IP Multicast

Compendium

Introduction

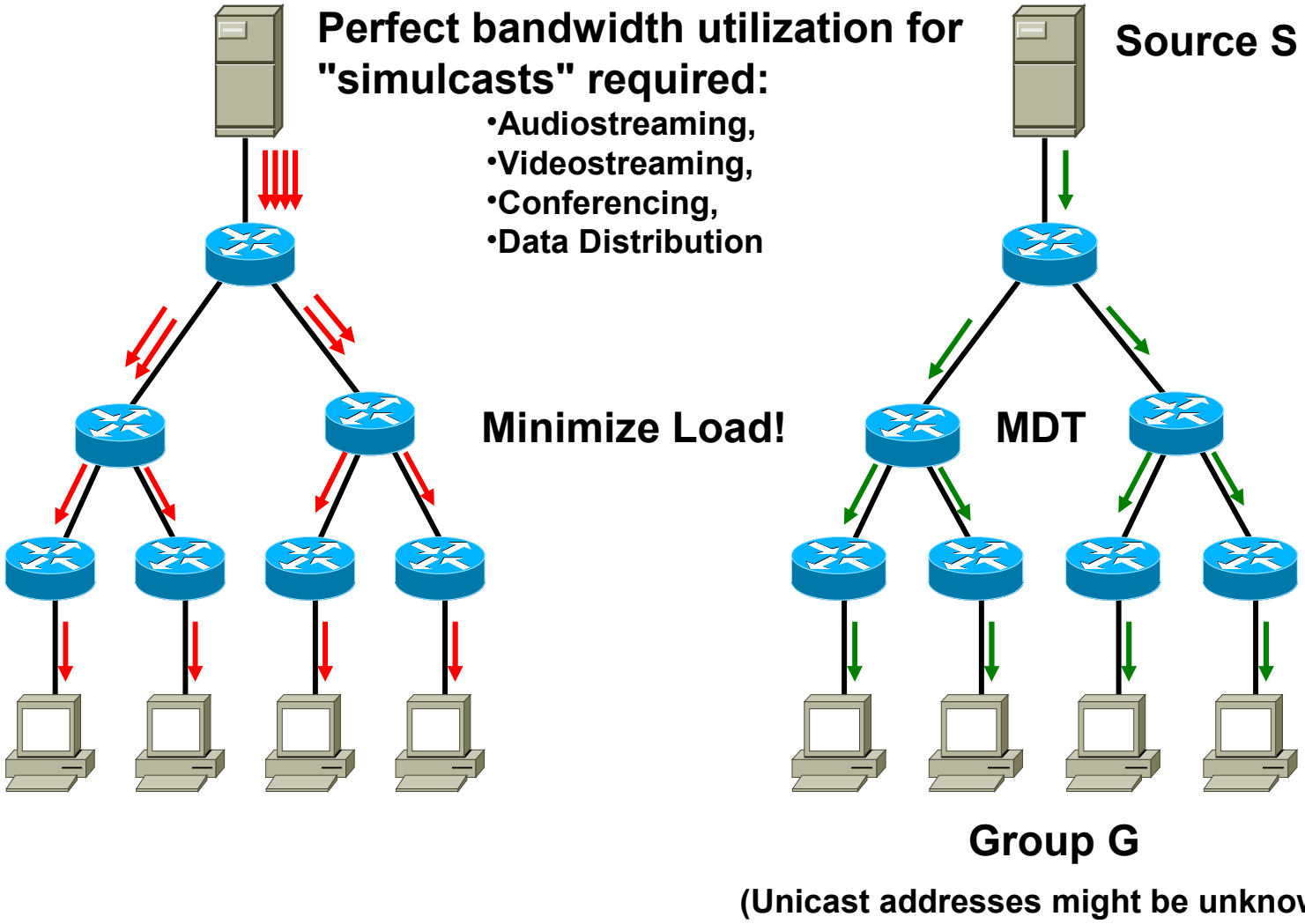
New IP Applications

- Corporate Broadcasts
- Distance Learning/Training
- Video Conferencing
- Whiteboard/Collaboration
- Multicast File Transfer
- Multicast Data and File Replication
- Real-Time Data Delivery for Financial Applications
- Video-On-Demand
- Live TV and Radio Broadcast to the Desktop
- Multicast Games

Multicast Models

- One-to-many
  - One host is multicast source, other hosts are receivers
  - Simplest and most important type
  - Might only be jitter sensitive (voice/video)
- Many-to-many
  - Hosts are both senders and receivers
  - All hosts are in same multicast *group*
  - Might be delay sensitive (bidirectional communication forbids more than 0.5 sec delays)
- Flexible variants
  - Many-to-one (implosion problem!)

Unicast vs. Multicast



Facts

- Developed in the late 1980s
  - First used 1992 during IETF Conference
- Building block for QoS
  - RSVP and RTP
- UDP based
  - No Congestion Avoidance!
  - Packet drops occur!
- Classification based on distribution trees
  - Shortest Path Trees
  - Shared Trees

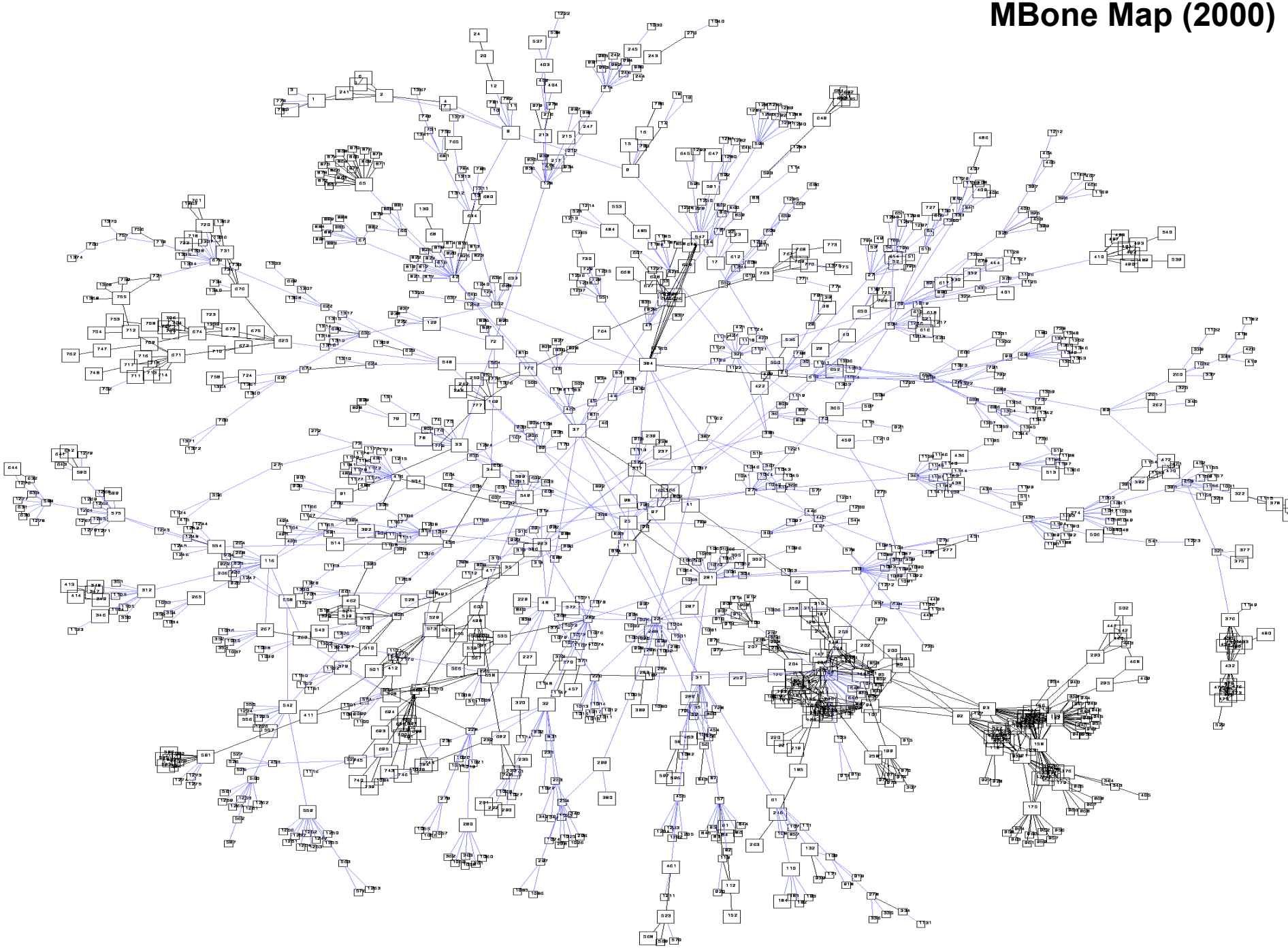
How IP Multicast Works...

- Sources don't care at all!
  - Simply send multicast packets to the first-hop router
- First-hop router
  - Forwards multicast packets into the multicast-tree
- Intermediate routers
  - Determines upstream interface (to first-hop router) and downstream interfaces (RPF check)
- Last-hop routers
  - Are leafs of this tree
  - Receive users registration via IGMP
  - Communicate group membership to upstream routers

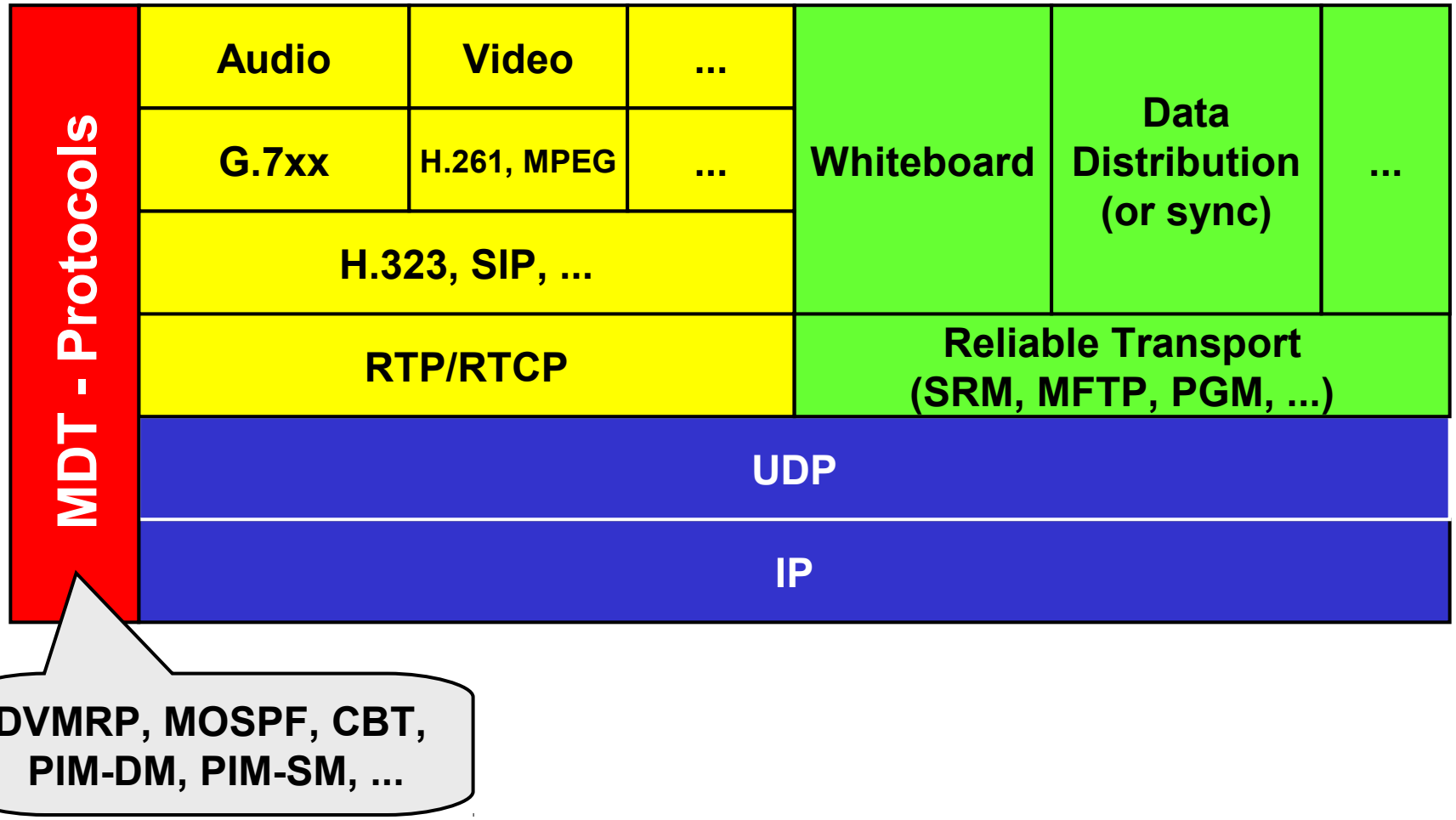
The Mbone

- World-wide multicast backbone
  - Based on tunnels
  - Playground for experiments
- Rich Mbone toolset
  - Session Directory (SDR)
  - Visual Audio Tool (VAT)
  - Robust Audio Tool (RAT)
  - Video Conferencing Tool (VIC)
  - Whiteboarding Tool (WB)

MBone Map (2000)



Integrated Multicast



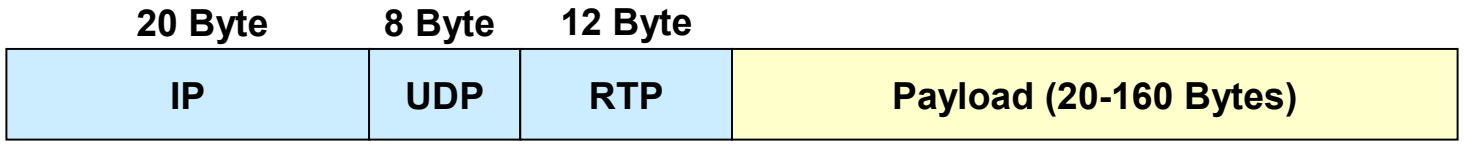
Realtime Protocols

Audio and Video

- Are typically transported by RTP/RTCP
- Feedback mechanism very important
  - For maintaining multicast distribution tree (MDT)
  - For applications to switch codecs when bandwidth becomes scarce

Realtime Transmission

- Real Time Transport Protocol (RTP)
  - Connectionless environment
  - Payload type identification and sequence numbering
  - Time-stamping and delivery monitoring
- RTP Control Protocol (RTCP)
  - Provides feedback on current network conditions
  - Helps with lip synchronization and QoS management, etc



RTP Facts

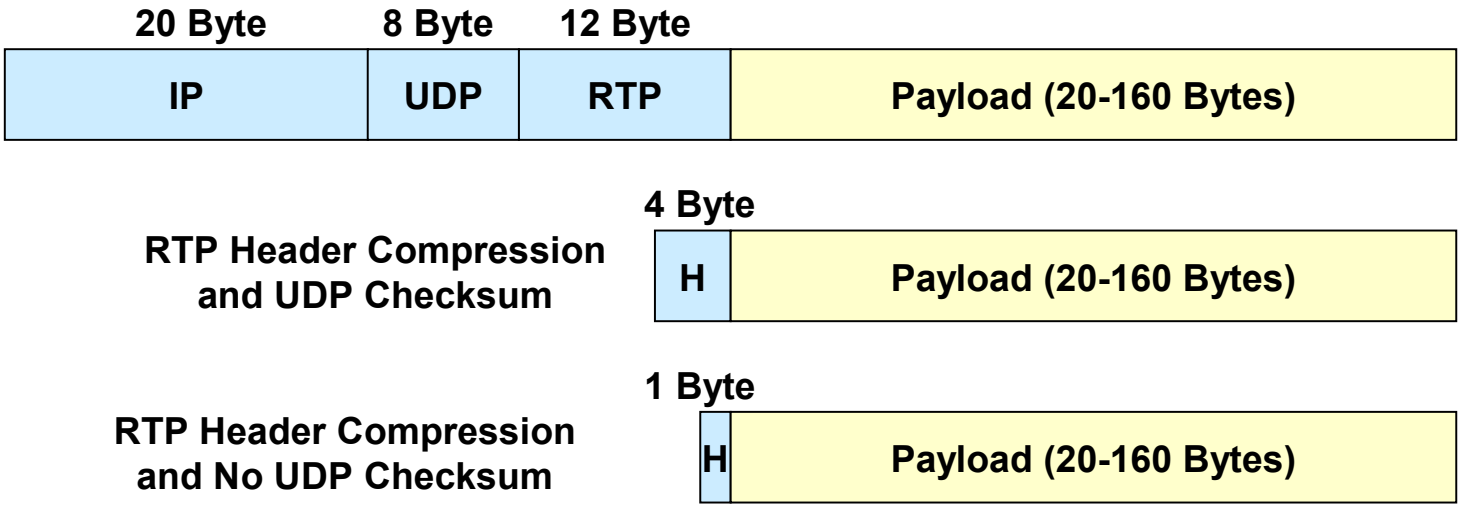
- RTP does NOT provide:
  - Reliable packet delivery
  - QoS
  - Prevent out-of-order delivery
- RTP uses *mixers*
  - Special relays to combine separate video streams into one video stream
  - Also care for synchronization
  - Optionally re-encode an original stream to meet link-specific bandwidth requirements

RTCP Facts

- Sent by RTP receivers
  - RTCP provides feedback for RTP senders *and other receivers!*
  - Sent to same multicast group!
- RTP sender (=multicast source) uses RTCP information to
  - Log group activity
  - Measure QoS conditions
- Other RTP receivers learn total RTCP utilization
  - Try to keep total utilization below 5% of network bandwidth

RTP Compression

- Simple substitution principle
  - Only point-to-point !
  - Not CPU intensive !
  - Might be memory greedy



Realtime Streaming Protocol

- RTSP = "Internet VCR remote control protocol"
- Efficient delivery of streamed multimedia over IP networks
  - Client-Server based
  - Large-scale audio/video on demand
  - VCR-style control functionality
- Also uses RTP for delivery
- RFC 2326

Multicast Addresses

Reserved Class D Addresses

- IANA reserved range 224.0.0.0 to 224.0.0.255 to be *local scope:*
  - 224.0.0.1 = all multicast systems on subnet
  - 224.0.0.2 = all routers on subnet
  - 224.0.0.4 = all DVMRP routers
  - 224.0.0.5 = all OSPF routers
  - 224.0.0.6 = all OSPF designated routers
  - 224.0.0.9 = all RIPv2 routers
  - 224.0.0.10 = all (E)IGRP routers
  - 224.0.0.13 = all PIMv2 routers

