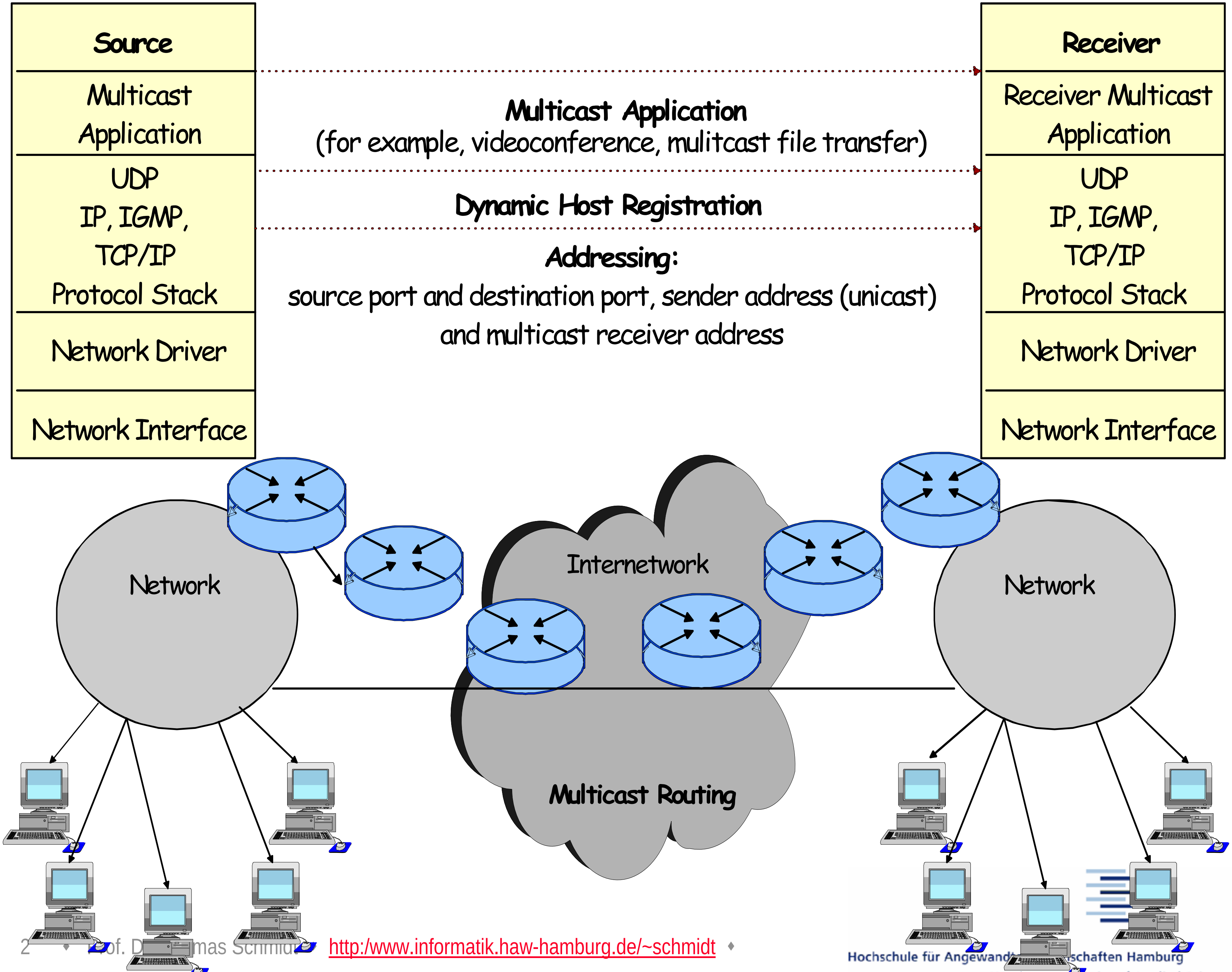
Multicast Routing

- The Problem of IP Mcast Routing
- Routing Algorithms
- ASM Routing Protocols
- SSM Routing
- Properties of Multicast Distribution Trees
- Efficiency versus Deployment Complexity

Some graphics originate (in part) from cisco

Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences


Source
Multicast Application
UDP
IP, IGMP,
TCP/IP
Protocol Stack
Network Driver
Network Interface
Multicast Application
(for example, videoconference, mulitcast file transfer)
Dynamic Host Registration
Addressing:
source port and destination port, sender address (unicast) and multicast receiver address
Receiver
Receiver Multicast Application
UDP
IP, IGMP,
TCP/IP
Protocol Stack
Network Driver
Network Interface
Network
Internetwork
Multicast Routing
Network

Multicast Routing

Unicast IP-Routing

- Guides IP-Datagrams stepwise to *one* receiver
- Routing decision on *where* to forward packet to
- Solely based on *destination* address
- Adapts to Router topology, *never* to IP-Packets

⇒ Multicast turns Routing upside down

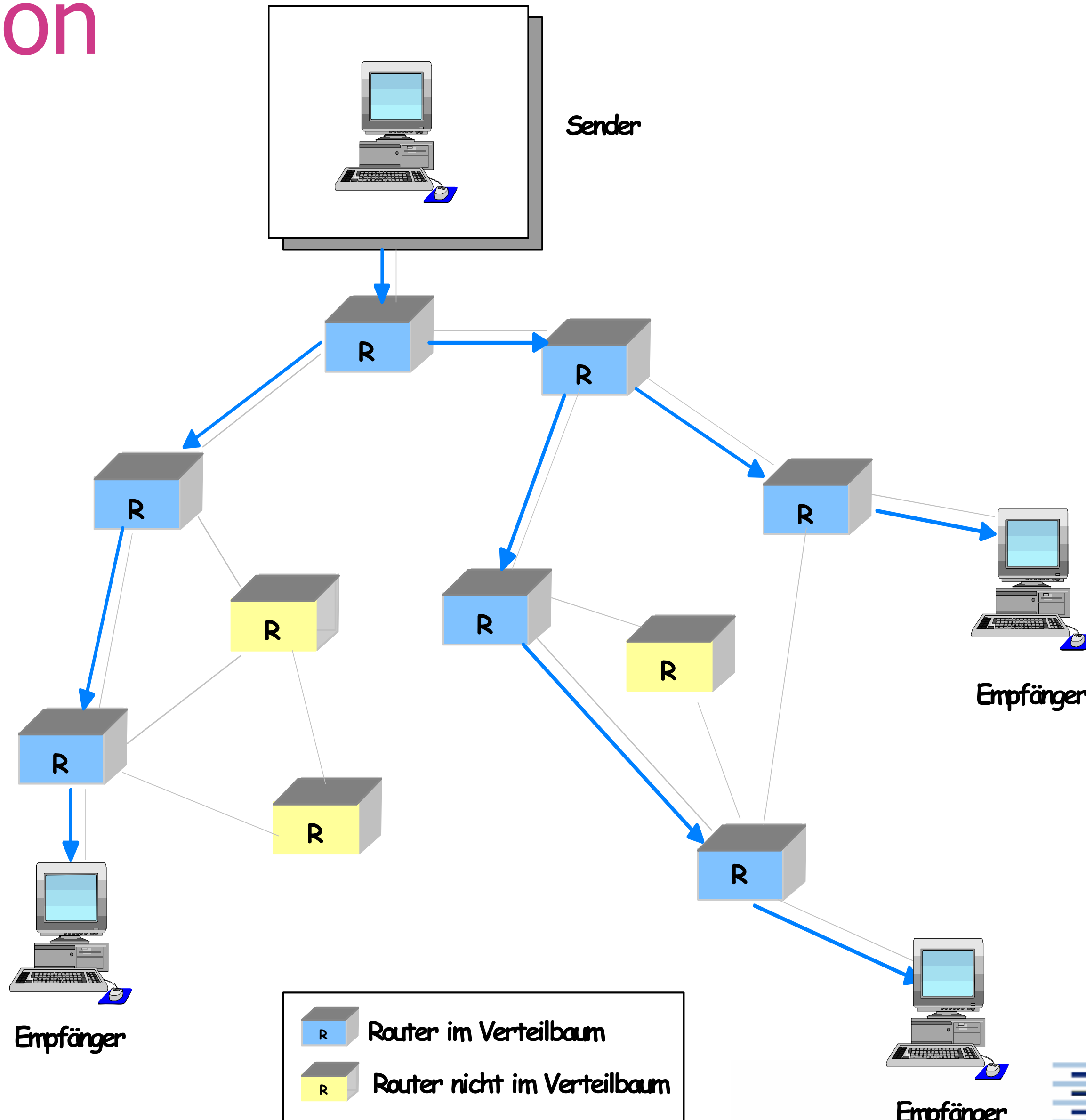
Multicast Routing (2)

IP Multicast - Routing is receiver initiated:

- Guides mcast-Datagrams according to a distribution tree
- Duplicates Datagrams
- Based on *Source* address
- Changes according to group dynamics
- Uses ‚Reverse' Paths

Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Distribution Tree



Receiver Initiated Routing

- Group initiation by sender results in distribution tree
- Two types of distribution trees:
    - Source Specific Tree originating at sender (S,G) or
    - Shared Tree originating at Rendezvous Point (*,G) (serving a group of senders)
- Calculation of Routing Information stimulated by receiver
    - A receiver adds/removes branches to/from distribution tree
- Unicast routing tables usable (requires symmetric routing!)
- Forwarding Algorithm: Reverse Path Forwarding

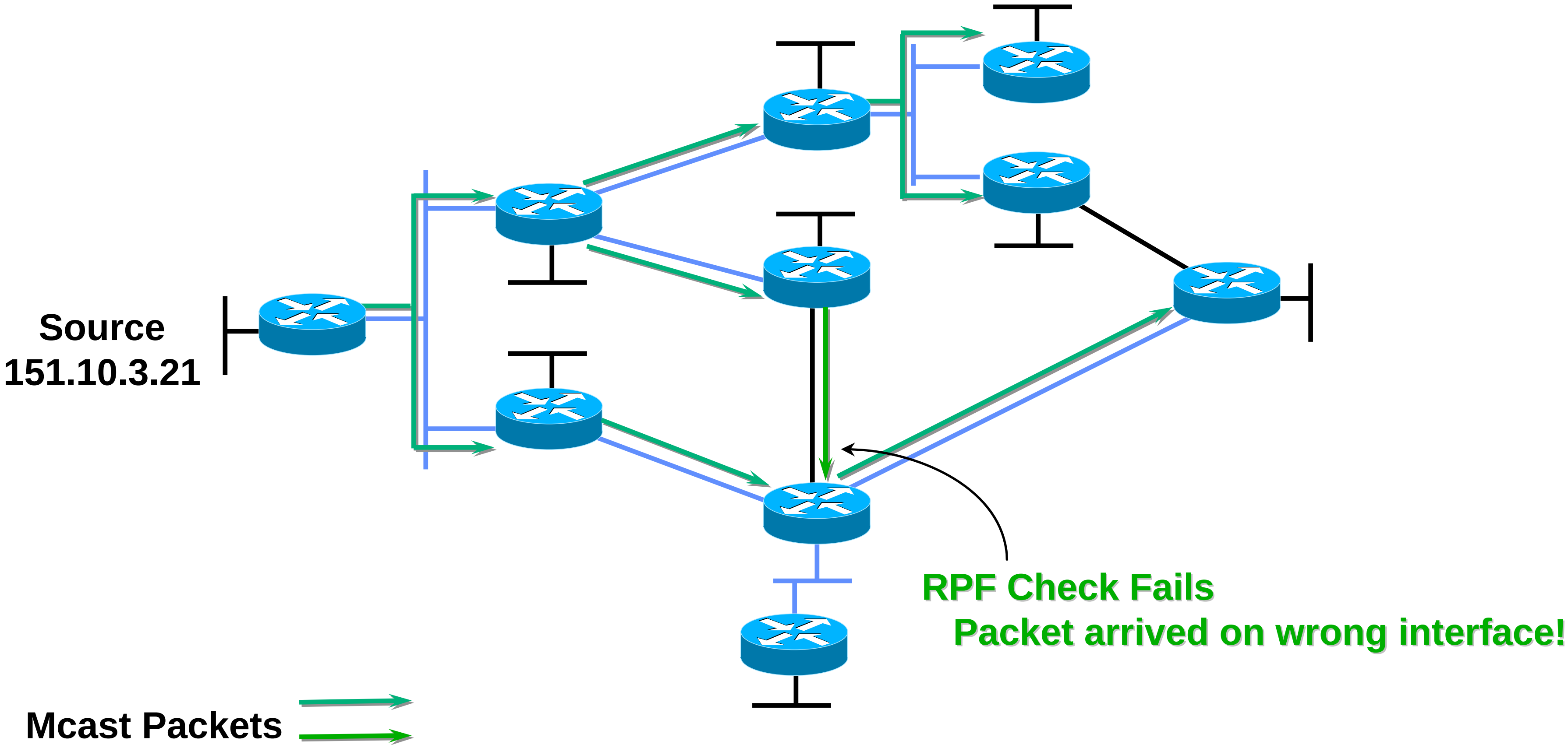
Reverse Path Forwarding (RPF)

A Router forwards a packet only, if it was received on the proper route to source.

RPF Check:

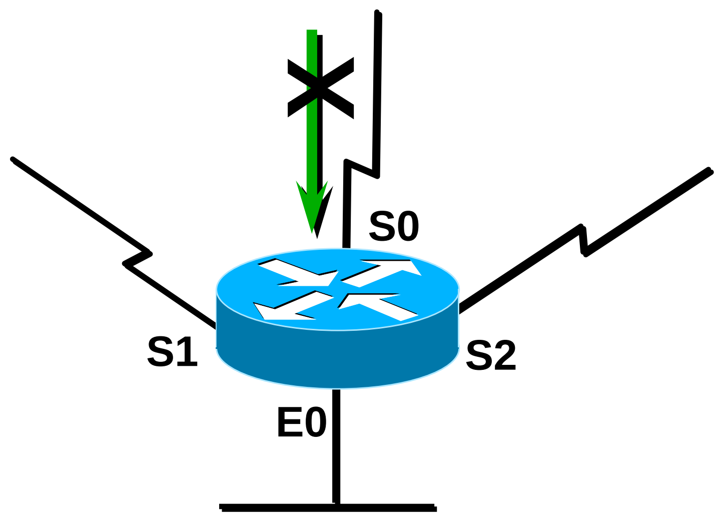
- active routing table searched for *source*-address
- Packet transmitted, if received on the interface foreseen as source address destination
- Packet discarded otherwise

RPF Check


Source
151.10.3.21
RPF Check Fails
Packet arrived on wrong interface!
Mcast Packets

RPF Check: Failure

Multicast Packet from Source 151.10.3.21



RPF Check Fails!

