

Internet2 Multicast Workshop

University of British Columbia
Vancouver, BC
May, 2004

Acknowledgements

Greg Shepherd

Beau Williamson

Marshall Eubanks

Bill Nickless

Caren Litvanyi

Patrick Dorn

Leonard Giuliano

Alan Crosswell

University of Oregon

Cisco Systems

Juniper Networks

Columbia University



Engineering Workshops

Contents

- Overview
- Multicast on the LAN
- Source-Specific Multicast (SSM)
- Any-Source Multicast (ASM)
 - Intra-domain ASM
 - Inter-domain ASM
- Troubleshooting Methodology
- Making the Case for Multicast

Overview

The Basic Idea

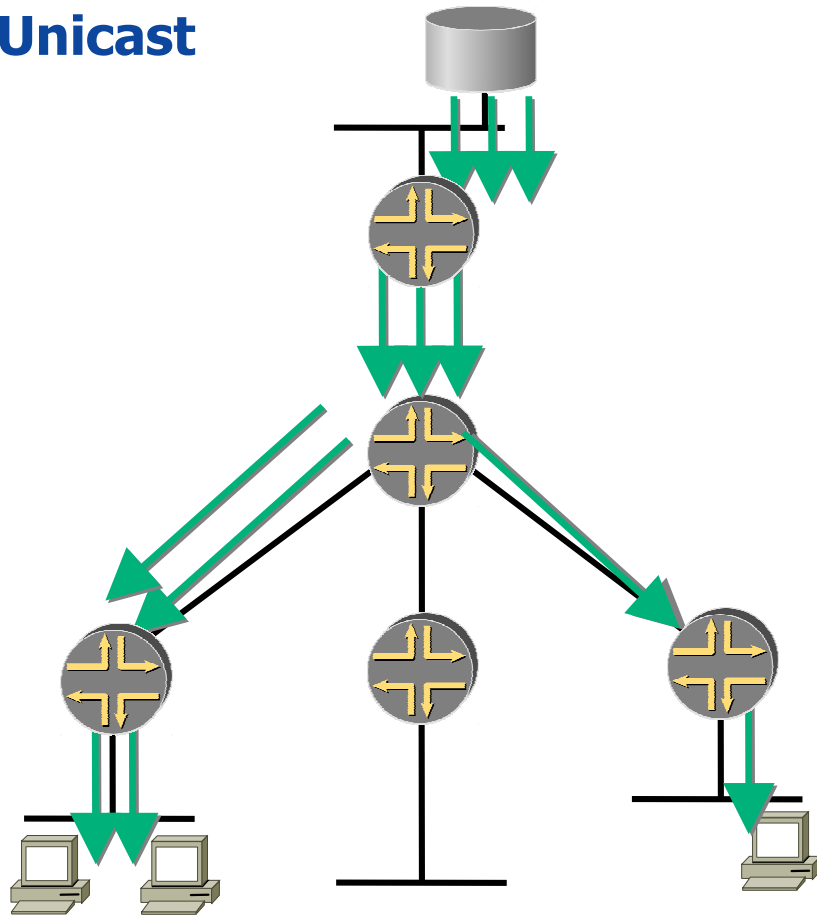
Rather than sending a separate copy of the data for each recipient, the source sends the data only once, and routers along the way to the destinations make copies as needed.

Unicast does mass mailings;
multicast does chain letters.