Other Class D Addresses

- Global scope: 224.0.1.0 to 238.255.255.255
  - Internet-wide dynamically allocated multicast applications
  - Typically Mbone applications
- Administratively scoped: 239.0.0.0 to 239.255.255.255
  - Locally administrated multicast addresses (like RFC 1918 addresses)
  - Organization-local scope: 239.192.0.0/14
  - Site-local scope: 239.255.0.0/16

Static Group Address Assignment for Interdomain Multicast

- Temporary method to allow Internet content providers to assign static multicast addresses
  - For inter-domain purposes
- Group range 233.x.x.0 to 233.x.x.255
  - x.x contains AS number
  - Remaining low-order octet used for group assignment within AS

SSM Addressing

- For globally known sources and source-specific distribution trees
  - Across domains
- Group range: 232.0.0.0/8
  - 232.0.0.0 to 232.255.255.255

Dynamic Multicast Addressing

- Method of SDR (Mbone)
  - Sessions announced over well-known multicast groups (e.g. 224.2.127.254)
  - Address collisions detected and resolved at session creation time via lookup into an SDR cache
  - Not scalable
- Multicast Address Set-Claim (MASC)
  - Hierarchical concept
  - Extremely complex garbage-collection problem
  - Under development

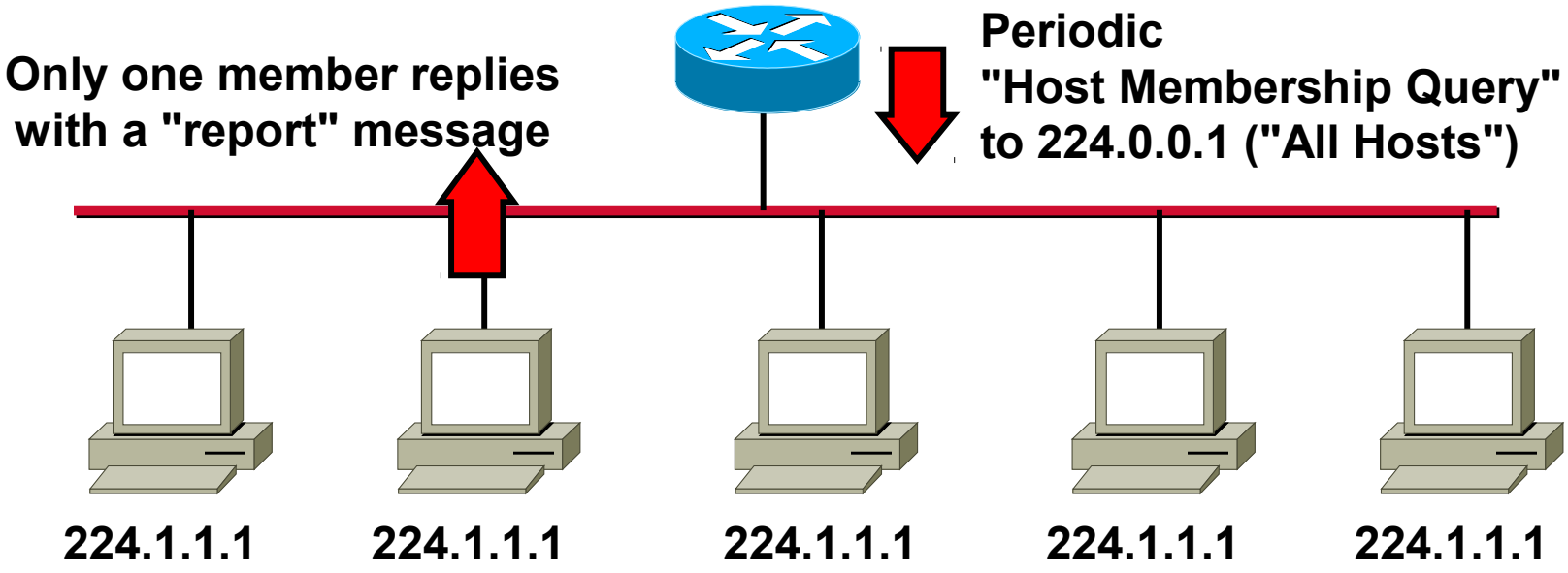
IGMP

Internet Group Membership Protocol

- Used (mainly) by hosts
  - To tell designated routers about desired group membership
  - Supported by nearly all operating systems
- IGMP Version 1
  - "I want to receive (*, G)"
  - Silly: Leaving group only by being silent...
  - Specified in RFC 1112 (old)
- IGMP Version 2
  - Also: "I do not want to receive this any longer"
  - Specified in RFC 2236 (current)
- IGMP Version 3
  - "I want to receive (S, G)"
  - DR can directly contact source
  - Still under development

IGMP

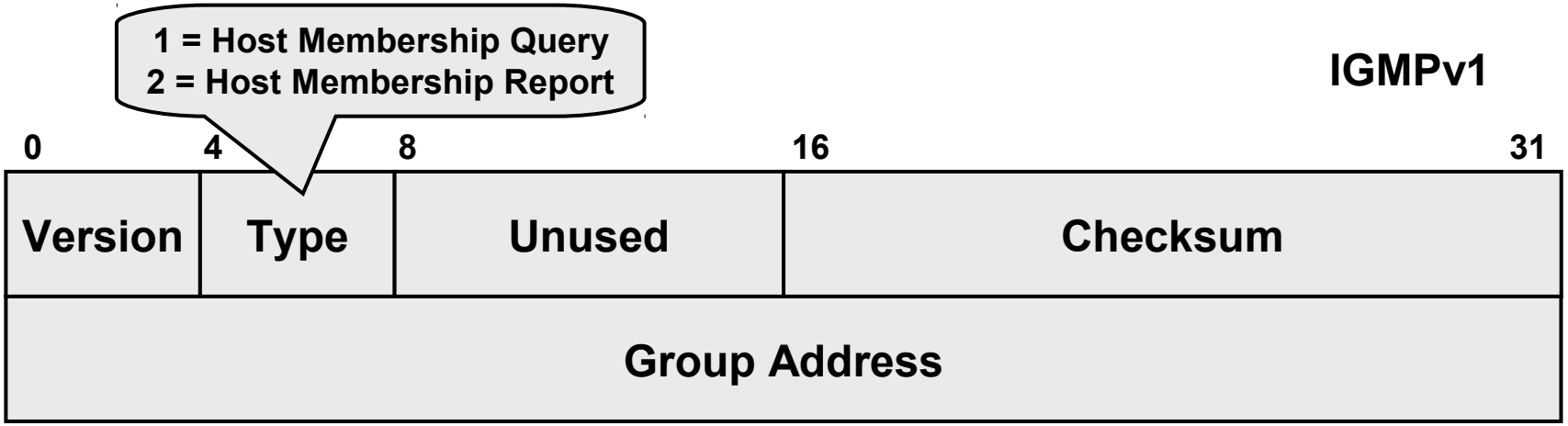
- DR send every 60-120s Host Membership queries to 224.0.0.1
  - Telling all active groups to local receivers
- Interested hosts send IGMP "report"
  - With destination address = group address !
  - Countdown-based, TTL=1



Other Important Differences

- IGMPv1
  - Does not elect designated query router
    - Task for multicast routing protocol (different mechanisms implemented)
    - Often results in multiple queriers on a single multiaccess network
  - Makes general queries only
    - Contain listing of all active groups
- IGMPv2 (backwards compatible with IGMPv1)
  - Router with lowest IP address becomes IGMP querier on this LAN segment
  - General queries specify "Max Response Time"
    - Maximum time within a host must respond
  - Allows for group-specific query
    - To determine membership of a single group

IGMP Protocol Details



IP Protocol Number = 2

IGMPv2

Type	Max Response Time	Checksum
Group Address		

IGMPv3

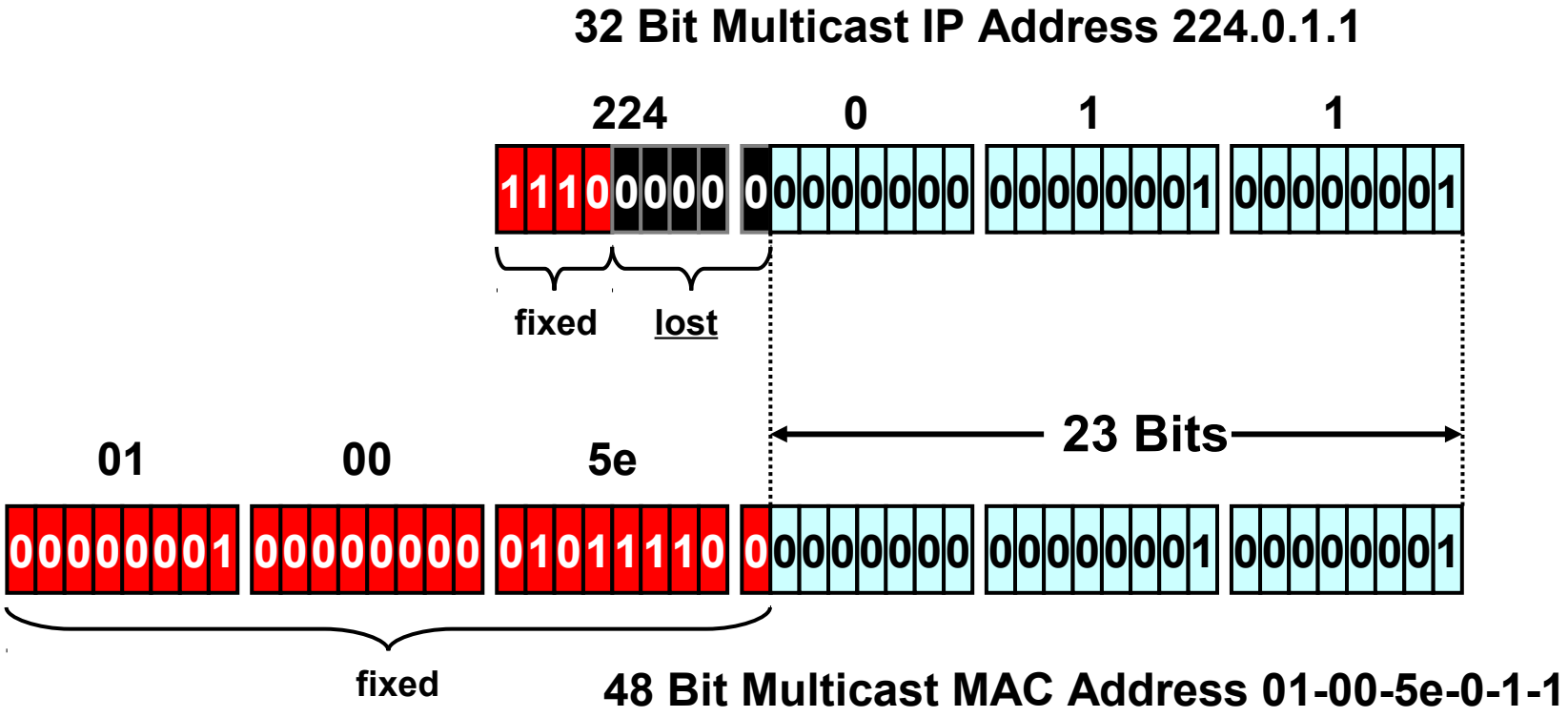
- Hosts could even send a list of sources
  - Either (S, G) or [(S1, S2, ..., Sn), G]
- Advantages:
  - Leaf routers can build a source distribution tree without RPs
  - LAN switches, which would do IGMP snooping

Layer 2 Multicast

L2/L3 Address Mapping

- Switches should also perform L2 multicast for efficient multicast delivery
  - Address mapping required
- Strange solution standardized:
  - 23 low-order bits of multicast IP address is mapped into 23 low-order MAC address bits
  - MAC prefix is always "01-00-5e"
  - 5 bits of IP address are lost !!!

Address Mapping to Ethernet



- MAC prefix "01-00-5e" indicates L3-L2 mapping
- Only 23 bits had been reserved for Ethernet: 32:1 Overlapping!

Multicast Switching

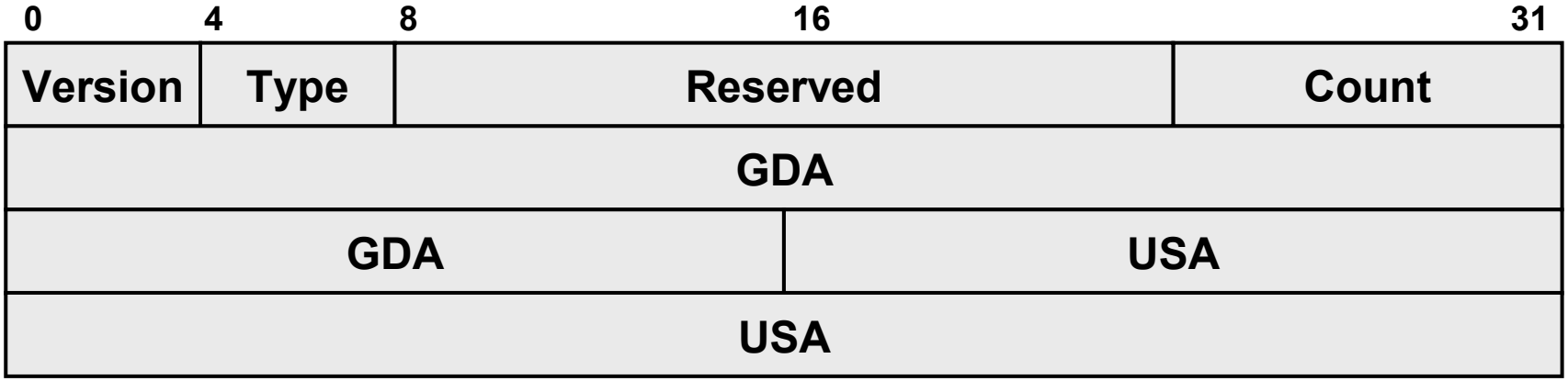
- Normal switches flood multicast frames through every port
  - No entries in CAM table (how to learn?)
  - Waste of LAN capacity
- Some switches allow to configure dedicated multicast ports
  - Not satisfying because users want to join and leave dynamically over any port

Multicast Switching Solutions

- Cisco Group Management Protocol (CGMP)
  - Simple but proprietary
  - For routers and switches
- IGMP snooping
  - Complex but standardized
  - Also proprietary implementations available
  - For switches only
- GARP Multicast Registration Protocol (GMRP)
  - Standardized but not widely available
  - For switches and hosts
- Router-port Group Management Protocol (RGMP)
  - Simple but Cisco-proprietary
  - For routers and switches

CGMP

- Sent by routers to switches
  - Destination address 0100.0cdd.dddd
- Message contains
  - Type field (join or leave)
  - MAC address of IGMP client (host)
  - Multicast MAC address of group
- Now switch can create multicast table
- Low CPU overhead



CGMP – Notes (HIDDEN)

- Supported by wide range of routers and switches
- Conflicts with IGMP snooping
- How to learn about all receivers in spite of the report suppression mechanism?
  - Good question...

IGMP Snooping

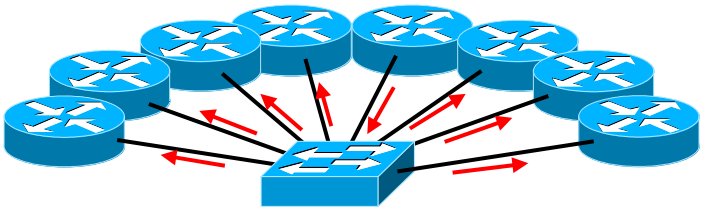
- Switches must decode IGMP
  - Which traffic should be forwarded to which ports?
  - Read IGMP membership reports and leave messages
  - Either by NMP (slow) or by special ASICs
- The CAM table must allow multiple port entries per MAC address!
  - Also the CPU port (e.g. 0) must be added!
  - Upon high mc-traffic load the CPU gets overloaded!
  - Special ASICs might differentiate IGMP from data traffic to improve performance

GARP Multicast Registration Protocol

- IEEE 802.1p GARP (Generic Attribute Registration Protocol) extended for IP multicast
  - Runs on hosts and switches
- Pro-active processing:
  - Hosts must also join to switch using GMRP
  - Switch configures CAM table and notifies other switches
- Incoming mc-traffic can be efficiently switched

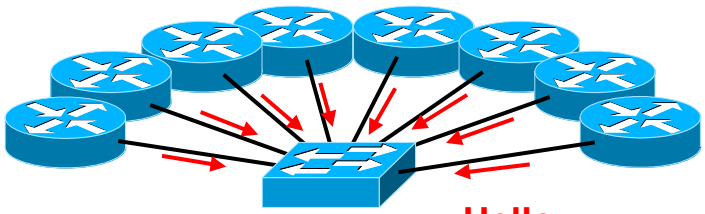
Switch/Router Problems

- Any switch connected to multiple routers must forward *all* multicast traffic to *all* routers!
  - Since routers don't send IGMP membership reports
  - Routers might get lots of unneeded packets!
- Using RGMP a router can tell a switch all multicast groups the router manages
  - Router-only switched topologies only!

