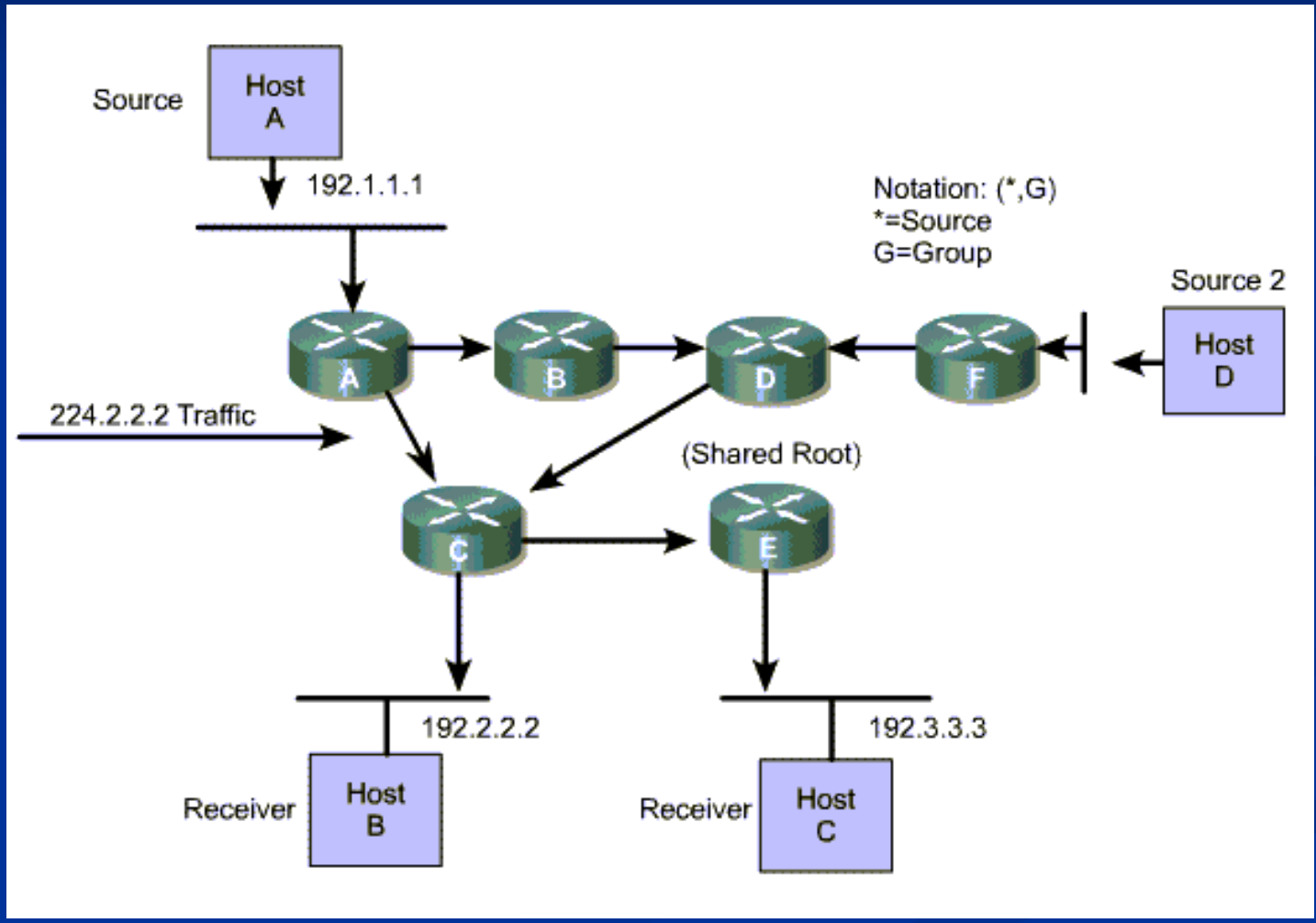
Example: < 192.2.2.2, 224.1.1.1 >



Shared Tree

- Common tree for all sources of multicast group
- Some router is designated to act as root
  - Obviously called Rendezvous Point (RP)
- Denoted as < *,G >
- Single tree for all multicast groups
  - saves resources
- Not optimal
- Source has to send multicast packets to the RP