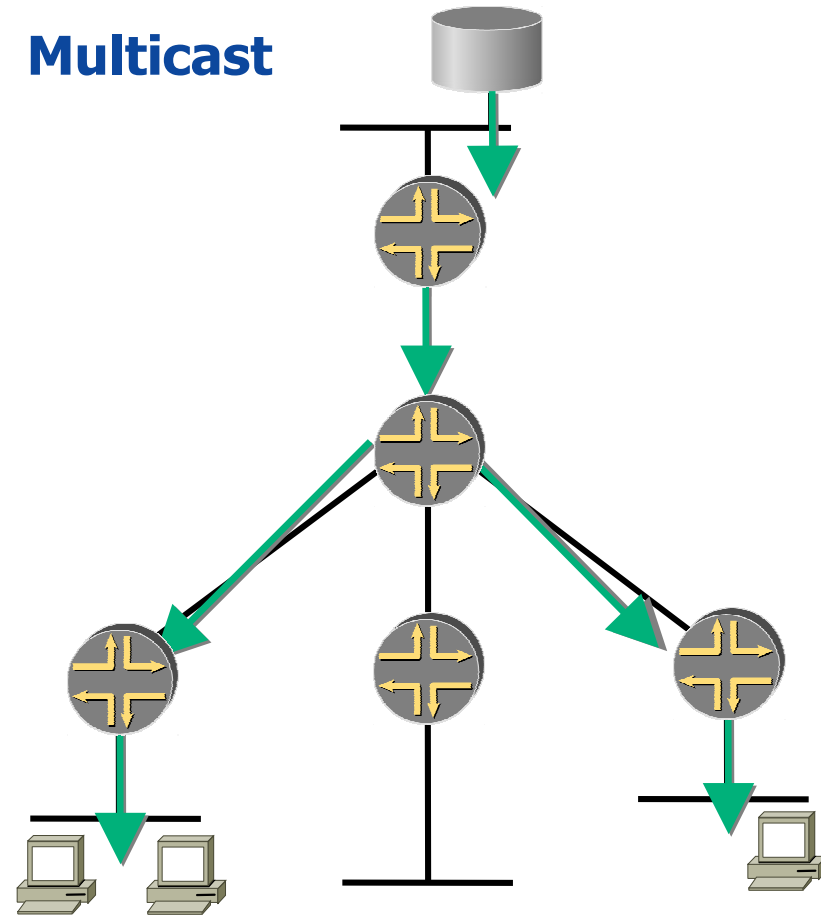


Unicast vs. Multicast

Unicast



Multicast



Some Uses for Multicast

- Any application with multiple receivers
 - one-to-many or many-to-many
- Live video distribution
- Collaborative groupware
- Periodic data delivery - “push” technology
 - stock quotes, sports scores, magazines, newspapers
 - advertisements



Some More Uses for Multicast

- Server/web site replication
- Reducing network/resource overhead
 - more efficient to establish multicast tree rather than multiple point-to-point links
- Resource discovery
- Distributed interactive simulation
 - war games
 - virtual reality



What Happened to Multicast?

- By 1995, multicast seemed well on its way to adoption.
 - The MBone (Multicast backBone) had been set up and was growing.
 - Audiocasts and Videocasts of meetings, seminars, etc., were fairly routine.
 - Serious interest was coming from industry.
- So why isn't it ubiquitous now ?
 - The hype got ahead of the technology!
 - The original technology was not suitable for adoption throughout the Internet. Basic parts had to be re-engineered.
 - This took from 1997 to early 2001.



The MBone

- The original multicast network was called the MBone. It used a simple routing protocol called DVMRP (Distance Vector Multicast Routing Protocol).
- As there were only isolated subnetworks that wanted to deal with DVMRP, the old MBone used tunnels to get multicast traffic between DVMRP subnetworks.
 - i.e., the multicast traffic was hidden and sent between the subnetworks via unicast.
- This mechanism was simple, but required manual administration and absolutely could not scale to the entire Internet.
- Worse, DVMRP requires substantial routing traffic behind the scenes and this grew with the size of the MBone.
 - Thus, the legend grew that multicast was a bandwidth hog.

