

Introduction to Multicast Protocols and Applications

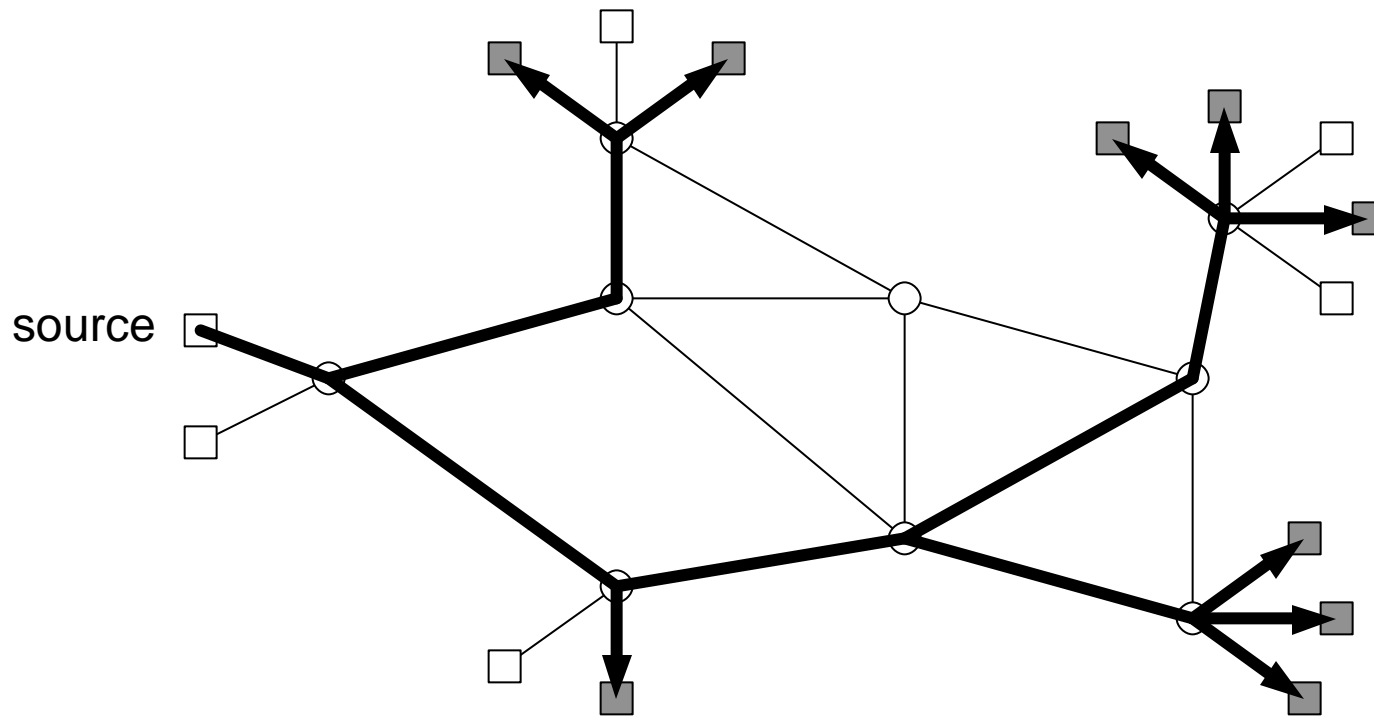
Kevin Almeroth

UC--Santa Barbara

<http://www.cs.ucsb.edu/~almeroth>

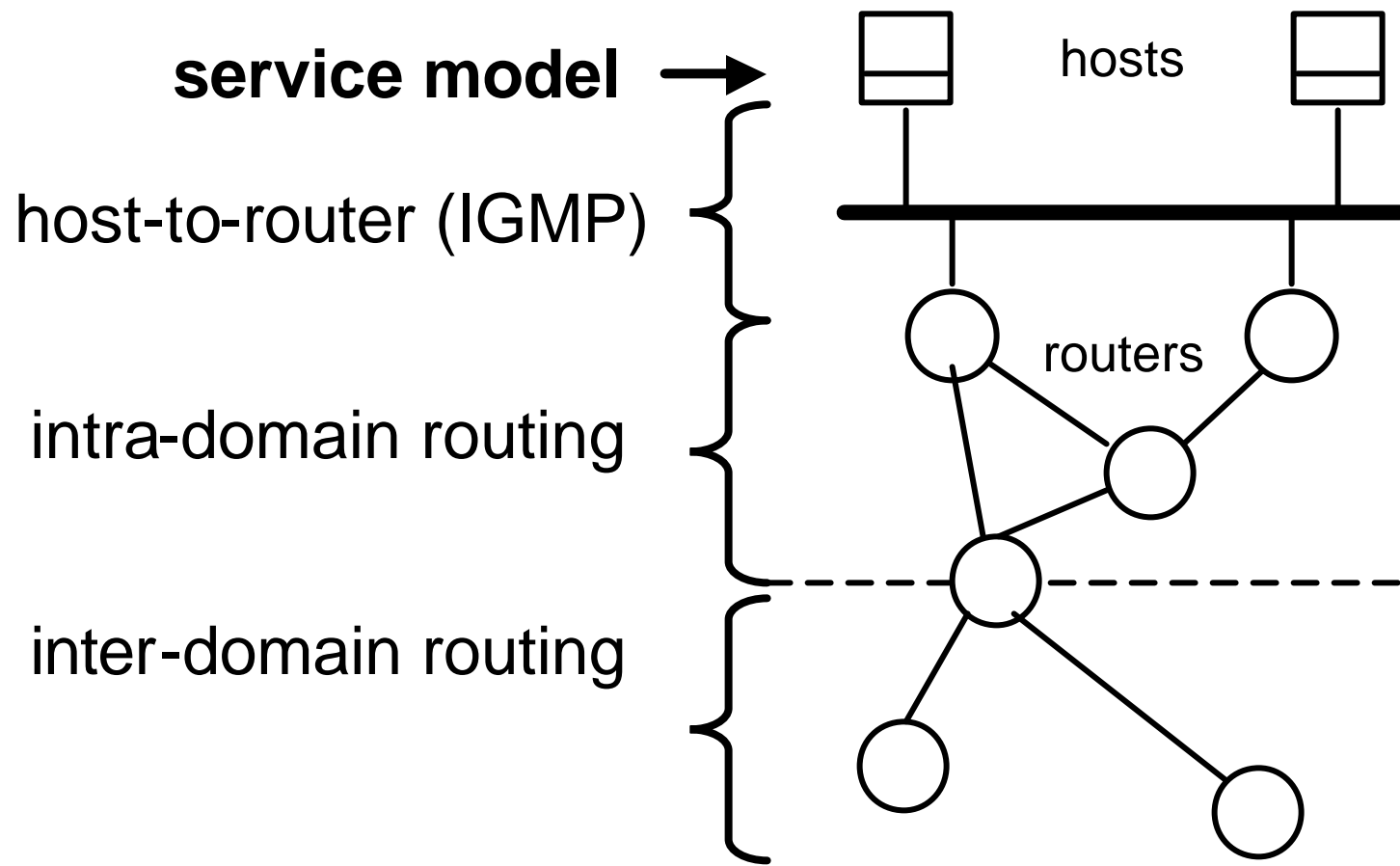
almeroth@cs.ucsb.edu

Multicast Routing (and Functions)



routing (path determination) [but in the reverse direction]
packet forwarding and possibly replication
dynamic membership---path pruning/grafting

Components of the IP Multicast Architecture



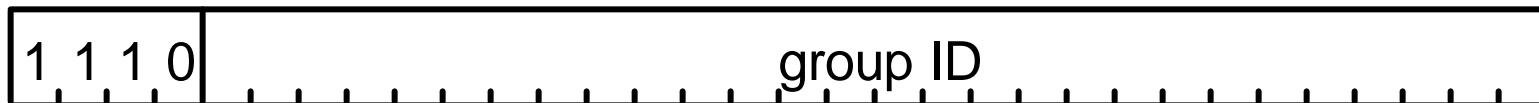
Original IP Multicast Service Model (RFC-1112)

- each group identified by a single IP address
- groups may be of any size
- members of groups may be located anywhere in the Internet
- members of groups can join and leave at will
- senders need not be members

analogy: each multicast address is like a radio frequency, on which anyone can transmit, and to which anyone can tune-in.

IP Multicast Addresses

Class D IP addresses:



in “dotted decimal” notation: 224.0.0.0 — 239.255.255.255

two administrative categories:

- “well-known” multicast addresses, assigned by IANA
- “transient” multicast addresses, assigned and reclaimed dynamically, e.g., by “sdr”, GLOP, etc

IP Multicast Service — Sending

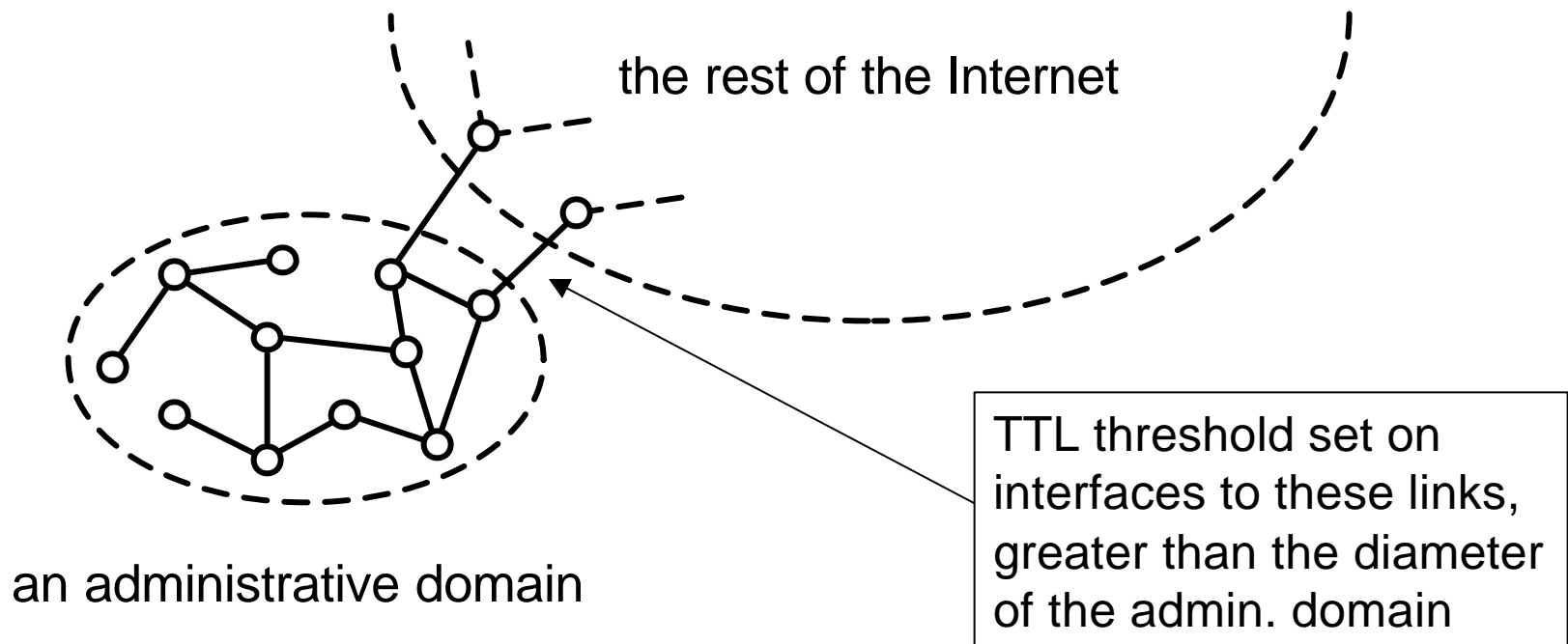
- **uses normal IP-Send operation**, with an IP multicast address specified as the destination
- must provide sending application a way to:
 - specify outgoing network interface, if >1 available
 - specify IP time-to-live (TTL) on outgoing packet
 - enable/disable loopback if the sending host is a member of the destination group on the outgoing interface

IP Multicast Service — Receiving

- two new operations:
 - Join-IP-Multicast-Group (group-address, interface)
 - Leave-IP-Multicast-Group (group-address, interface)
- receive multicast packets for joined groups via normal IP-Receive operation