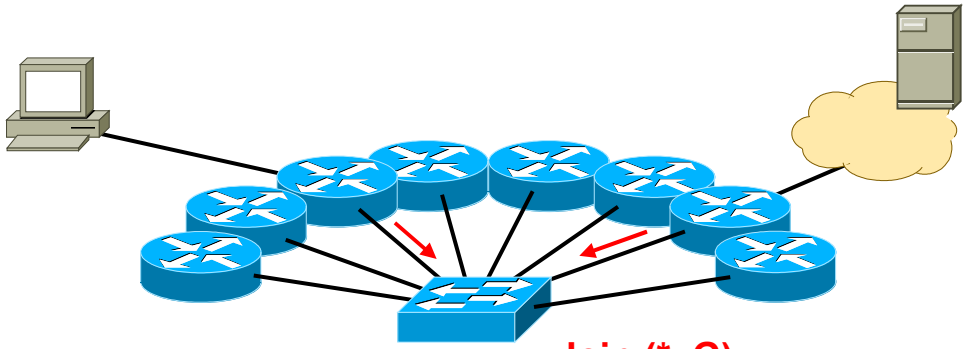


RGMP Details

- Routers periodically send hello messages to the switch
  - Switch learns about existence of routers
- Routers send RGMP (*, G) joins for groups they belong to
- Well-known address 224.0.0.25
- Restrictions:
  - Not all routers need to support RGMP
  - No directly connected sources allowed



Hello



Join (*, G)

Session Information

Session Information

- Potential receivers must be informed about multicast sessions
  - Sessions are available before receiver launches application
  - Might be announced via well-known multicast group address
  - Or via publicly available directory services
  - Or via web-page or even E-Mail

SDR (1)

- Mbone session description protocol and transport mechanism
  - Used by sources for assigning new multicast addresses
  - Checks sdr cache to avoid conflicts
  - Creates a session and sends its description via sdr announcements (224.2.127.254)
- Anybody can announce a session
  - Source is part of the session description
- Announcement frequency
  - Ratio number of session / available BW = const
  - Typically 5-10 minutes
  - Late join latency problem avoided by caching

SDR (2)

- RFC 2327 only specifies variables but no transport mechanism
  - Session Announcement Protocol (SAP, RFC 2974)
  - Session Initiation Protocol (SIP, RFC 2543)
  - Real Time Streaming Protocol (RTSP, RFC 2326)
  - E-mail (MIME/SDR) and also web pages

Security

- Receiver identification
  - Generally not needed except for security and feedback mechanisms (QoS)
  - Provided by RTCP
  - Applications might use unicast return messages
- Multicast flows from the sender and from receivers may be encrypted for security reasons
  - If receivers are not known to the sender, the encryption may be done only one way

Multicast Routing Basics

Multicast Routing Basics

- Opposite function than traditional unicast routing:
  - Unicast routing calculates the path to the destination of the packet
  - Multicast routing calculates the path to the origin of the packet
- Basic algorithm: Reverse Path Forwarding (RPF)
  - Prevents forwarding loops
  - Ensures shortest path from source to receivers

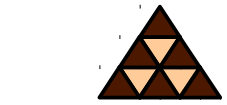
In Other Words...

- Multicast routing:
  Which is best path to the source?
- Prevent multicast storms: Tree!
- Routers do
  "Reverse Path Forwarding" (RPF)
  - Forwards a multicast packet only if received on the upstream interface to the source
  - Check source IP address in the packet against routing table to determine upstream interface

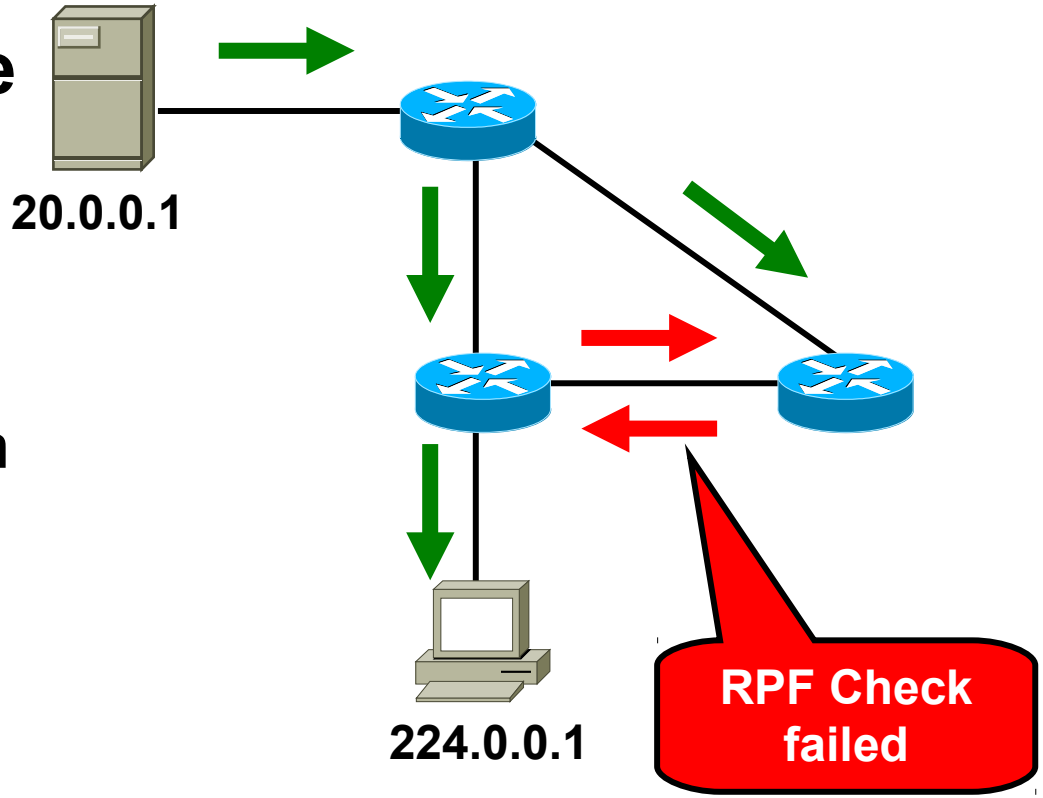
RPF Check

- Router forwards multicast packet only if it was received on the upstream interface to the source
  - Then this packet is already on the distribution tree
- Utilizes unicast routing table to determine the nearest interface to the source
  - RPF check fails: packet is silently discarded
  - RPF check succeeds: packet is forwarded according OIL

RPF Check

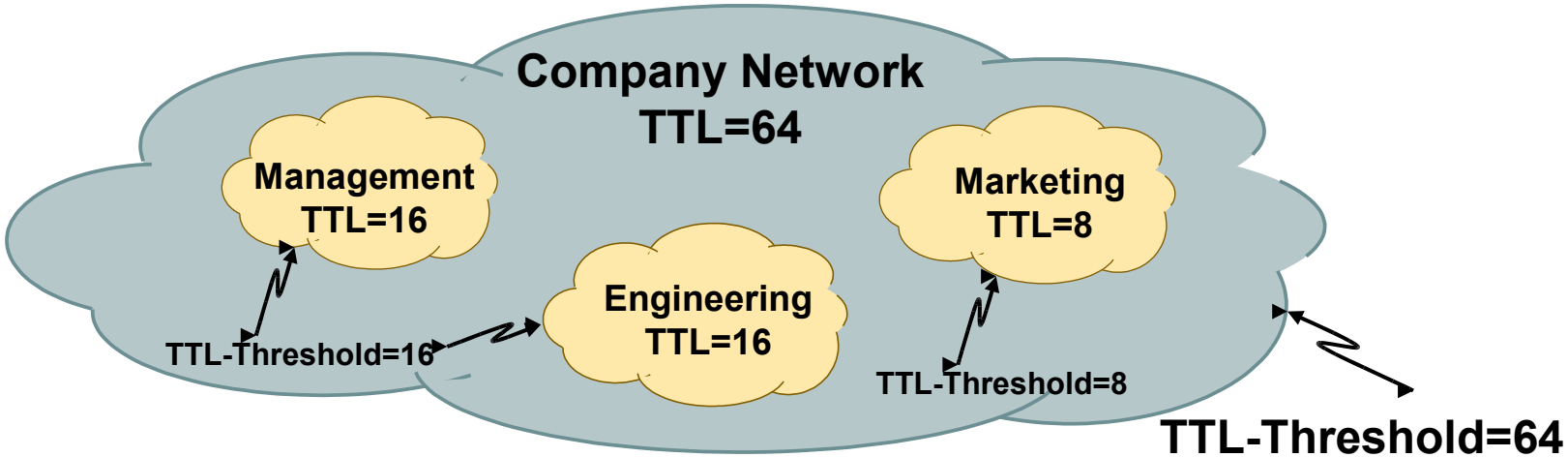


- RPF Check prevents duplicate forwarding
- Look one step ahead
  - Determine if outgoing link is on upstream path for the next router
  - Avoids any duplicates



Multicast Scoping using TTL

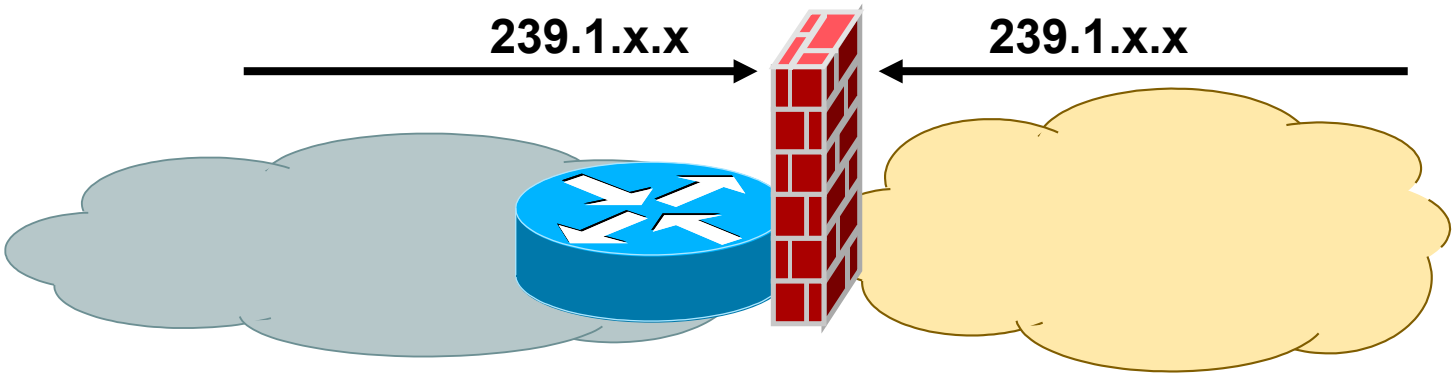
- Packet's TTL is decremented by 1 when packet arrives at incoming interface
- Then the packet is forwarded according OIL which also contains TTL thresholds per interface
  - May be configured to limit the forwarding of multicast packets with TTL>threshold
  - Default threshold = 0 (no threshold)



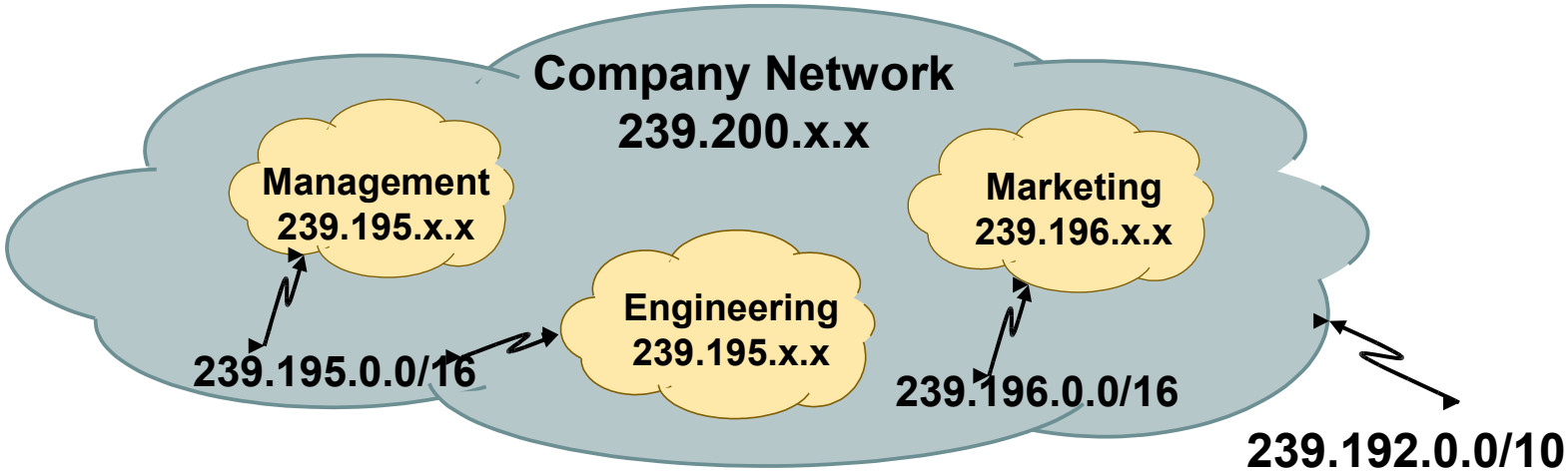
Multicast Scoping using Addresses

- Scoping via TTL thresholds relies on the TTL configurations
  - Might be unknown or unpredictable
- Administrative boundaries can be created using address scoping
  - Traffic which does not match the ACL cannot pass this interface
  - In both directions!

Administrative Boundaries



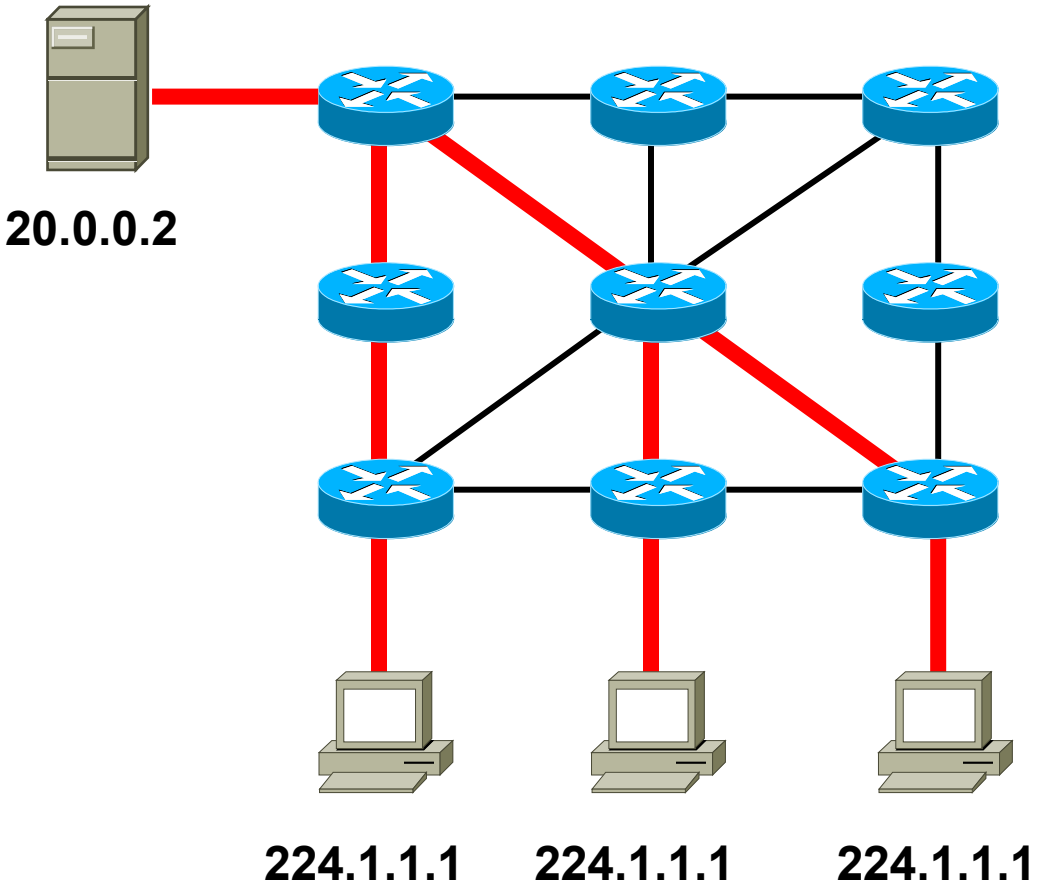
Serial0: Administrative boundary
for all 239.1.0.0/16 packets



Shortest Path Tree (1)

Also called "Source Distribution Tree" or "Source (-based) Tree"

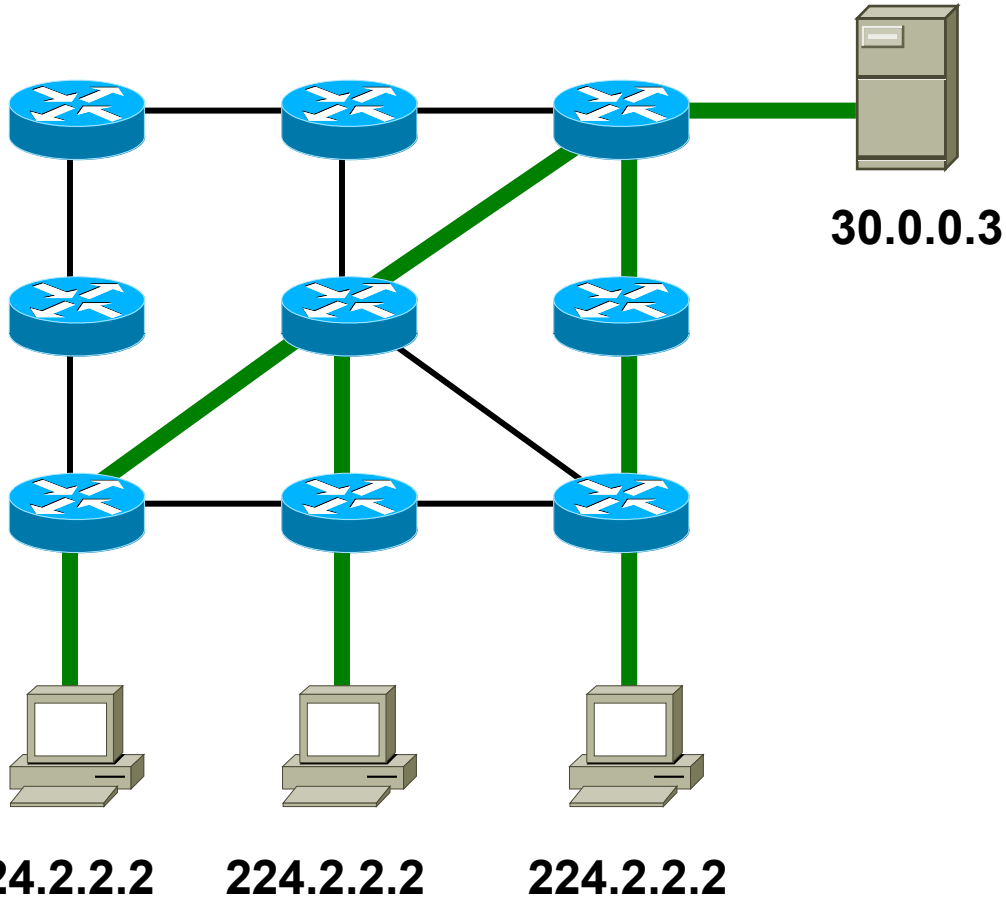
(S, G) = (20.0.0.2, 224.1.1.1)