Unicast Route Table	
Network	Interface
151.10.0.0/16	S1
198.14.32.0/24	S0
204.1.16.0/24	E0

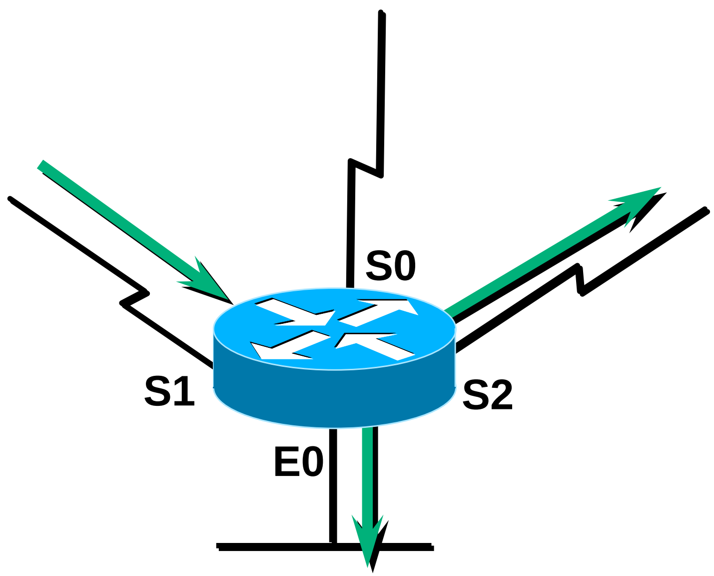
Packet Arrived on Wrong Interface!

Discard Packet!

Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

RPF Check: Success

Multicast Packet from Source 151.10.3.21



RPF Check Succeeds!

Unicast Route Table	
Network	**Interface**
151.10.0.0/16	S1
198.14.32.0/24	S0
204.1.16.0/24	E0

Packet Arrived on Correct Interface!

Forward out all outgoing interfaces.
(i. e. down the distribution tree)

Any Source Multicast (ASM)

How to construct distribution tree to reach all receivers?

Two classes of algorithms:

Dense Mode

- Push Model
- Flooding and Pruning

Sparse Mode

- Pull Model
- Directional traffic only
- Rendezvous Points

Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

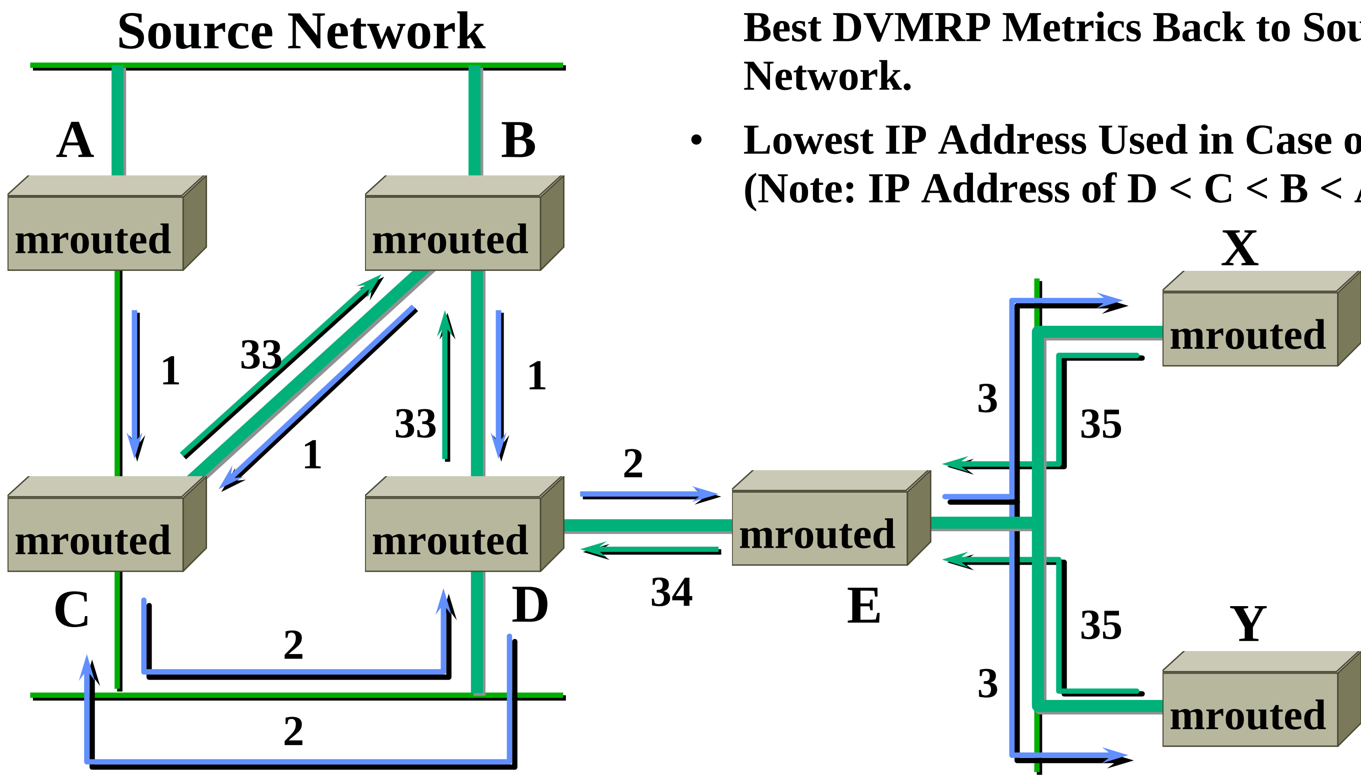
Distance Vector Multicast Routing Protocol (DVMRP)

- Oldest IP mcast Routing Protocol (v1: RFC 1075, Deering)
- Destination based Distance Vector Protocol
- Dense Mode Protocol
- Generates source-specific shortest path trees
- Currently V3 allows for Mcast tunnelling
- Operates on RIP bases (as Unicast Routing Protocol)
- Transmits Subnetmasks
- ∞ = 32 Hops, sometimes 16

Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

DVMRP Distribution Tree: Construction

- **Truncated Broadcast Trees Are Built using Best DVMRP Metrics Back to Source Network.**
- **Lowest IP Address Used in Case of a Tie. (Note: IP Address of D < C < B < A)**

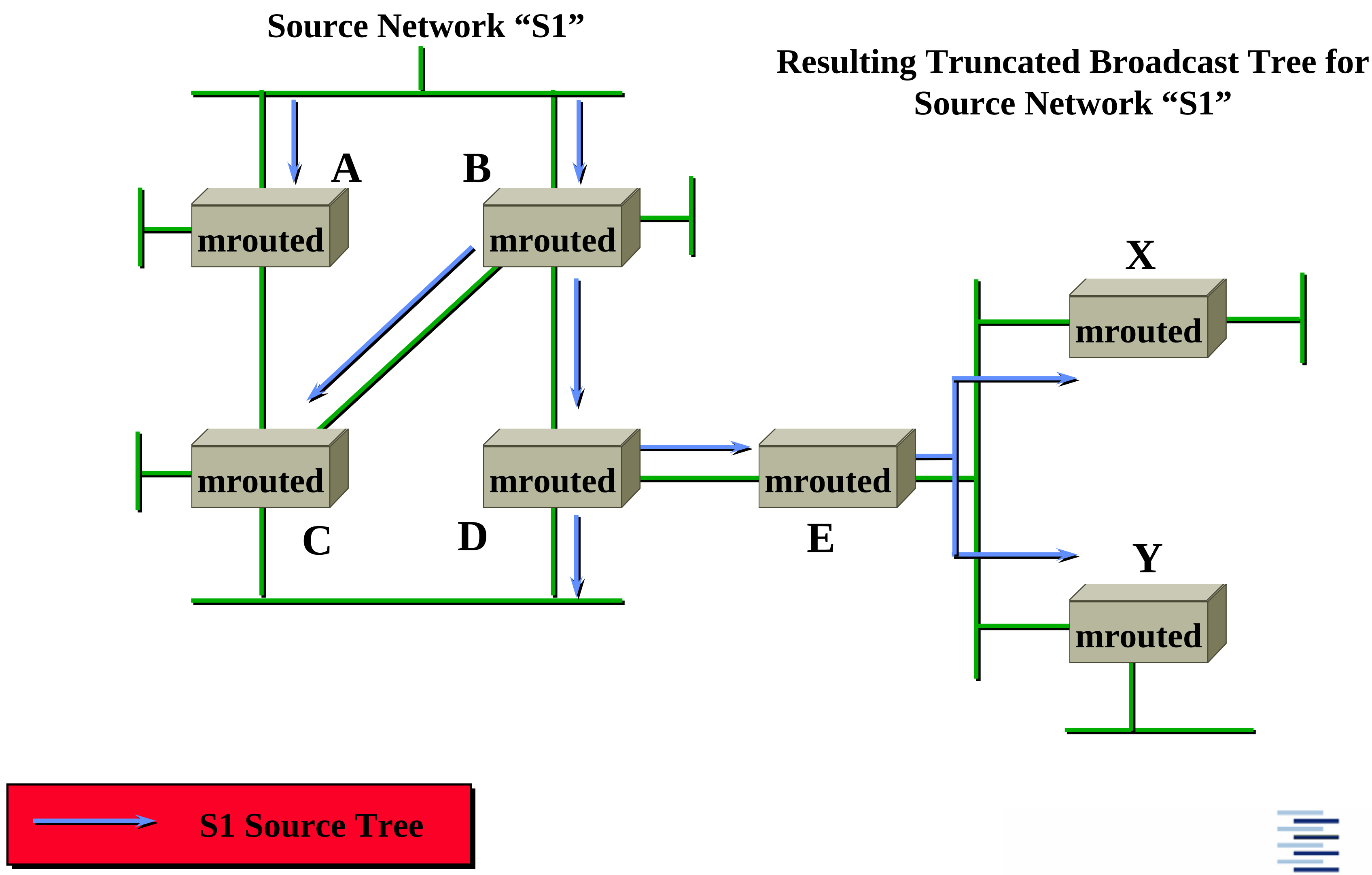


n — **Route for source network of metric "n"**

m — **Poison reverse (metric + infinity) sent to upstream "parent" router. Router depends on "parent" to receive traffic for this source.**

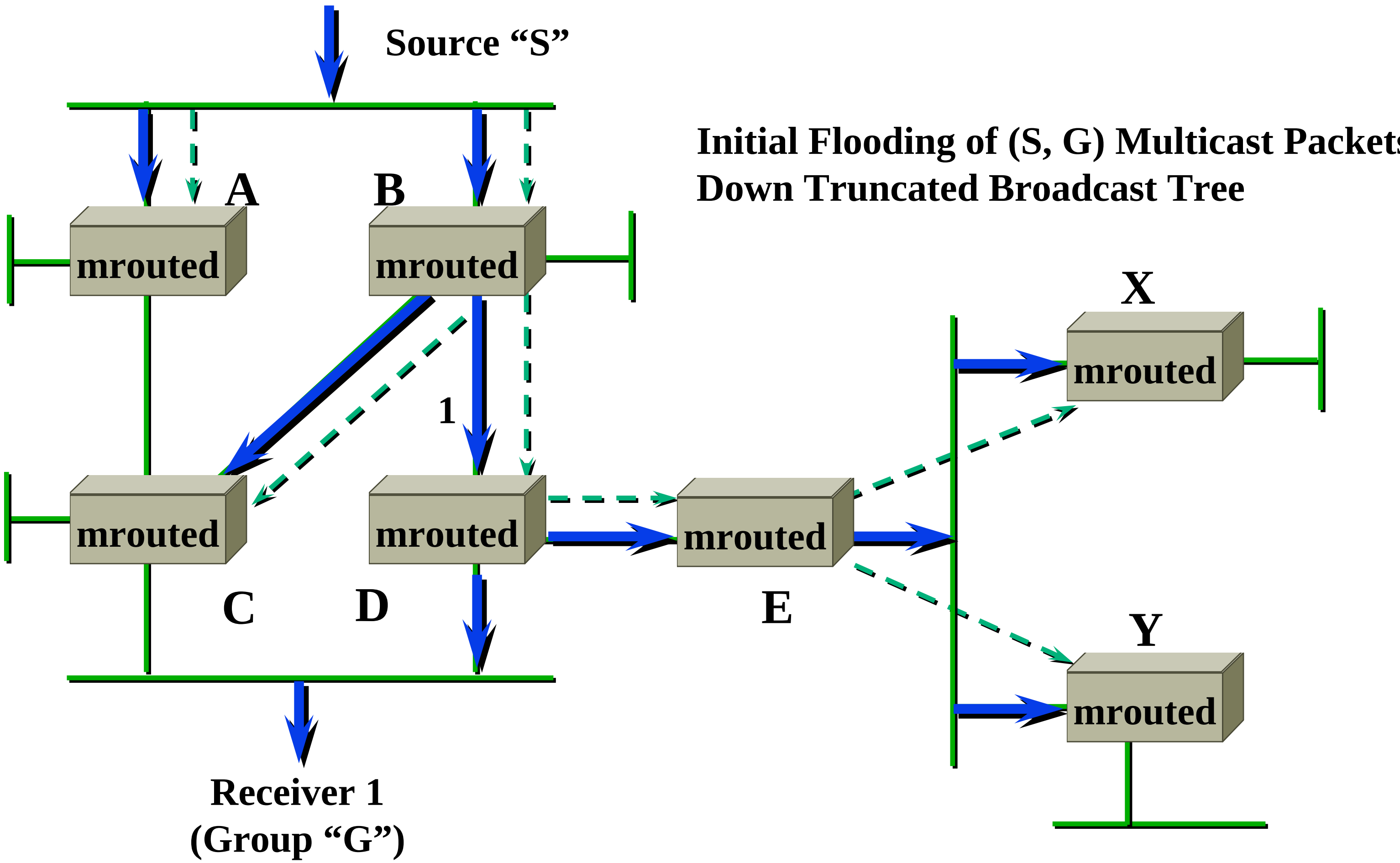
Resulting Truncated Broadcast Tree for Source Network

