Bob’s Multicasting Overview

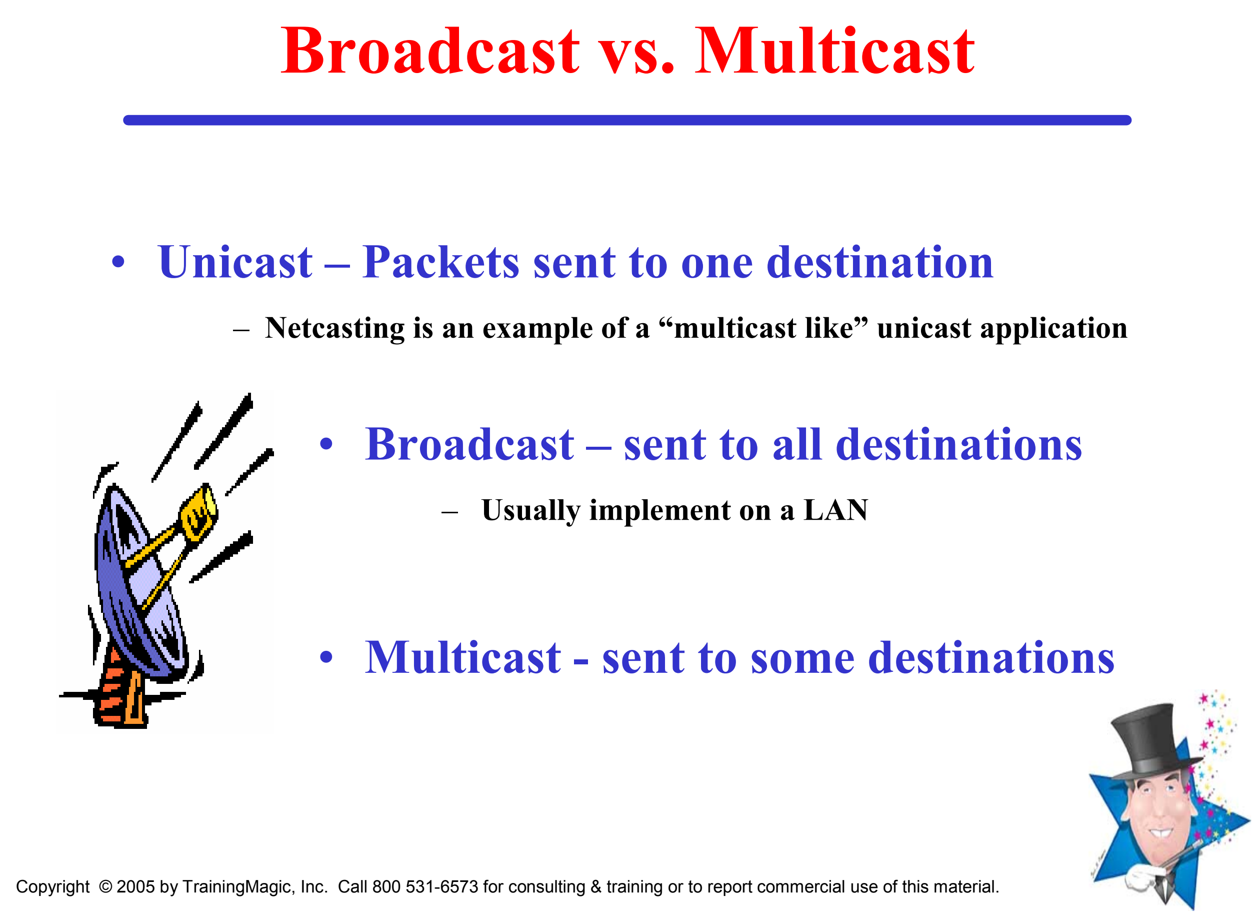
An overview of Multicasting Concepts & Technology

Multicasting – Objectives

- Define multicasting
- Describe the operation of RPF
- Explain the role of IGMP in group memberships
- Distinguish between shared trees and source trees
- Contrast RIP & DVMRP
- Explain the difference between PIM-DM & PIM-SM
- Identify the role of MBGP in multicasting

Why Multicast?