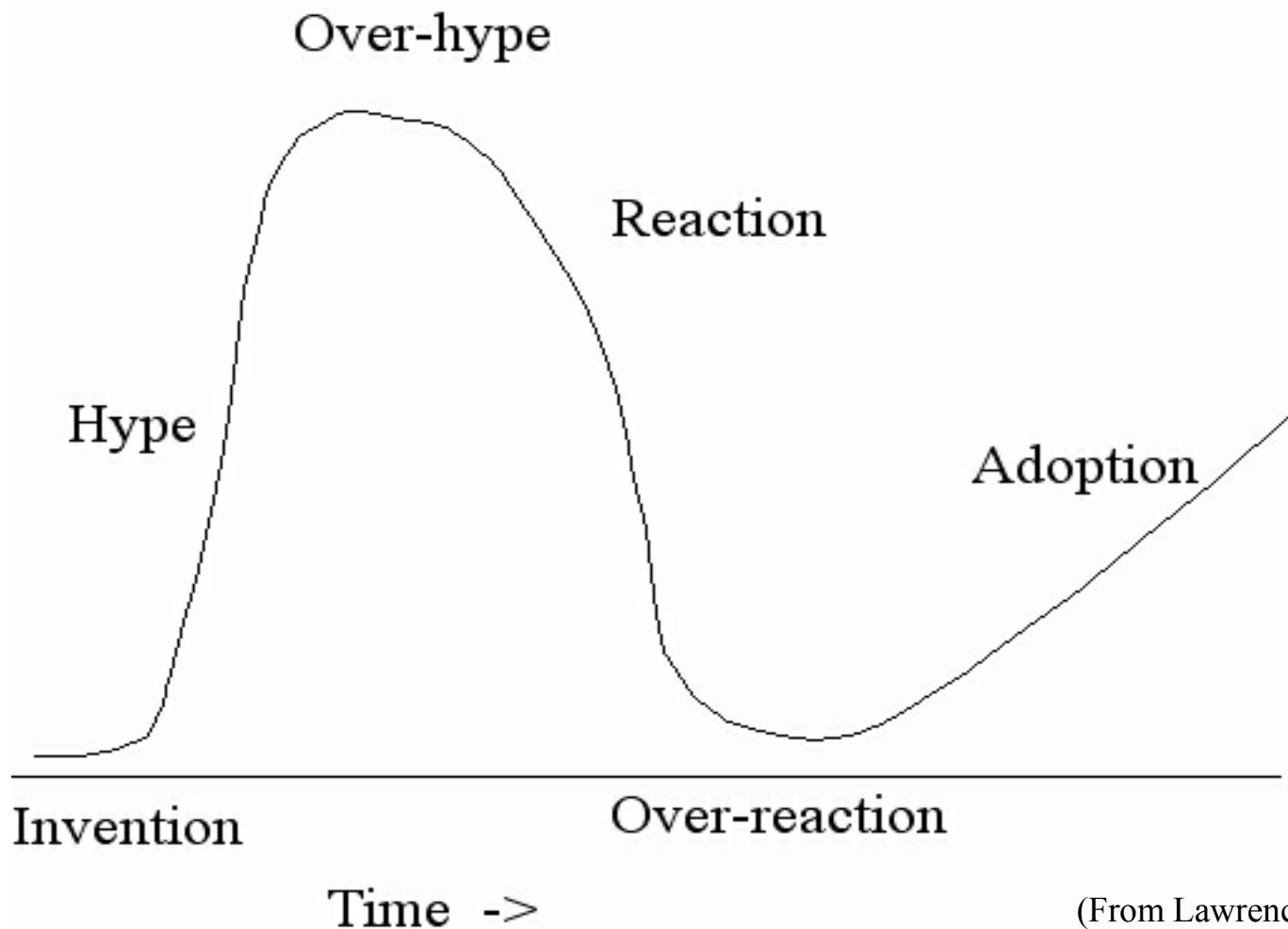


Multicast Grows Up

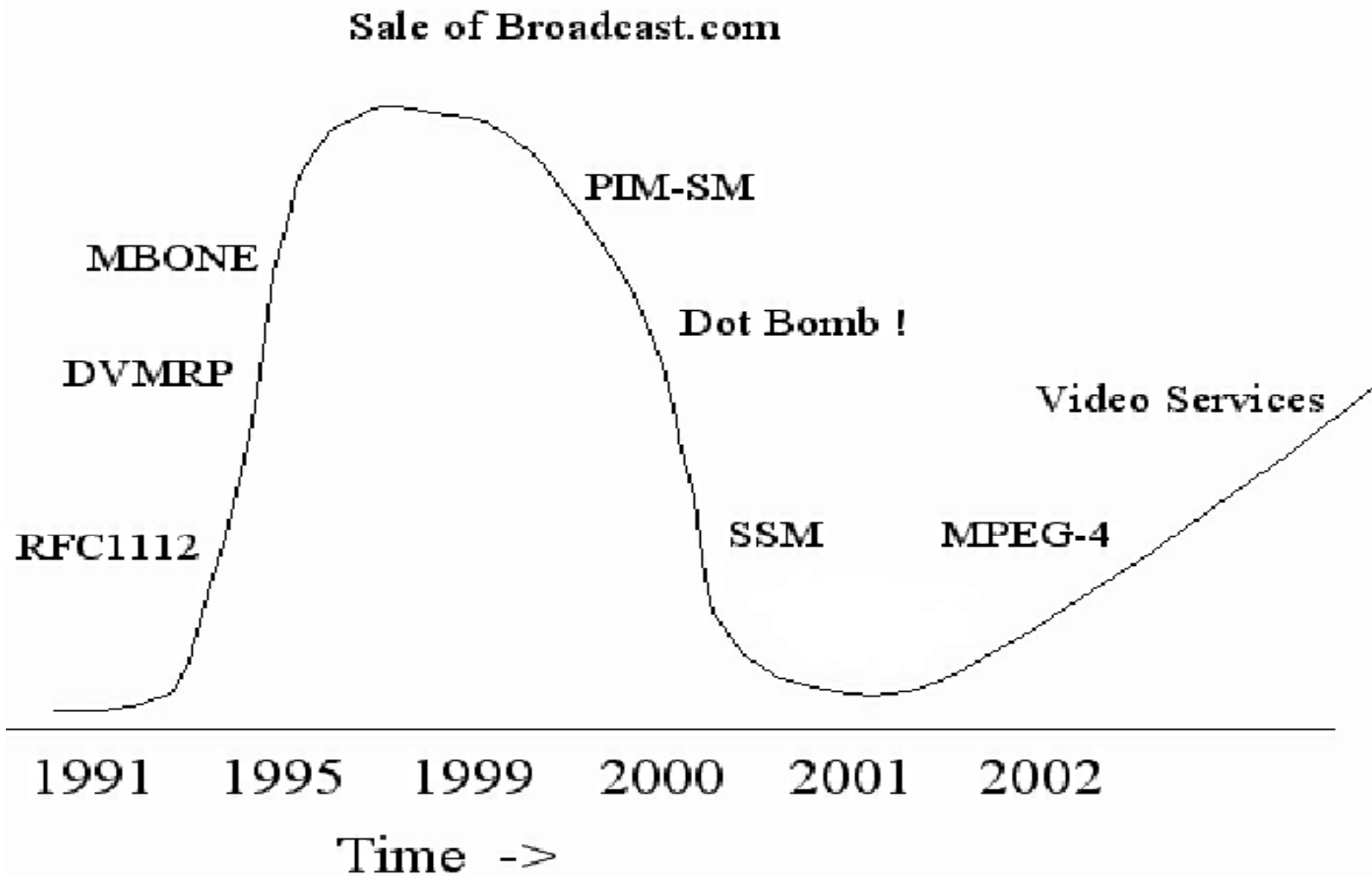
- Starting about 1997, the building blocks for a multicast-enabled Internet were put into place.
 - An efficient modern multicast routing protocol, Protocol Independent Multicast – Sparse Mode (PIM-SM), was deployed.
 - The mechanisms for multicast peering were established, using an extension to BGP called Multiprotocol BGP (MBGP), and peering became routine.
 - The service model was split into:
 - a many-to-many part (e.g., for videoconferencing): Any-Source Multicast (ASM), and
 - a one-to-many (or “broadcast”) part: Source-Specific Multicast (SSM).
- By 2001, these had completely replaced the old MBone.
- This path is not unusual for new technology...