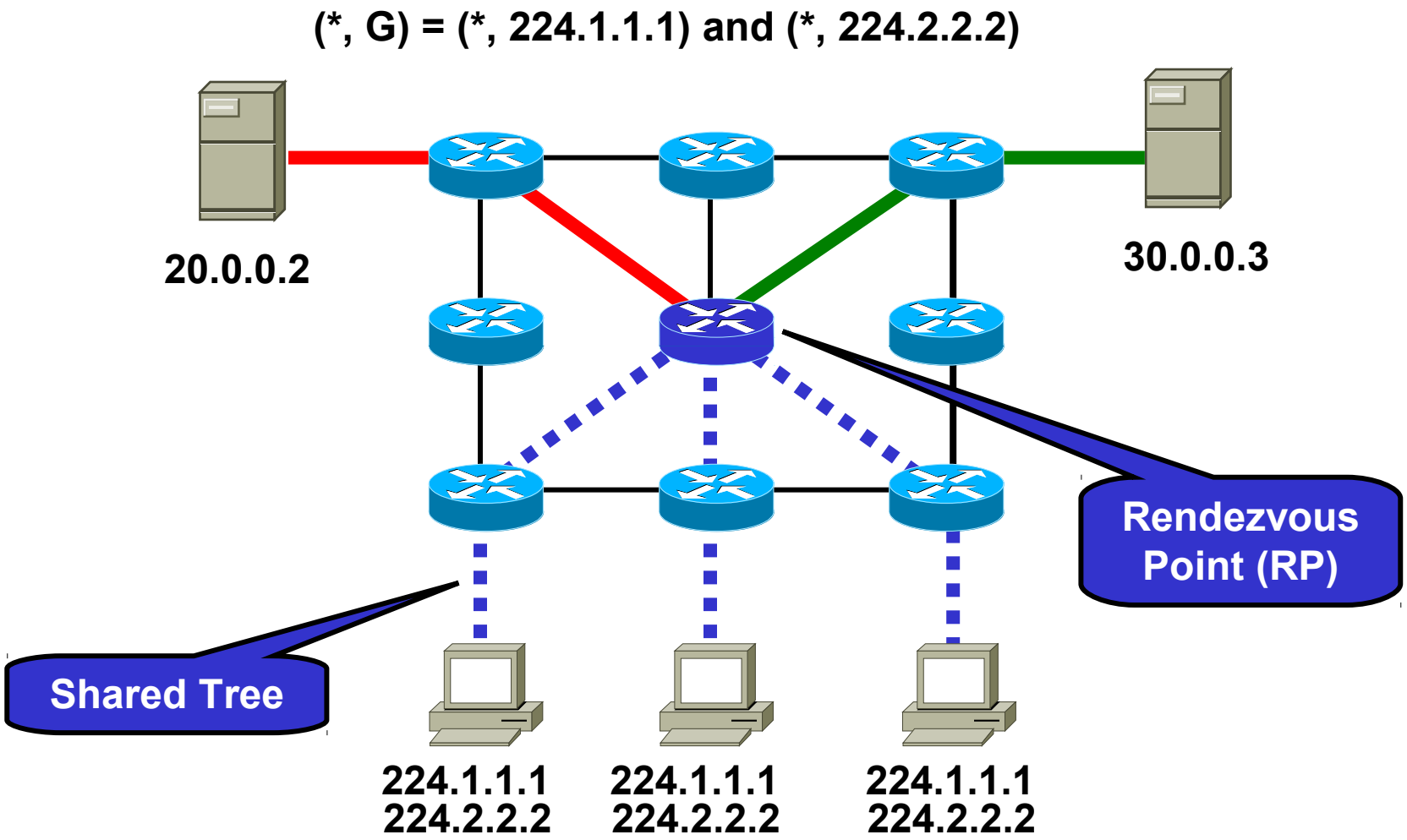
Shortest Path Tree (2)

Also called "Source Distribution Tree" or "Source (-based) Tree"

(S, G) = (30.0.0.3, 224.2.2.2)



Shared Tree


(*, G) = (*, 224.1.1.1) and (*, 224.2.2.2)
20.0.0.2
30.0.0.3
Rendezvous Point (RP)
Shared Tree
224.1.1.1
224.2.2.2
224.1.1.1
224.2.2.2
224.1.1.1
224.2.2.2

Multicast Routing Protocols

Multicast Protocol Types

- Dense Mode: Push method
  - Initial traffic is flooded through whole network
  - Branches without receivers are pruned (for a limited time period only)
- Sparse Mode: Pull method
  - Explicit join messages
  - Last-hop routers pull the traffic from the RP or directly from the source

Multicast Protocols Overview

- **DVMRP** — Distance Vector Multicast Routing Protocol
- **MOSPF** — Multicast OSPF
- **PIM-DM** — Protocol Independent Multicast – Dense Mode
- **PIM-SM** — Protocol Independent Multicast – Sparse Mode
- **CBT** — Core Based Trees

...and others...

What is what?

- DVMRP
- MOSPF
- PIM-DM
- PIM-SM
- CBT

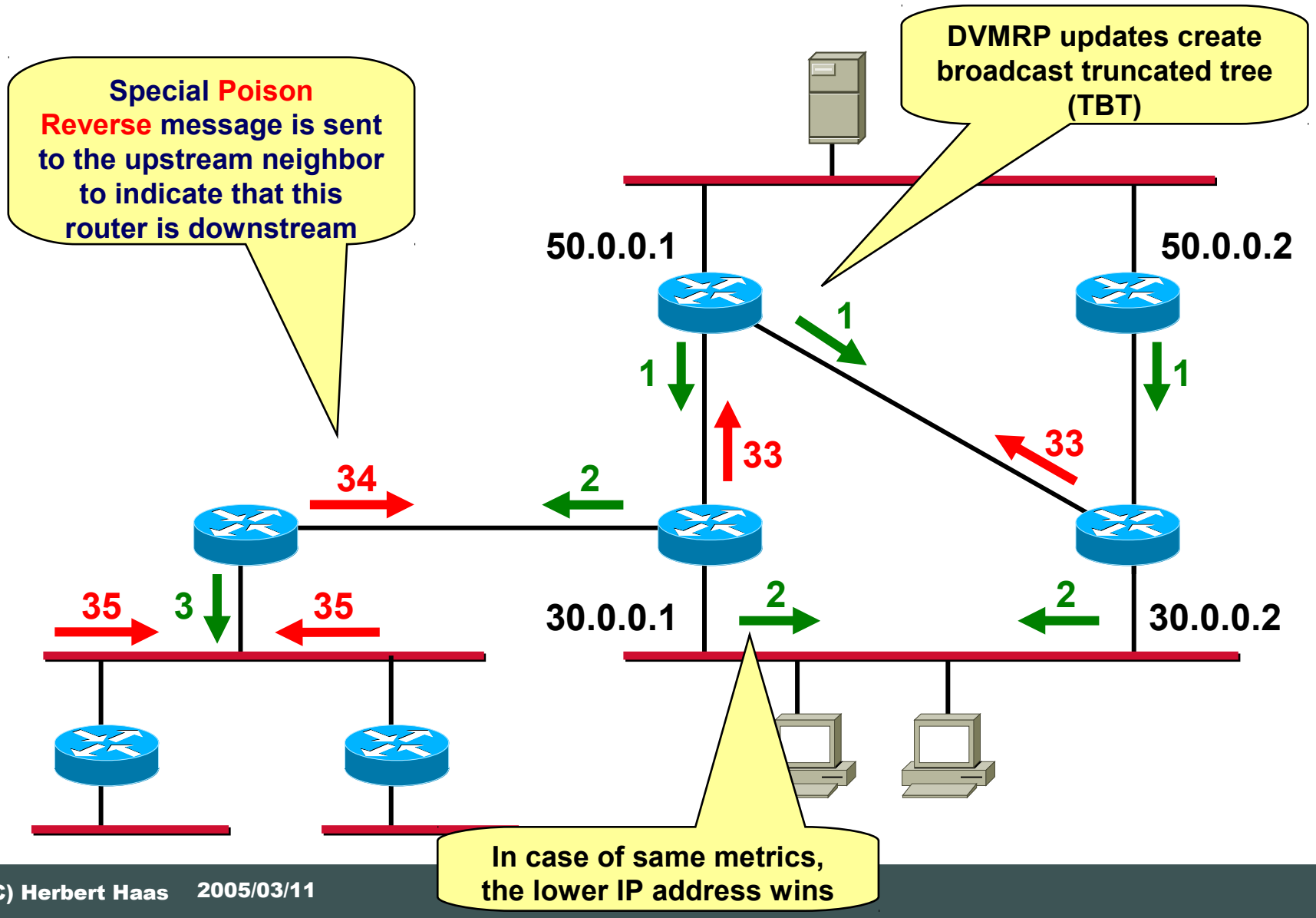
Dense Mode

Sparse Mode

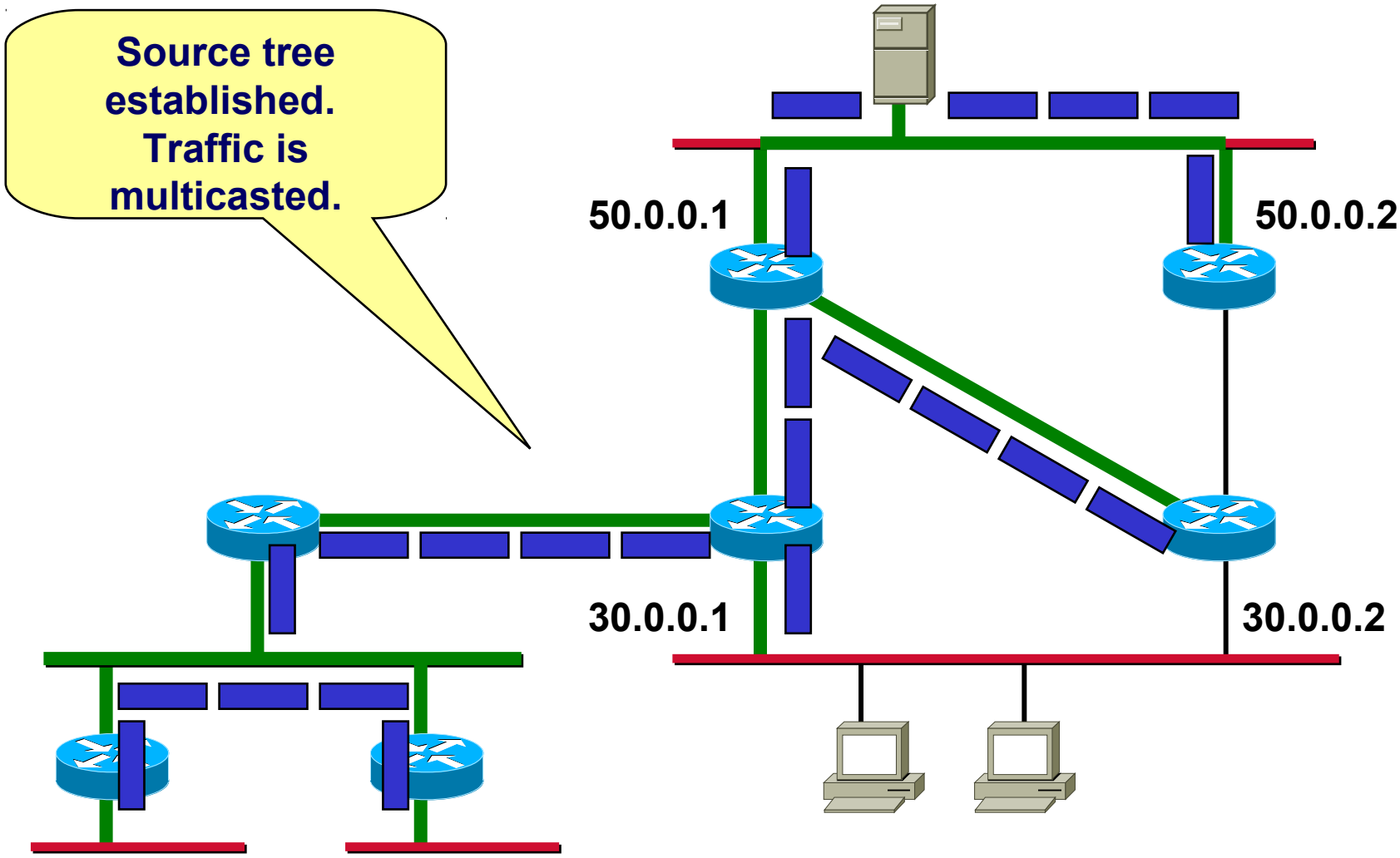
DVMRP – Facts

- Dense mode protocol (Prune and Graft)
- Distance Vector announcements of networks
  - Similar to RIP but classless
  - Infinity = 32 hops
- Creates Truncated Broadcast Trees (TBTs)
  - Each source network in the DVMRP cloud produces its own TBT
  - Source Tree principle

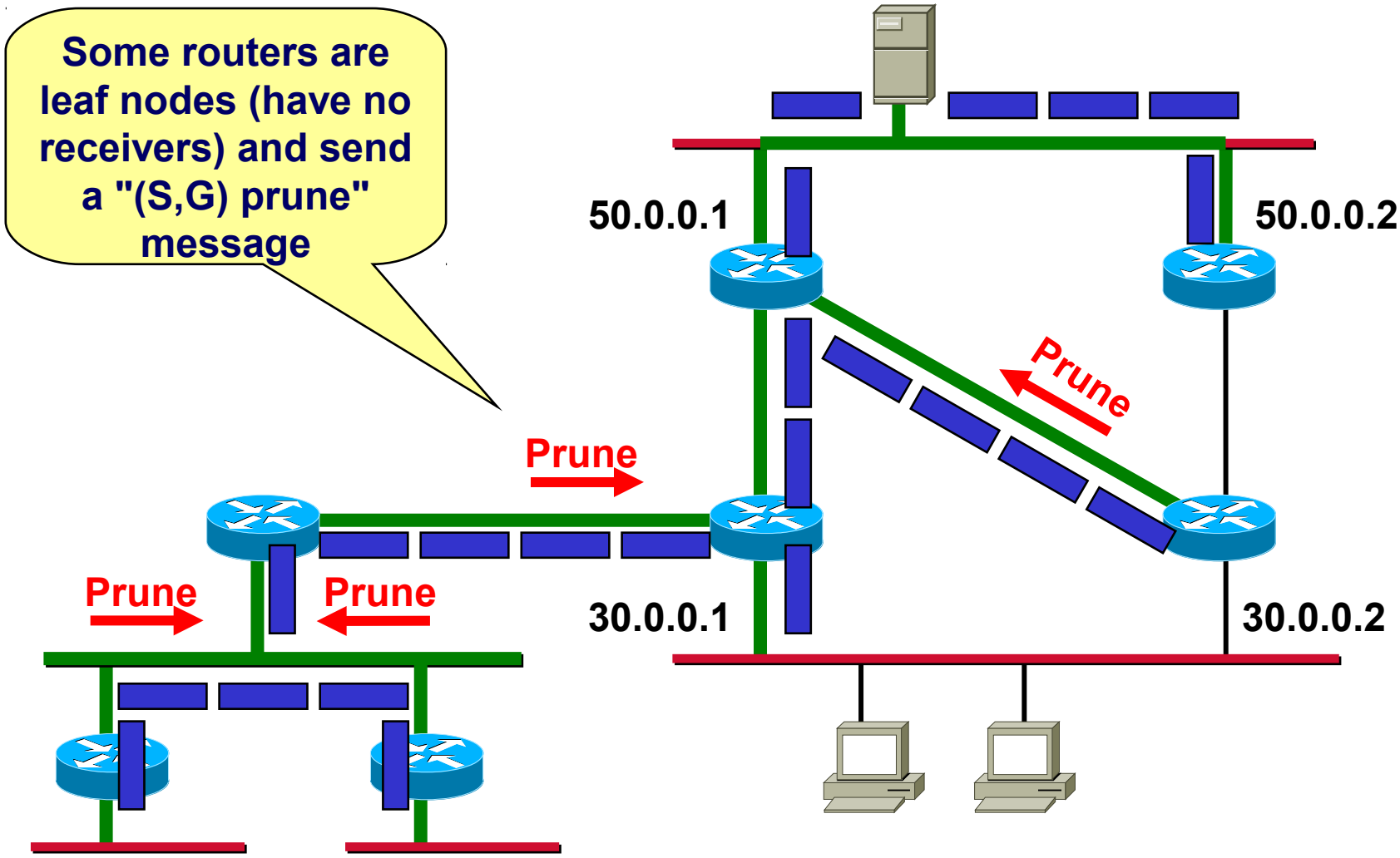
DVMRP – Flood


Special Poison Reverse message is sent to the upstream neighbor to indicate that this router is downstream
DVMRP updates create broadcast truncated tree (TBT)
50.0.0.1
50.0.0.2
30.0.0.1
30.0.0.2
1
1
1
33
33
34
2
35
3
35
2
2
In case of same metrics, the lower IP address wins