Multicast Scope Control: TTL Boundaries

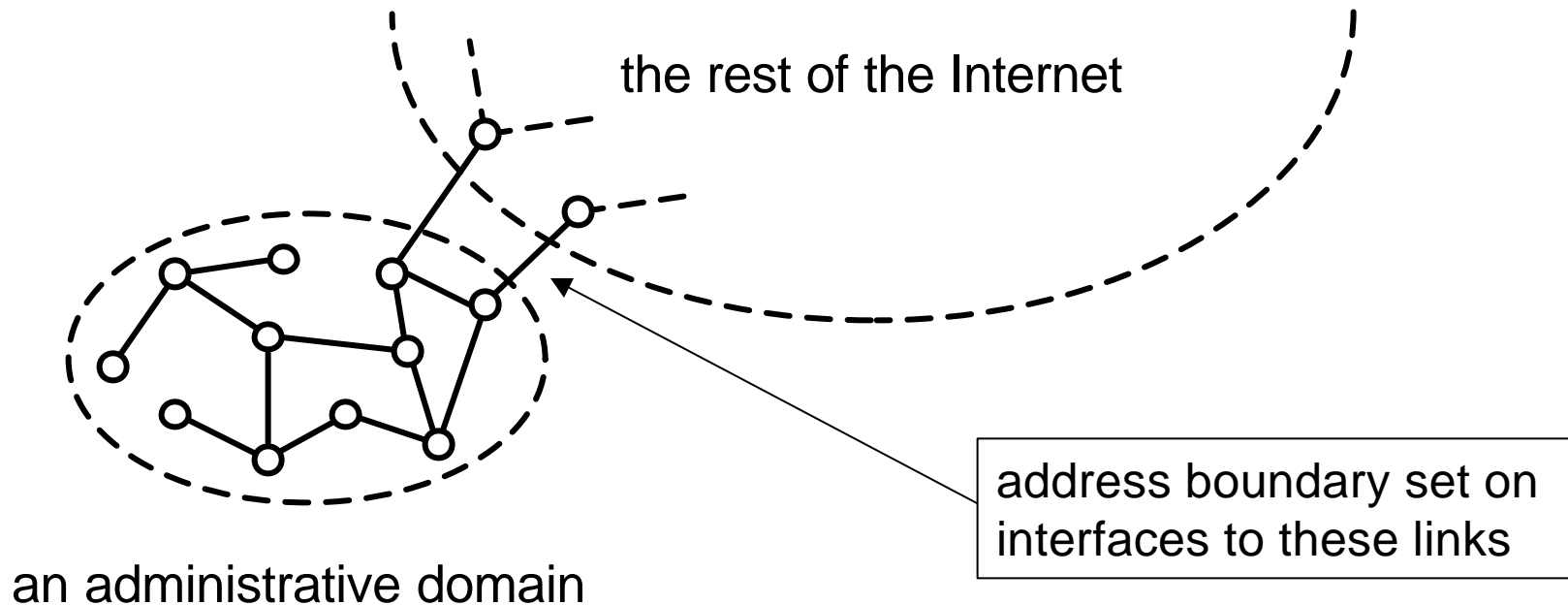
to keep multicast traffic within an administrative domain, e.g., for privacy or resource reasons



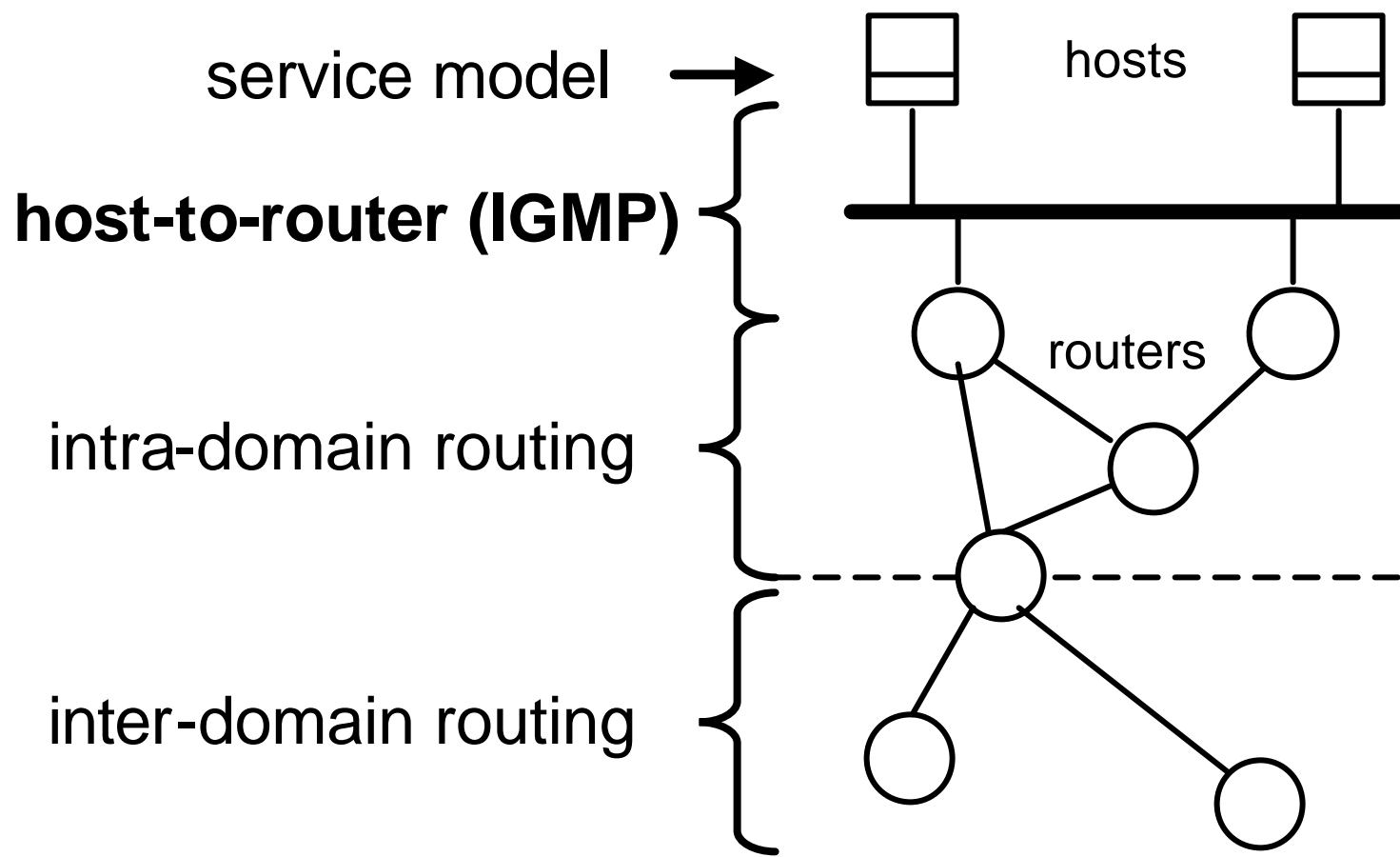
Multicast Scope Control: Administratively-Scoped Addresses

a better way to keep multicast traffic within an administrative domain (new since RFC 1112)

- uses address range 239.0.0.0 — 239.255.255.255
- supports overlapping (not just nested) domains



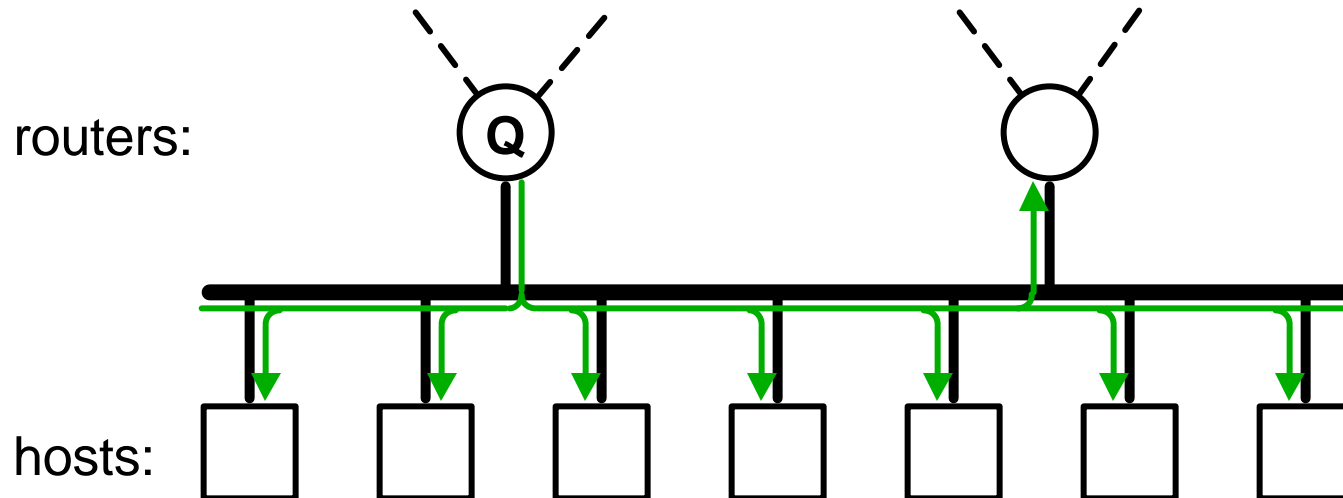
Components of the IP Multicast Architecture



Internet Group Management Protocol (IGMP)

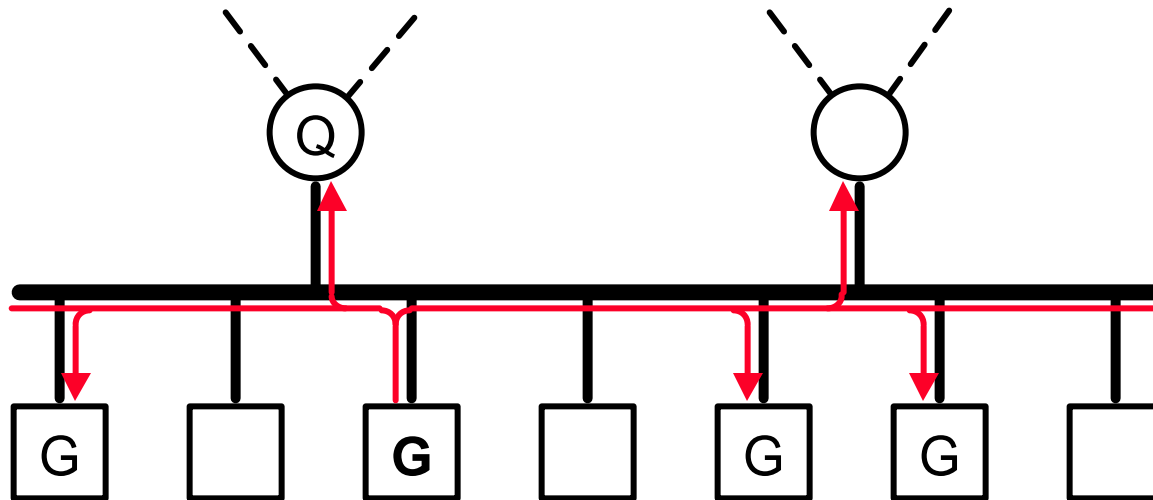
- the protocol by which hosts report their multicast group memberships to neighboring routers
 - RFC-1112 specifies version 1, the original Standard
 - RFC-2236 specifies version 2, the most widely used
- occupies similar position and role as ICMP in the TCP/IP protocol stack

How IGMP Works



- on each link, one router is elected the “querier”
- querier periodically sends a Membership Query message to the all-systems group (224.0.0.1), with TTL = 1
- on receipt, hosts start random timers (between 0 and 10 seconds) for each multicast group to which they belong

How IGMP Works (cont.)



- when a host's timer for group G expires, it sends a Membership Report to group G, with TTL = 1
- other members of G hear the report and stop their timers
- routers hear all reports, and time out non-responding groups

How IGMP Works (cont.)