The Life Cycle of New Technologies in General

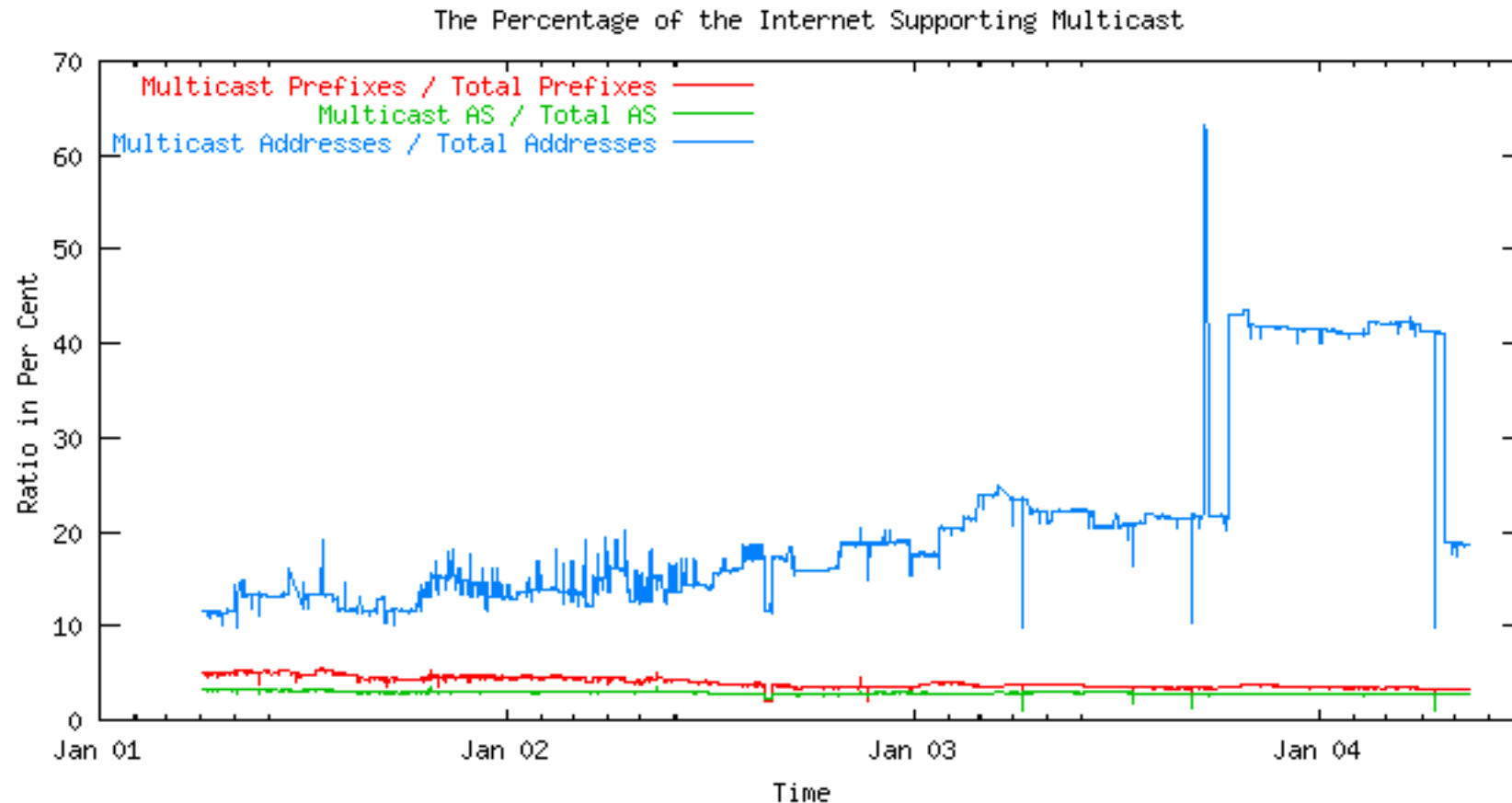


The Life Cycle of Multicast in Particular



Deployment of Multicast

Multicast Technologies has been monitoring the state of multicast since April 2001...