- **Audio/Video Conferencing**
  - SEC compliant stockholder briefings
  - Company meetings – remote offices/employees
- **Training**
  - Corporate wide
  - Promotional
- **Push content/software**
- **Entertainment**
  - A novelty or the future?



Broadcast vs. Multicast

- Unicast – Packets sent to one destination
  - Netcasting is an example of a “multicast like” unicast application
- Broadcast – sent to all destinations
  - Usually implement on a LAN
- Multicast - sent to some destinations

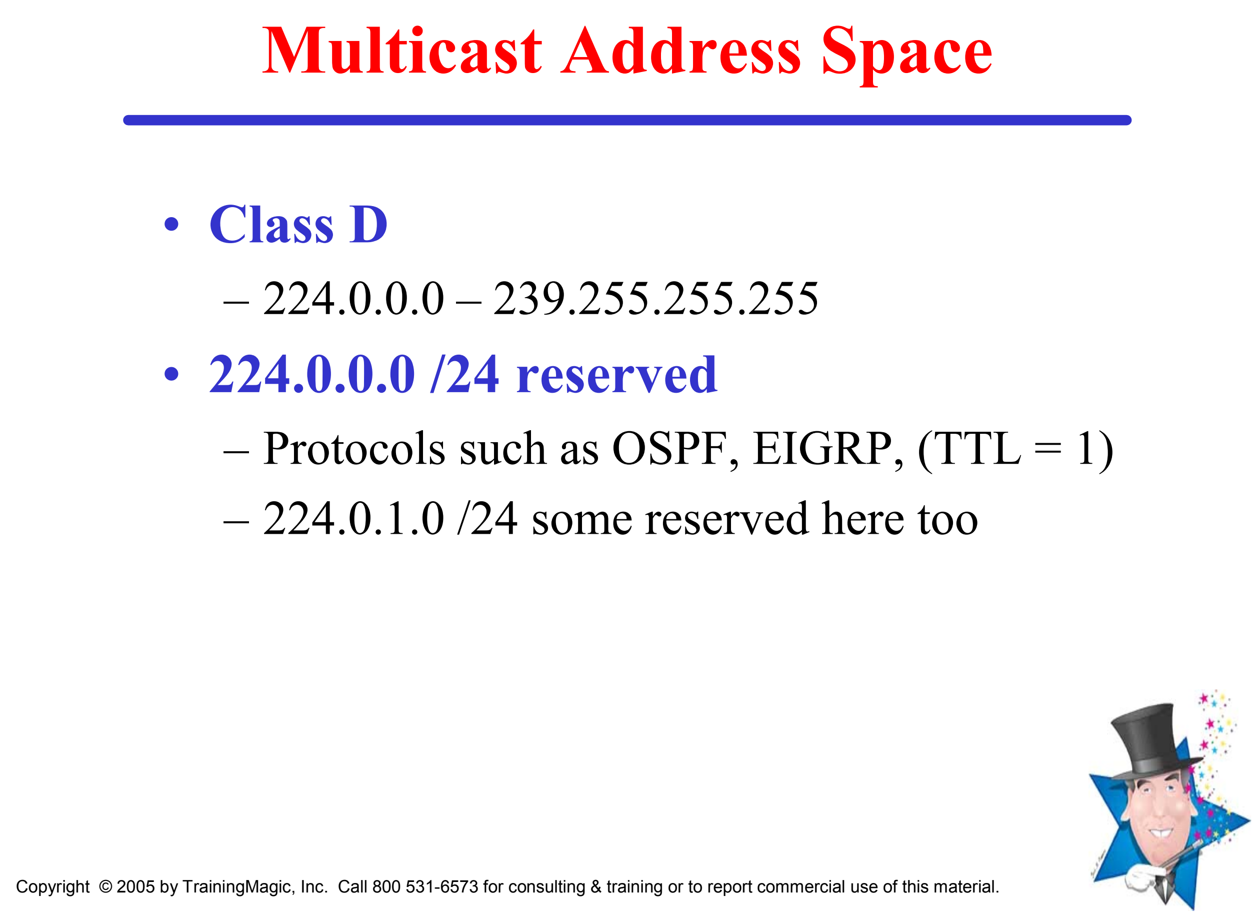
Broad cast vs. Multicast

There are three ways to design multipoint networking applications: unicast, broadcast, and multicast.