Example: < *, 224.2.2.2 >



Multicast Routing

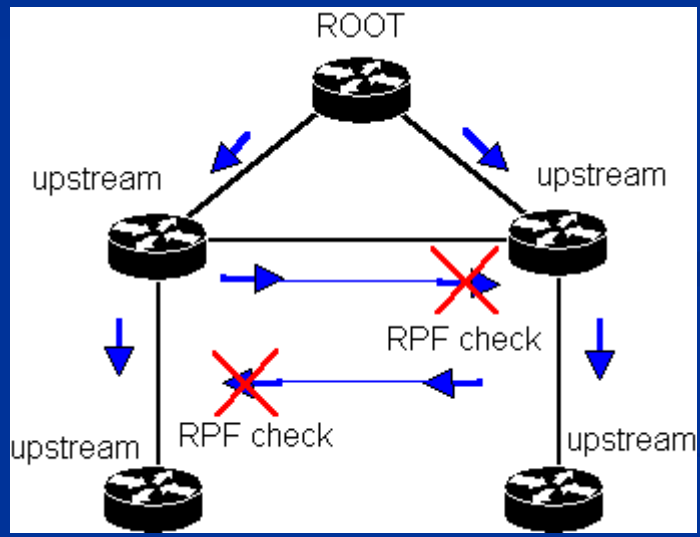
Multicast routing principles

- Routing based on SOURCE address
- Packet routed along distribution tree still away from multicast source

Reverse Path Forwarding (RPF)

- Used to route multicast away from source
- Interfaces to forward multicast to are determined using normal routing table
- Interfaces classified as “upstream” and “downstream” according to position in distribution tree
- RPF check: Only multicast packets arriving from upstream interface are forwarded next, others are dropped
  - Even with RPF check, some segments may receive multiple multicast packet copies (but packets don’t circulate in loops)

RPF check pitfalls



Function of multicast routing protocols

- For every distribution tree, routers need to know upstream and all downstream interfaces
- If multiple routers reside on the same segment, one of them has to be chosen to route multicast traffic to shared segment
  - Based on their distances to the root

Multicast routing table

Built using RPF check and/or multicast routing protocols

For every <*,G> or <S,G> tree, one multicast routing table entry is maintained with the following information:

- single upstream interface
  - Interface normally used to send unicast packets to distribution tree root (multicast source/RP)
  - Determined using normal unicast routing table
    - If multiple equal-cost path exist, router chooses the one with lowest next-hop address
- List of downstream interfaces
  - Interfaces to network segments where multicast group receivers reside
    - end-stations or multicast routers

Multicast routing table

- <*,G> entries:
  as many entries as number of active multicast groups
- <S,G> entries:
  number of multicast groups multiplied by number of sources contributing to every group

Some multicast routing protocols use <*,G> tree first and switch to <S,G> tree if an extensive traffic comes from particular source

- <S,G> tree entry always preferred over <*,G> entry

Multicast Routing Protocols

Multicast routing protocols classification

- Dense Mode
  - Assumes that multicast group receivers are located on most network segments (i.e. densely distributed)
  - DVMRP, PIM-DM, MOSPF
- Sparse Mode
  - Assume that multicast group receivers are located only on some network segments (i.e. sparsely distributed)
  - PIM-SM, CBT

Dense mode

- Assumes receivers on all network segments
- By default, multicast traffic distributed to all network segments
- Router with no receivers of multicast group on it's downstream interfaces can ask to "prune" multicast tree branch
  - if a new multicast receiver registers, router can ask to "graft" the branch again immediately
- After some time (typically 3 min), upstream router will timeout prune request and continue to send multicast traffic to all branches
  - flood-and-prune cycle repeats periodically

Sparse mode

- Assumes receivers only on sparsely distributed network segments
- By default, multicast traffic is not forwarded
- When multicast receiver registers with router, router sends join message in the direction of tree root
- Routers on the path add interface hearing join request to the list of downstream interfaces for the tree
  - After few minutes, router timeouts the join request and stops sending multicast traffic
  - Downstream routers need to resend join message periodically
- Only routers on the distribution tree must deal with multicast
  - More suitable for WAN environment than dense mode protocols

Forwarding of multicast routing protocol control messages

- Prune and Join/Graft messages always flow from leaves to the root of the tree
- Root of every distribution tree known to all routers
  - (source or RP)
- Since distribution tree is always the shortest paths tree from tree root, router can simply send control messages out of interface used by unicast routing to reach the root

Protocol independent multicast (PIM)

- Most widely used today
- Independent on specific unicast routing protocol used in the network
  - but uses it's (unicast) routing table
- Multicast routed using RPF check
  - + prune messages to avoid unnecasery duplicates
- In fact, PIM is not a routing protocol
  - Does not send nor receiver routing updates

Dense mode multicast routing protocols

Distance Vector Multicast Routing Protocol (DVMRP)