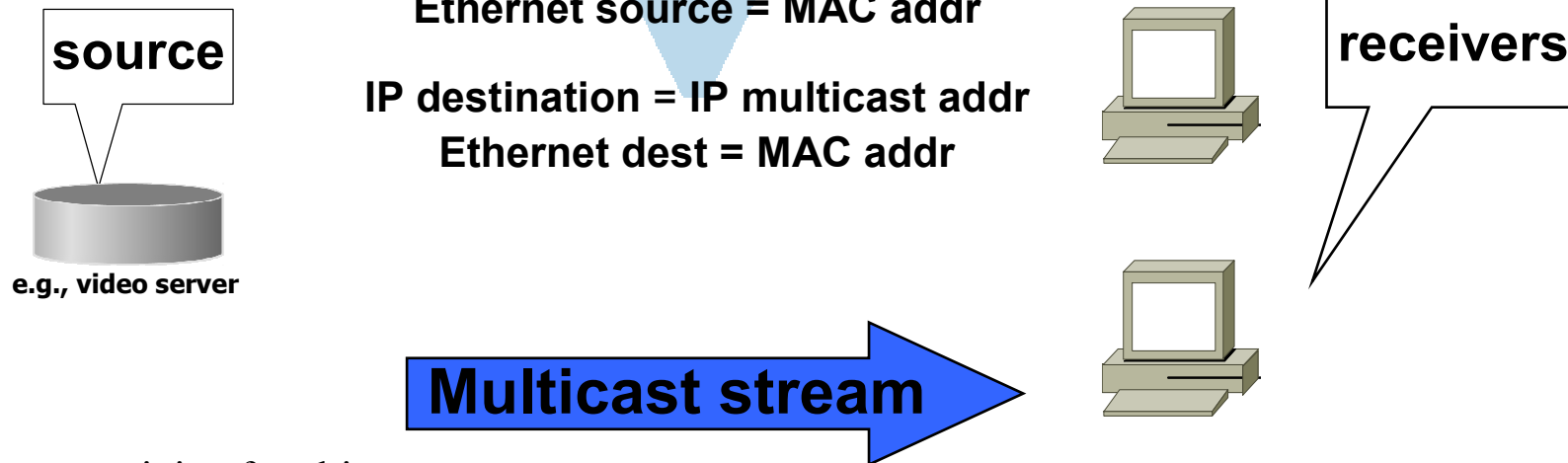


From <http://www.multicasttech.com/status/>



Engineering Workshops

Multicast Technology: the basics



- *source* = origin of multicast stream
- *multicast address* = an IP address in the Class D range (224.0.0.0 – 239.255.255.255), used to refer to multiple recipients. A multicast address is also called a *multicast group* or *channel*.
- *multicast stream* = stream of IP packets with multicast address for IP destination address.
 - (S,G) = (source, group) reference
 - All multicast uses UDP packets
- *receiver(s)* = recipient(s) of multicast stream

Multicast Protocol Summary

- IGMP - Internet Group Management Protocol is used by hosts and routers to tell each other about group membership.
- PIM-SM - Protocol Independent Multicast-Sparse Mode is used to propagate forwarding state between routers.
- MBGP - Multiprotocol Border Gateway Protocol is used to exchange routing information for inter-domain RPF checking.
- MSDP - Multicast Source Discovery Protocol is used to exchange active source information between RPs.



(S,G) notation

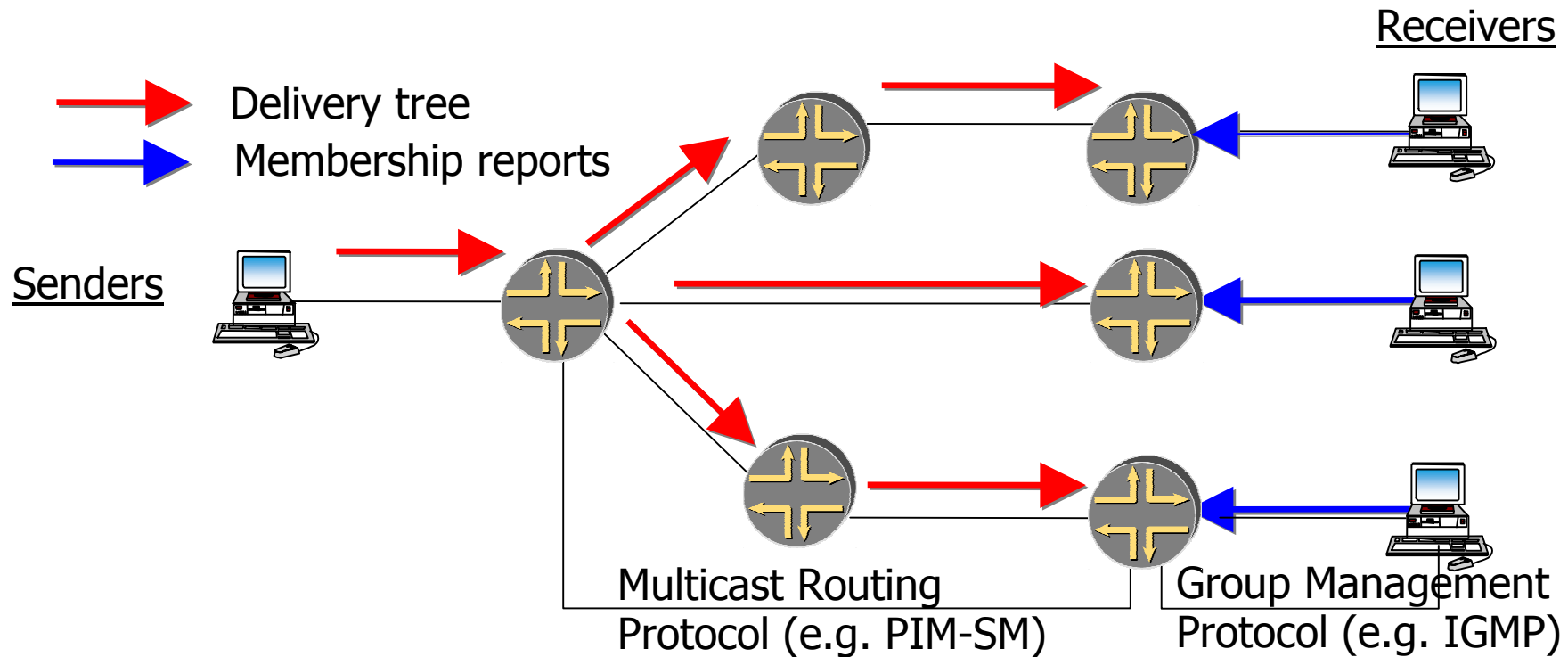
- For every multicast source there must be two pieces of information: the source IP address, S, and the group address, G.
 - These correspond to the sender and receiver addresses in unicast.
 - This is generally expressed as (S,G).
 - Also commonly used is (*,G) - every source for a particular group.