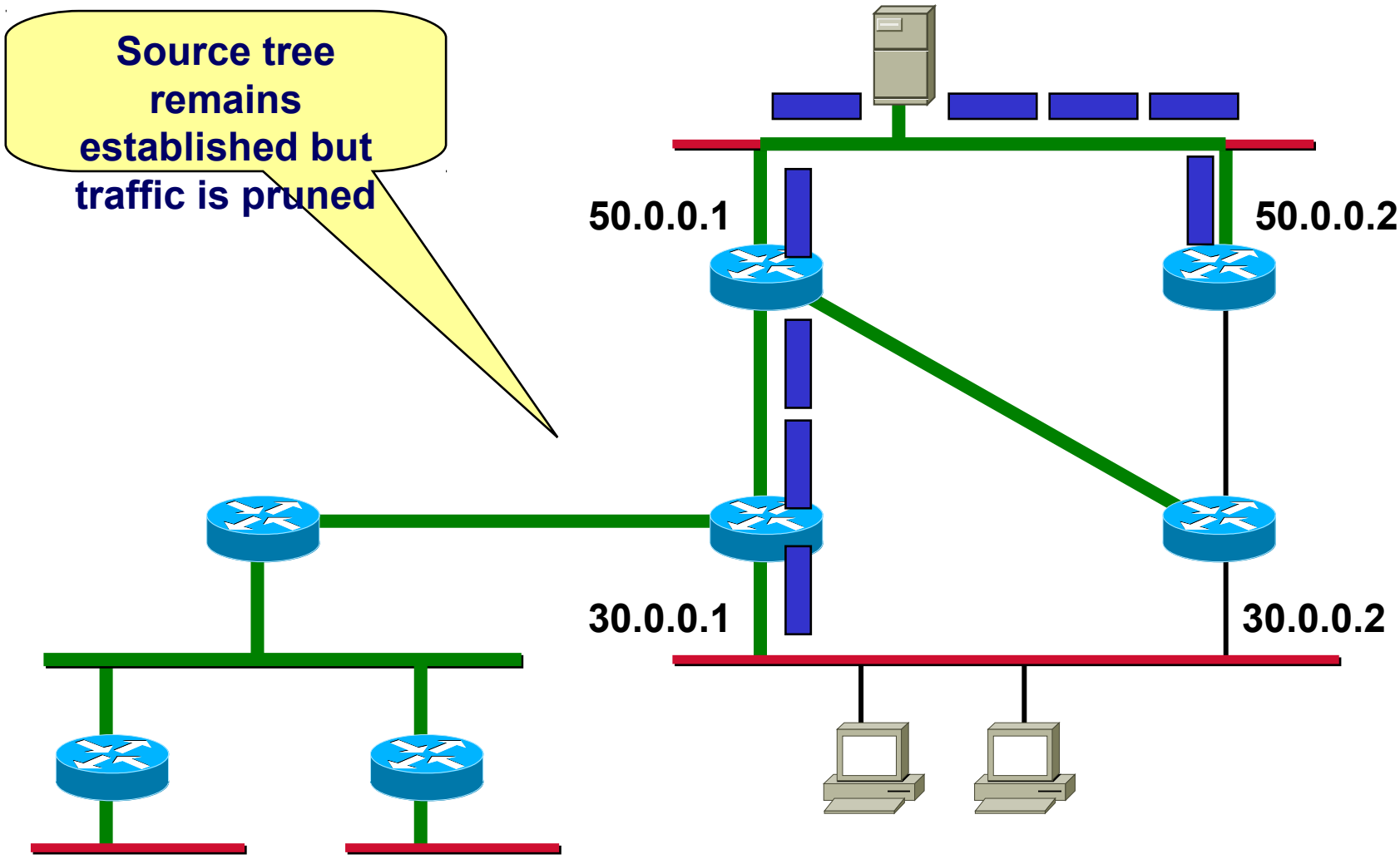
DVMRP – Source Tree



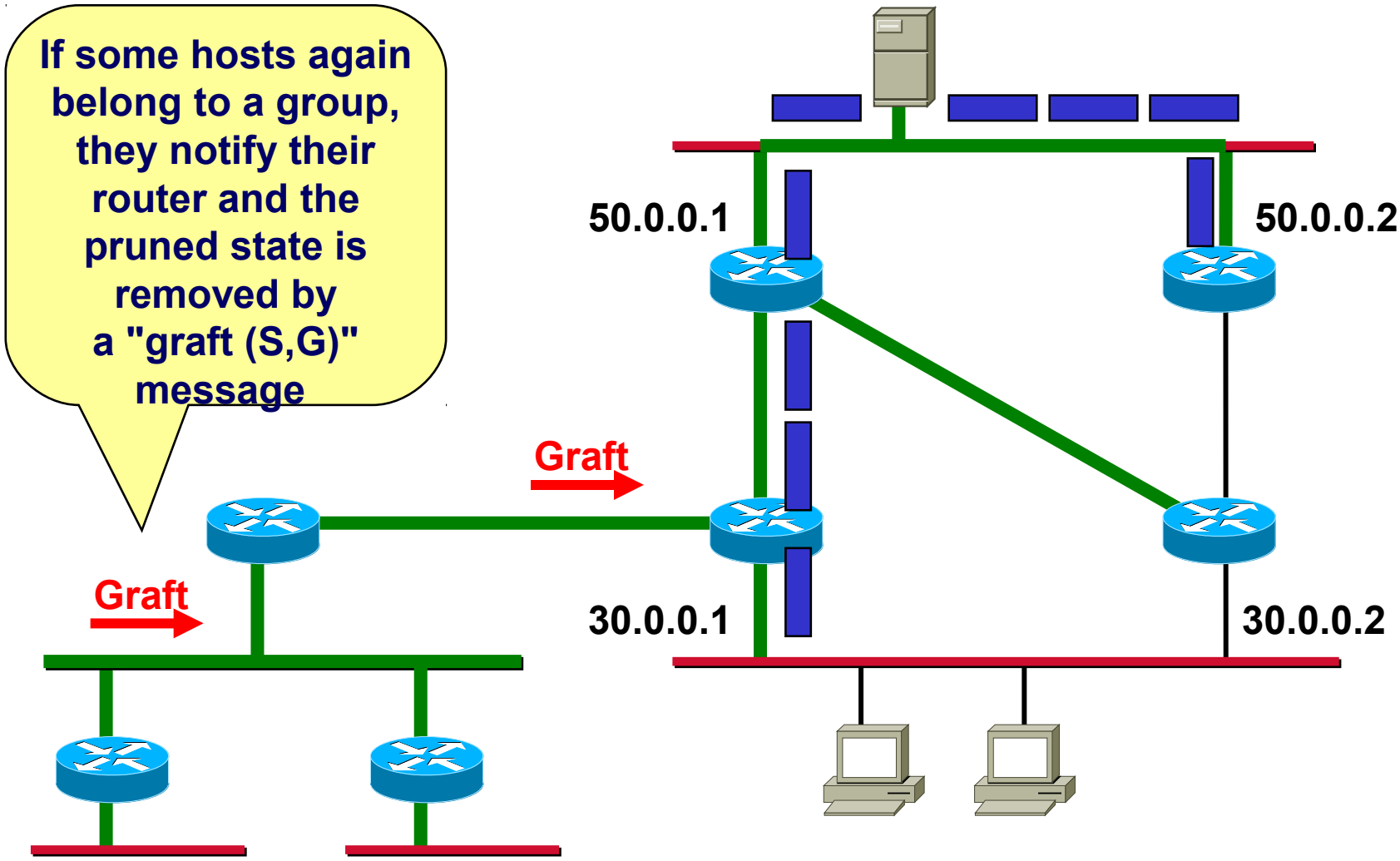
DVMRP – Prune


Some routers are leaf nodes (have no receivers) and send a "(S,G) prune" message
50.0.0.1
50.0.0.2
Prune
Prune
Prune
Prune
30.0.0.1
30.0.0.2

DVMRP – TBT


Source tree remains established but traffic is pruned
50.0.0.1
50.0.0.2
30.0.0.1
30.0.0.2

DVMRP – Graft


If some hosts again belong to a group, they notify their router and the pruned state is removed by a "graft (S,G)" message
Graft
Graft
50.0.0.1
50.0.0.2
30.0.0.1
30.0.0.2

DVMRP Facts

- Significant scaling problems
  - Slow Convergence (RIP-like)
  - Significant amount of multicast routing state information stored in routers
  - No support for shared trees
  - Maximum number of hops < 32
- Used in the MBONE
  - Today worldwide available and accessible
  - Virtual network through IP tunnels

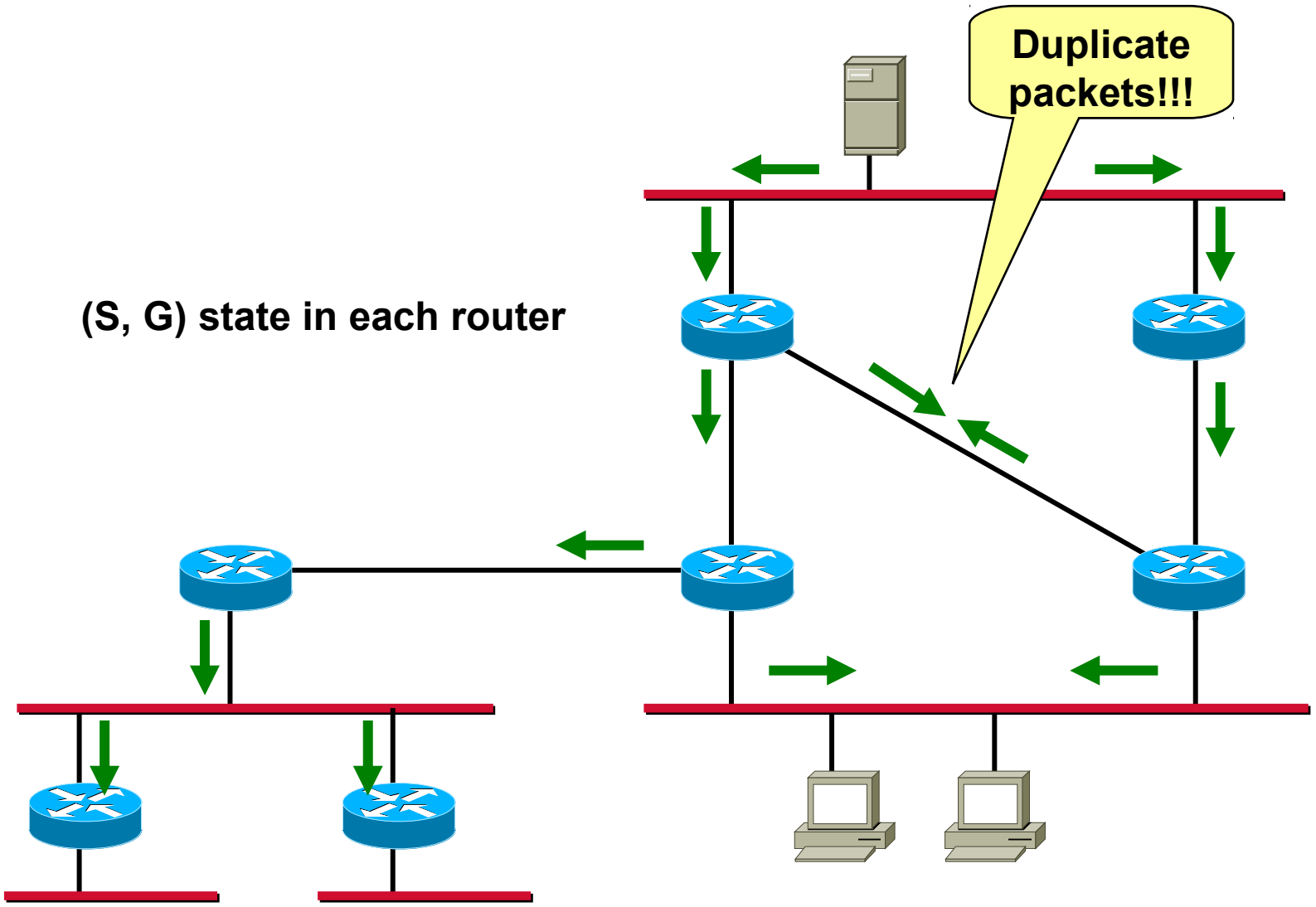
MOSPF

- Useful only in OSPF domains
- Include multicast information in OSPF link states
  - Group Membership LSAs flooded throughout OSPF routing domain
  - Each router knows complete network topology!
  - MOSPF Area Border Routers (MABR) would improve performance
- Significant scaling problems
  - Dijkstra algorithm run for EVERY multicast (SNet, G) pair!
  - Only a few (S,G) should be active
  - No shared tree support
- Not used

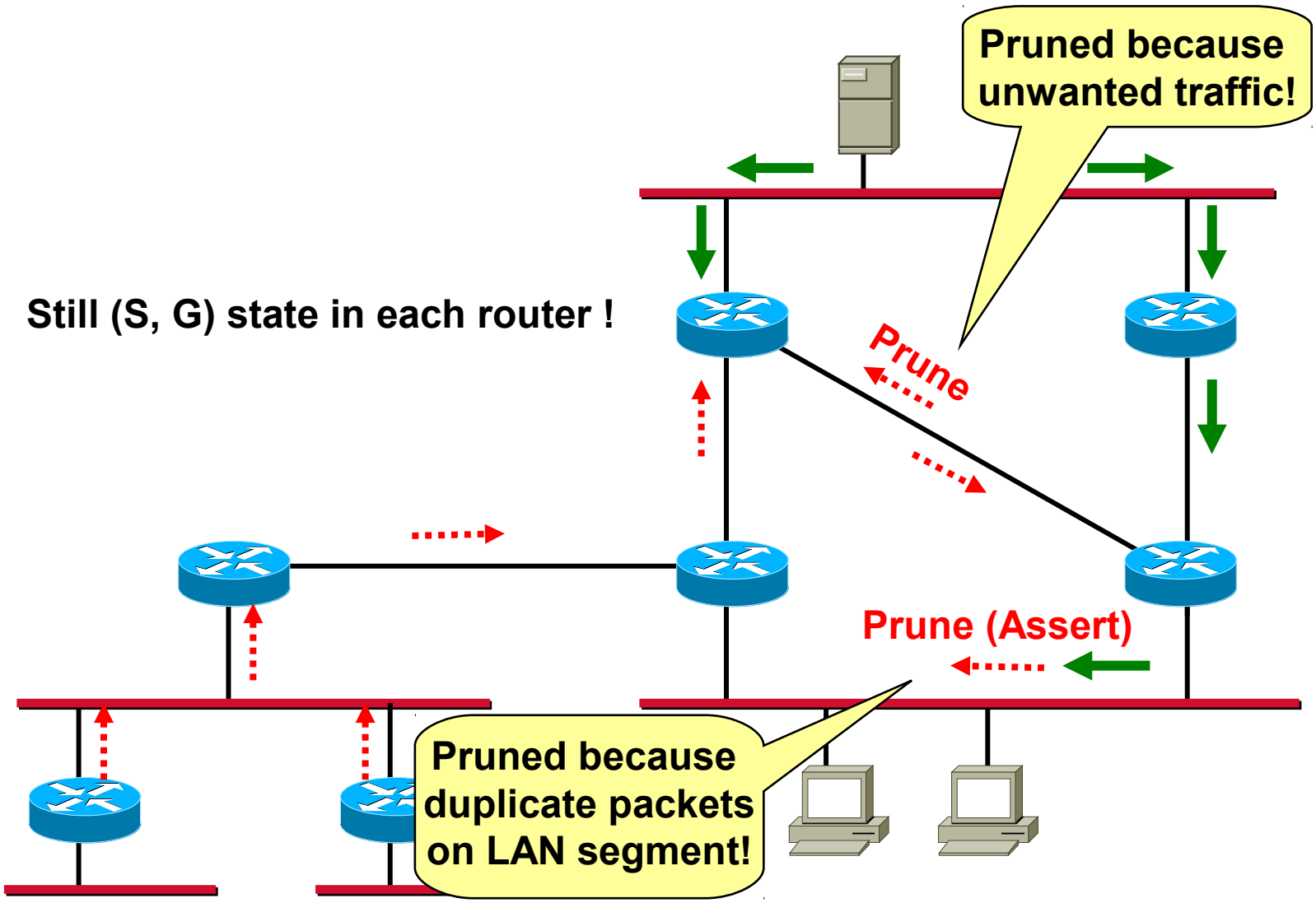
PIM-DM

- Protocol Independent
  - Utilizes any underlying unicast routing protocol
- Similar to DVMRP but
  - No TBT because no dedicated multicast protocol in use
  - Instead: RPF, flood and prune is performed
- For small networks only
  - Every router maintains (S, G) states
  - Initial flooding causes duplicate packets on some links
- Easy to configure
  - Two command lines
  - Useful for small trial networks

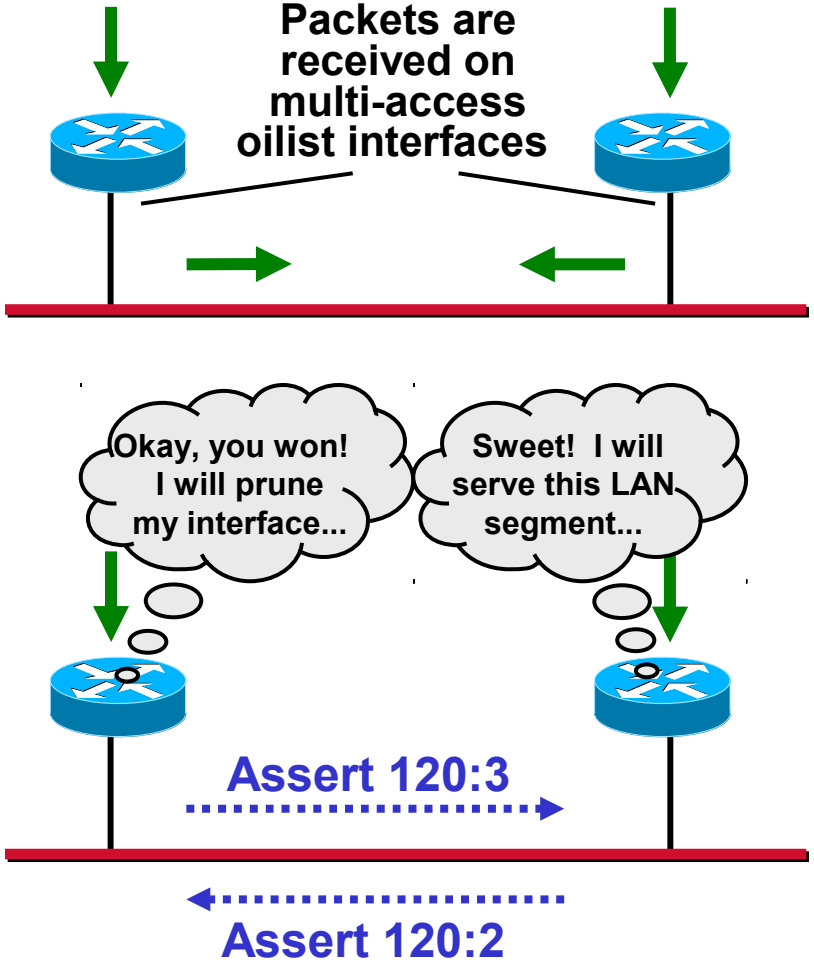
PIM-DM: Initial Flooding



PIM-DM: Pruning



PIM-DM: Assert Mechanism



- Each router receives the same (S, G) packet through an interface listed in the oilist
  - Only one router should continue sending
- Both routers send "PIM assert" messages
  - To compare administrative distance and metric to source
- If assert values are equal, the highest IP address wins