- First publicly used multicast routing protocol
- Router copies multicast packets to all interfaces except those leading to multicast source ("reverse-path flooding")
- Multicast tree branch can be pruned if no receiver is present
  - Prune request has limited lifetime
  - After expiration, multicast flooding continues
    - Needed for newly started receivers
- Upstream interface determined using it's own unicast distance routing protocol (similar to RIP)
  - Hop-count metric, maximum 32 hops
  - periodic updates every 60 secs
- Today used only to bridge between different multicast routing protocols
  - But is more and more replaced with MBGP

DVMRP Problem

Because DVMRP uses it's own unicast routing protocol, multicasts are routed using routes (shortest paths) independent of routes used for unicast traffic

- It can bring a lot of unexpected problems

Protocol Independent Multicast Dense Mode (PIM-DM)

- Floods multicast from source into all networks
- Distribution tree branch may be pruned if no receivers are located there
  - Every 3 minutes flooding refreshed
    - (and possibly pruned again)
- Effective only in special conditions:
  - Receivers and senders not too far
  - Few senders, many receivers
  - Intensive and constant multicast traffic flow

Multicast OSPF (MOSPF)

- Extension of OSPF
  - Limited to single OSPF routing domain
- Multicast group membership information for individual network segments distributed in Link State Updates
- Every router calculates distribution tree for every <S,G> pair independently
  - Suitable when not too many < S,G > pairs active at the same time

Sparse mode multicast routing protocols

Protocol Independent Multicast Sparse Mode (PIM-SM)

- Suitable for sparsely distributed receivers and intermittent multicast flows
- Uses RP as distribution tree root
- Last-hop router explicitly joins distribution tree
  - Sends join message toward RP
- RP for every multicast group configured manually in every router or learnt from multicast announcement
  - PIM bootstrap process distributes group-to-RP mapping
- Single tree for a group requires less state information maintained, but results to suboptimal routing

PIM-SM source registration

- Distribution trees are unidirectional
- Sources (first-hop routers) are upstream to RP
- First-hop routers register with RP when source sends multicast packet
  - RP sends Join request toward particular source in response to Register-start message
  - Before distribution tree is extended to source, first-hop router encapsulates multicast packets and sends them to RP
  - When RP hears multicast packets from particular source delivered using standard forwarding (along just finished branch of distribution tree), it sends Register-stop message to the first-hop router
    - First-hop router stops sending encapsulated multicast packets to RP and starts to transmit along distribution tree

PIM SM: switchover to SPT

- Initially, shared tree with RP as root is used
  - Needed to learn about multicast sources
- If predefined threshold is exceeded, last-hop router initiates switch to source tree
  - sends join toward particular multicast source
  - prunes specific <S,G> pair from shared tree

PIM sparse-dense mode

PIM can operate in sparse mode for some groups and in dense mode for the others

- Router uses sparse mode if it knows RP for multicast group
- Otherwise, dense mode is used

PIM neighbor discovery

PIM neighbor table created using PIM Hello messages (multicast with TTL=1)

- Pruning state maintained for every neighbor
- Supports election of Designated router
  - Act as IGMP Querier for LAN segment
  - In PIM-SM, issues Join message when multicast group receiver registers on LAN segment via IGMP

PIM Messages

Version 1: IGMP headers

Version 2: separate protocol

- (103, carried in IP packets)

- Hello
- Register,Register-stop (SM)
- Join/Prune
- Graft, Graft-ACK (DM)
- Assert
- Bootstrap, Cadidate-RP-Advertisement

Core-Based Tree (CBT)

- Single (shared) tree for every group
  - "Core" router is tree root
- Distribution tree is **bidirectional**
  - Member senders starts distribution of their traffic starting with first-hop router
  - Non-member senders tunnels multicast traffic to Core router
- First-hop routers explicitly ask to join distribution tree
  - State information built in routers when join request is acknowledged (by core router or some router at "core" tree) and acknowledgement sent back to the requestor
- Still under development, three incompatible versions

Internet Multicast Backbone (MBONE)

- Not all Internet routers support multicasting
- MBONE=experimental Internet multicast backbone

Multicasting in multi-AS environment

- PIM-SM preferred
  - PIM-DM Flooding not suitable for WAN
  - ISPs wants to run their own RPs, independent of RPs of other ISPs
    - MSDP protocol to distribute information about active sources between RPs in different AS-es
- MBGP can propagate separate routes for multicast
  - Used for RPF checks
  - Multicast traffic between AS-es may use other routes than unicast traffic

Links for lab

- MGEN User's guide: http://computing.ee.ethz.ch/sepp/mgen-3.0-mo/
- MGEN homepage http://computing.ee.ethz.ch/sepp/mgen-3.0-mo.html
- http://pf.itd.nrl.navy.mil/project/showfiles.php?group_id=16396&release_id=18343