IP Multicast building blocks

- The SENDERS send without worrying about receivers
 - Packets are sent to a multicast address (RFC 1700)
 - This is in the class D range (224.0.0.0 - 239.255.255.255)
- The RECEIVERS inform the routers what they want to receive
 - done via Internet Group Management Protocol (IGMP), version 2 (RFC 2236) or later
- The routers make sure the STREAMS make it to the correct receiving networks.
 - Multicast routing protocol: PIM-SM



Essential IP Multicast Protocols



- Group Management Protocol - enables hosts to dynamically join/leave multicast groups. Receivers send group membership reports to the nearest router.
- Multicast Routing Protocol - enables routers to build a delivery tree between the sender(s) and receivers of a multicast group.

Multicast Routing

- Multicast routing can be thought of as the reverse of unicast forwarding.
 - Unicast forwarding is concerned with where the packet is going.
 - Multicast routing is concerned with where the packet will be coming from.
- Multicast paths to receivers form a “tree”. The tree is built (or torn down) from the receiver back toward the source.

PIM-SM

Protocol Independent Multicast - Sparse Mode

- The core multicast protocol: builds and tears down multicast trees
- draft-ietf-pim-sm-v2-new-06.txt obsoletes RFC 2362; bootstrap router removed from PIM spec
- Explicit join: assumes that not everyone wants the data
- Uses externally-provided reachability table to build forwarding topology.
- Each router maintains an outgoing interface list (OIL) for each (S,G) and (*,G) for which it has downstream listeners. Multicast packets received from a given source for a given group are sent out only on the interfaces specified in the appropriate OIL.

Multicast Distribution Trees

- The path taken by multicast data is called a tree.
 - Routing loops are not allowed, so there is always a unique series of branches between the root of the tree and the receivers.
 - The tree is built based on RPF reachability information.