- note that, in normal case, only one report message per group present is sent in response to a query
 - (routers need not know who all the members are, only that members exist)
- query interval is typically 60—90 seconds
- when a host first joins a group, it sends one or two immediate reports, instead of waiting for a query

IGMP Version 2

- changes from version 1:
 - new message and procedures to reduce “leave latency”
 - standard querier election method specified
 - version and type fields merged into a single field
- backward-compatible with version 1
- is currently a Proposed Standard
- widely implemented already

IGMP Version 3

- still at the design stage, but advancing rapidly
- changes from version 2:
 - extension of service interface and protocol to enable hosts to:
 - listen to only a specified set of hosts sending to a group
 - listen to all but a specified set of hosts sending to a group
 - additional protocol to inform a source host when no one is listening, to suppress unnecessary first hop transmission
- to be backward-compatible with versions 1 & 2

IP Multicast Meets Bridged LANs

- LANs are no longer just rings and “yellow hoses”!
- classic Ethernet bridges forward all multicasts to all segments, in case any receivers are there.
- current ways to do better:
 - IGMP Snooping

IGMP Snooping

- bridges look inside received multicast frames for:
 - IGMP Reports, to learn in which direction(s) group members reside
 - IGMP Queries, DVMRP Probes, MOSPF Hellos, PIM Hellos to learn in which direction(s) multicast routers reside
- multicast data packets forwarded only towards group members and multicast routers.
- IGMP Report suppression done “per branch” rather than “per LAN”

Problems with IGMP snooping

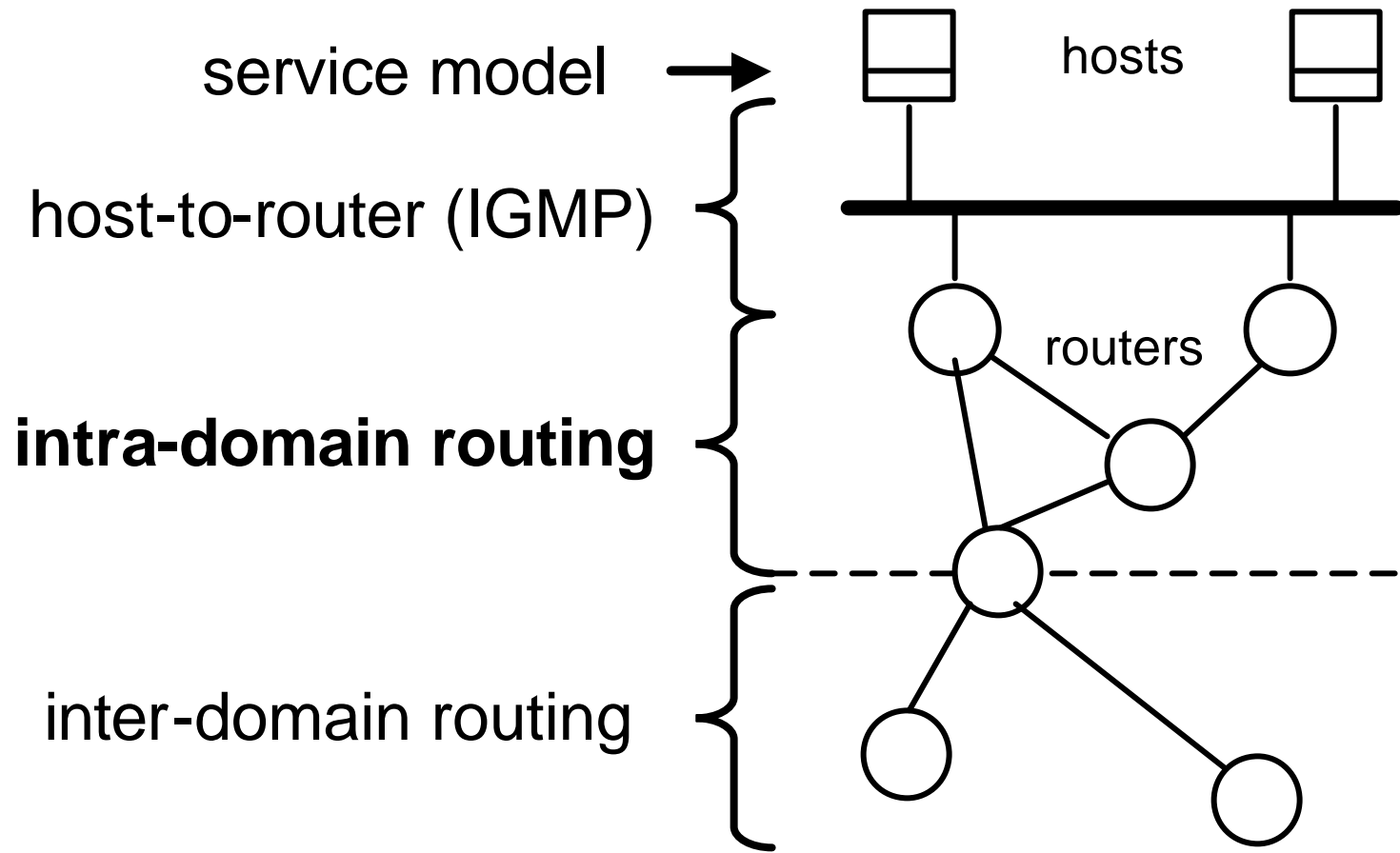
- doesn't work for non-IP multicasts
- stops working if new multicast routing protocol deployed
- performance cost of snooping inside of every multicast frame

For More Information on IGMP

- Specifications
 - IGMPv1: RFC 1112
 - IGMPv2: RFC 2236
 - IGMPv3: draft-ietf-idmr-igmp-v3-*.txt
- WWW page
 - <http://www.ietf.org/html.charters/idmr-charter.html>
- Mailing list
 - Subscribe to: idmr-request@cs.ucl.ac.uk

Intra-Domain Multicast Routing Protocols

Components of the IP Multicast Architecture



The First Intra-Domain Routing Protocol: DVMRP

Distance-Vector Multicast Routing Protocol

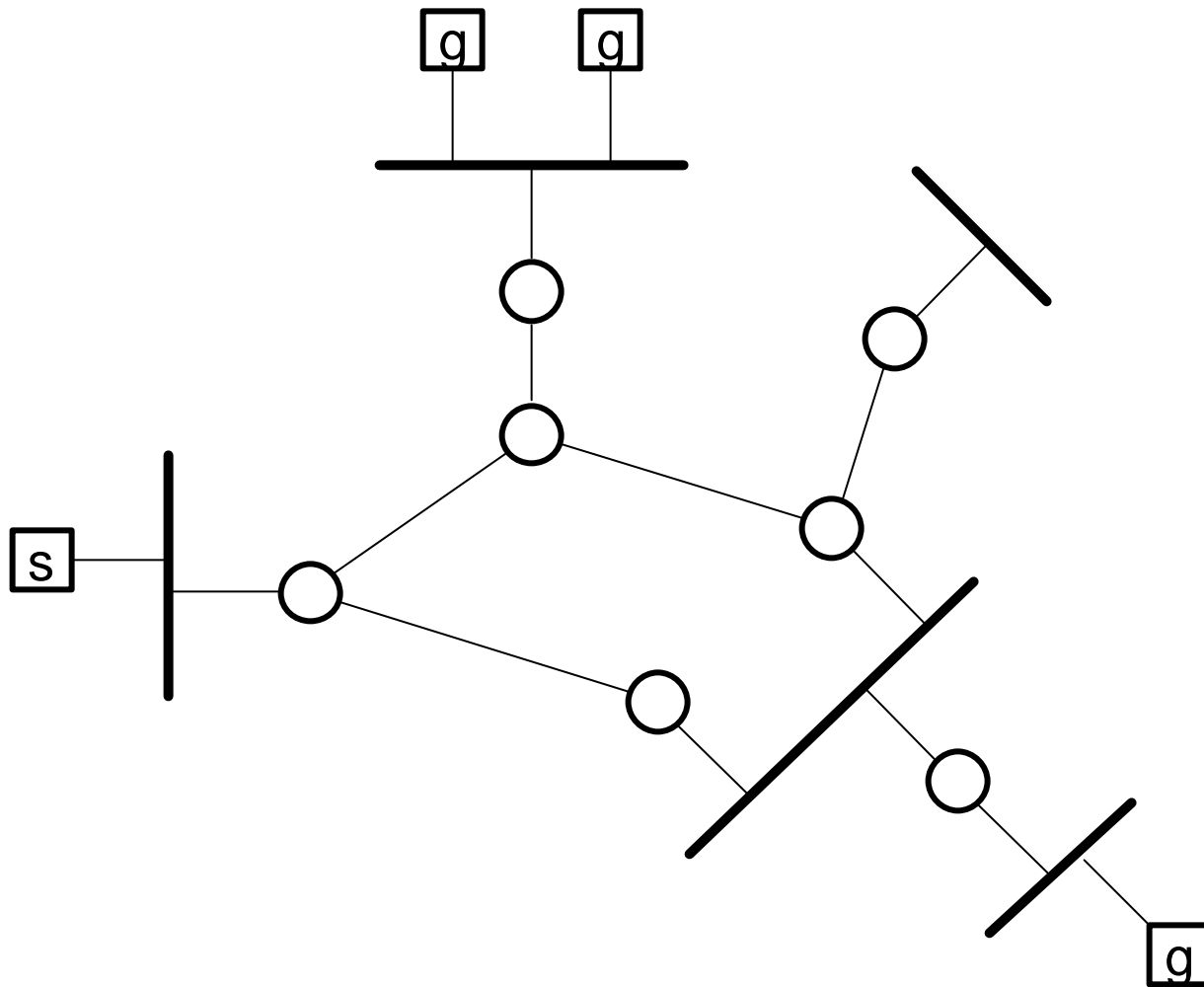
DVMRP consists of two major components:

- (1) a conventional distance-vector routing protocol (like RIP) which builds, in each router, a routing table like this:

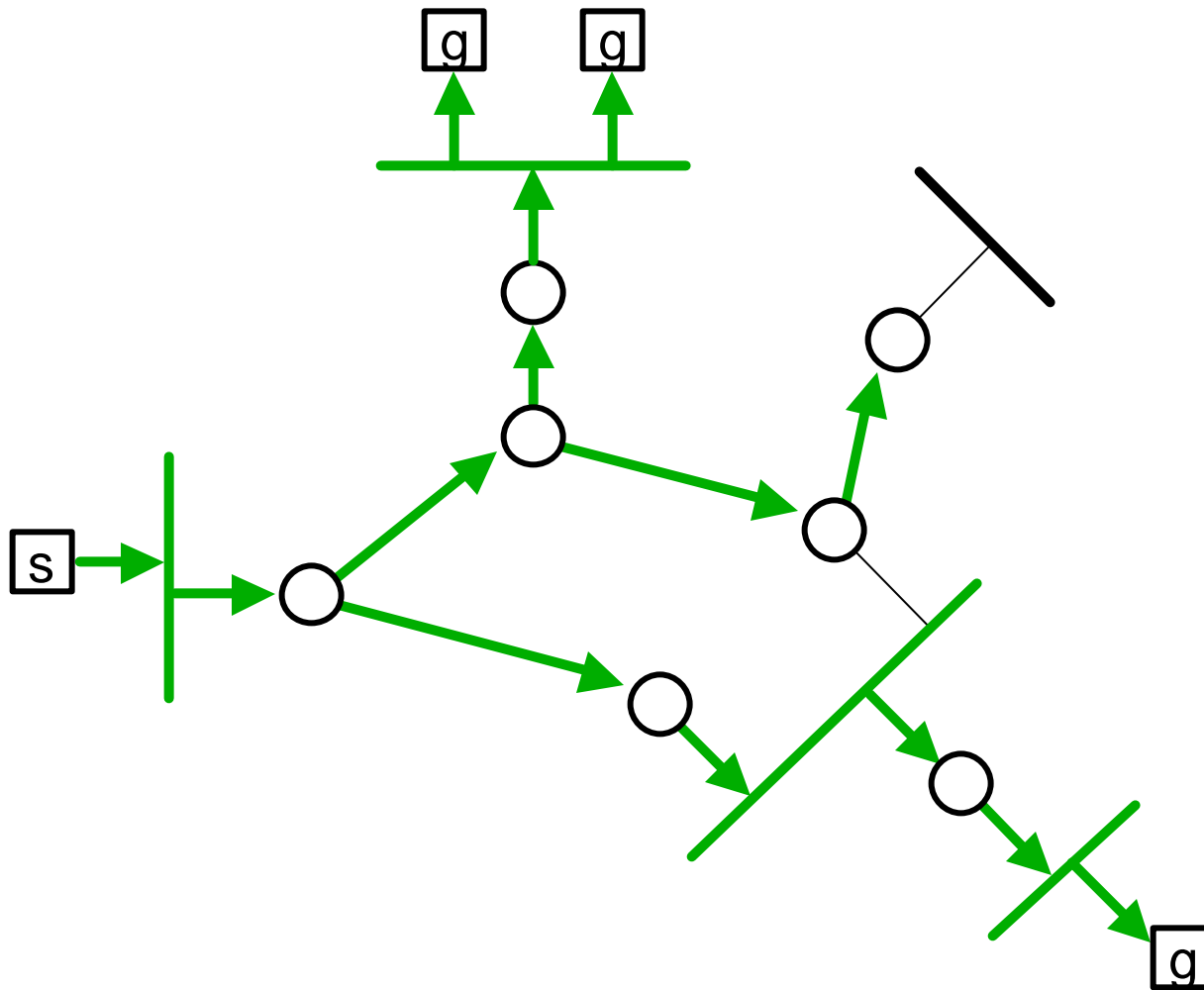
subnet	shortest dist	via interface
a	1	i1
b	5	i1
c	3	i2
...