•*Unicast*. With a unicast design, applications can send one copy of each packet to each member of the multicast group. This technique is simple to implement, but it has significant scaling restrictions if the group is large. In addition, it requires extra bandwidth, because the same information has to be carried multiple times -- even on shared links.

•*Broadcast*. In a broadcast design, applications can send one copy of each packet and address it to a broadcast address. This technique is even simpler than unicast for the application to implement. However, if this technique is used, the network must either stop broadcasts at the LAN boundary (a technique that is frequently used to prevent broadcast storms) or send the broadcast everywhere. Sending the broadcast everywhere is a significant usage of network resources if only a small group actually needed to see the packets.

•*Multicast*. With a multicast design, applications can send one copy of each packet and address it to the group of computers that want to receive it. This technique addresses packets to a group of receivers rather than to a single receiver, and it depends on the network to forward the packets to only the networks that need to receive them.



Multicast Address Space

- **Class D**
  - 224.0.0.0 – 239.255.255.255
- **224.0.0.0 /24 reserved**
  - Protocols such as OSPF, EIGRP, (TTL = 1)
  - 224.0.1.0 /24 some reserved here too

Multicast Address Space

Unlike Class A, B, and C IP addresses, the last 28 bits of a Class D address have no structure. The multicast group address is the combination of the high-order 4 bits of 1110 and the multicast group ID. These are typically written as dotted-decimal numbers and are in the range 224.0.0.0 through 239.255.255.255. Note that the high-order bits are 1110. If the remaining bits in the first octet are 0, this yields the 224 portion of the address.

The set of hosts that responds to a particular IP multicast address is called a host group. A host group can span multiple networks. Membership in a host group is dynamic, hosts can join and leave host groups.

The Internet Assigned Numbers Authority (IANA) assigns some multicast group addresses. These multicast group addresses are called permanent host groups and are similar in concept to the well-known TCP and UDP port numbers. Address 224.0.0.1 means “all systems on this subnet,” and 224.0.0.2 means “all routers on this subnet.” Groups in the range of 224.0.0.xxx are always sent with a TTL of 1. Groups in the range of 224.0.1.xxx are reserved for protocol operations and sent with normal TTLs.

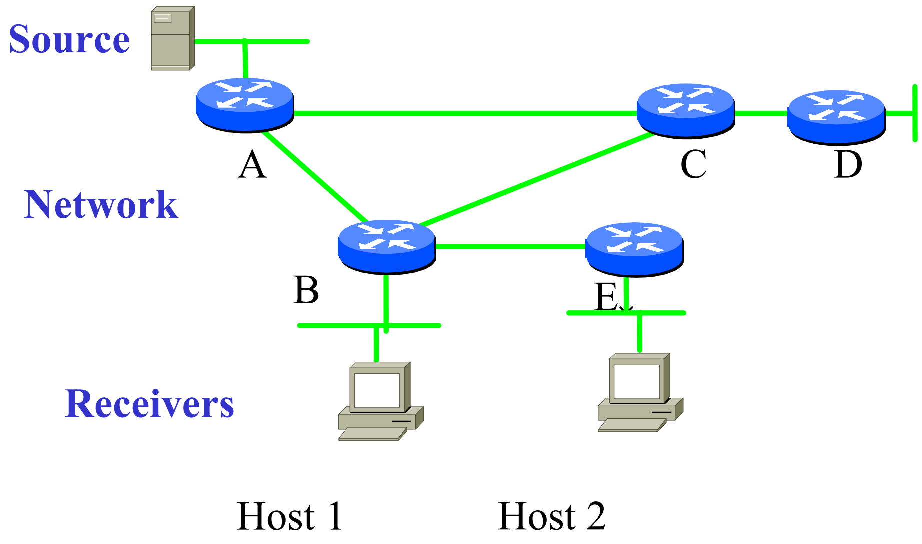
The Time To Live (TTL) value defines scope and limits distribution IP multicast. The interface is assigned a TTL value depending on the multicast scope of the interface. These values are: 0=host, 1=network, 32=same site, 64=same region, 128=same continent, 255=unrestricted. The packet must have a TTL greater than the interface TTL or it is discarded

However, this is no longer recommended as a reliable mechanism to limit the scope a packet can travel.