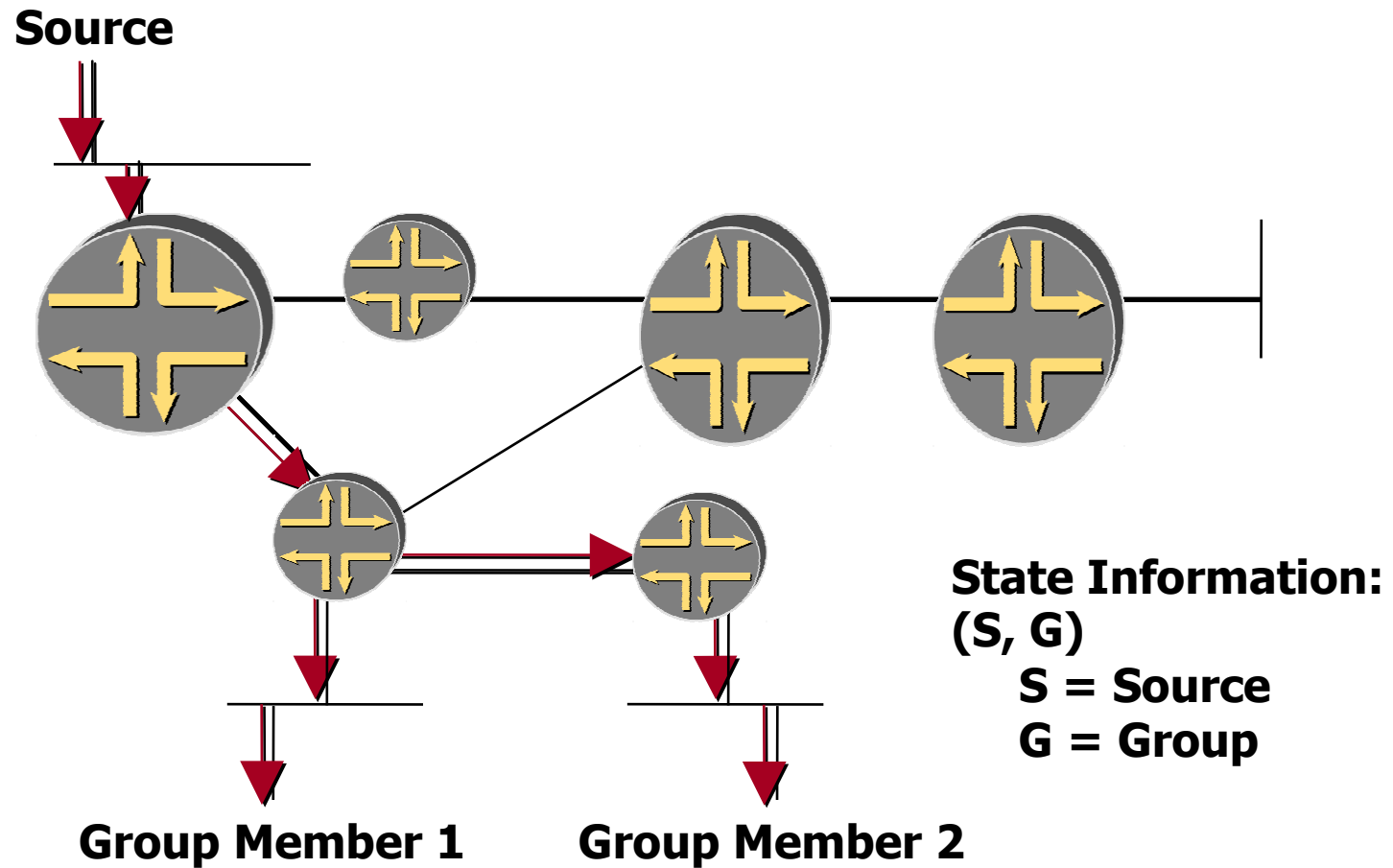
Reachability

- When a *unicast packet* shows up on an interface, the *destination address* is looked up in the unicast forwarding table to determine where the router should send the packet next.
- When a *multicast (S,G) Join* shows up on an interface, the *source address, S*, is looked up in the reverse-path forwarding (RPF) table to determine how the router should join the source-based forwarding tree. This information is used to build the multicast forwarding tree.
- The unicast forwarding table and the RPF table contain the same kind of information — **unicast routes, or reachability information** — and may in fact be the same table.
- The point of having separate tables is to enable separate policies and paths for unicast forwarding and RPF. You need MBGP, IS-IS, or static mroutes to do this.
- Once the multicast forwarding tree is built, multicast forwarding works similarly to unicast forwarding.