DVMRP Distribution Tree



Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

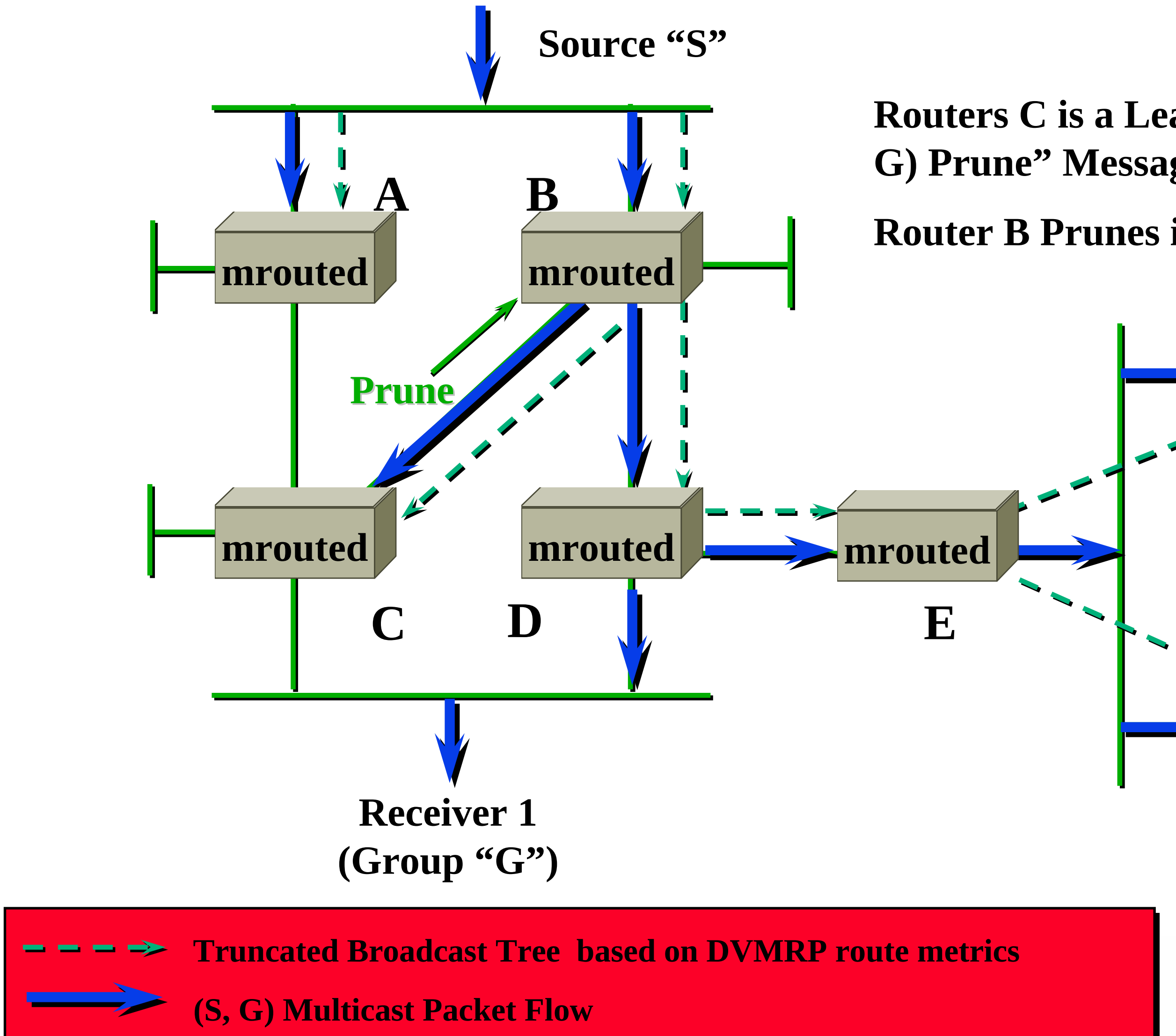
DVMRP Flood & Prune



Truncated Broadcast Tree based on DVMRP route metrics

(S, G) Multicast Packet Flow

DVMRP Flood & Prune



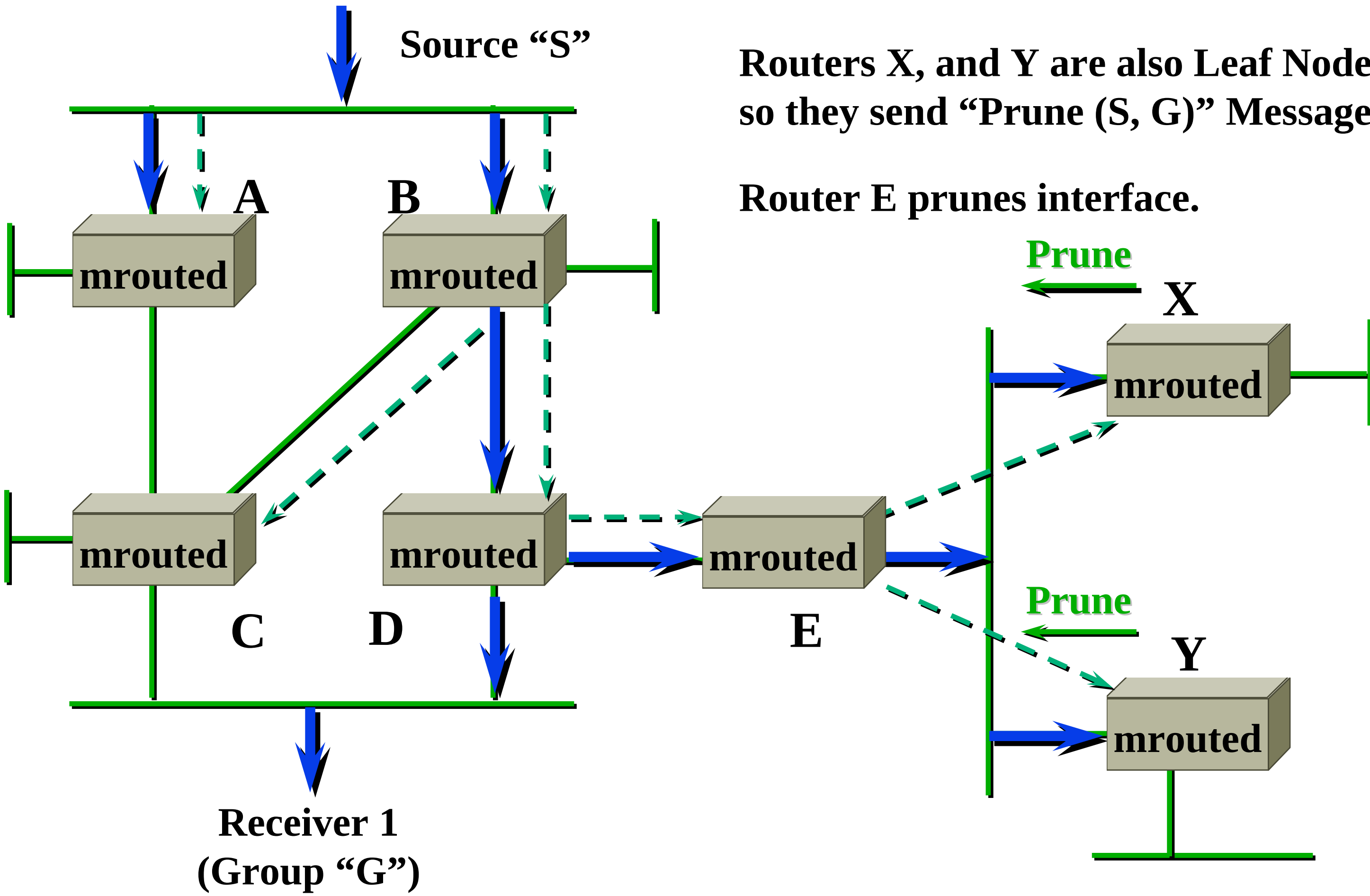
Routers C is a Leaf Node so it sends an "(S, G) Prune" Message

Router B Prunes interface.

DVMRP Flood & Prune

Routers X, and Y are also Leaf Nodes so they send “Prune (S, G)” Messages

Router E prunes interface.

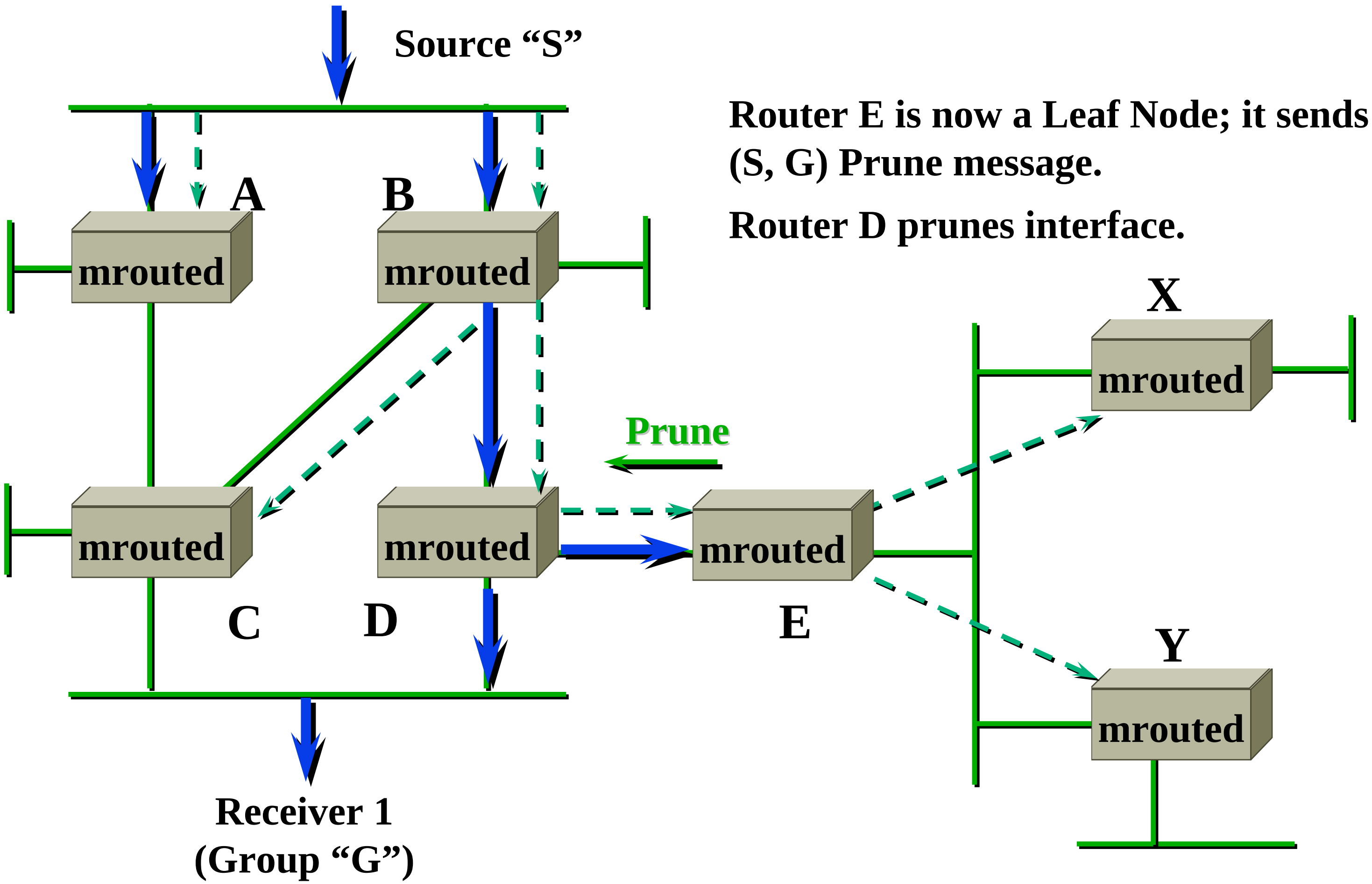


Truncated Broadcast Tree based on DVMRP route metrics

(S, G) Multicast Packet Flow

Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

DVMRP Flood & Prune



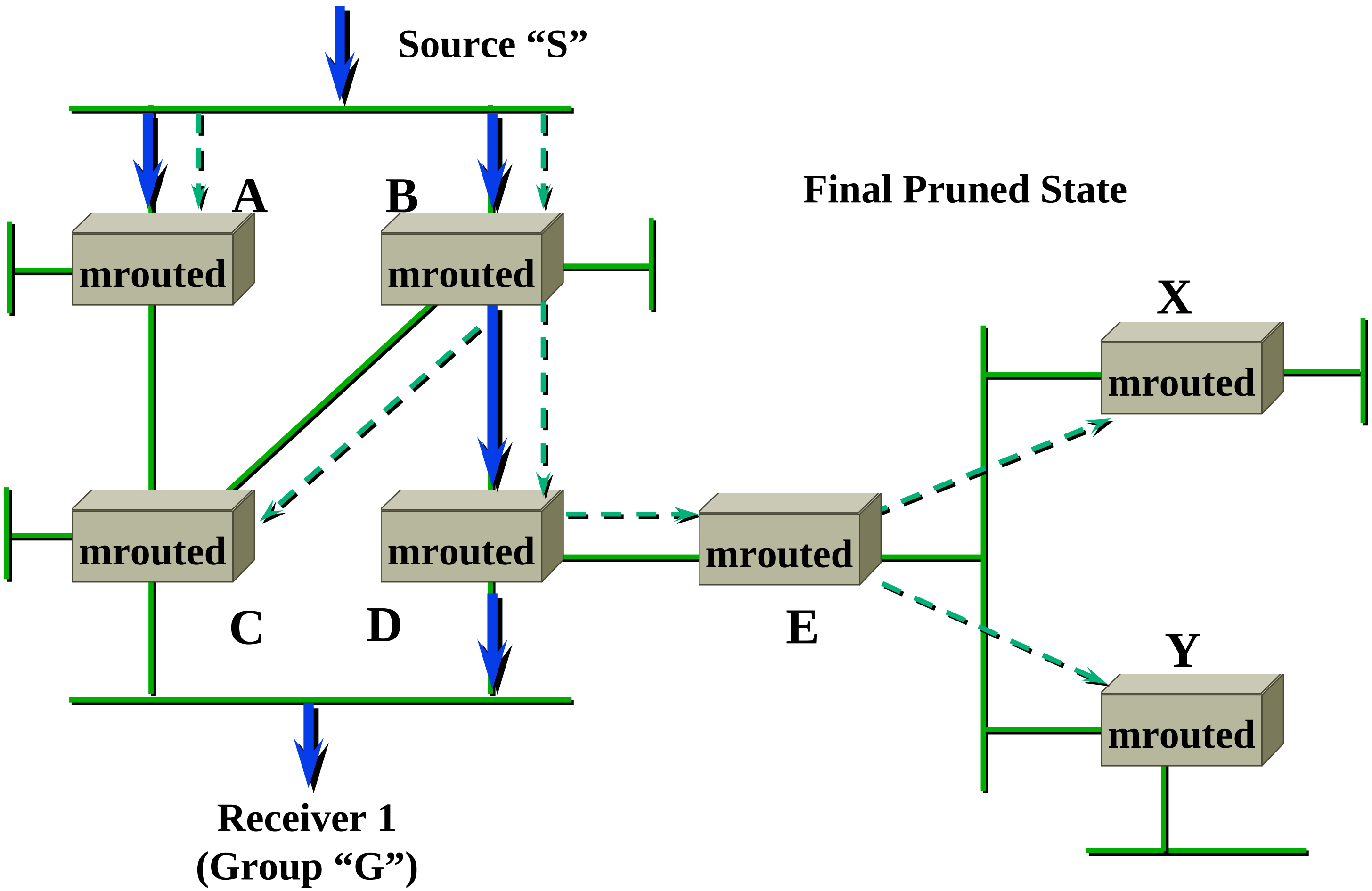
Router E is now a Leaf Node; it sends an (S, G) Prune message.

Router D prunes interface.