Multicast Address Space

- **233.0.0.0 /24**
  - Assigned to ISP's via ASN e.g. AS 3356 -> 233.13.28.0 /24
- **239.0.0.0 /8**
  - Private address range
- **TTL value limits distribution**

Multicast Network Components



Multicast Network Components

To support IP multicast, the sending and receiving nodes, intermediate routers and the network infrastructure between them must be multicast-enabled.

Multicast depends on the network to forward the packets to only those networks and hosts that need to receive them, therefore controlling network traffic and reducing the amount of processing that hosts have to do. The network must be able to build packet distribution trees that allow sources to send packets to all receivers. These trees ensure that only one copy of a packet exists on any given network.

End node hosts must have IP multicast application software such as video conferencing and must be able to support IP multicast transmission and reception in the TCP/IP protocol stack.

Reverse Path Forwarding (RPF)

- Earliest multicast protocol
- Packets are forwarded away from source
  - Receive incoming multicast packet
  - Check source IP address
  - Forward packets if they came from the addresses used to send unicast packets to source
  - If not discard
  - If so, forward on outgoing interfaces
- Creates source-based shortest path tree
- Protocol defunct – concept not!

Reverse Path Forwarding

Reverse Path Forwarding was one of the first protocols to be used in Multicast routing. RPF creates a source-based shortest path tree using the packets source address to prevent routing loops.

RPF requires the router to have a current database that reflects the topology of the network and the shortest path to each node. When a packet arrives, the router checks the source address and looks up the interface that leads to the source address. If the packet arrived on the interface that leads to the source, then the packet is forwarded to all of the other interfaces. If the packet did not arrived on the interface that leads to the source, then the packet is discarded.

A Tale of Two Protocols

- **Group Membership Protocol**
  - **Enables hosts to dynamically join/leave multicast groups.**
  - **Membership info is communicated to nearest router**
- **Multicast Routing Protocol**
  - **Enables routers to build a delivery tree between the sender(s) and receivers of a multicast group**

A Tale of Two Protocols

IGMP is used to enroll hosts into a multicast group.

Multicast routing protocols (RPF, DVMRP, MOSPF, PIM-DM, PIM-SM, and MBGP) enable routers to build a delivery tree between the senders and the receivers of a multicast group.

IGMP

- **Internet Group Messaging Protocol**
- **THE group membership protocol**
- **IGMPv1**
  - Tell router you want to receive the multicast stream
  - no leave process
  - Windows 95
- **IGMPv2**
  - Allows explicit leave messages
  - Modern versions of Windows, UNIX, Linux
- **IGMPv3**
  - Work in progress

IGMP

The Internet Group Management Protocol (IGMP) is used by IP hosts to report their host group memberships to any immediately-neighboring multicast routers. IGMP relies on Class D IP addresses for the creation of multicast groups and is defined in RFC 1112. IGMP dynamically registers individual hosts in a multicast group with a Class D address. Hosts identify group memberships by sending IGMP messages, and traffic is sent to all members of that multicast group. Under IGMP, routers listen to IGMP messages and periodically send out queries to discover which groups are active or inactive on particular LANs. Routers communicate with each other by using one or more protocols to build multicast routes for each group. IGMP may be used, symmetrically or asymmetrically, between multicast routers.

IGMP has two steps. First, when a host joins a new multicast group it sends an IGMP message to the group's multicast address declaring its intention to join the group. This message is then passed on to all other multicast networks in the network. Second, local multicast routers poll group members to ensure that they are still active. After not responding to several polls a member is dropped.

IGMP Messages

0 3	4 7	8 15	16 31
Version	Type	Code	Checksum
Group Address			

- **Version is 1, 2, or 3 (Under development)**
- **Type**
  - Membership query
  - Membership Report
  - Leave group
- **Code contains the maximum delay**
  - Default is 10 seconds for poll response

IGMP Messages

IGMP messages are used by routers to establish group memberships in an immediately connected subnetwork. The following rules apply for IGMP v1: 1.A host sends an IGMP "report" for joining a group

2.A host will never send a report when it wants to leave a group.