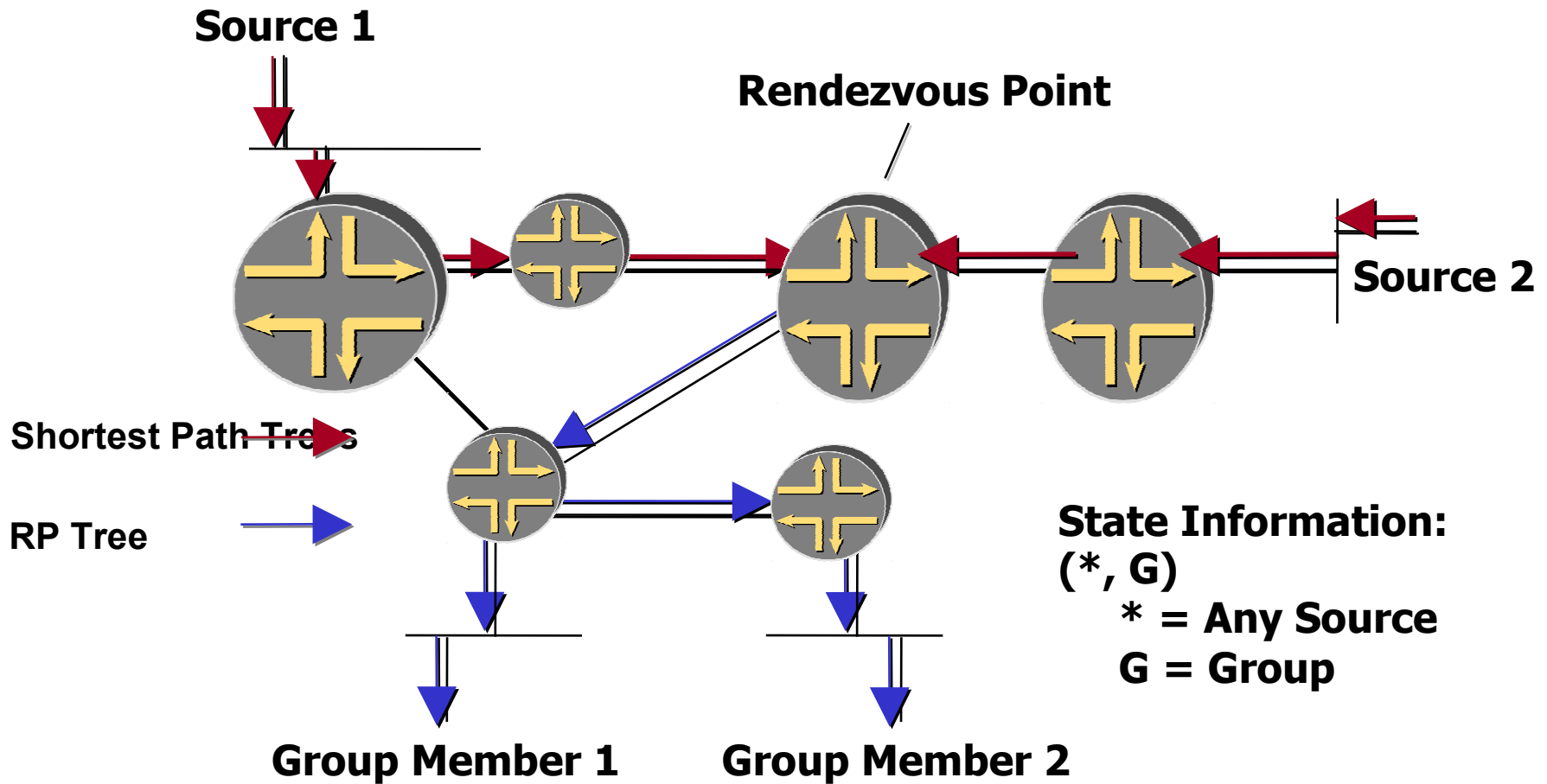
Multicast Distribution Trees

- A shortest path tree (SPT) is a tree rooted in a multicast source. An SPT is sometimes called a source tree.
- A rendezvous point tree (RPT) is a tree rooted in a multicast rendezvous point (RP). An RPT is sometimes called a shared tree.
- In the original multicast service model, a connection between a source and a receiver is first set up by building an RPT from the receiver back to the RP, and an SPT from the RP back to the source. Once data starts flowing to the receiver, an SPT is built directly from the receiver back to the source.

Shortest Path Tree



Rendezvous Point Tree



Multicast Distribution Trees Compared

- Shortest Path Tree
 - More resource-intensive; requires more state $\rightarrow n(S \times G)$
 - You get optimal paths from source to all receivers, which minimizes delay
 - Best for one-to-many distribution
- Rendezvous Point Tree
 - Uses less resources; requires less memory $\rightarrow n(G)$
 - You may get suboptimal paths from source to all receivers, depending on topology
 - The RP itself and its location may affect performance
 - Best for many-to-many distribution
 - Necessary for in-band source discovery



Multicast Addressing

- IPv4 Multicast Group Addresses
 - 224.0.0.0–239.255.255.255
 - Class D Address Space
 - High order bits of 1st Octet = “1110”
 - Source sends to group address
 - Receivers receive traffic sent to group address

CIDR Address Notation

- The multicast address block is 224.0.0.0 to 239.255.255.255
- It is cumbersome to refer to address blocks in the above fashion. Address blocks are usually described using “CIDR notation”
 - This specifies the start of a block, and the number of bits THAT ARE FIXED.
 - In this shorthand, the multicast address space can be described as 224.0.0.0/4 or, even more simply, as 224/4. The fixed part of the address is referred to as the prefix, and this block would be pronounced "two twenty four slash four."
 - Note that the LARGER the number after the slash, the LONGER the prefix and the SMALLER the address block.



Multicast Addressing

- RFC 3171
- <http://www.iana.org/assignments/multicast-addresses>
- Examples of Reserved & Link-local Addresses
 - 224.0.0.0 - 224.0.0.255 reserved & not forwarded
 - 224.0.0.1 - All local hosts
 - 224.0.0.2 - All local routers
 - 224.0.0.4 - DVMRP
 - 224.0.0.5 - OSPF
 - 224.0.0.6 - Designated Router OSPF
 - 224.0.0.9 - RIP2
 - 224.0.0.13 - PIM
 - 224.0.0.15 - CBT
 - 224.0.0.18 – VRRP
 - 239.0.0.0 - 239.255.255.255 Administrative Scoping
- “Ordinary” multicasts don’t have to request a multicast address from IANA.



Scoping

- TTL value defines scope and limits distribution
 - IP multicast packet must have $TTL > \text{interface TTL}$ or it is discarded
 - No longer recommended as a reliable scoping mechanism
- Administratively Scoped Addresses – RFC 2365
 - 239.0.0.0–239.255.255.255
 - Private address space
 - Similar to RFC 1918 unicast addresses
 - Not used for global Internet traffic
 - Used to limit “scope” of multicast traffic
 - Same addresses may be in used in different sub-networks for different multicast sessions
 - Examples
 - Site-local scope: 239.253.0.0/16
 - Organization-local scope: 239.192.0.0/14



Multicast Address Allocation

- For a long time, this was a sore spot. There was no way to claim or register a Multicast Class D address like unicast address blocks can be registered.
 - For temporary teleconferences, this is not such a problem, but it does not fit well into a broadcast model.
- Now, there are two solutions :
 - For SSM, addresses don't matter, as the broadcast address is really unique as long as the (S,G) pair is unique.
 - For ASM, there is "GLOP".

Multicast Addressing

GLOP addresses

- Provides globally available private Class D space
- 233.x.x/24 per AS number
- RFC 2770

How?

- Insert the 16-bit AS number into the middle two octets of the 233/8
- Online GLOP calculator:
www.shepfarm.com/multicast/glop.html
- If you have an AS, you have multicast addresses.