Truncated Broadcast Tree based on DVMRP route metrics

(S, G) Multicast Packet Flow

DVMRP Flood & Prune



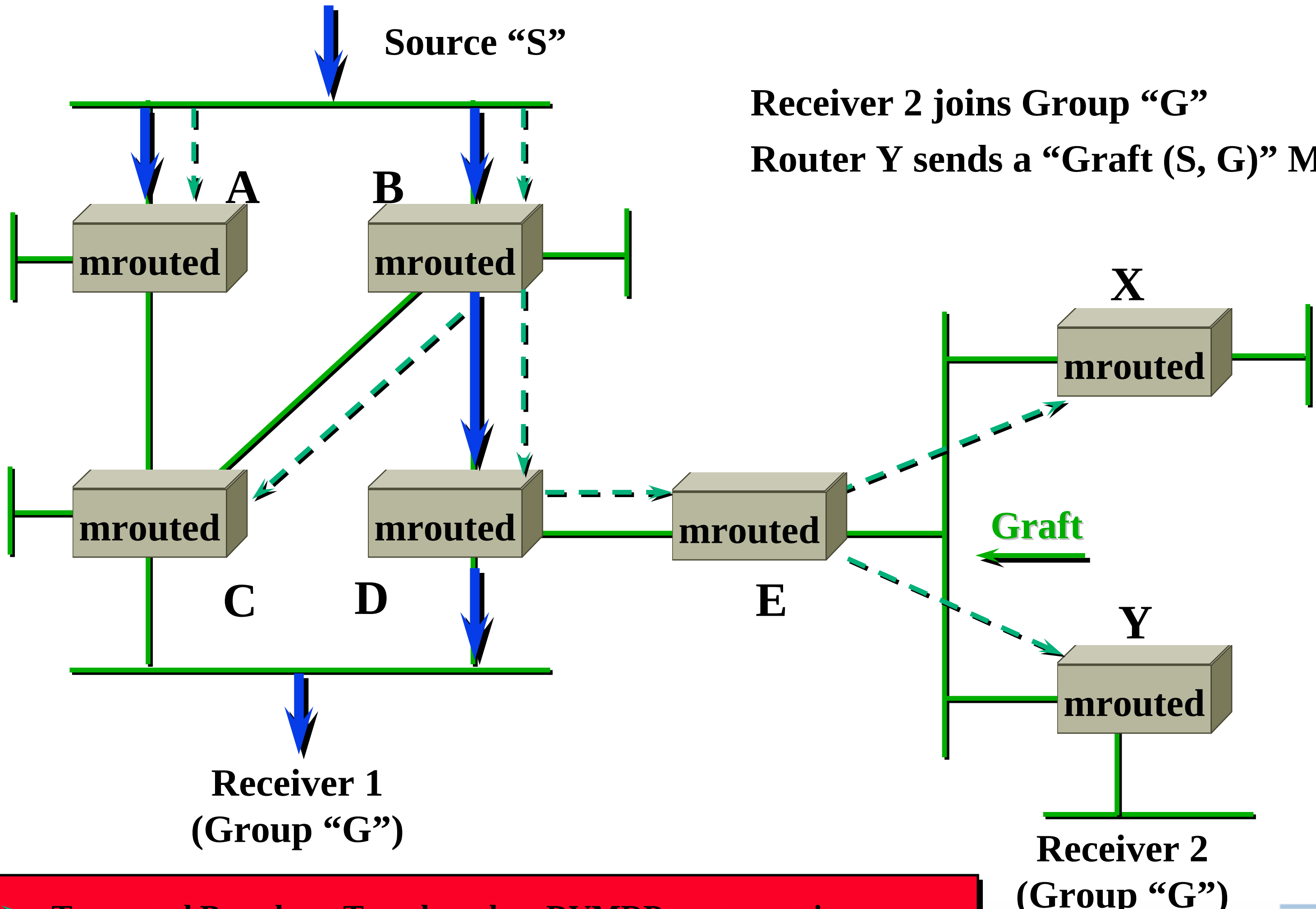
Truncated Broadcast Tree based on DVMRP route metrics

(S, G) Multicast Packet Flow

New Receiver in DVMRP: Grafting

Receiver 2 joins Group “G”

Router Y sends a “Graft (S, G)” Message



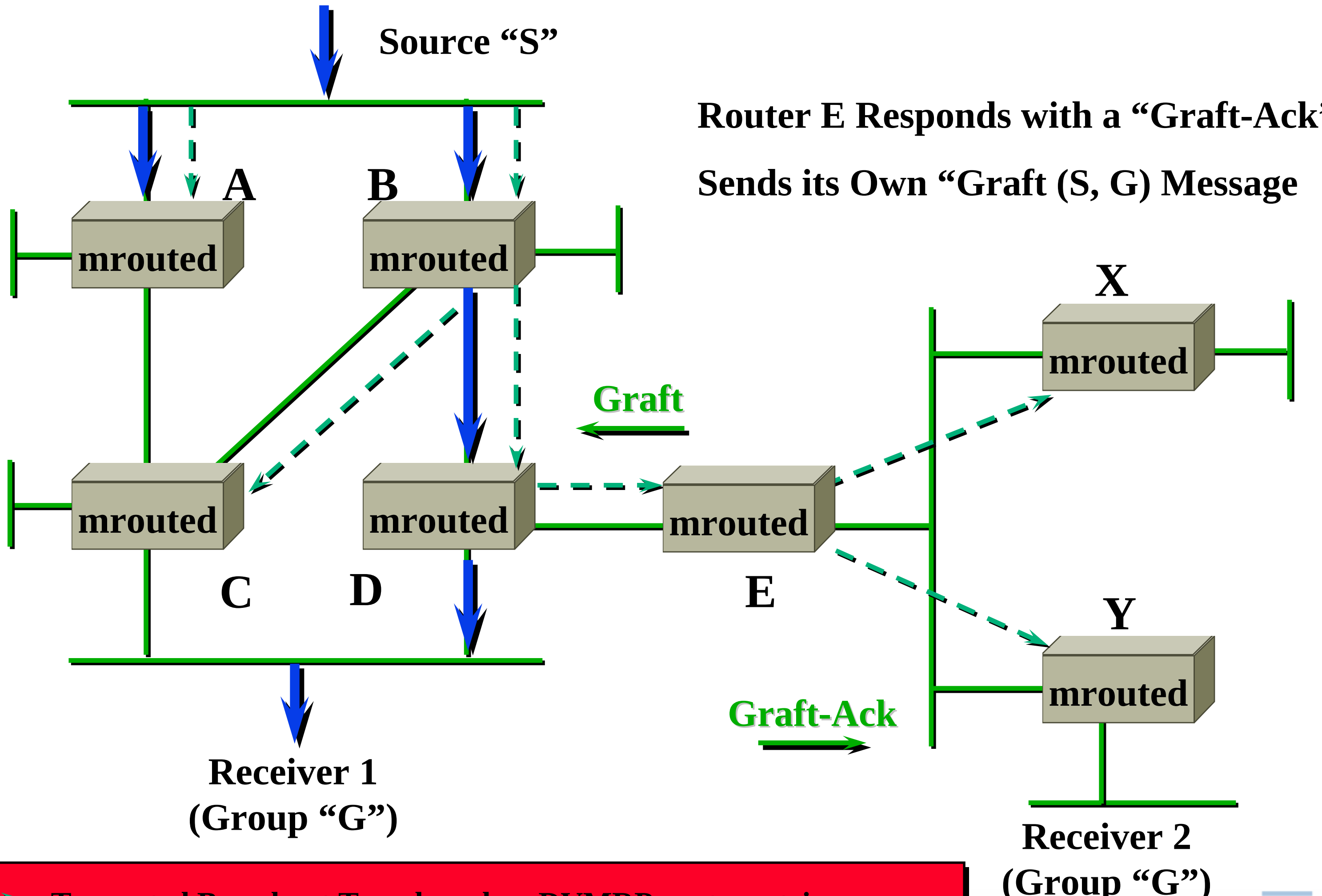
Truncated Broadcast Tree based on DVMRP route metrics

(S, G) Multicast Packet Flow

DVMRP Grafting

Router E Responds with a “Graft-Ack”

Sends its Own “Graft (S, G) Message



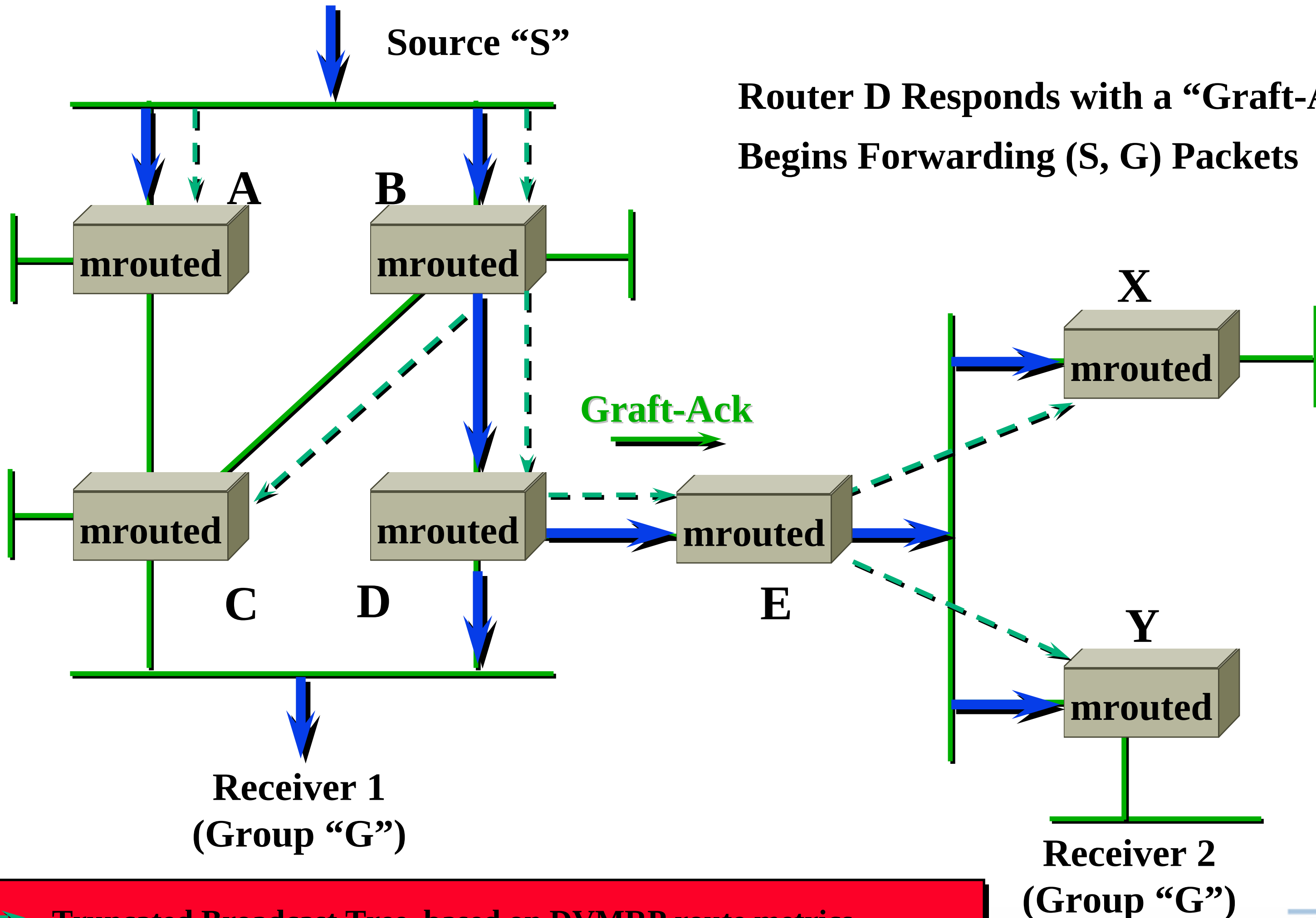
Truncated Broadcast Tree based on DVMRP route metrics

(S, G) Multicast Packet Flow

DVMRP Grafting

Router D Responds with a “Graft-Ack”

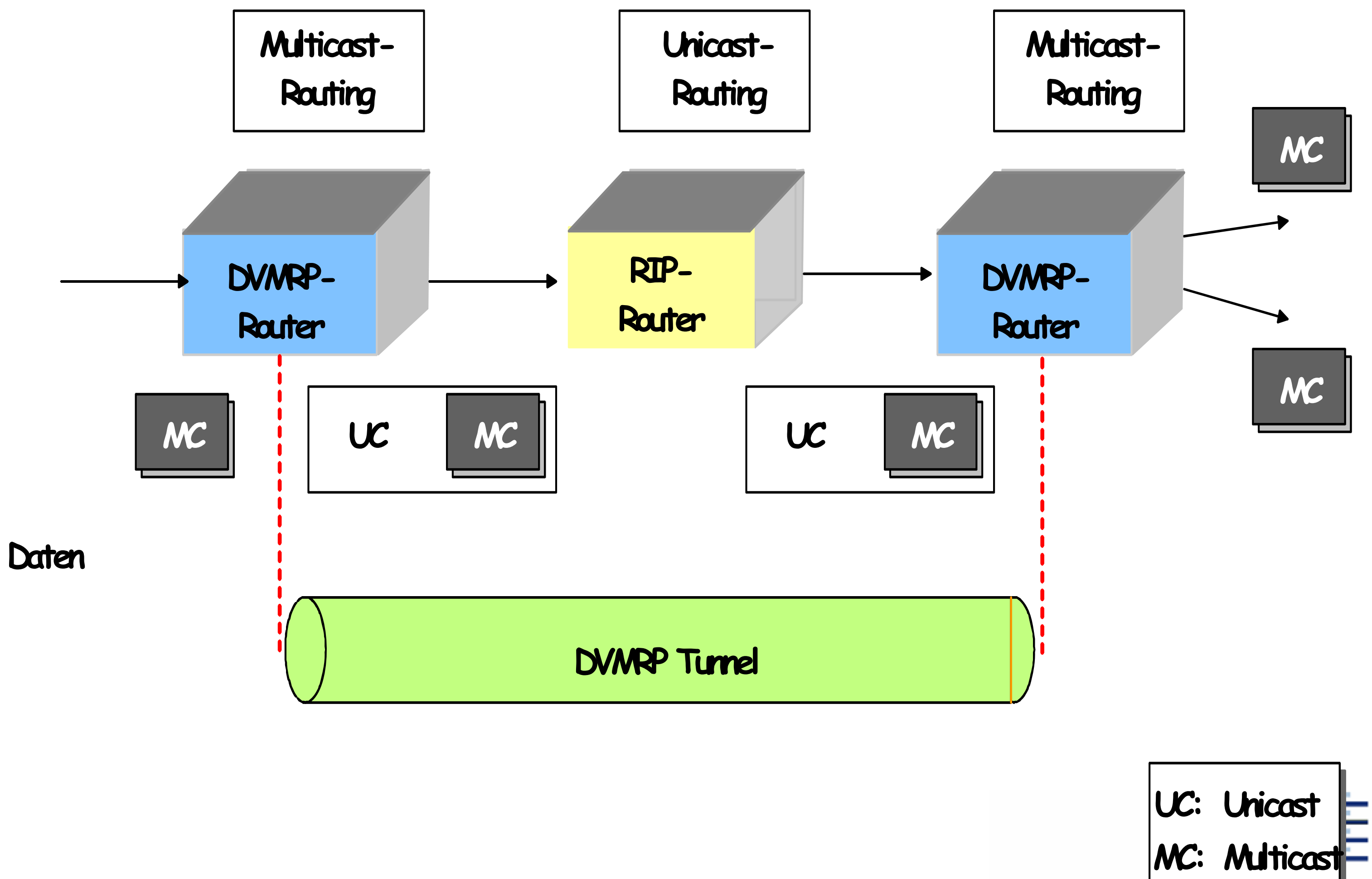
Begins Forwarding (S, G) Packets



Truncated Broadcast Tree based on DVMRP route metrics

(S, G) Multicast Packet Flow

DVMRP Tunnelling



Multicast Open Shortest Path First (MOSPF)