3.Multicast routers send IGMP queries (to the all-hosts group address: 224.0.0.1) periodically to see whether any group members exists on their subnetworks. If no response is received for a particular group after a number of queries, the router assumes that there there is not any group member on the network (physically connected to a particular interface of the router). It should be noted that the TTL field of query messages is set to 1 so that the queries do not get forwarded to other subnetworks.

Based on the reports a router receives from the hosts it can decide whether to forward a multicast packet on a particular interface or not.

IGMP Version 2 includes a few extensions the most important of which is the explicit leave messages for faster pruning.

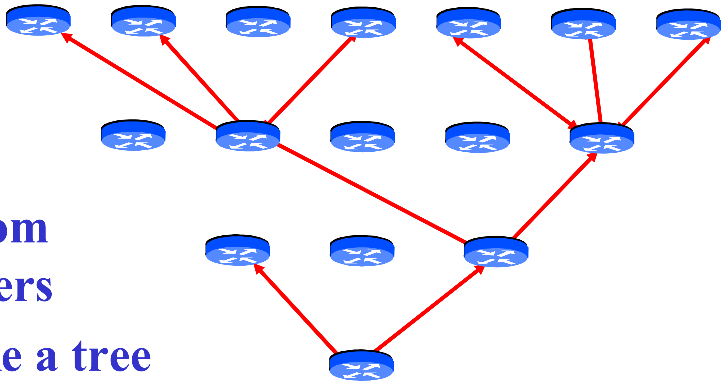
IGMP Snooping

- **Limits multicast flow in a switched LAN**
  - **Security issues – everyone on LAN can see your data**
  - **Multicast data can degrade LAN performance**
  - **IGMP snooping stops multicast traffic going to networks without receivers**
- **Performance issues**
  - **CPU can be weighed down by inspecting all multicast traffic to identify IGMP messages**

IGMP Snooping

A layer-2 switch supported IGMP snooping can passively snoop on IGMP Query, Report and Leave (IGMP version 2) packets transferred between IP Multicast Routers/Switches and IP Multicast hosts to learn the IP Multicast group membership. It checks IGMP packets passing through it, picks out the group registration information, and configures multicasting accordingly. Without IGMP snooping, multicast traffic is treated in the same manner as broadcast traffic - it is forwarded to all ports. With IGMP snooping, multicast traffic of a group is only forwarded to ports that have members of that group. IGMP Snooping generates no additional network traffic, allowing you to significantly reduce multicast traffic passing through your switch – but there is a performance penalty.

Path Distribution Trees



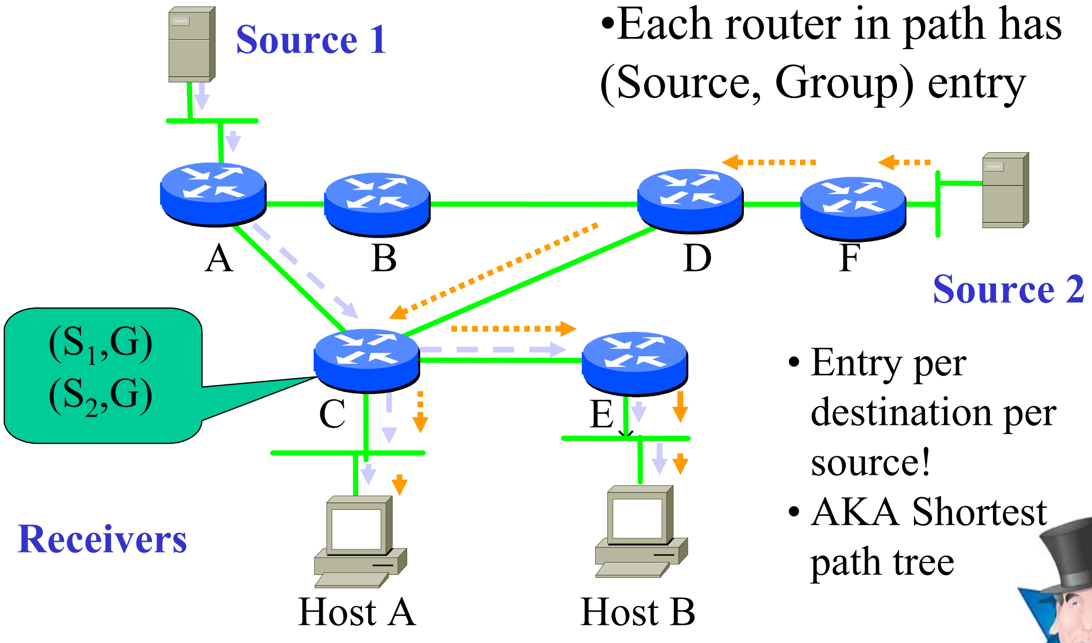
- Two types:
  - Shared tree
  - Source tree
- Traffic flows from source to receivers
- Traffic looks like a tree
- Tree are loop free
- Messages replicate when tree branches