- (2) a protocol for determining how to forward multicast packets, based on the routing table and routing messages of (1)

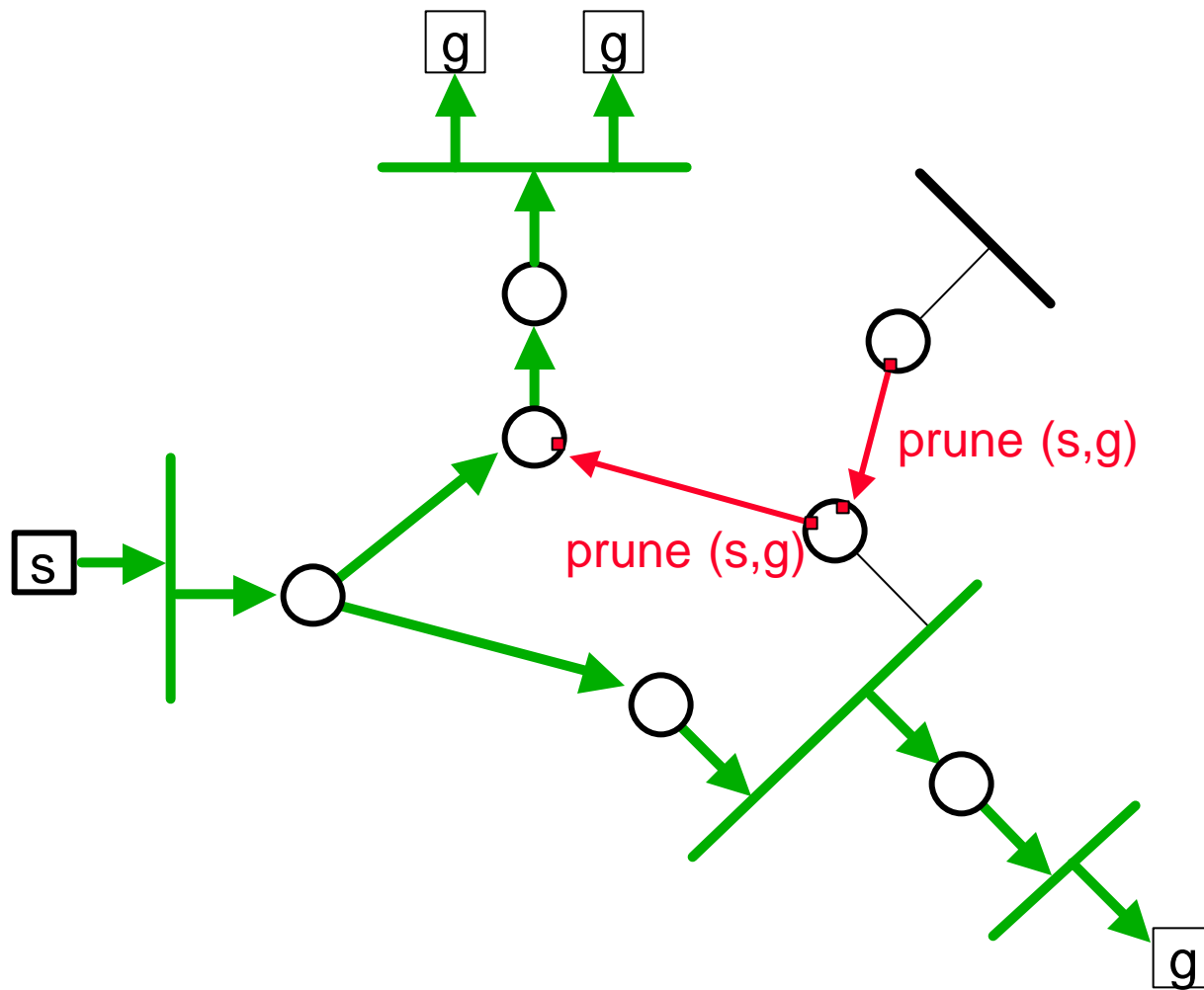
Example Topology



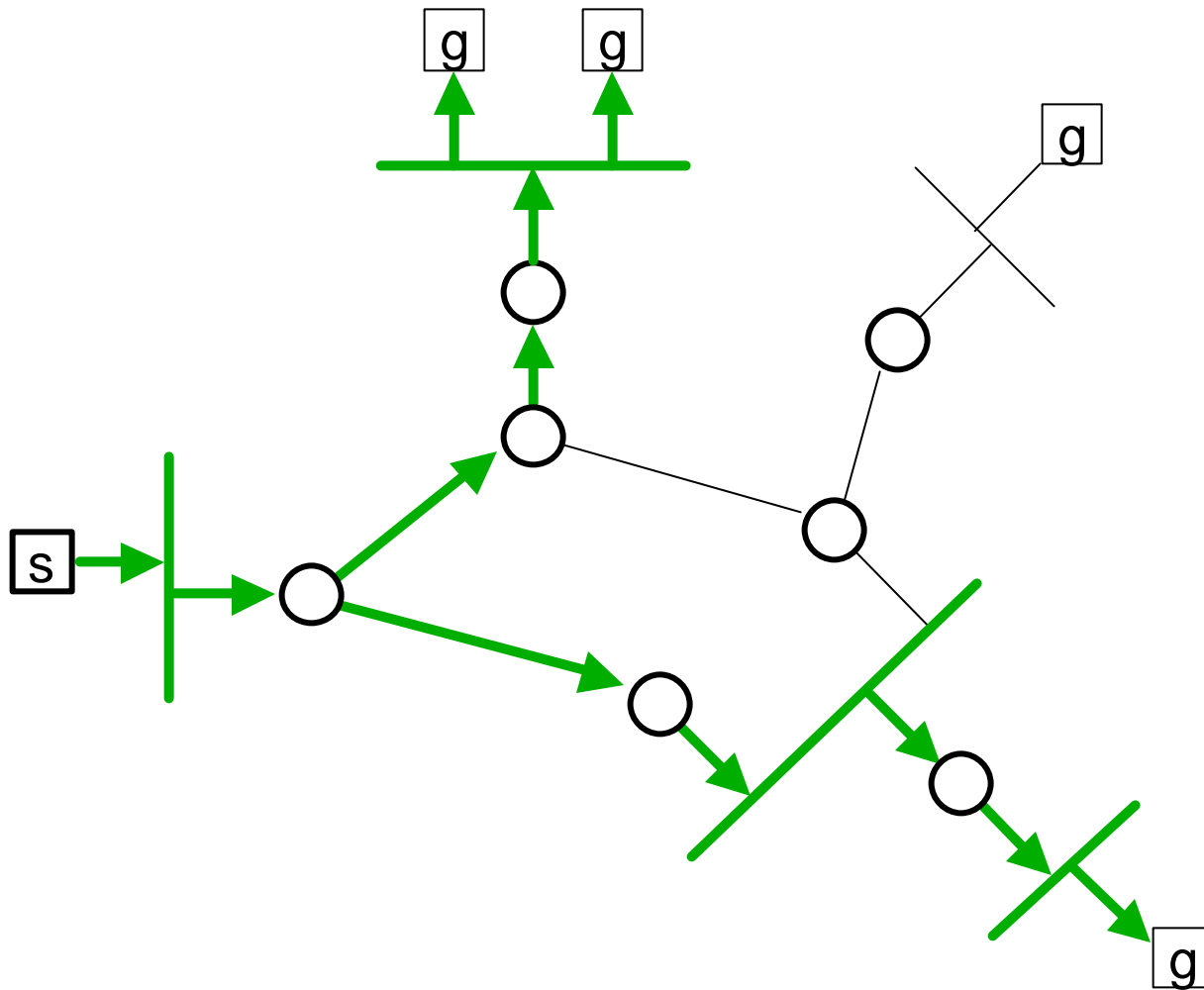
Phase 1: Truncated Broadcast



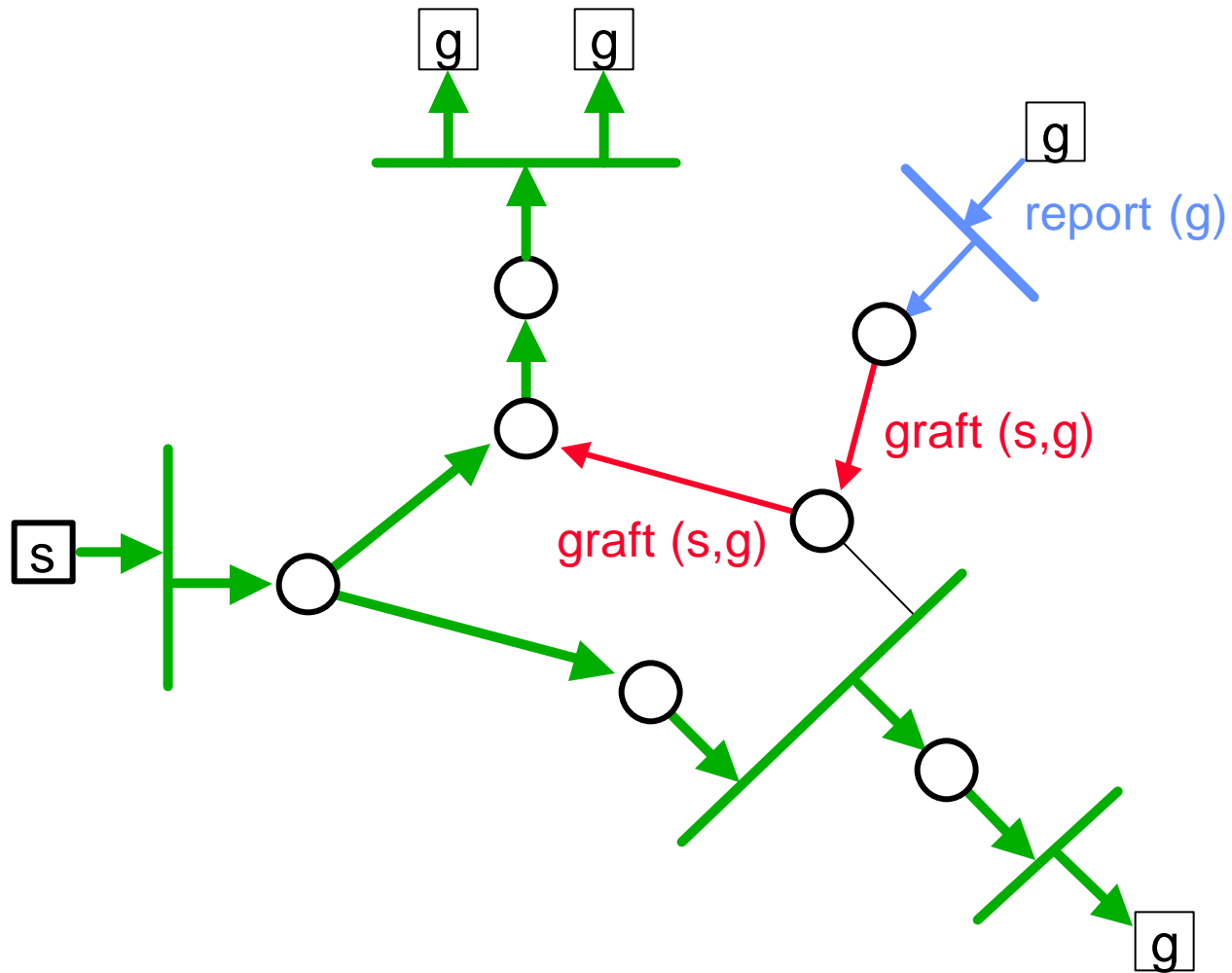
Phase 2: Pruning



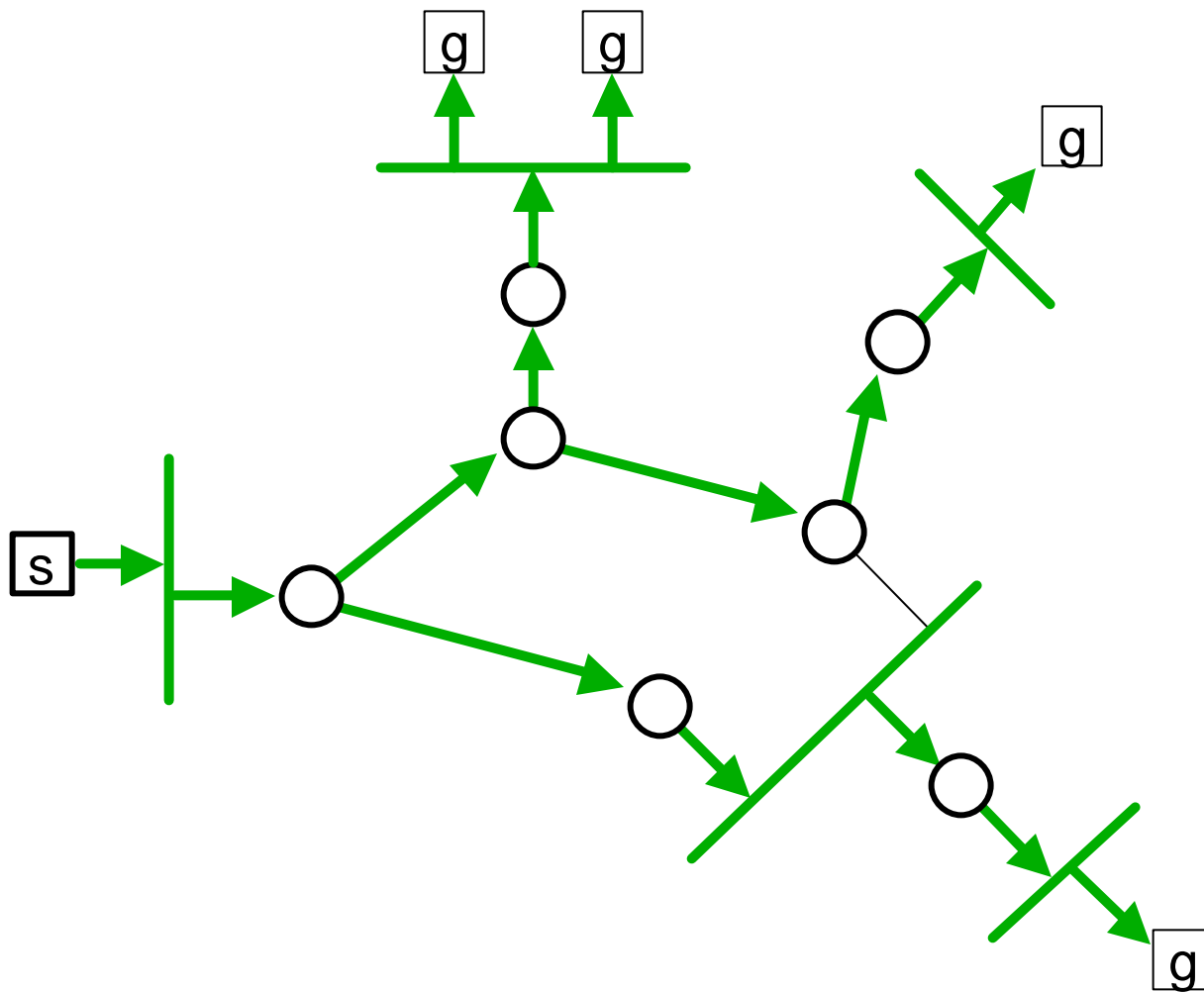
Steady State



Grafting on New Receivers



Steady State after Grafting



Distinguishing Properties of IP Multicast Routing Protocols

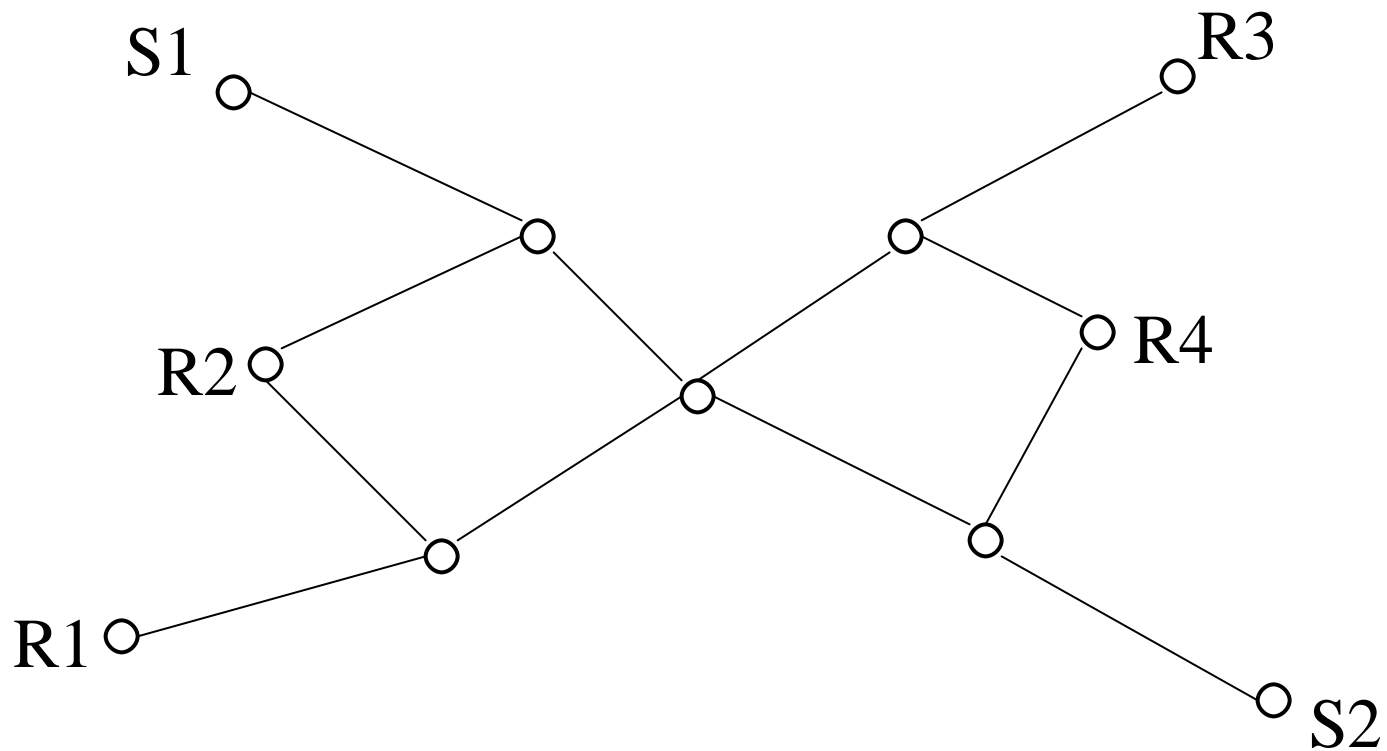