Core Based Trees (CBT)

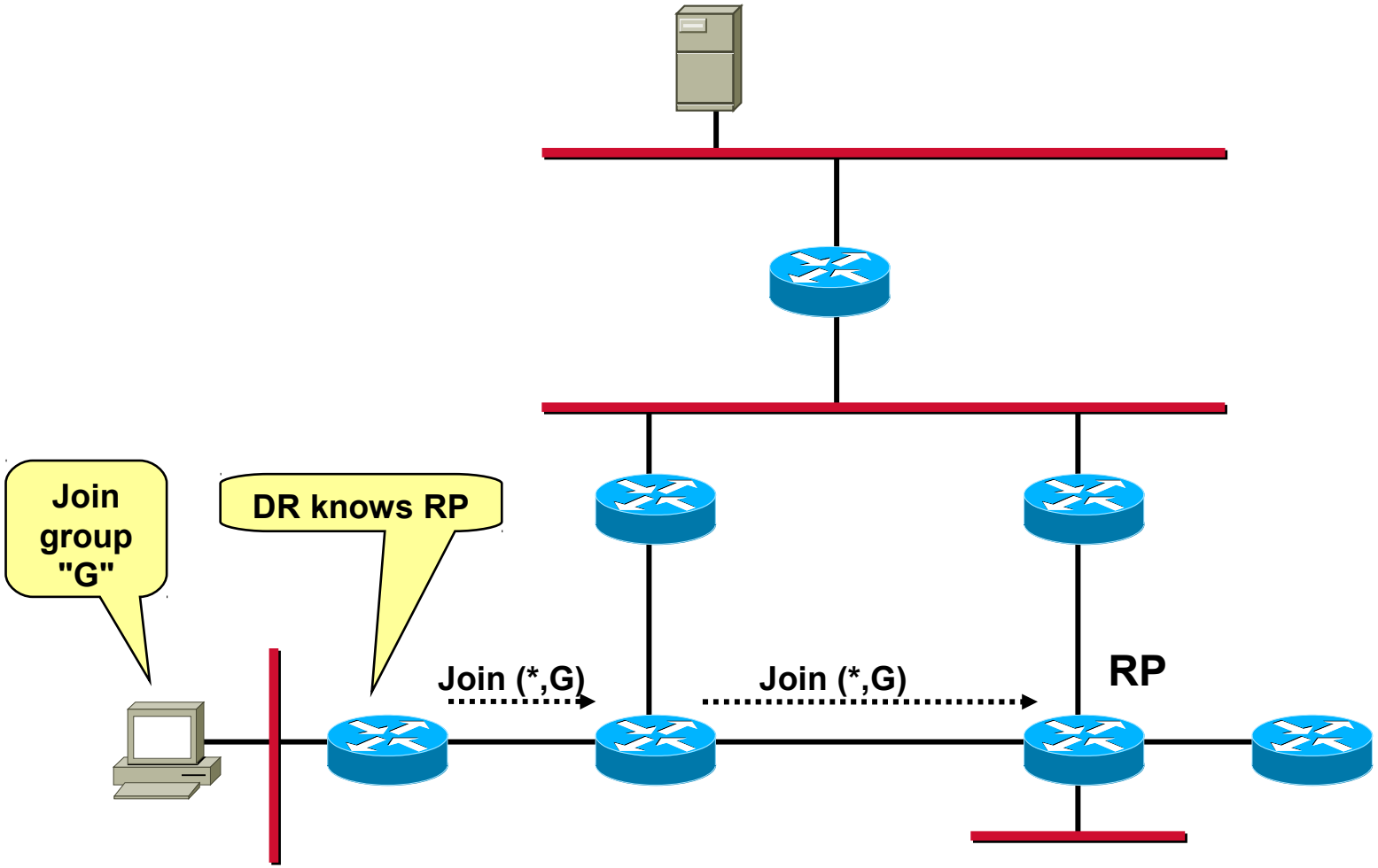
We do not waste time with CBT !!!

Let's go directly to PIM-SM...

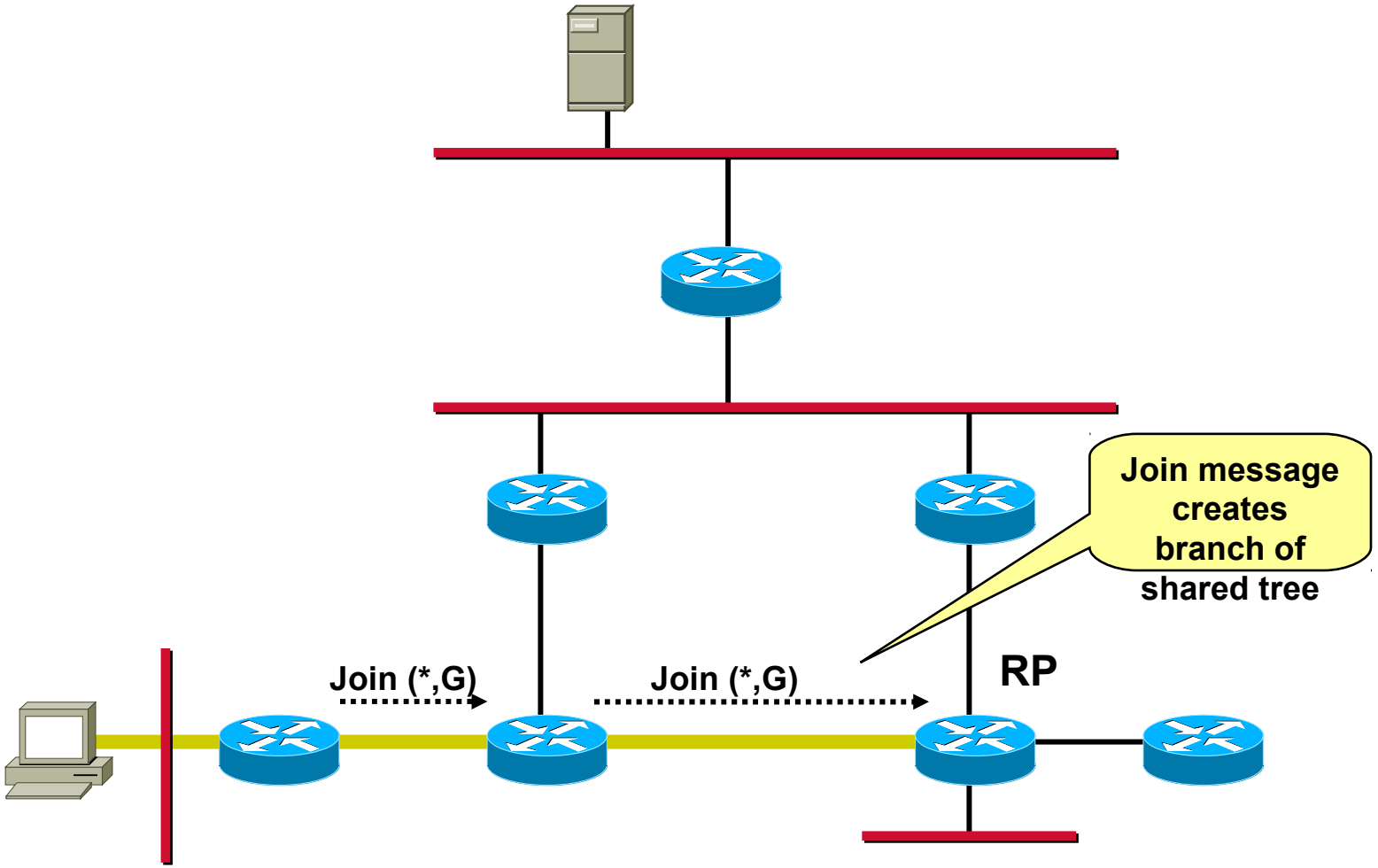
PIM-SM

- Protocol Independent
  - Utilizes any underlying unicast routing protocol
- Supports both source and shared trees
- Uses a Rendezvous Point (RP)
  - Sources are registered at RP by their first-hop router
  - Groups are joined by their local designated router (DR) to the shared tree, which is rooted at the RP
- Best solution today
  - Optimal solution regardless of size and membership density
- Variants
  - Bidirectional mode (PIM-bidir)
  - Source Specific Multicast (SSM)