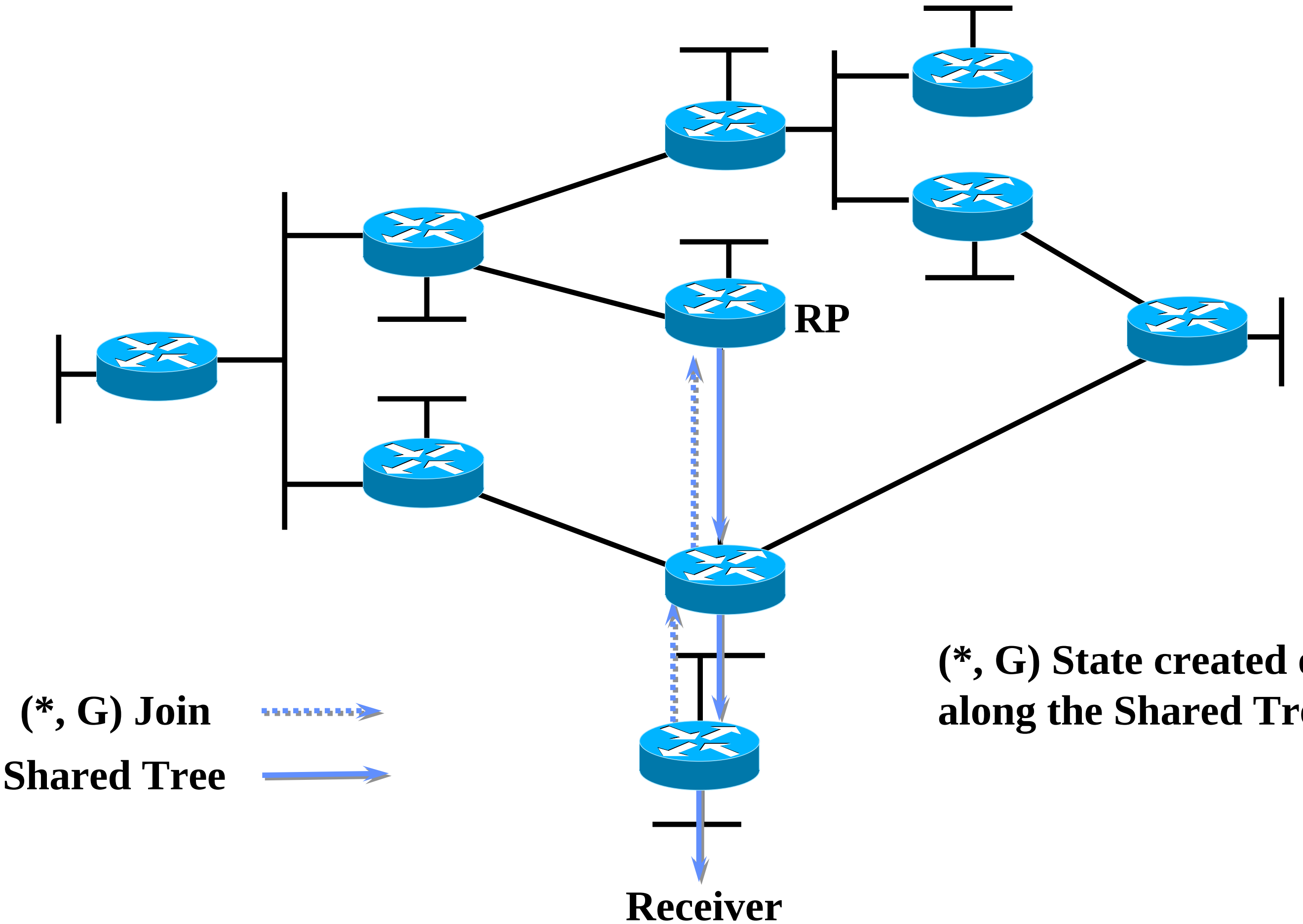
- Extends OSPF for Multicast Routing
- Destination based link state protocol (dense mode)
- Distribution of link states (OSPF)
- Group member link states flooded
- Every router learns a complete topology and calculates shortest path tree
- MOSPF corresponds to OSPF-Unicast-Routing

Protocol Independent Multicast (PIM)

- *Protocol independence:*

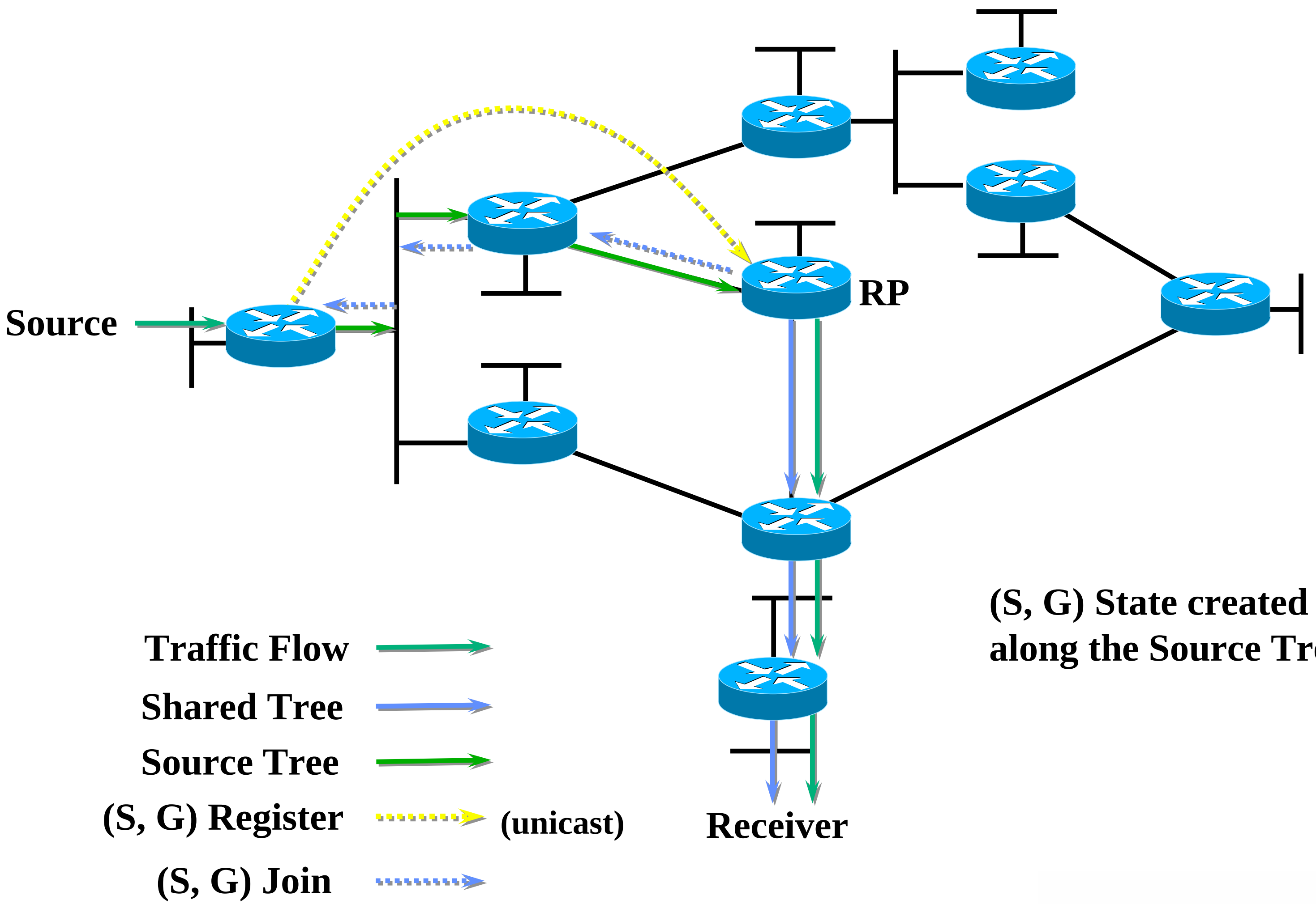
  works with all underlying Unicast Routing Protocols
- Dense und Sparse Mode PIM (RFC …, current RFC 4601 08/'06)
- Dense Mode PIM floods & prunes (as DVMRP)
- Sparse Mode PIM uses Rendezvous Points (RP)
  - Constructs a shared distribution tree centred at RP
  - Efficient for widely distributed groups
  - Favoured for wide area networks – problem: inter-RP signalling
  - Now widely implemented

PIM SM Tree Joins



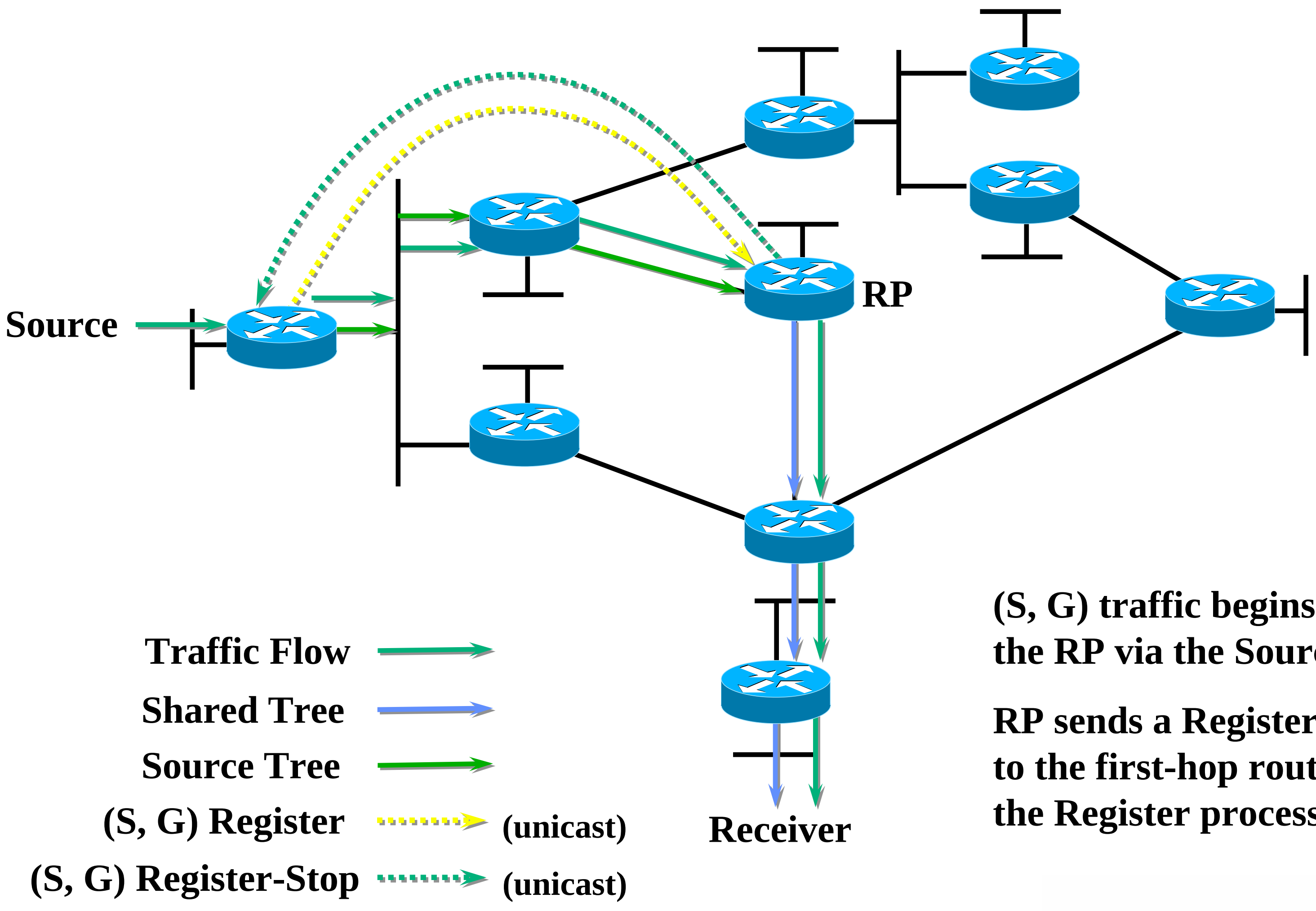
(*, G) State created only along the Shared Tree.

PIM SM Sender Registration



(S, G) State created only along the Source Tree.

PIM SM Sender Registration

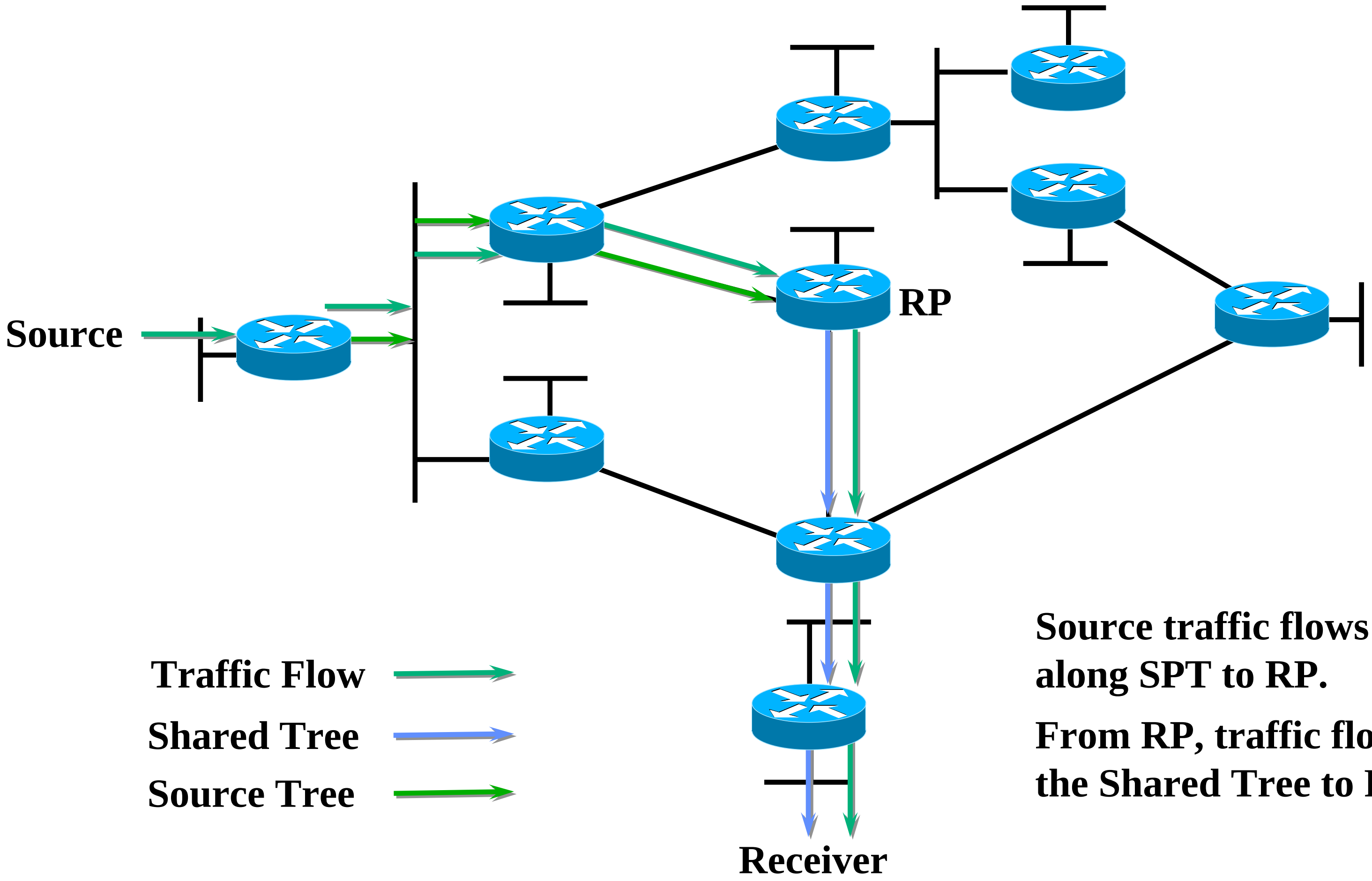


(S, G) traffic begins arriving at the RP via the Source tree.