- (1) how delivery trees are established between multicast senders and receivers
 - broadcast data, then prune
 - broadcast membership
 - use one or more “meeting places”

Distinguishing Properties of IP Multicast Routing Protocols (cont.)

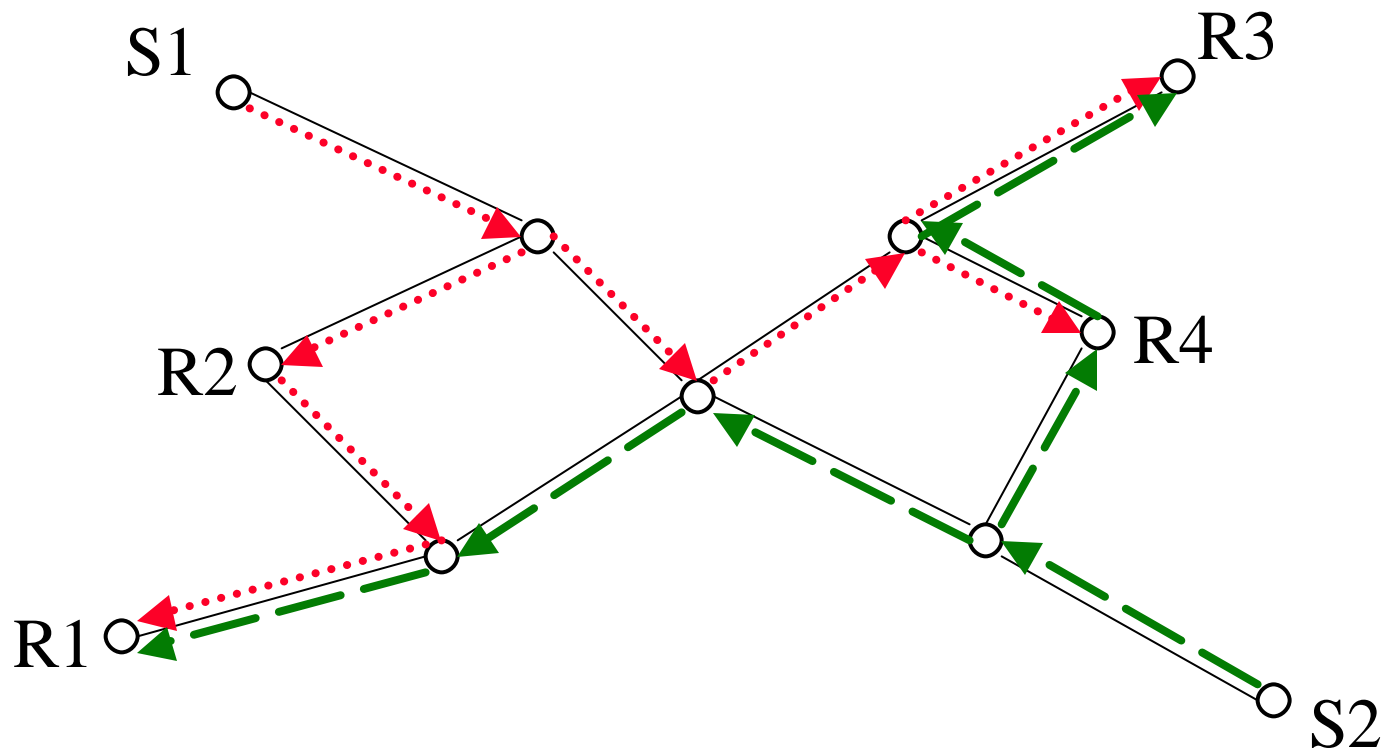
(2) types of delivery trees

- unidirectional, per-source, per-group trees
- unidirectional, per-group trees, shared by all sources
- bidirectional, per-group trees, shared by all sources

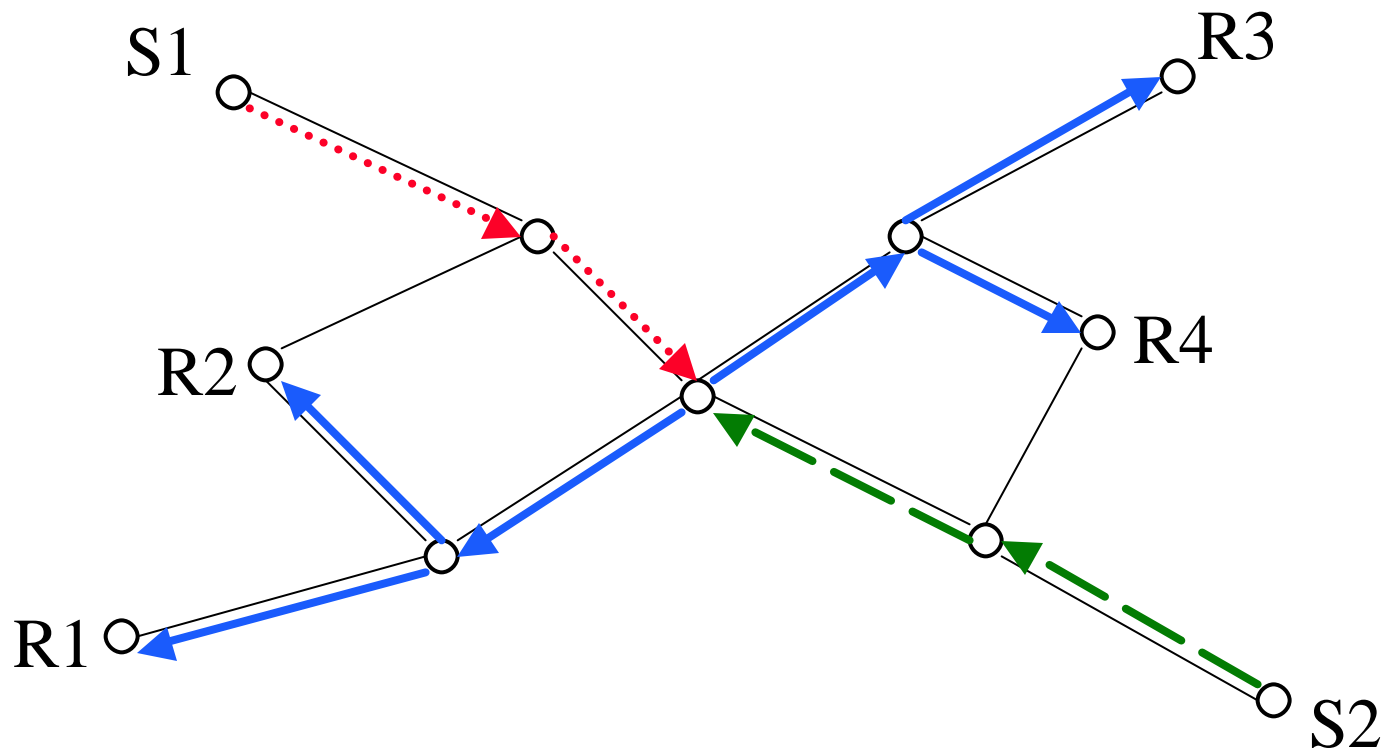
Topology to Illustrate Types of Delivery Trees