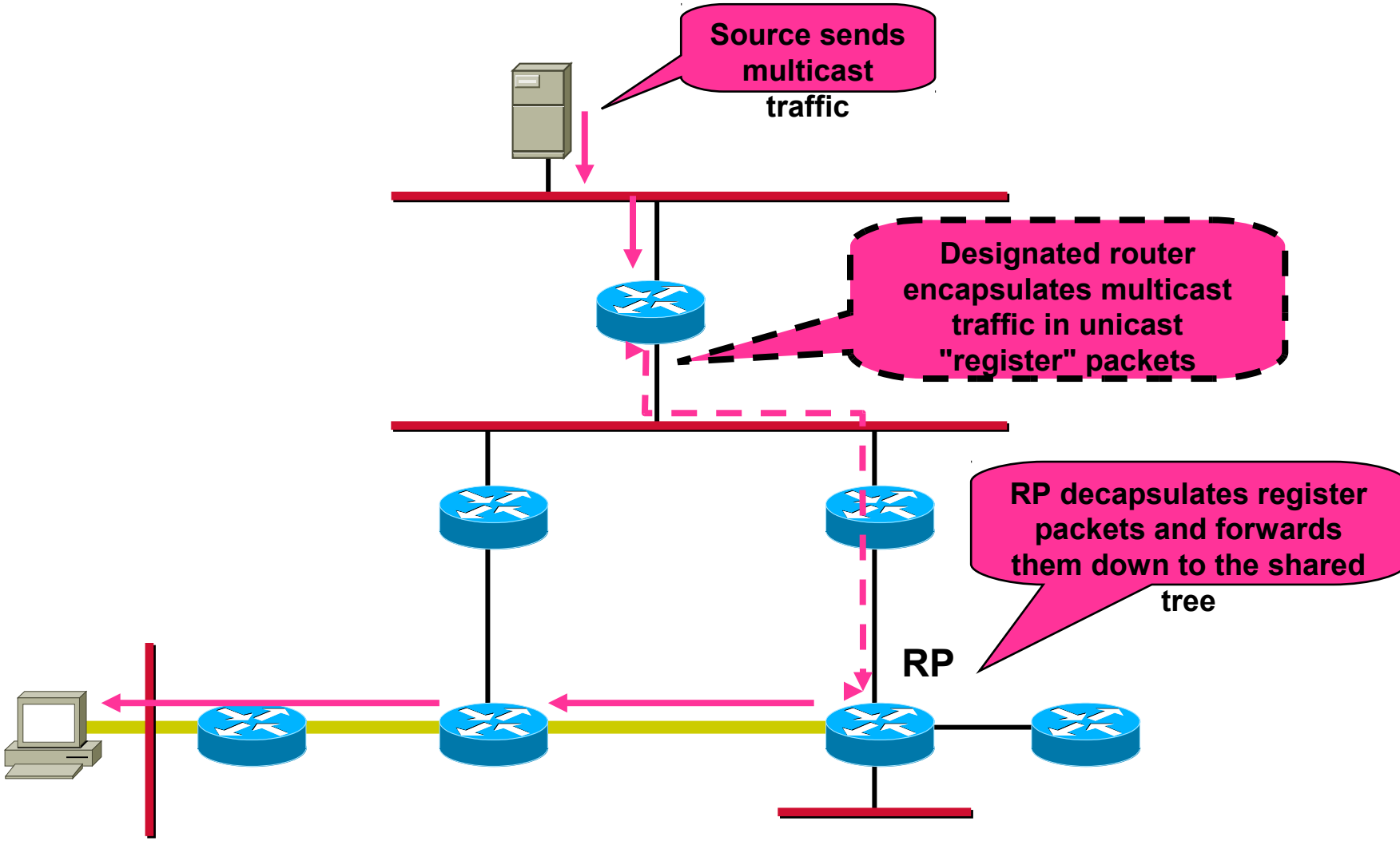
PIM-SM / User becomes active



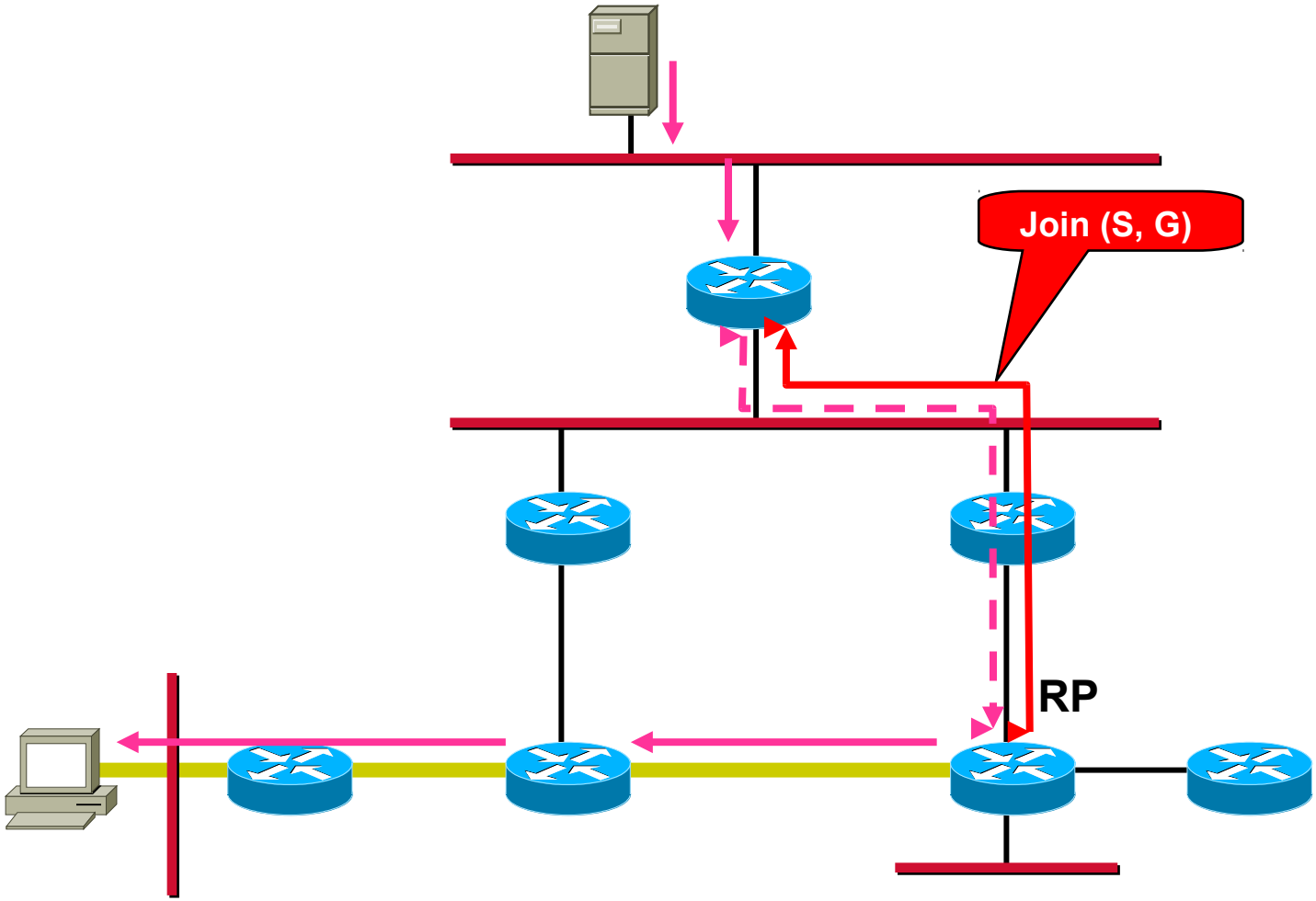
PIM-SM / Create Shared Tree



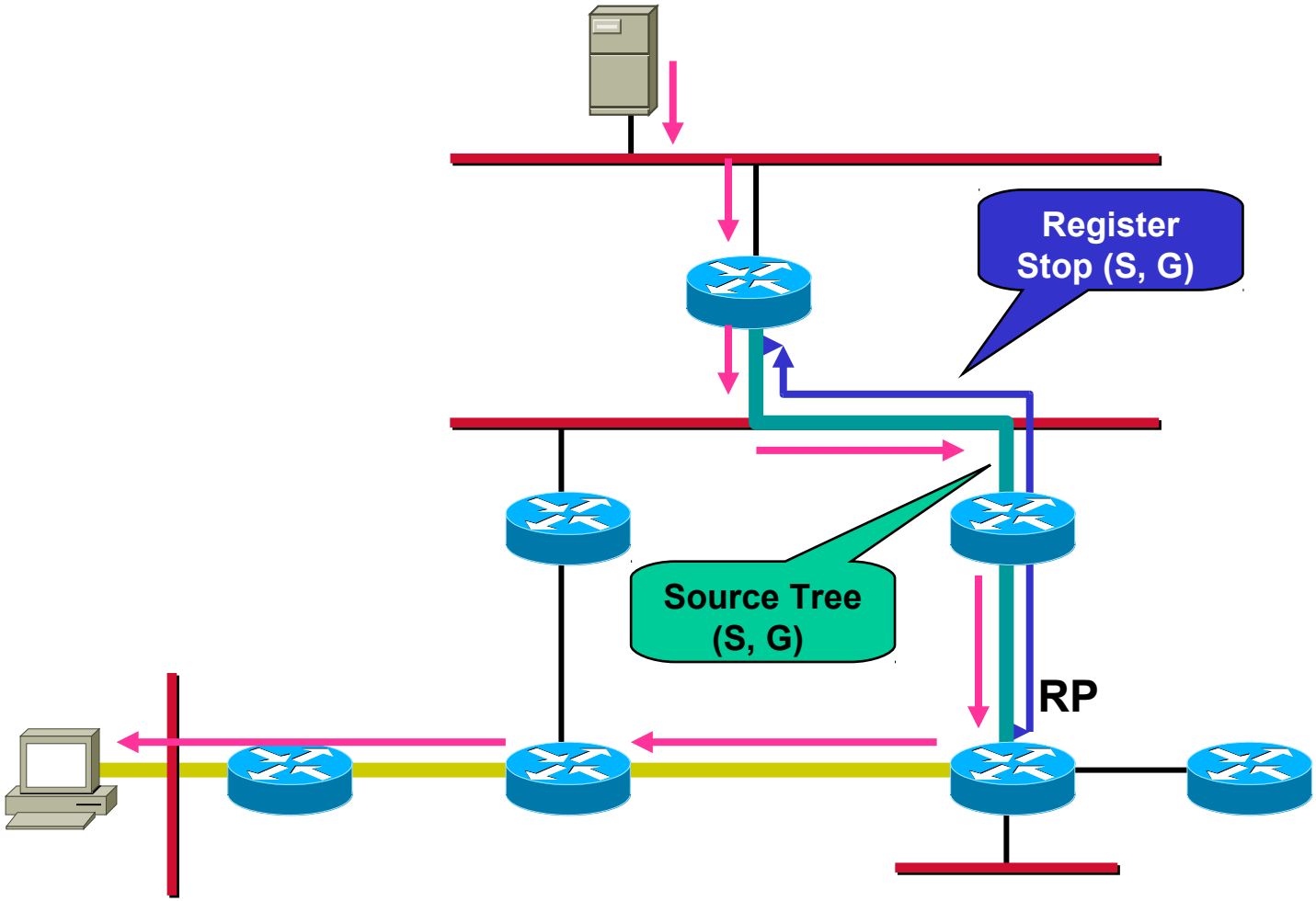
PIM-SM / Register Source



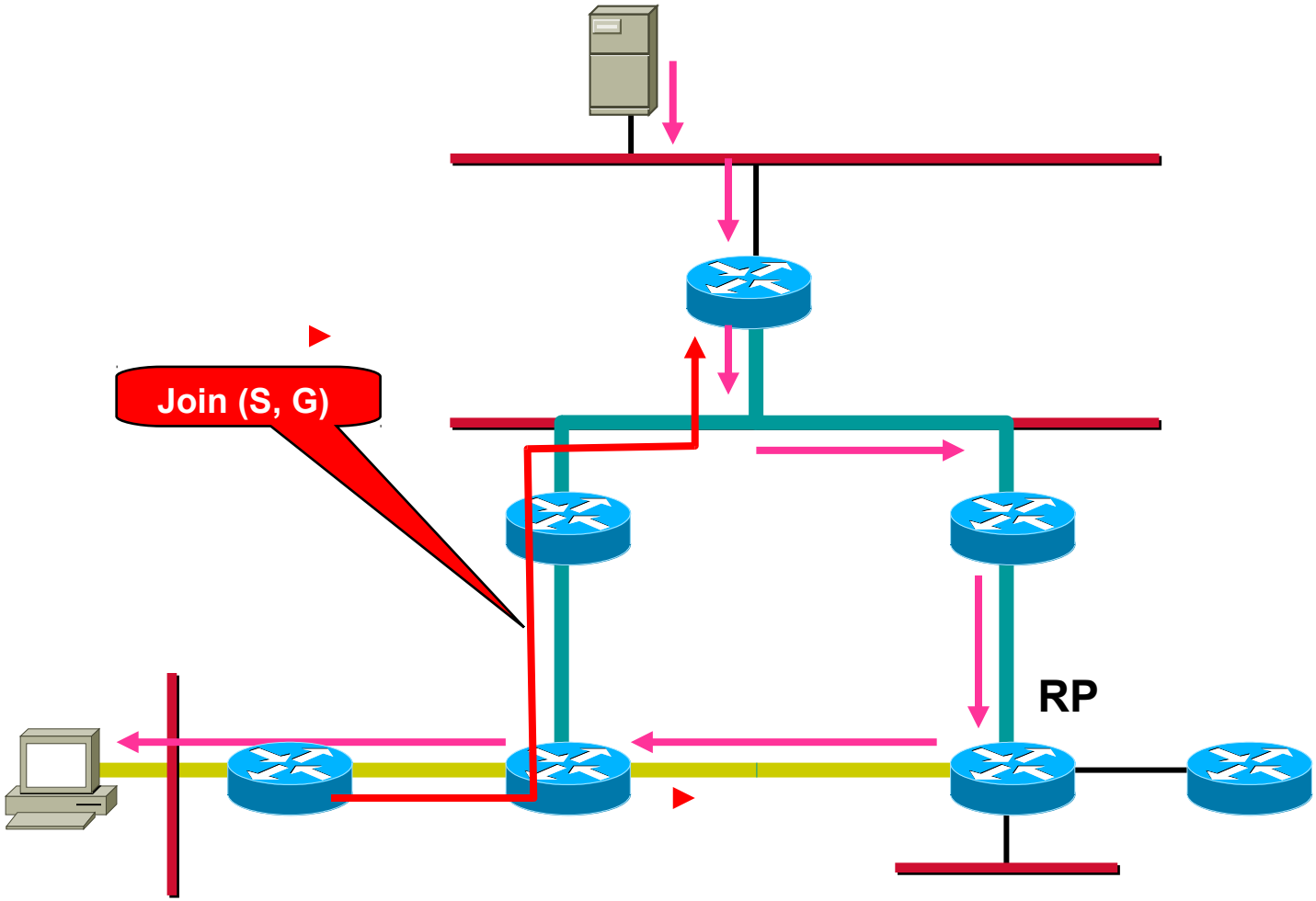
PIM-SM / Create Source Tree



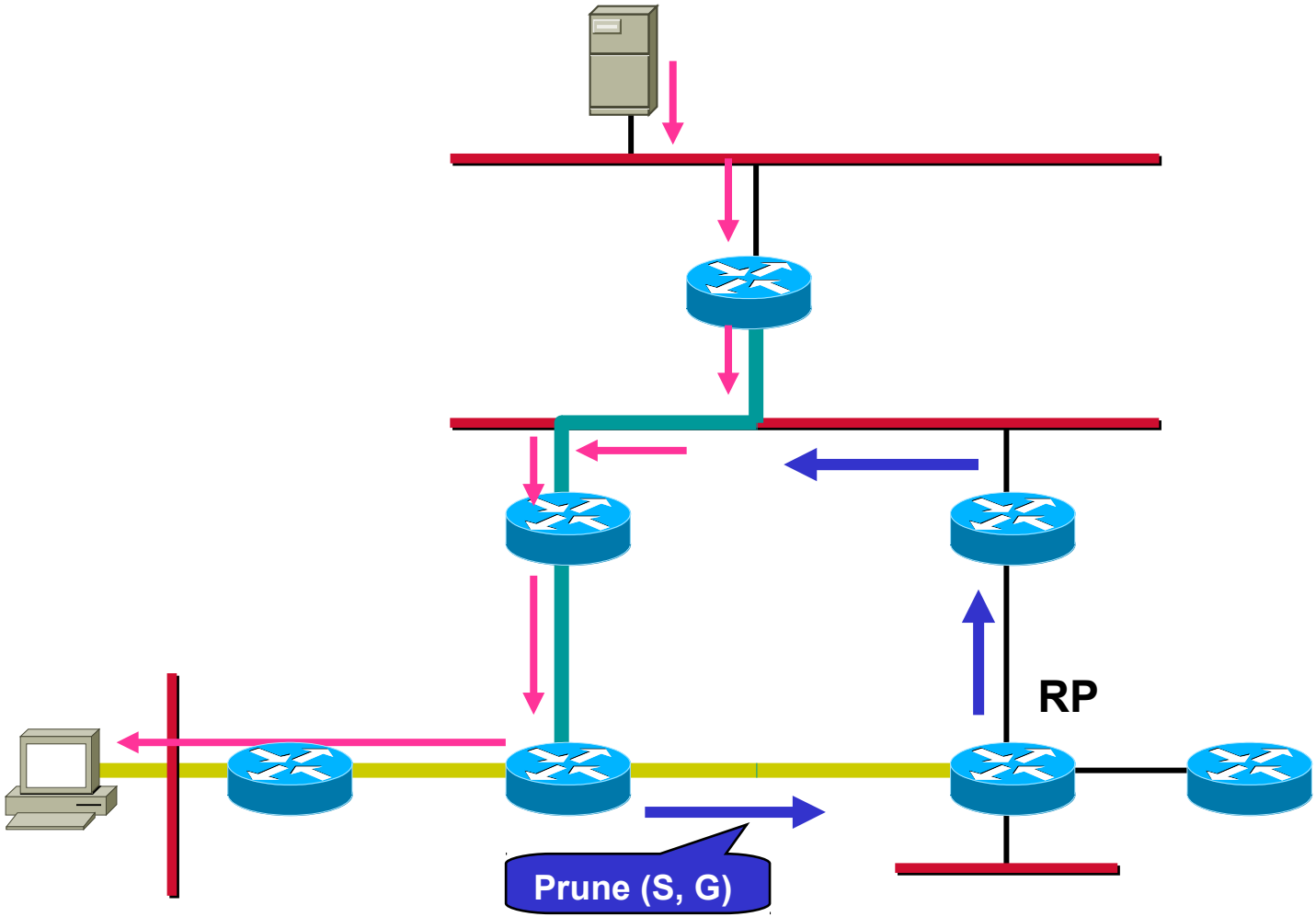
PIM-SM / Create Source Tree



PIM-SM / Switchover



PIM-SM / Pruning



PIM-SM Summary

- Now we learned:
  - PIM-SM can also create SPT (S, G) trees
  - But in a much more economical way than PIM-DM (fewer forwarding states)
- PIM-SM is:
  - Efficient, even for large scale multicast domains
  - Independent of underlying unicast routing protocols
  - Basis for inter-domain multicast routing used with MBGP and MSDP

Addendum: Bidir-PIM

- Less routers states
  - Only one (*, G) for multiple sources
  - No (S, G)
  - Same tree for traffic from sources toward RP and from RP to receivers
  - Trees may scale to an arbitrary number of sources
- Now bidirectional groups
  - Coexist with traditional unidirectional groups
  - All routers must recognize them (via PIMv2 flags)
  - Dedicated bidir RP required
- Designated Forwarder (DF) required
  - No register packets anymore
  - Knows best unicast route to RP
  - DF needed on any link between participant and RP

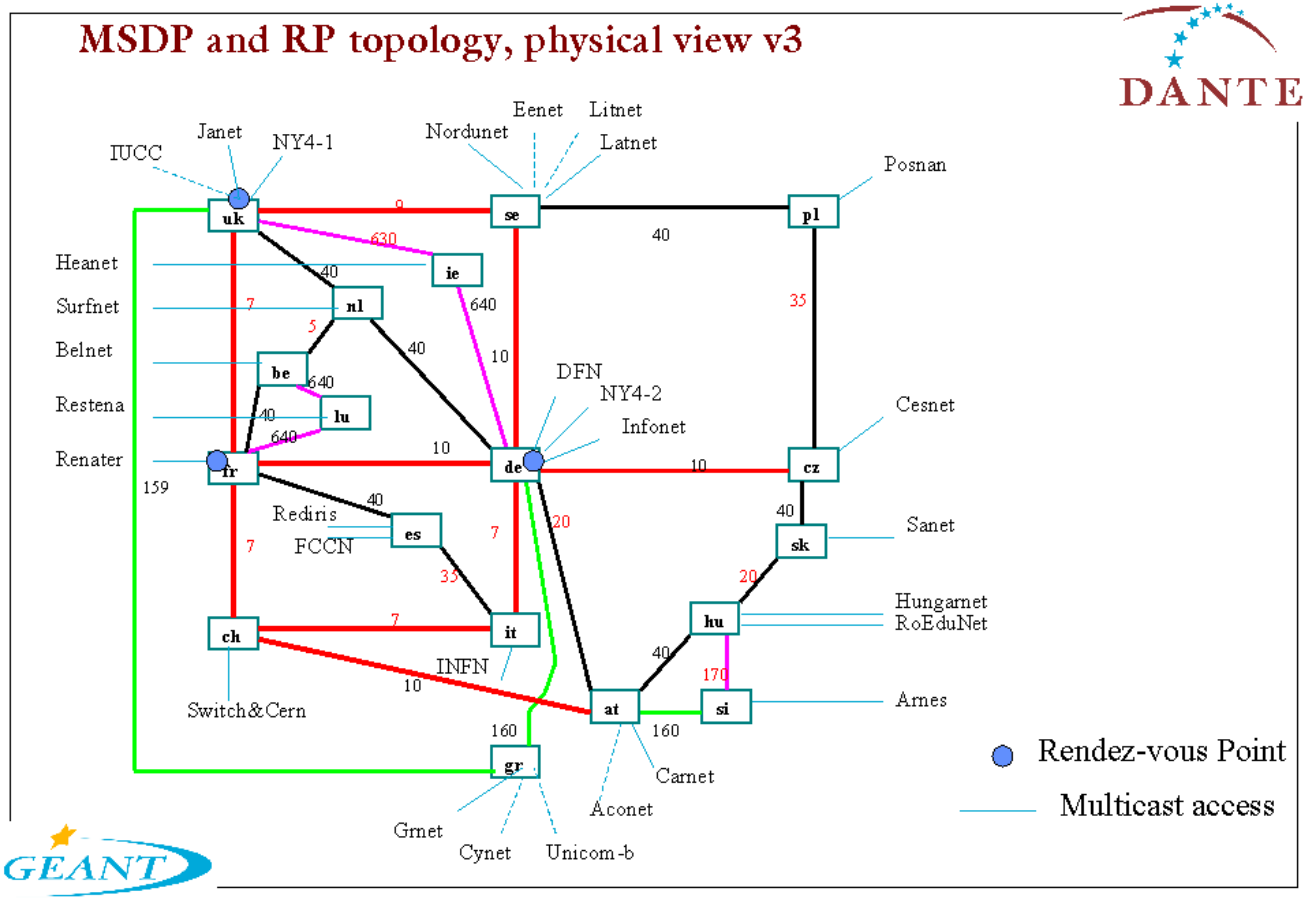
Addendum: PIM-SS

- Source-Specific Multicast (SSM)
  - Much simpler when sources are well known
- Immediate shortcut receiver to source
  - No need to create shared tree
  - DR sends (S, G) join directly to source
  - No MSDP needed for finding sources
- IGMPv3 needed!
  - Or IGMPv3 lite
  - Or URL Rendezvous Directory (URD)

SSM – Notes

- Take care that no shared tree uses the same group address
  - SSM protocols cannot avoid address collisions
  - Register/Join packets to 232/8 should be filtered

Inter-domain Multicast Routing



BGP Mcast Extensions

- Border Gateway Multicast Protocol (BGMP)
  - Supports global, scalable inter-domain multicast
  - Only disadvantage: Far from completion!
- MBGP/MSDP as intermediate solution
  - MBGP communicates multicast RPF information between AS's
  - MSDP distributes active source information between PIM-SM domains

Note

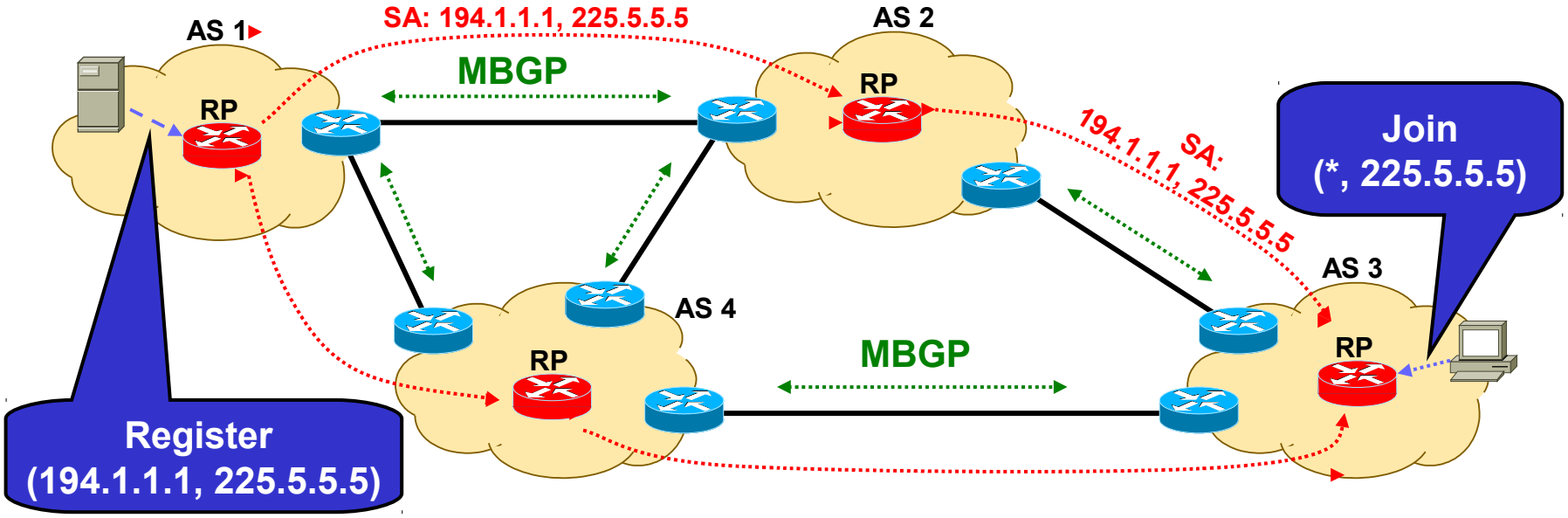
- ISPs often want to use a separate multicast topology
  - But PIM relies on underlying unicast routing protocol
  - Reverse path might be different
- MBGP creates multicast database
  - Filled with multicast NLRIs=(S, G)
- PIM-SM supposes one (closed) administrative multicast domain
  - MSDP sessions between RPs to interconnect multiple domains
  - Similar to eBGP (TCP)

MSDP

- MSDP peering from source RP to
  - Border routers
  - Other AS's RP
- If MSDP peer is a RP and has a (*, G) entry
  - This means there exists some interested receiver
  - Then a (S, G) entry is created an a shortcut to the source is made
  - Furthermore the receiver itself might switchover to the source