RP sends a Register-Stop back to the first-hop router to stop the Register process.

Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

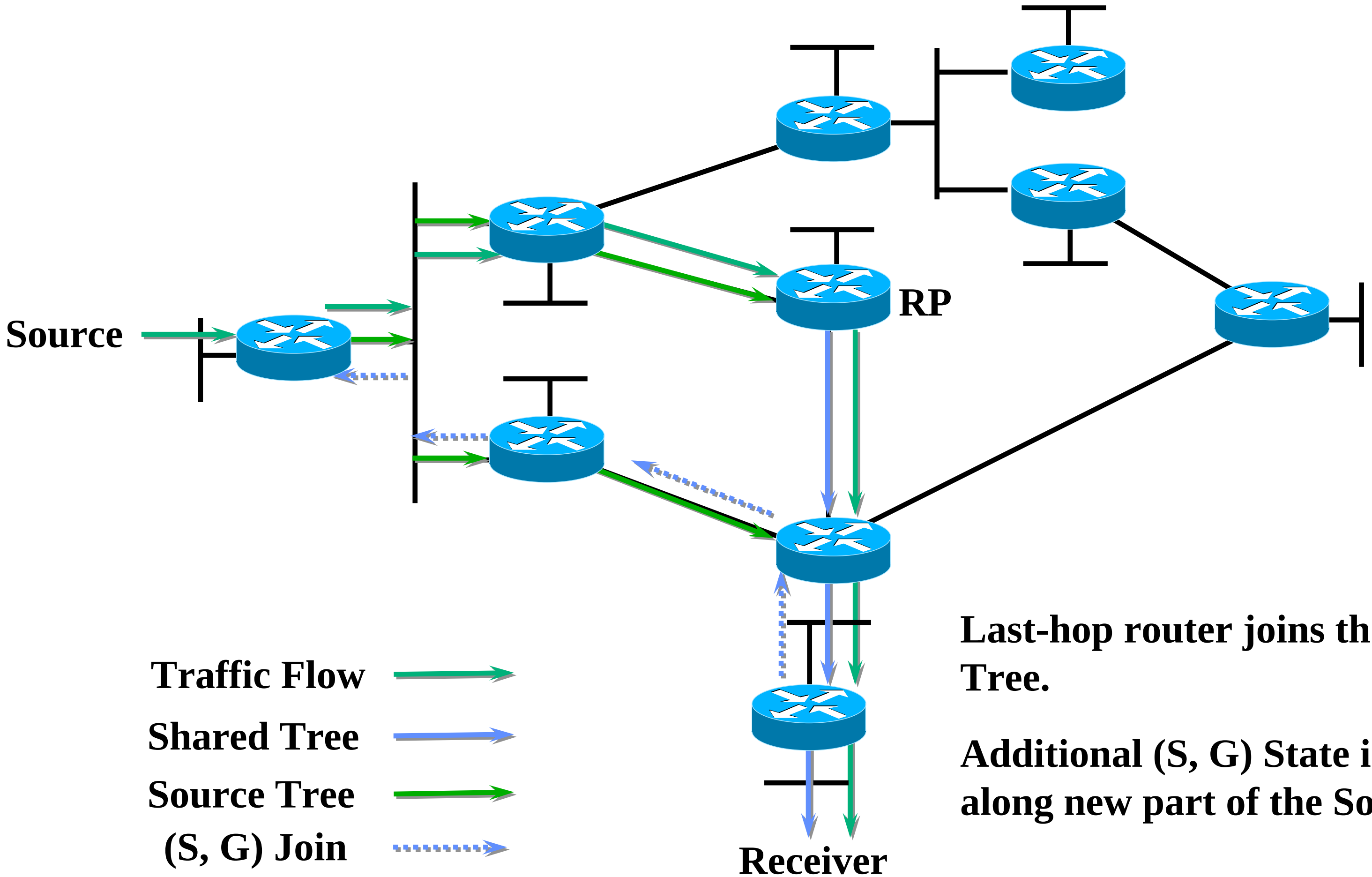
PIM SM Sender Registration



Source traffic flows natively along SPT to RP.

From RP, traffic flows down the Shared Tree to Receivers.

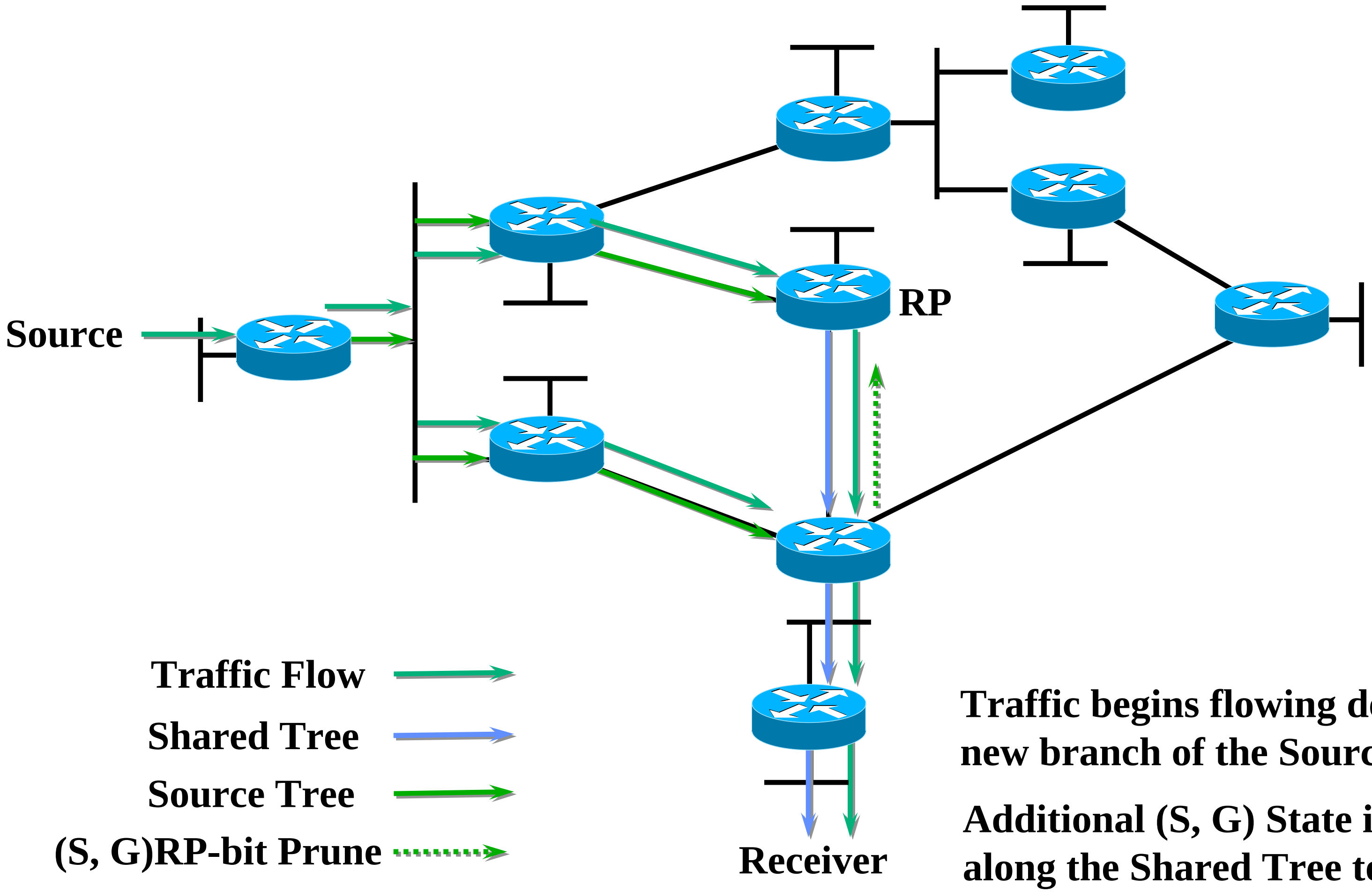
PIM SM Short Cut



Last-hop router joins the Source Tree.

Additional (S, G) State is created along new part of the Source Tree.

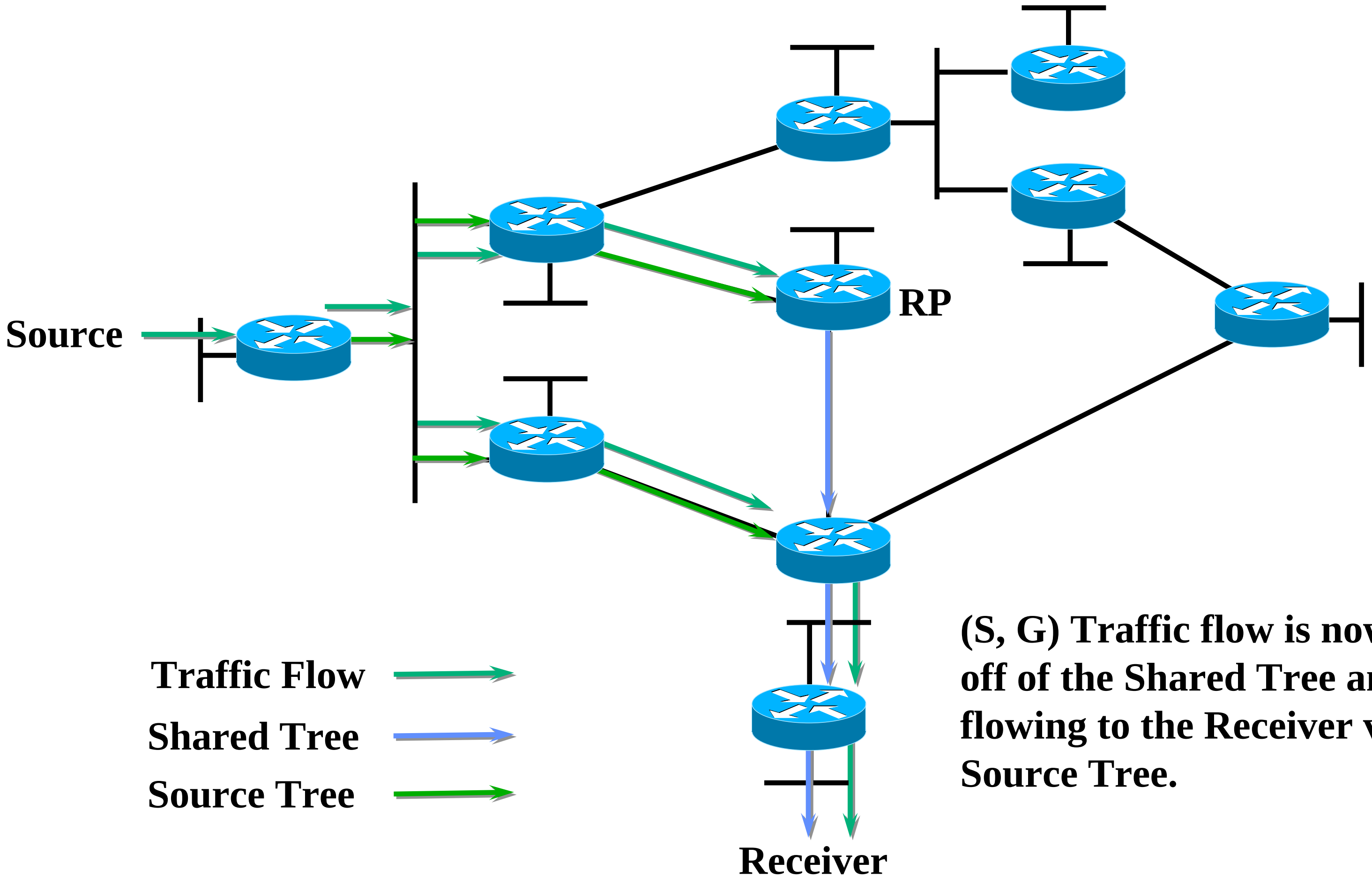
PIM SM Short Cut



Traffic begins flowing down the new branch of the Source Tree.