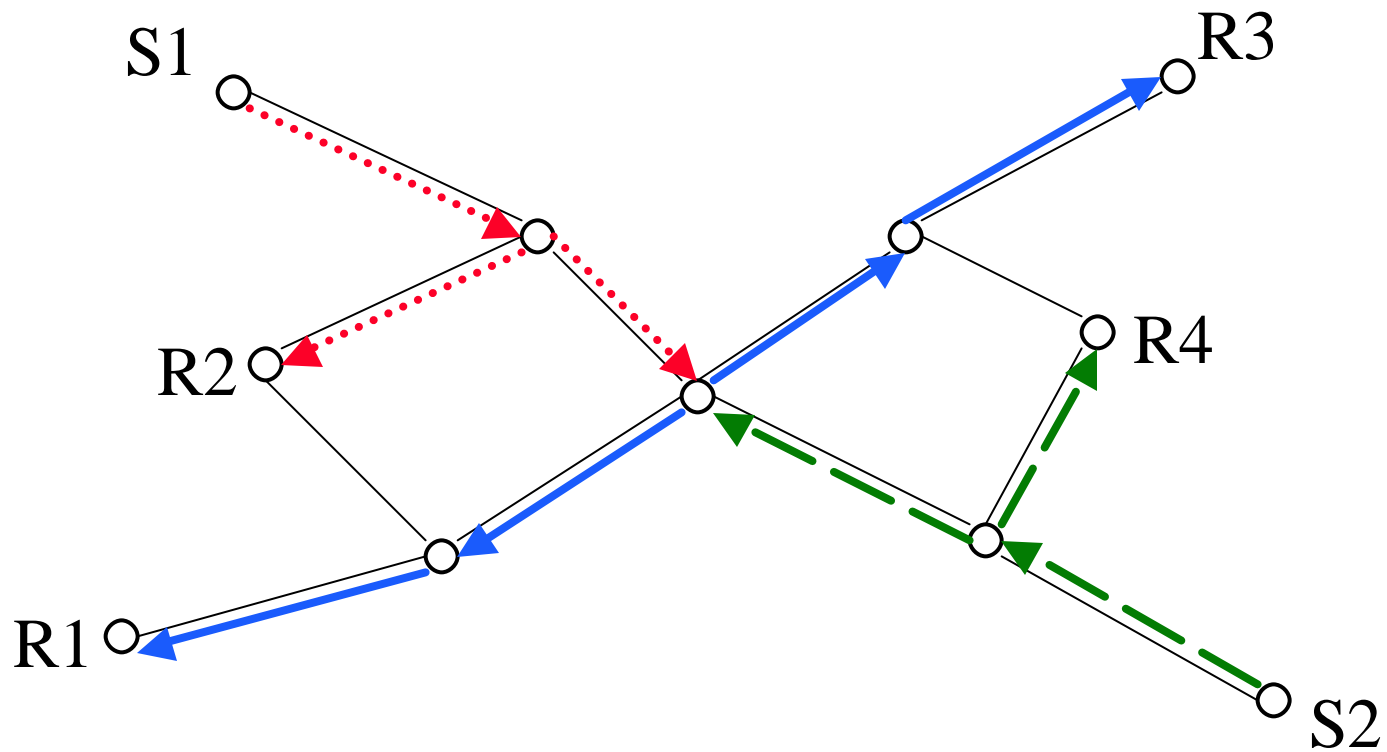
Unidirectional Tree, One Tree Per Source



Unidirectional Tree, Shared by All Sources



Bi-directional Tree, Shared by All Sources



Distinguishing Properties of IP Multicast Routing Protocols (cont.)

(3) topology database (routing table)
construction

- build own routing table
- use unicast routing table

Current IP Multicast Routing Protocols

DVMRP — Distance-Vector Multicast Routing Protocol

broadcast-and-prune,
unidirectional per-source trees,
builds own routing table

MOSPF — Multicast Extensions to Open Shortest-Path
First Protocol

broadcast membership,
unidirectional per-source trees,
uses unicast routing table

Current IP Multicast Routing Protocols (cont.)

PIM-DM — Protocol-Independent Multicast, Dense-Mode

broadcast-and-prune,
unidirectional per-source trees,
uses unicast routing table

PIM-SM — Protocol-Independent Multicast, Sparse-Mode

uses meeting places (“rendezvous points”),
unidirectional per-source or shared trees,
uses unicast routing table

CBT — Core-Based Trees

uses meeting places (“cores”),
omnidirectional shared trees,
uses unicast routing table

Multicast Routing: PIM

Protocol Independent Multicast (PIM)

- “Protocol Independent”
 - does not perform its own routing information exchange
 - uses unicast routing table made by any of the existing unicast routing protocols
- PIM-DM (Dense Mode) - similar to DVMRP, but:
 - without the routing information exchange part
 - differs in some minor details
- PIM-SM (Sparse Mode), or just PIM - instead of directly building per-source, shortest-path trees:
 - initially builds a single (unidirectional) tree per group , shared by all senders to that group
 - once data is flowing, the shared tree can be converted to a per-source, shortest-path tree if needed

PIM Protocol Overview

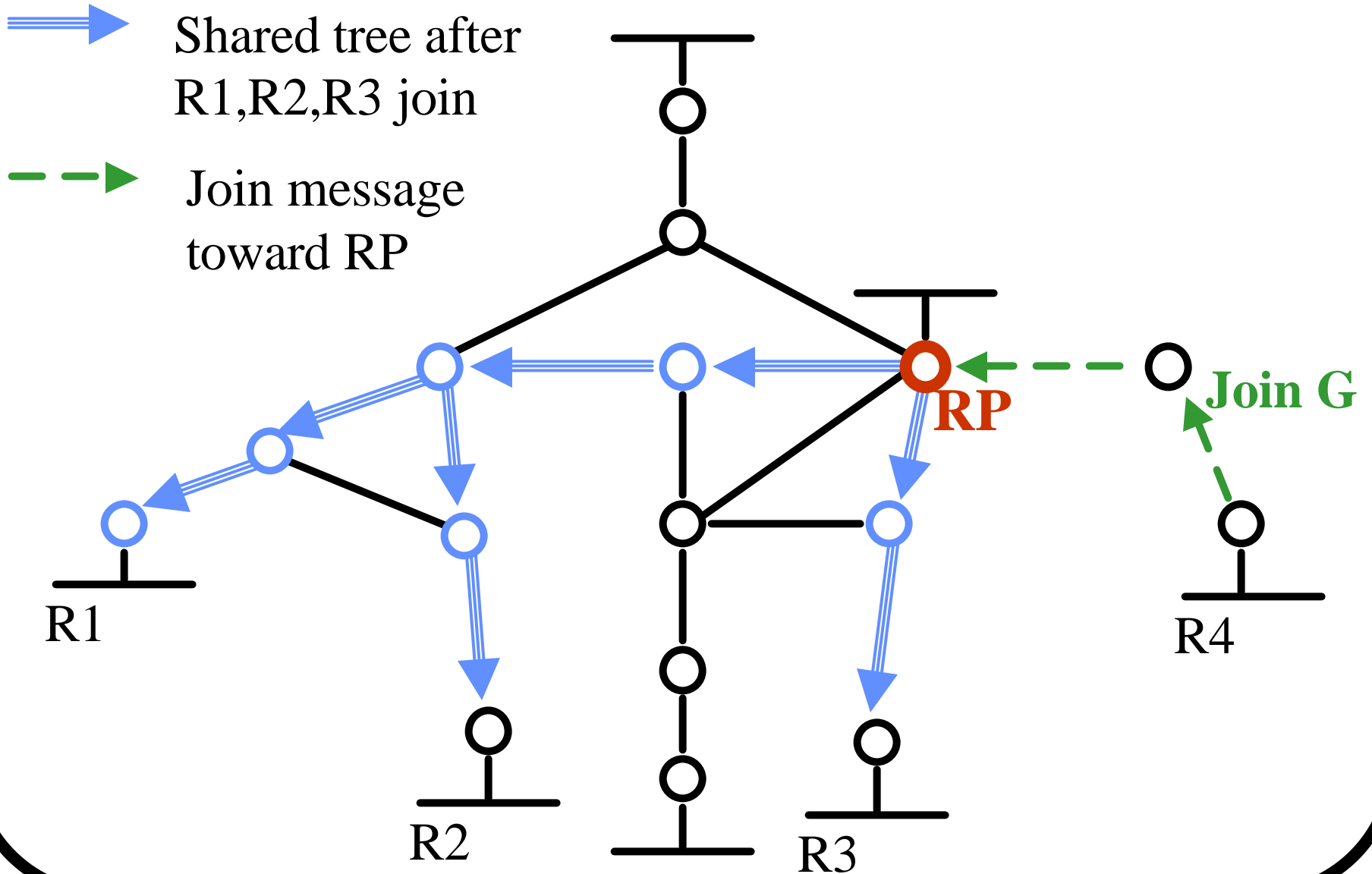
basic protocol steps

- routers with local members send Join messages towards a Rendezvous Point (RP) to join shared tree
- routers with local sources encapsulate data to RP
- routers with local members may initiate data-driven switch to source-specific, shortest-path tree

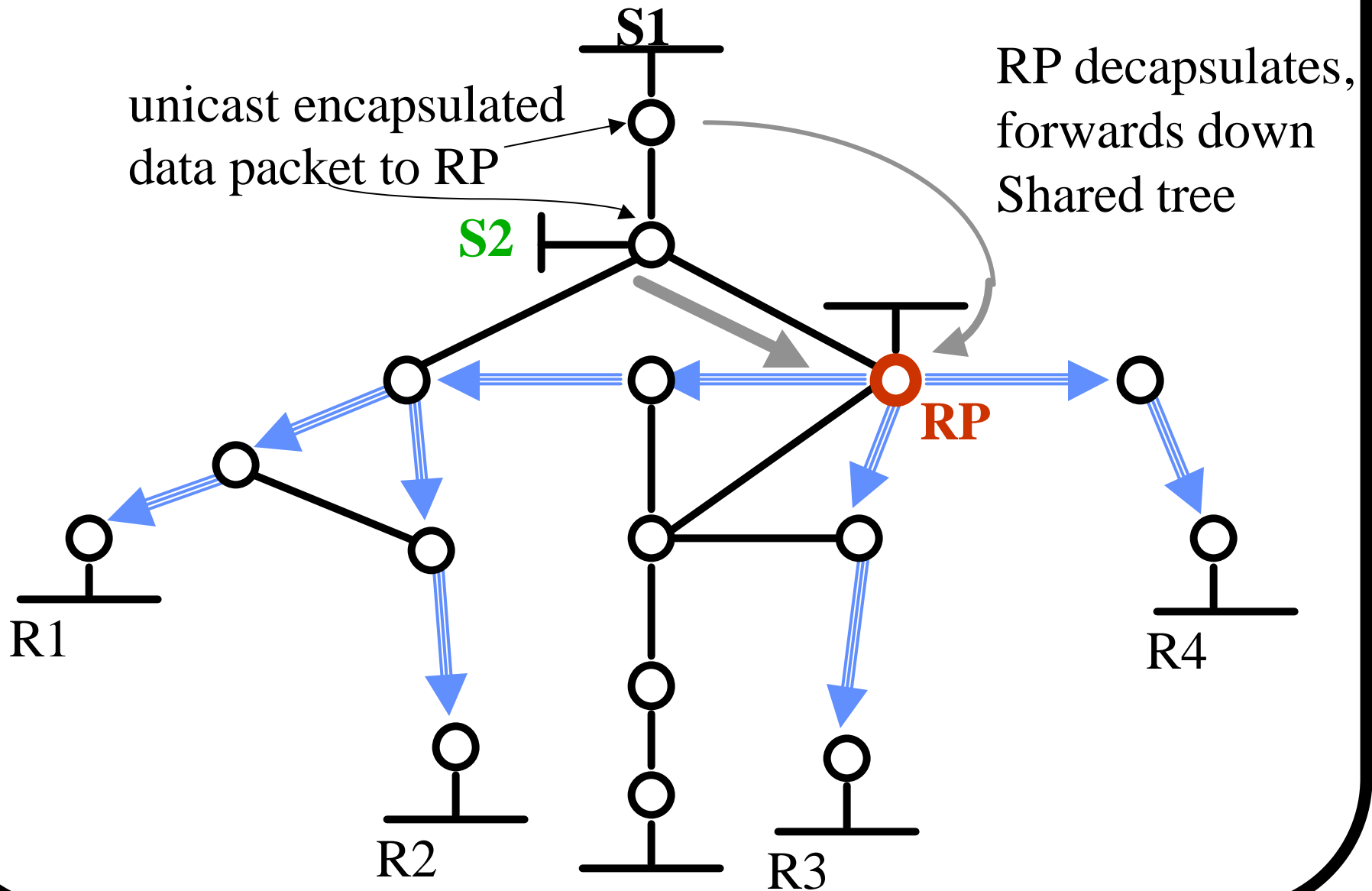
IETF PIM WG started in Aug'98 to standardize PIM