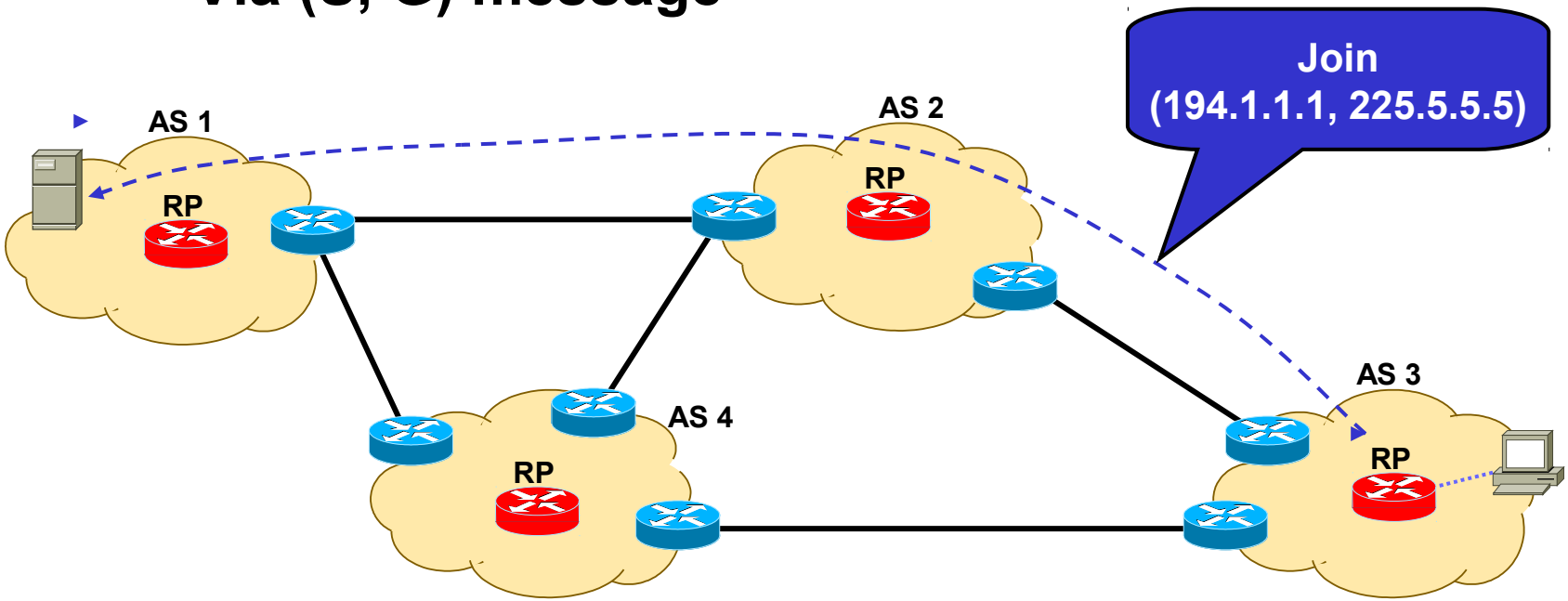
MBGP/MSDP (1)

- ASs establish multicast peering using MBGP
  - Via special Multicast RPF NLRI types
  - Used by PIM-SM to send (S, G) joins
- MSDP tells all RPs about *active* sources
  - Using Source Active (SA) messages
  - Containing (S, G) information



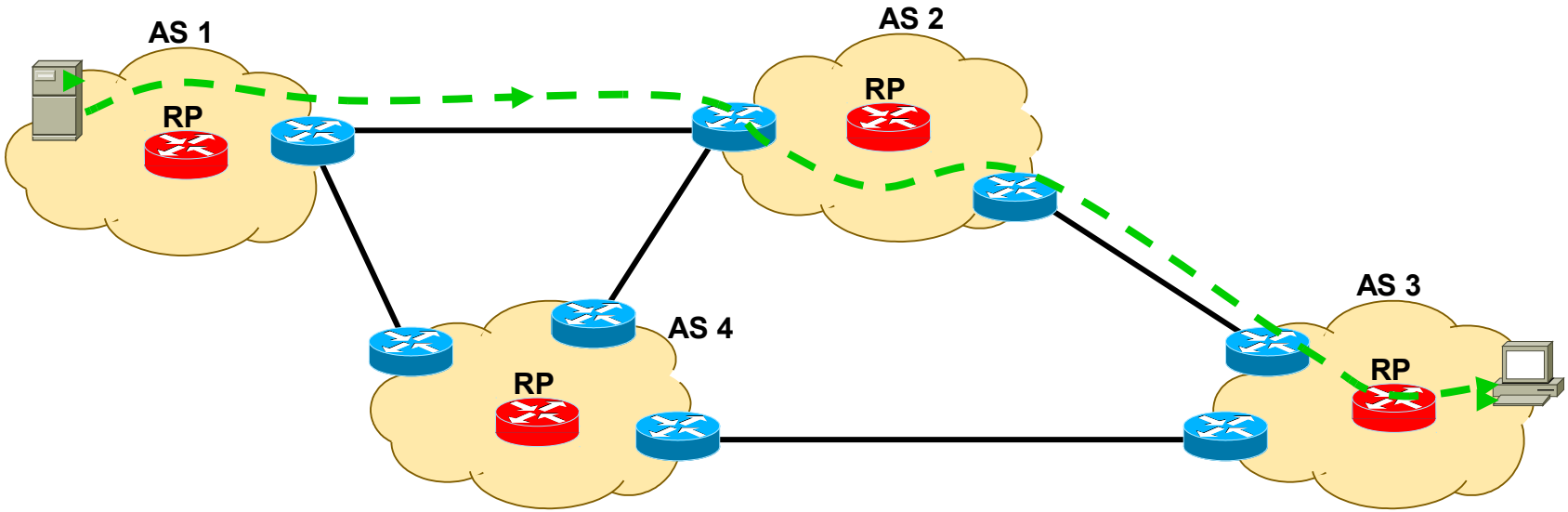
MBGP/MSDP (2)

- Receiver joined local RP
  - Via (*, G) message
- Local RP joins source directly
  - Via (S, G) message



MBGP/MSDP (3)

- Multicast traffic flows directly from the source to the receiver
  - Along a SPT downstream (to perhaps multiple receivers)
- Note: DRs and intermediate routers are omitted for simplicity!



Reliable Multicast

What is this? Who needs it?

- Reliable transmission means: no single bit gets lost over MDT !!!
- Traditional multicast can't guarantee that—and doesn't need to!
  - Audio and video does not bother
- But important for *data-based* applications
  - Whiteboarding
  - Efficient Usenet updates
  - Database synchronization
  - etc...
- Also real-time demands
  - Financial data delivery

Reliable Multicast (1)

- Remember: IP multicast is UDP based!
  - No guaranteed packet delivery!
  - No congestion control
  - Not intended for data transactions!
- RTP/RTCP only cares for
  - Duplicates
  - Sequence
- Reliable multicast requires UDP-based acknowledgements
  - TCP cannot do multicast by nature (too much overhead, state variables, buffers, timers, ...)
- Security issues for financial data delivery etc.!!!

Reliable Multicast (2)

- Guaranteed data delivery is provided by reliable multicast protocols
- Still UDP based *of course*
  - But ACKs are additionally implemented: *Feedback loop*
  - Data recovery mechanisms
  - Congestion control mechanisms

Feedback Loop

- Either performed by the *source*
  - End-to-end feedback loop (latency!)
  - Intermediate devices don't need to be multicast aware
  - Receivers send NACKs back to source
- Or *locally*
  - Hop-by-hop feedback loop
  - Intermediate "repair servers" cache packets for retransmissions
  - Nearest upstream server performs retransmission upon NACK
    - If not possible, NACK is sent to next upstream server

Optimizing Recovery

- One lost packet typically leads to a "NACK storm"
  - Sender must collapse all associated NACKs and retransmit only once
  - On a LAN only one receiver needs to send a NACK
  - (NACK suppression algorithm)
- Congestion-controlled retransmissions
  - Congestion is often cause of missing packets
  - Sender should retransmit when congestion is over
- Unidirectional links (e. g. satellite)
  - FEC against interferences
  - Redundant transmission against buffer overflows
    - Congestion control CRITICAL

Protocol Overview

- Reliable Multicast Protocol (RMP)
  - Token rotating scheme
- Reliable Multicast Transfer Protocol 2 (RMTP-2)
  - Relies upon "Top Node"
- Multicast File Transfer Protocol (MFTP)
  - Repair cycles
- Scalable Reliable Multicast (SRM)
  - Straight and simple
- Pragmatic General Multicast (PGM)
  - "Receivers self-help association"

RMP

- Useful for real-time, collaborative applications
- NACKs are sent to multicast address
  - Assures NACK suppression
  - Allows any member to perform retransmission
- Token rotation scheme
  - Owner of token may send ACK referring to recently received packets
  - Allows late joined members to inform about missing packets
- Retransmission to multicast group

RMTP

- Useful for bulk data distribution
- Hierarchically structured
- Periodic status messages:
  - Sent by leaf receivers to their designated receivers (DR)
  - Relayed via higher layer Designated Receivers up to the Sender
- Local retransmission and late joins possible
- Caching mechanisms in Designated Receivers

MFTP – 1. What is it?

- Useful for non-realtime bulk data distribution only
  - Developed by StarBurst Communications and Cisco Systems
  - Internet-draft February 1997
- Also includes diagnostic tools
  - Multicast ping (senders learn group population)
- Good scalability (thousands...)
- Flexible transport
  - Unicast, multicast, or broadcast dependent on number of receivers and medium

MFTP – 2. How does it?

- Server announces transmission and waits for receiver registration
  - Hereby learning population
  - Announcement contains filename and size
  - Well-known multicast group address for announcements
  - Registration suppression on LANs
- Then data is sent and NACKs collected
  - NACKs are collapsed, retransmission *afterwards*
  - Several retransmissions if necessary (slow but reliable)

MFTP – 3. Protocol Details

- File is sent in blocks
  - Some 1000 packets per block
  - Consists of Data Transmission Units (DTUs)
  - Source sends status request message after each block
- NACKs are sent after each block
  - Containing bit-map indicating bad DTUs
  - Unicast
- ACKs could be sent but are typically turned off to reduce traffic
  - Only one ACK at the session end is required

MFTP – 4. Three Group Models

- Closed groups
  - Members are known by source
  - Only those members may register
- Open limited groups
  - Unknown members
  - Source expects registration
- Unlimited groups
  - No registration expected

SRM

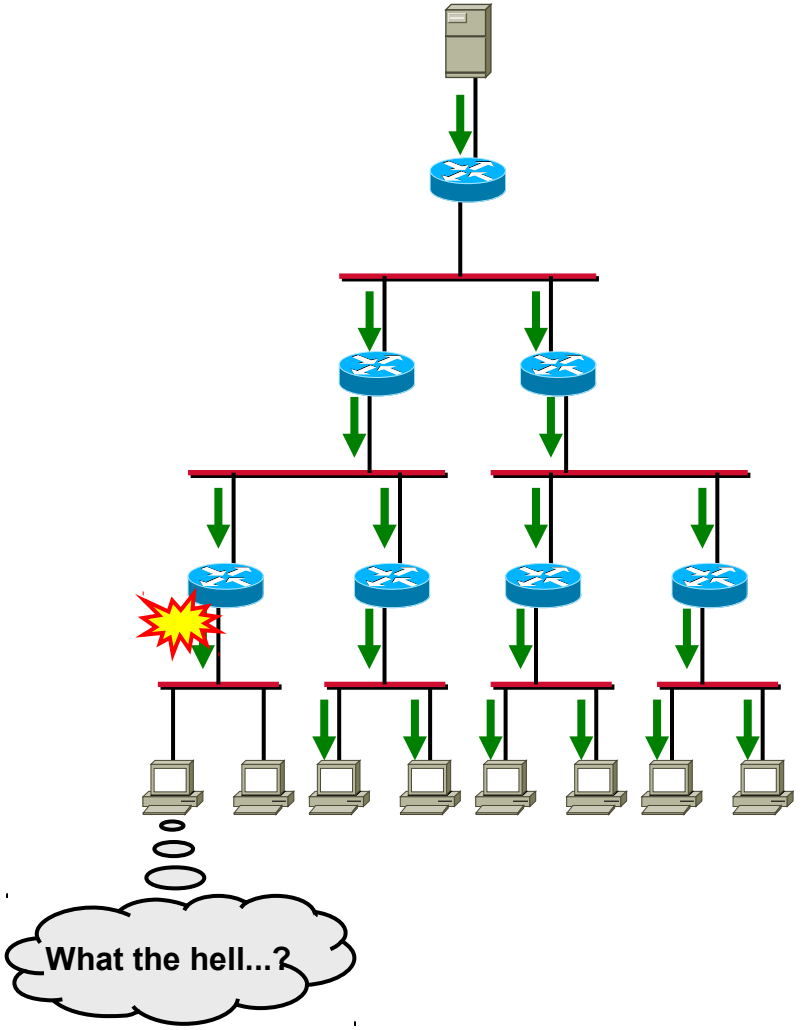
- For whiteboarding (wb) in Mbone and general data distribution
  - Does not care for ordered packet delivery
  - NACKs are sent to group
  - Both NACK and retransmission suppression
  - Two models: ALF and LWS
- Application Level Framing (ALF)
  - Data is uniquely identified by Source-ID and Page-ID
  - Time stamp, Sequence Number
  - Application must re-sequence
- Light-Weight Sessions (LWS)
  - Additional session messages as feedback loop
  - Ideal for conferencing applications

PGM

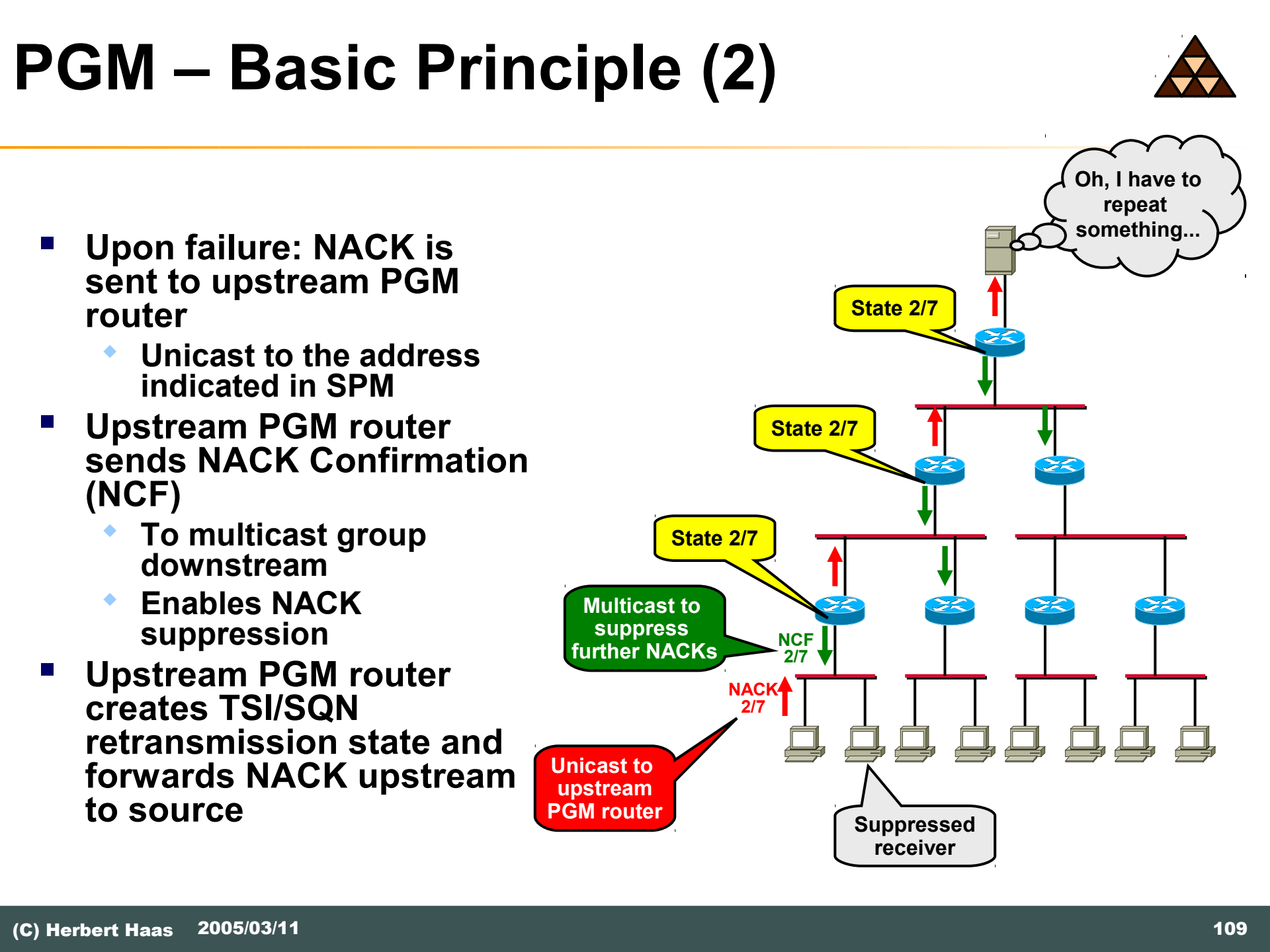
- Best known solution (Cisco)
  - Duplicate-free, ordered delivery
  - Several application-friendly features
  - Multiple senders and receivers
  - Independent of layer 3
  - Internet-Draft, January 1998
- Routers support *local* feedback loops
  - "PGM Assist features"

PGM – Basic Principle (1)

- Source sends ordered data (ODATA) containing
  - Transport session identifier (TSI)
  - Sequence number (SQN)
- Source sends also Source Path Messages (SPM)
  - Interleaved with ordered multicast data
  - Provides an upstream path
  - *Not shown in the picture*



PGM – Basic Principle (2)



- Upon failure: NACK is sent to upstream PGM router
  - Unicast to the address indicated in SPM
- Upstream PGM router sends NACK Confirmation (NCF)
  - To multicast group downstream
  - Enables NACK suppression
- Upstream PGM router creates TSI/SQN retransmission state and forwards NACK upstream to source

PGM – Options

- Late joining
  - Sources indicate whether lately joined receivers may request all missing data
- Time stamps
  - Receivers tell urgency of retransmissions
- Reception quality reports
  - Sent by receivers for congestion control
- Fragmentation
  - To confirm to MTU
- FEC
  - To reduce selective retransmissions

Summary

- Multicast routing requires creation of spanning trees
  - Avoid multiple packets
  - Avoid multicast storms
- Source-based and Shared trees
- Push and Pull methods
- IGMP to announce group membership
- Current favourite: PIM-SM
- Also reliable multicast solutions available
  - PGM is most important