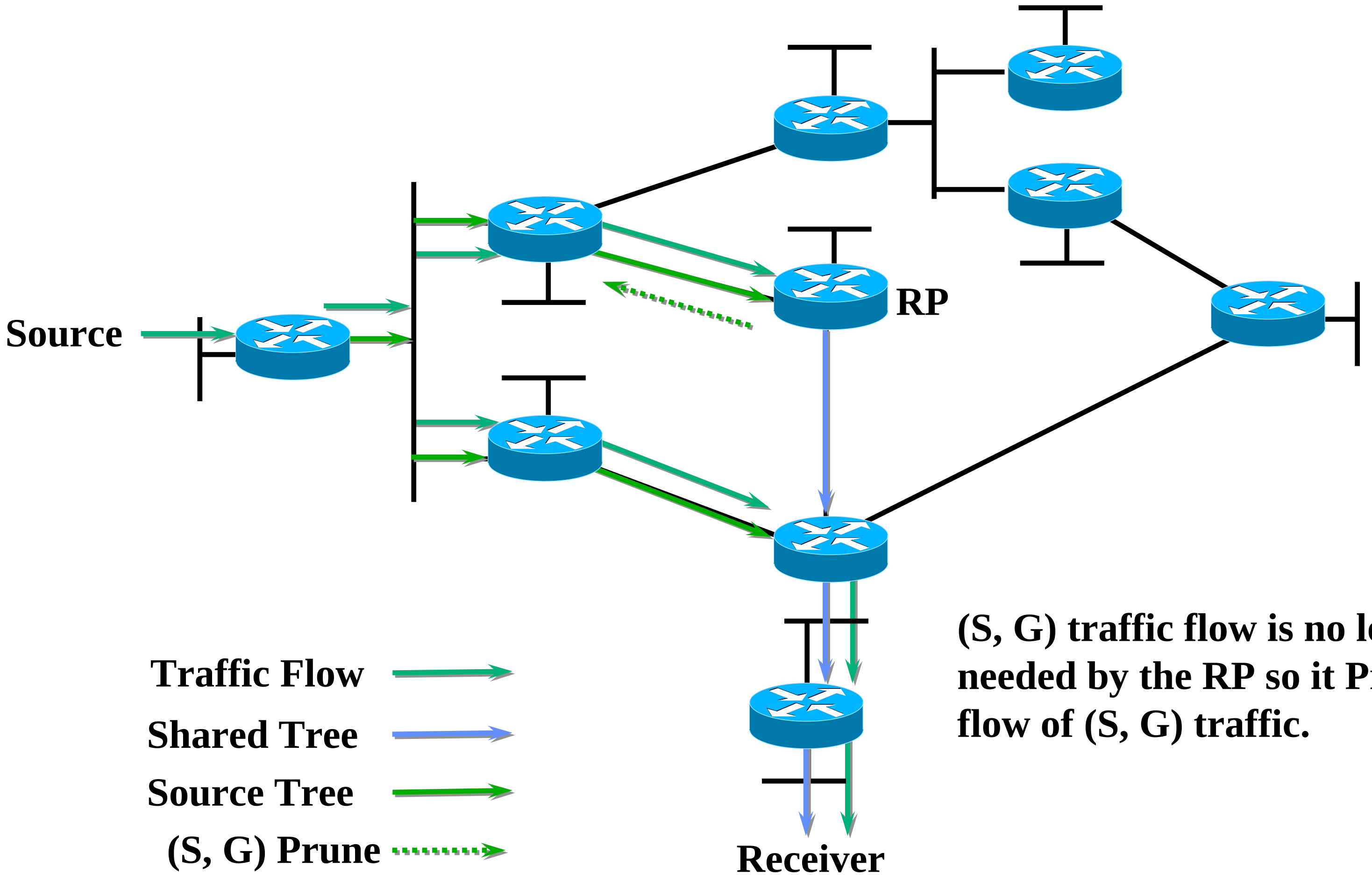
Additional (S, G) State is created along the Shared Tree to prune off (S, G) traffic.

PIM SM Short Cut



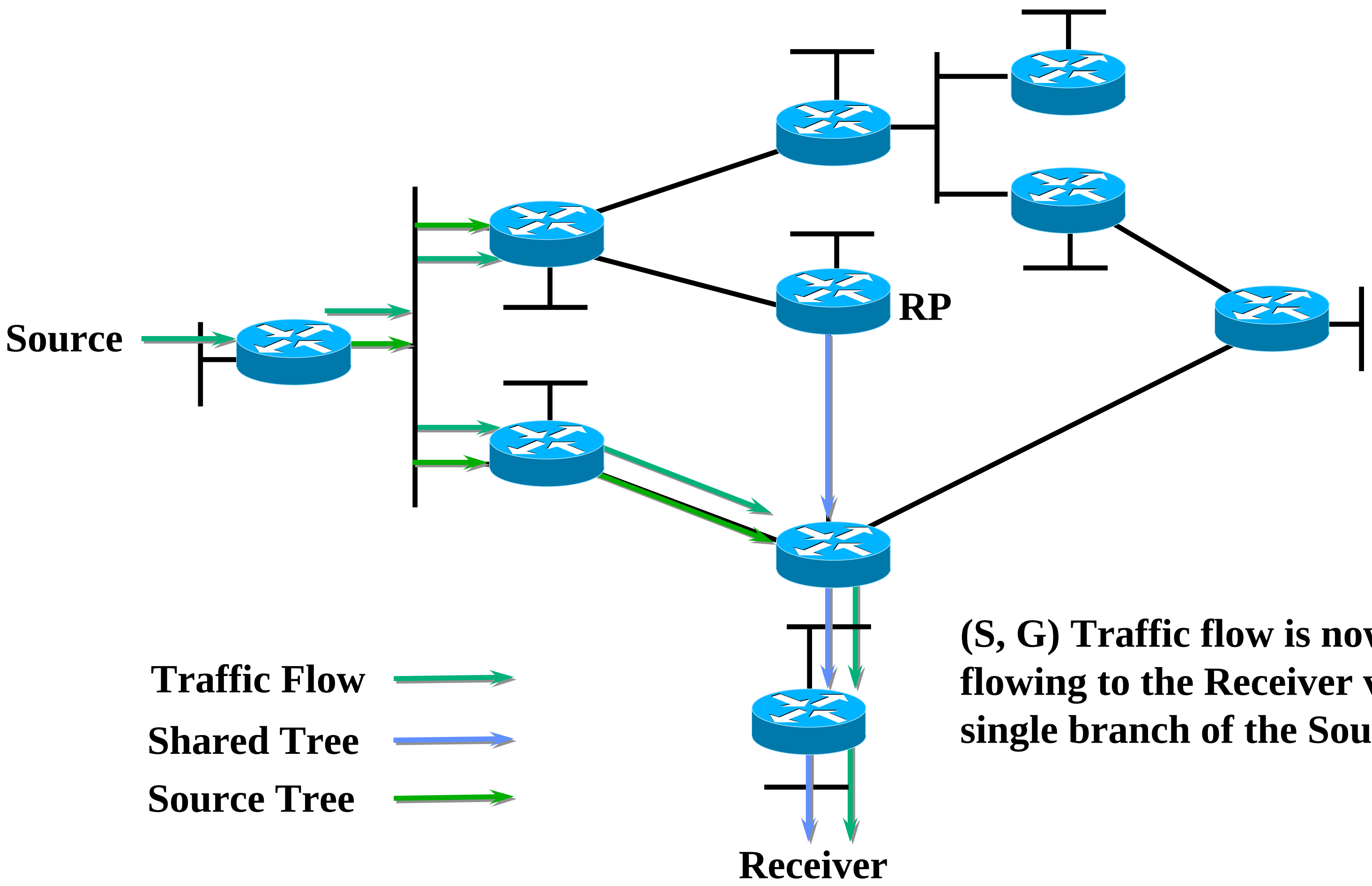
(S, G) Traffic flow is now pruned off of the Shared Tree and is flowing to the Receiver via the Source Tree.

PIM SM Short Cut



(S, G) traffic flow is no longer needed by the RP so it Prunes the flow of (S, G) traffic.

PIM SM Short Cut



(S, G) Traffic flow is now only flowing to the Receiver via a single branch of the Source Tree.

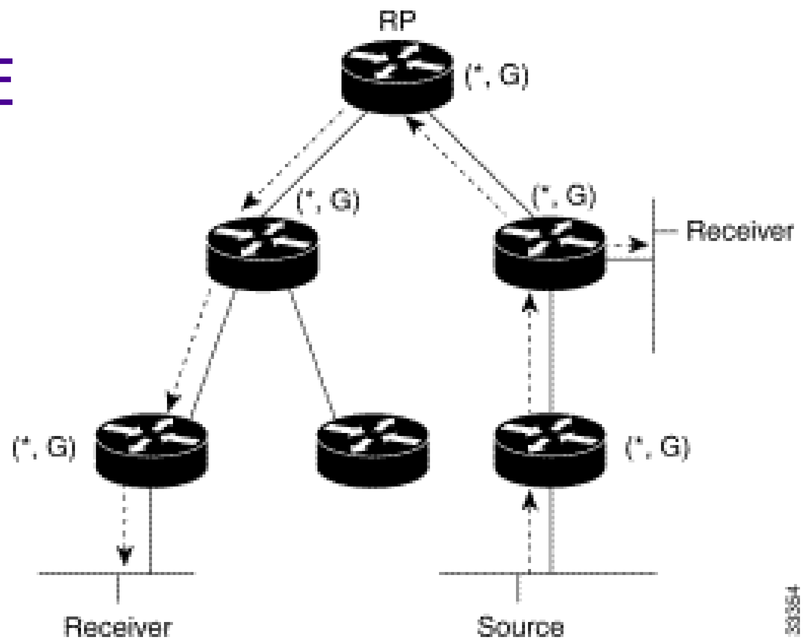
Bidirectional PIM

Mark Handley et. al.: RFC 5015

- Intra-domain protocol
- Selects (per Group) a “virtual” rendezvous point address (RPAs) – this may be an unused address on the rendezvous point link (RPL)
- Generates (RPA) a shared tree of designated forwarders (DFs): One router per link with best route to RPA
- Explores a domain by per group shared forwarding states: "NoInfo‟ or "Include"
- Decouples state management from data plane

Bidirectional PIM (2)

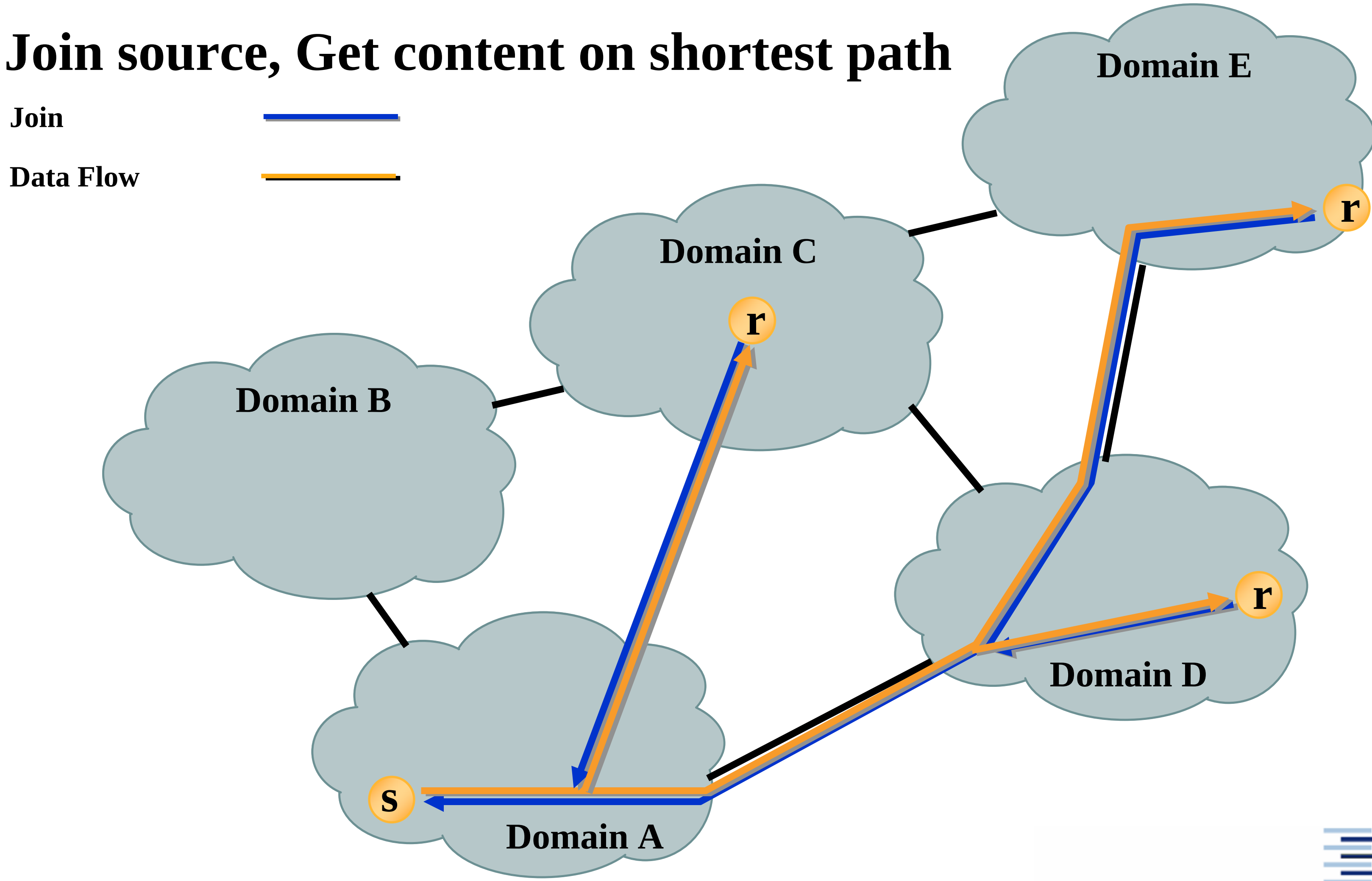
- Trees have RPA as virtual root, branch on RPL
- Group specific states are propagated by JOIN/PRUNE messages towards RPA
- Shared trees are operated bidirectionally
- Sources always forward upstream even without on-link receivers



Source Specific Multicast - SSM

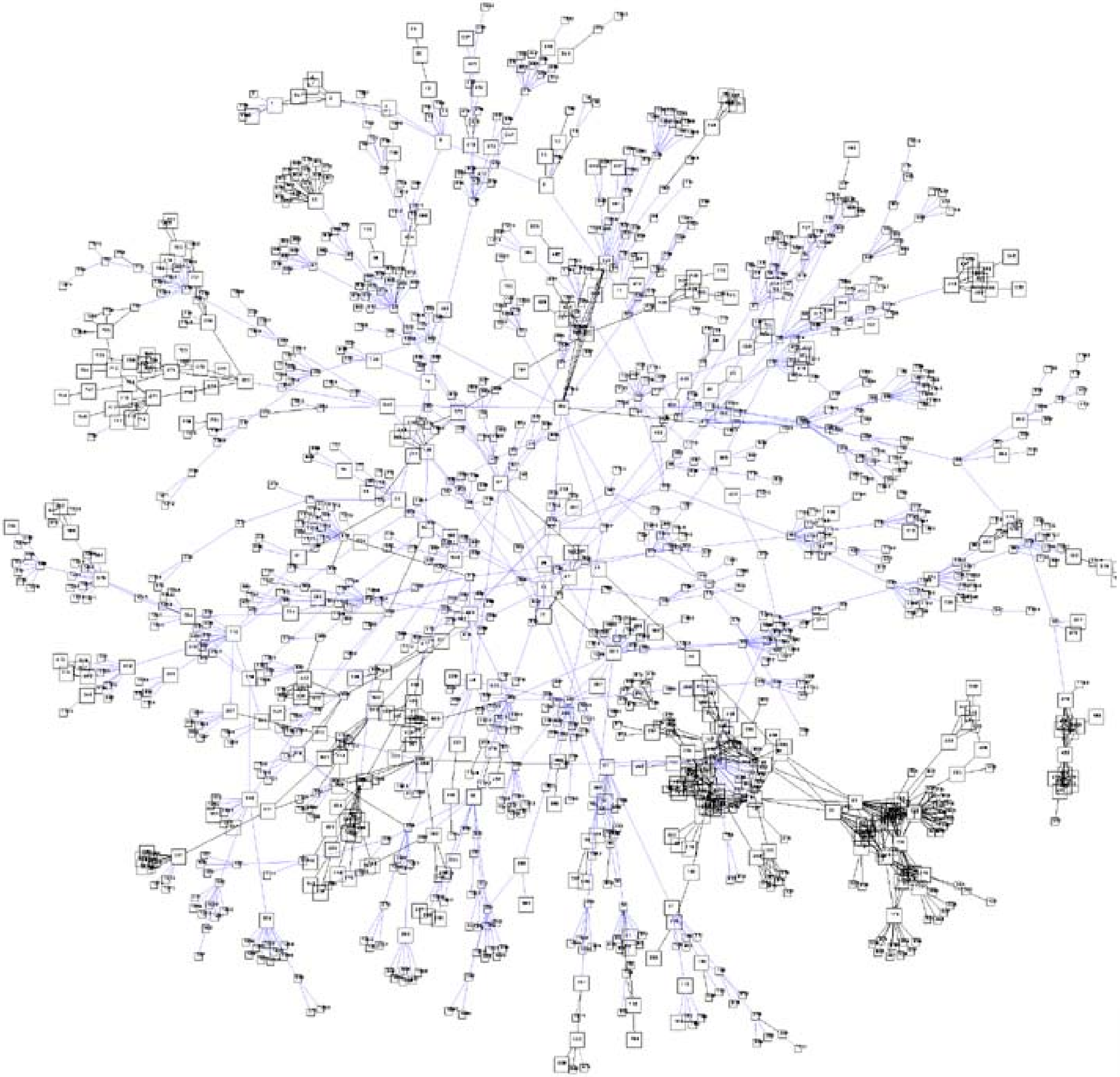
- Recently released (RFC 3569, RFC 4607 08/'06)
- Assumes source address known at receiver
  - Allows for source selection
  - Source discovery offline or via MSDP
- Receiver subscribes to (S,G) using IGMPv3/MLDv2
  - No state aggregation on shared trees
- Routing: PIM-SSM, a subset of PIM-SM (in RFC 4601)
  - Obsoletes rendezvous points & flooding
- Simpler, well suited for single source media broadcast or interdomain apps