- <http://www.ietf.org/html.charters/pim-charter.html>

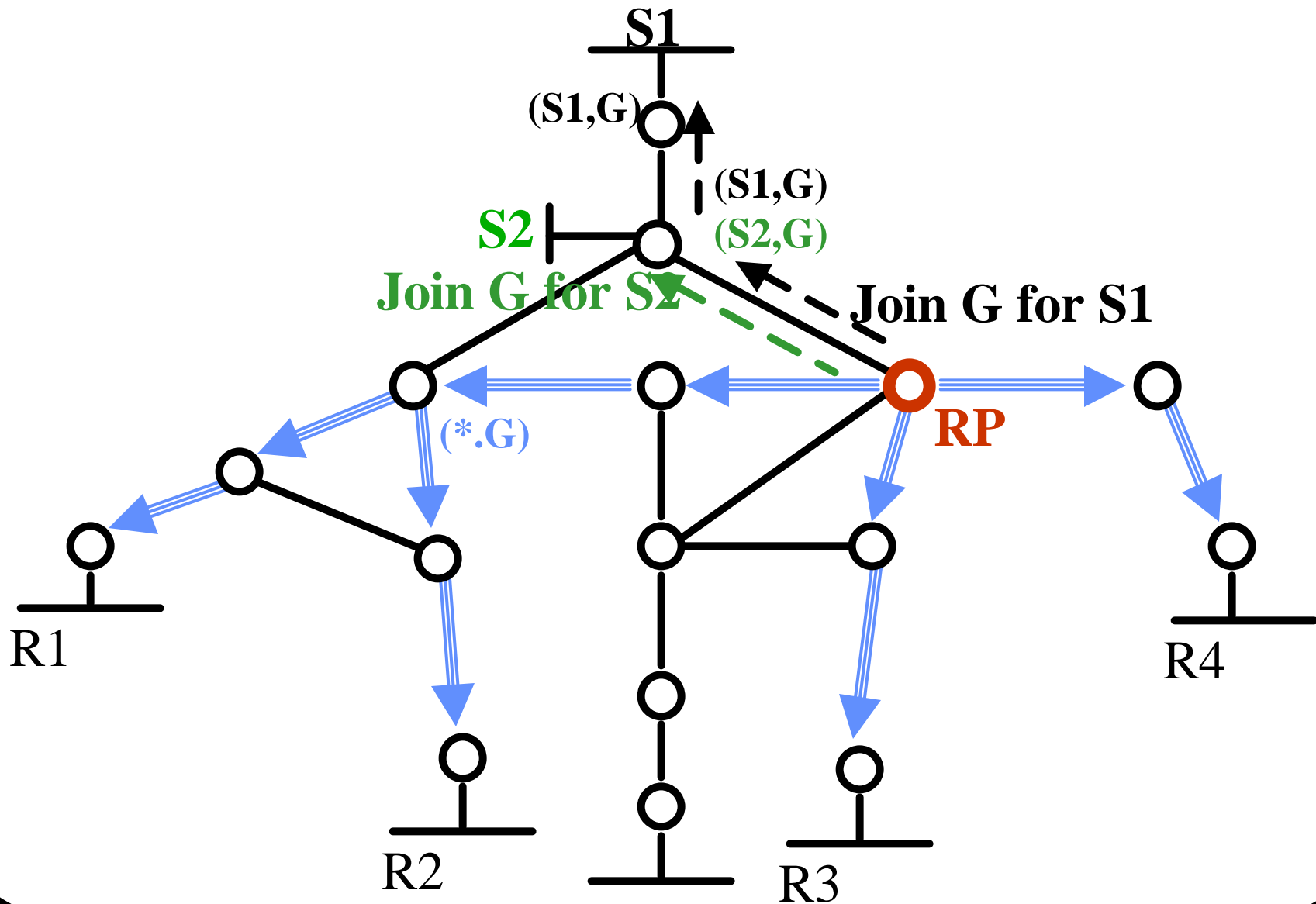
Phase 1: Build Shared Tree



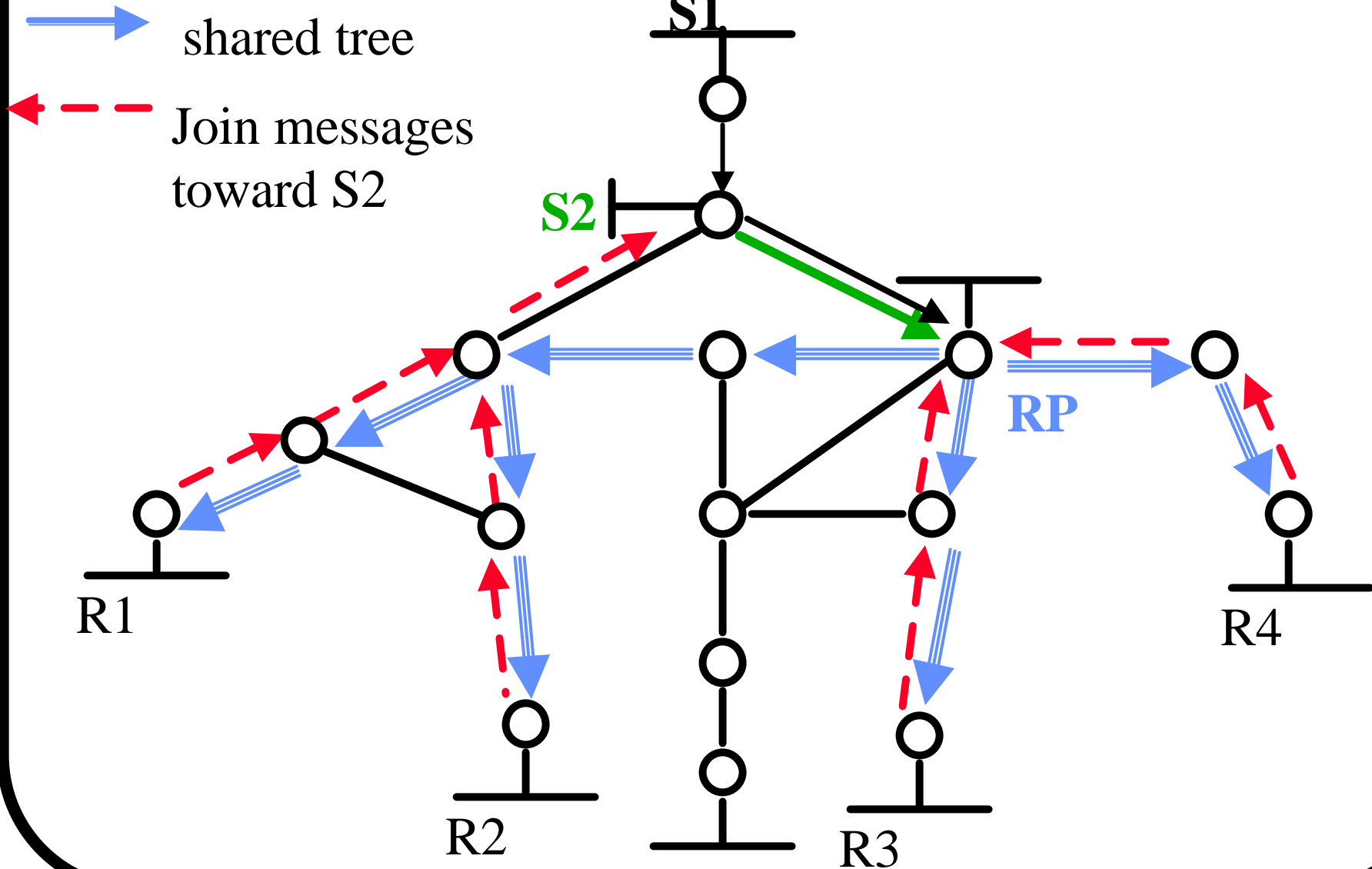
Phase 2: Sources Send to RP



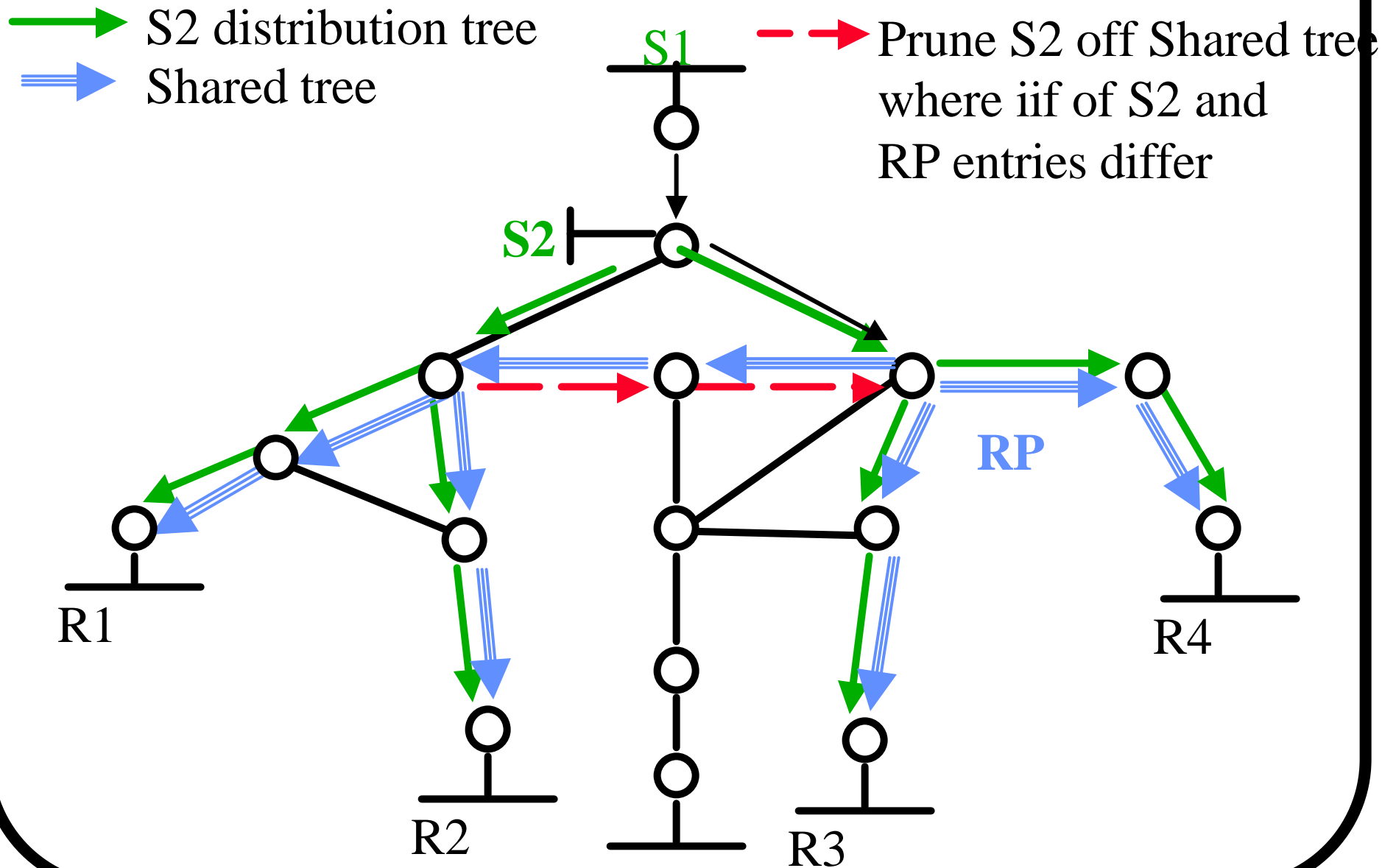
Phase 3: Stop Encapsulation



Phase 4: Switch to Shortest Path Tree



Phase 5: Prune (S2 off) Shared Tree

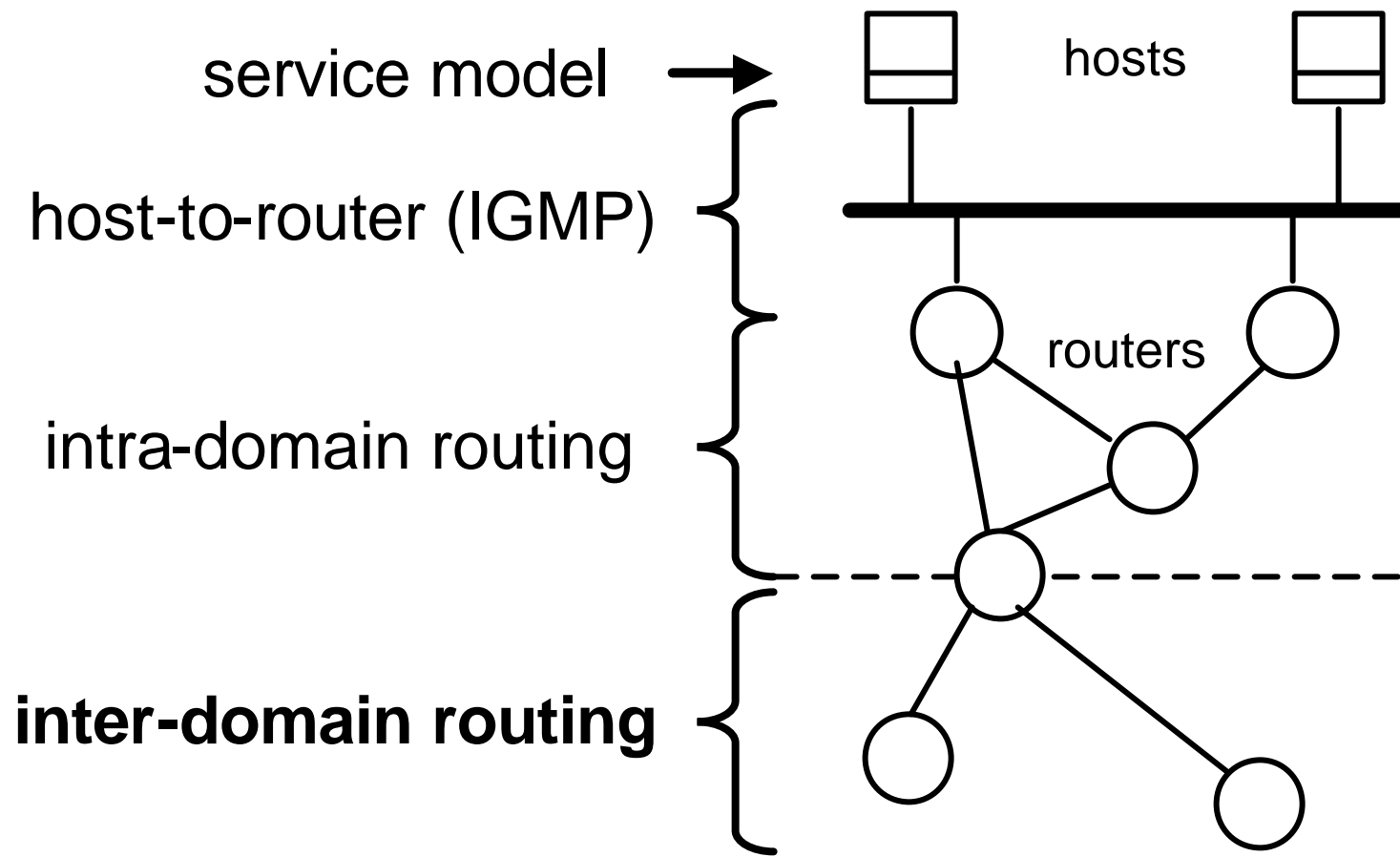


RP Mechanism

- end-systems only need multicast address to send or receive
- routers use algorithmic mapping of group address to RP from manageably-small set of RPs known throughout region
- consistent RP mapping and adaptation to failures is **CRITICAL**
 - all routers (within PIM region) must associate a single active RP with a multicast group
- optimal RP location not necessary

Inter-Domain Multicast Routing Protocols

Components of the IP Multicast Architecture



Historical Outline

- MBone Deployment – 1992
- Native Dense Mode – 1994
- Sparse Mode Development – 1994 (same time!)
- Inter-Domain Protocol Development – 1997
- Commercial Deployment – Beginning 1998
- Internet2 Deployment – 1998(vBNS) &1999(Abilene)
- New service models – 1999 (ongoing)

Need for “Inter-Domain” Multicast

- the MBone is one big, flat network
 - why is this bad?
- experience shows big, flat networks do not scale
 - every router knows the existence of every single other router/subnet in the topology (lots of state)
 - infrequent problems become frequent when the size of the network grows large (instability)
- scalability/instability are not the only problems
 - ISPs consider multicast protocols to be flawed

What Exactly is Needed?

- intra-domain routing protocols
- inter-domain route exchange protocol
- mechanism for connecting domains
 - mechanism for routing between backbones
 - consideration for “politics” between domains
- address allocation

What Exactly is Needed?

- **intra-domain routing protocols**
- inter-domain route exchange protocol
- mechanism for connecting domains
 - mechanism for routing between backbones
 - consideration for “politics” between domains
- address allocation

Intra-Domain Routing Protocols

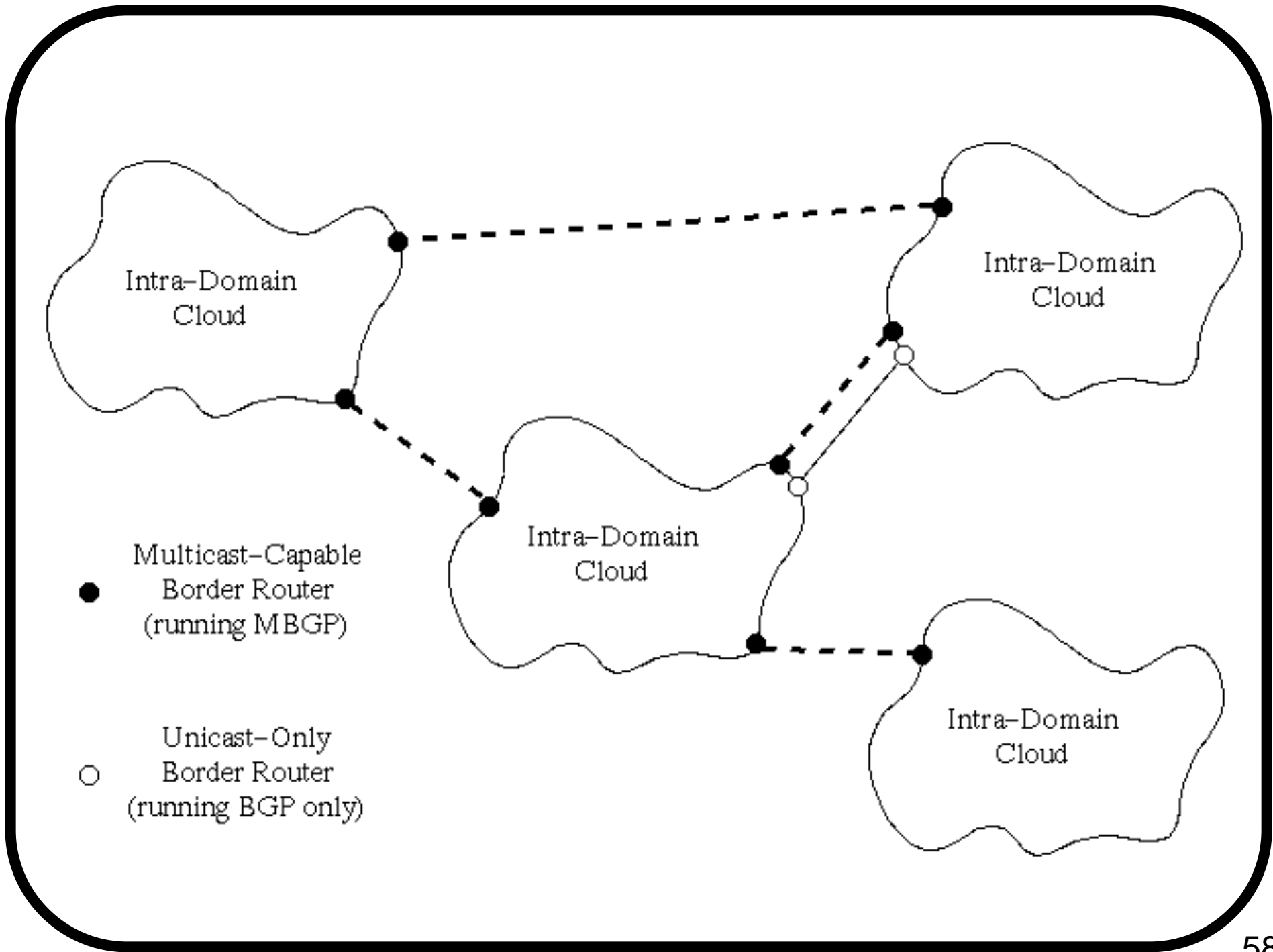