SSM Routing



Join source, Get content on shortest path
Join
Data Flow
Domain A
Domain B
Domain C
Domain D
Domain E
s
r
r
r
r

MBone



Visualisation of Multicast Group

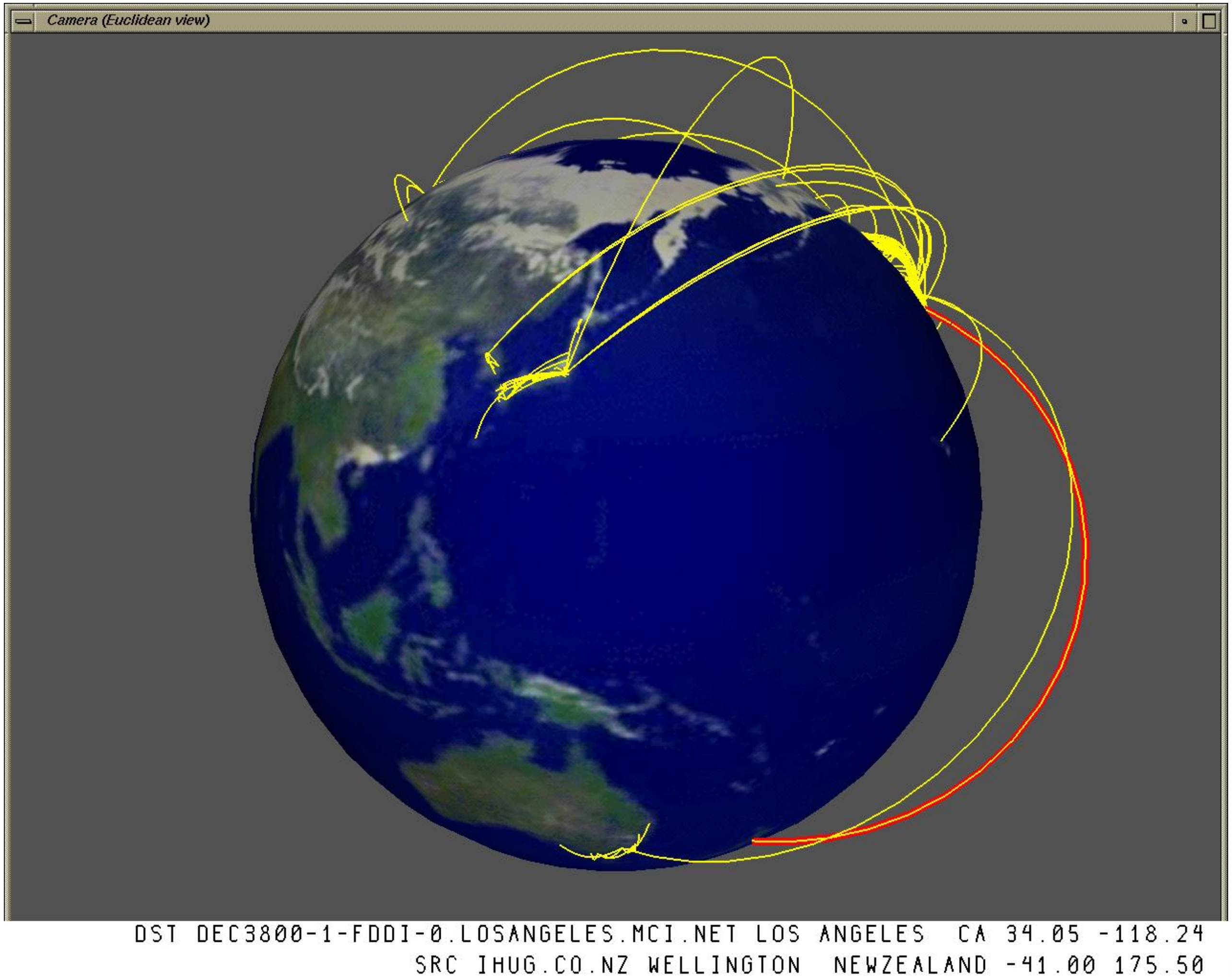


Image & Video by Tamara Munzer, Univ. of British Columbia

Efficiency of Multicast

- For m receivers
  - $L_M(m)$: Number of links in multicast SPT
  - $\langle L_U \rangle$: Average # of unicast hops between uniformly chosen end nodes, then clearly

$$L_M(m) < m * \langle L_U \rangle$$

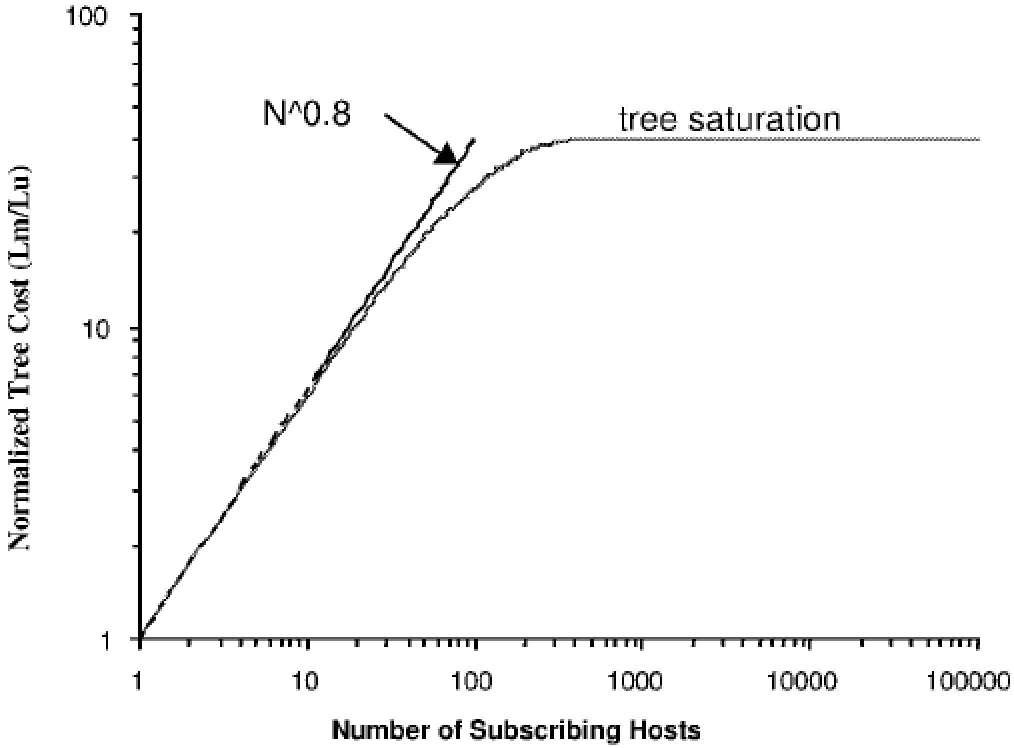
- Empirical Scaling Law (Chuang and Sirbu 1998/2001):

$$L_M(m) \approx \langle L_U \rangle * m^{0.8}$$

  - This means: multicast shortest path trees are of self-similar nature with many nodes of small, but few of higher degrees
  - Trees are shaped rather tall than wide

Chuang and Sirbu Scaling Law

- Empirical measurement on Internet & generated topologies
- Exponent found to be topology-independent
- Saturation due to exhaustive network exploration



Graphic from Chuang Sirbu (2001)

Efficiency of Multicast (2)

- Van Mieghem et al. (2001) proved that the Chuang and Sirbu scaling law cannot hold in general, but can be reasonably well approximated by

$$L_M(m) \approx < L_U > m^k, \; k = k(N) = \frac{var[L_U(N)]}{<L_U(N)>}$$

where *N* is the number of core nodes of the underlying network and $m \ll N$

- For the current Internet size ($N \approx 250.000$ core nodes) and moderate receiver numbers $m \ll N$:

$k \approx 0.8.$

Properties of Shortest Path Trees

- Assume: m multicast receivers are uniformly chosen out of N network nodes*, then

  If the link weights are iid., exponentially with mean 1, the Shortest Path Tree is a Uniform Recursive Tree

  - URTs are well studied self-similar trees
  - Relevant quantities can be derived analytically: Average hopcount, path weights, stability …
  - Allows to answer á priori deployment questions, e.g. cost efficiency of multicast …

* This assumption has been theoretically and empirically justified, cf. Van Mieghem 2006

Markov Discovery : Uniform Recursive Trees

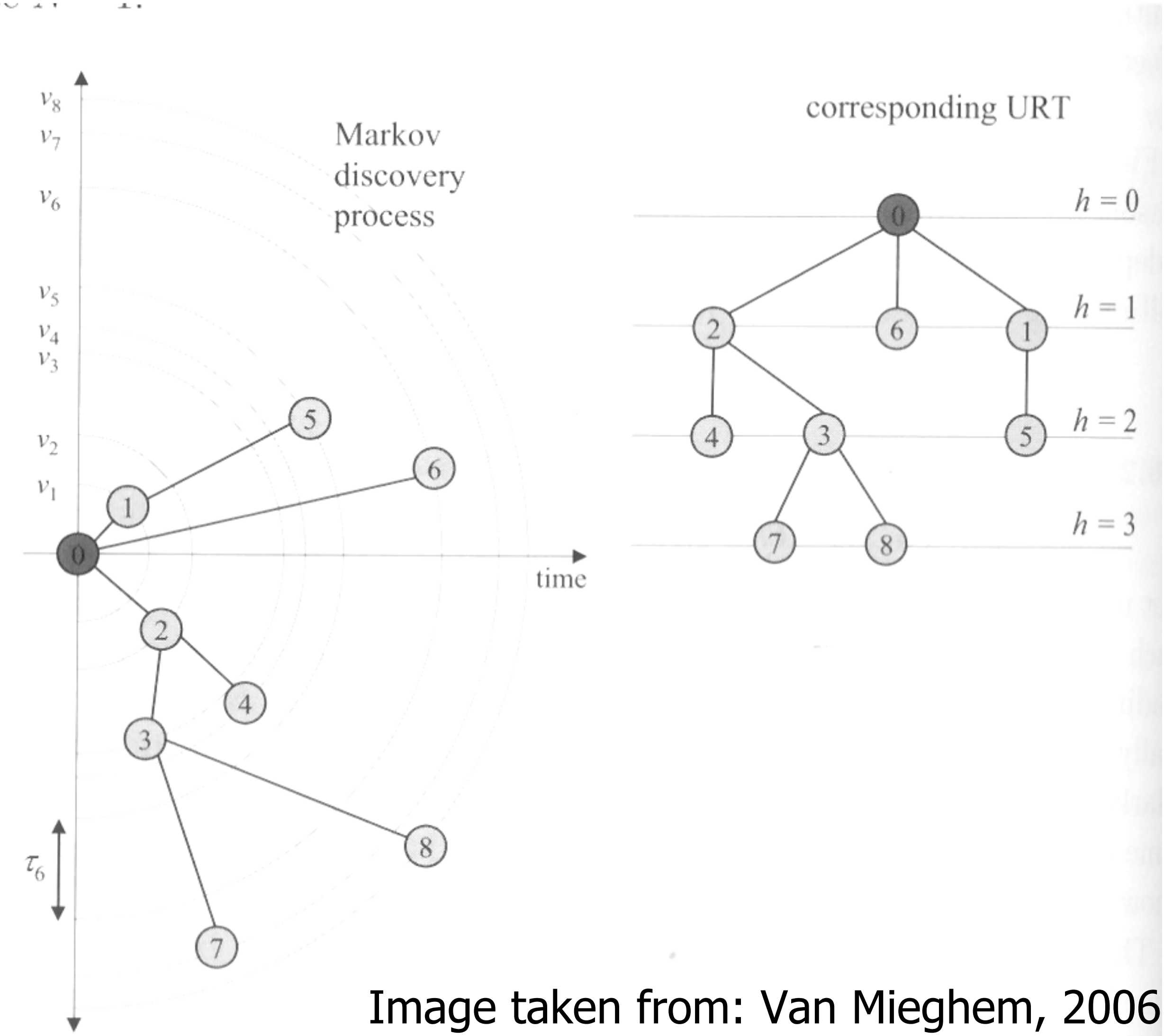


Image taken from: Van Mieghem, 2006

IP Mcast Deployment Issues

- Complexity versus Performance Efficiency
  - IP Multicast most efficient, but burdens infrastructure
- Provider Costs
  - Provisioning of knowledge, router capabilities & maintenance, Interdomain mcast routing problem
- Security
  - ASM simplifies DDoS-attacks
- End-to-End Design Violation?
  - Service complexity objects implementation at lower layer

Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

QoS for Multicasting

Resource ReSerVation Protocol (RSVP)

- RFC 2205
- Destination oriented Reservations
  - Sender pushes periodically PATH messages
  - Receiver answers with RESV packets
  - Router interpret these along the paths
- Soft-State-Concept: States time out
- Sender remains unsynchronised

Further Reading

- R. Wittmann, M. Zitterbart: *Multicast Communication,* Morgan Kaufmann, 2001
- www.ipmulticast.com
- www.rfc-editor.org
- ftp://ftpeng.cisco.com/ipmulticast.html
- J. Chuang and M. Sirbu: *Pricing Multicast Communication: A Cost-Based Approach*, Telecommunication Systems 17(3), 281 – 297, 2001.
- P. Van Mieghem: *Performance Analysis of Communication Networks and Systems*, Cambridge University Press, Cambridge, 2006.
- P. Van Mieghem, G. Hooghiemstra and R. van der Hofstad: *On the Efficiency of Multicast*, IEEE/ACM Trans. Netw. 9(6), pp. 719-732, 2001.