- Already exist
- Are either “broadcast-and-prune” or “explicit join”
 - Explicit join protocols are by far the most scalable and the most efficient
- “Best” protocol to-date seems to be PIM-SM
 - best is good tradeoff between efficiency/complexity
 - best is most widely available

What Exactly is Needed?

- intra-domain routing protocols
- **inter-domain route exchange protocol**
- mechanism for connecting domains
 - mechanism for routing between backbones
 - consideration for “politics” between domains
- address allocation

Inter-Domain Route Exchange

- Exchange multicast reachability between Autonomous Systems (AS)
 - Just like unicast routes are exchanged with BGP
 - Protocol is “Multiprotocol extensions to BGP” (RFC 2283)
 - Also known as “Multicast” BGP (MBGP)
 - Also known as BGP4+
- MBGP is available and deployed today.
 - Multiple vendors: Juniper, Cisco, Nortel, 3Com, IBM
- Allows congruent/different unicast/multicast topologies



What Exactly is Needed?

- intra-domain routing protocols
- inter-domain route exchange protocol
- **mechanism for connecting domains**
 - mechanism for routing between backbones
 - consideration for “politics” between domains
- address allocation

The Internet Solution

- Re-use existing protocols/solutions
 - Use PIM-SM in the inter-domain
- The challenge is to avoid “root dependencies”
 - A root/RP/core is one domain but no active group participants (sources or receivers) in the domain
 - Root dependencies can lead to political problems and inefficiencies

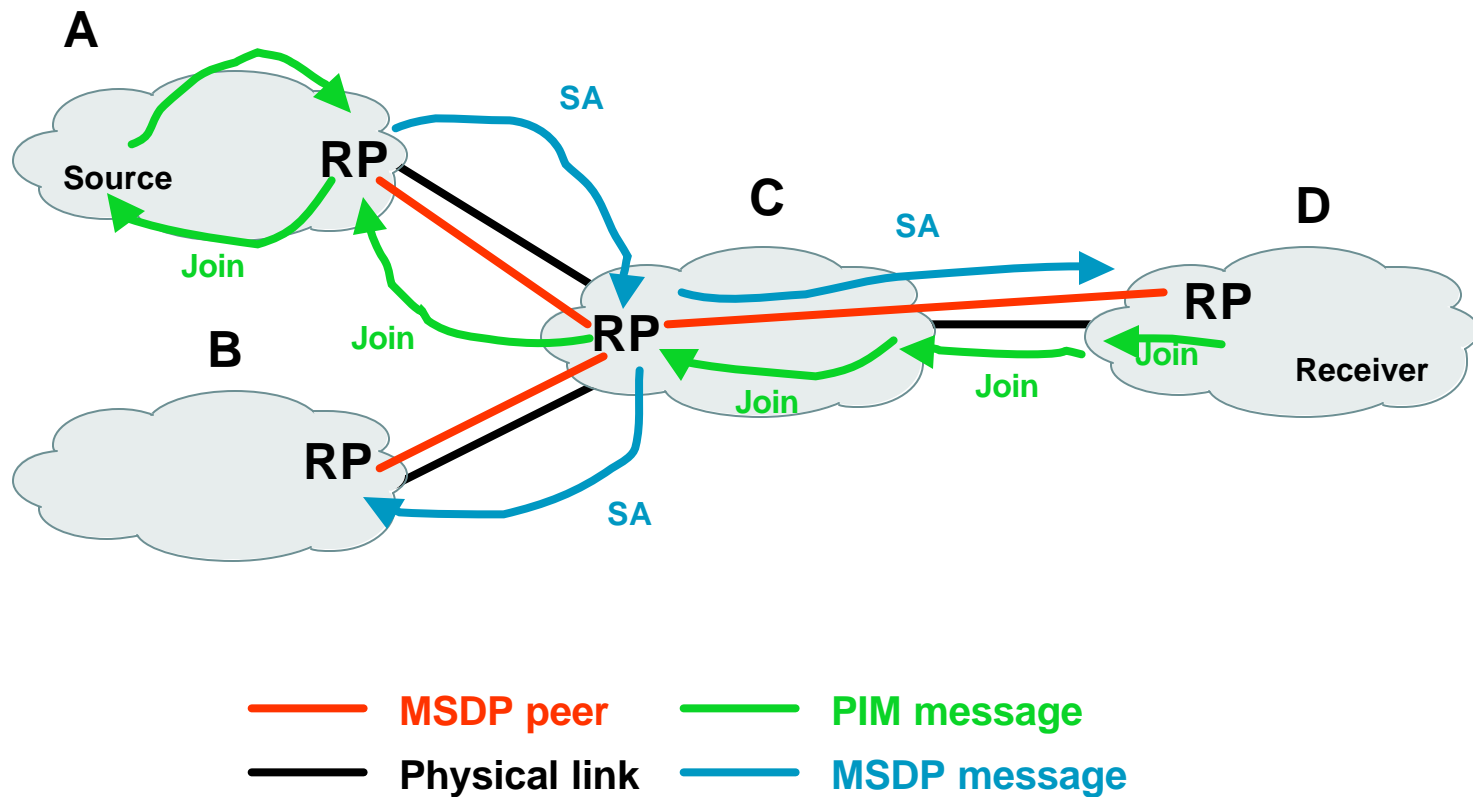
The Internet Solution (cont)

- The key: Establish a root/RP/core per domain
 - No “root dependencies”
- Remember the problem:
 - Connecting sources and receivers
 - Solution is to use Multicast Source Discovery Protocol (MSDP)
- MSDP is the last piece of the puzzle; is simple to implement; and yields an interim solution to inter-domain multicast

MSDP -- Basic Idea

- MSDP advertises multicast sources to other domains
- Other domains decide if group members are active and find a way to get the data
- “MSDP connects shared-trees together”

How MSDP works with PIM-SM



Done!!