Path Distribution Trees

IP multicast traffic flows from the source to the multicast group over a distribution tree that connects all of the sources in a group to all of the receivers in the group. A tree that is shared by all of the sources is called a shared-tree. A tree that creates a separate distribution tree for each source is called a source-tree. Depending on the Multicast service, trees could be one-way or bi-directional. These trees are loop-free. Messages are replicated only when the tree branches.

The distribution tree must be updated dynamically because members of multicast groups can join or leave at will. Branches with no listeners are discarded – this is called pruning.

Source Trees (S,G)



•Each router in path has (Source, Group) entry

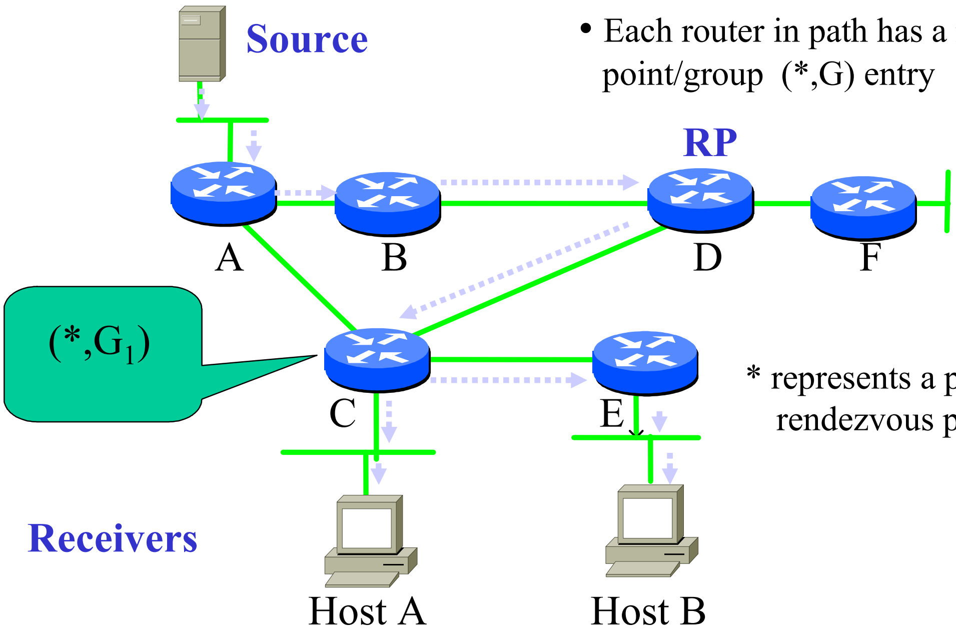
- Entry per destination per source!
- AKA Shortest path tree

Source Trees (S,G)

In multicast routing, every multicast entry in routing tables consists of a pair of entries, a Source entry and a multicast Group entry. The routers use these table entries to keep track of which interfaces to forward traffic and the interface where it receives the multicast traffic by the IP address of the source, and the IP multicast group address. These are often represented as (S,G).

Source-based distribution trees build an optimal shortest-path tree rooted at the source. Each source/group pair require their own table entries, so for groups with a very large number of sources, or networks that have a very large number of groups with a large number of sources in each group, the size of the ensuing routing table can stress the storage capability of routers.

Shared Trees (*,G)

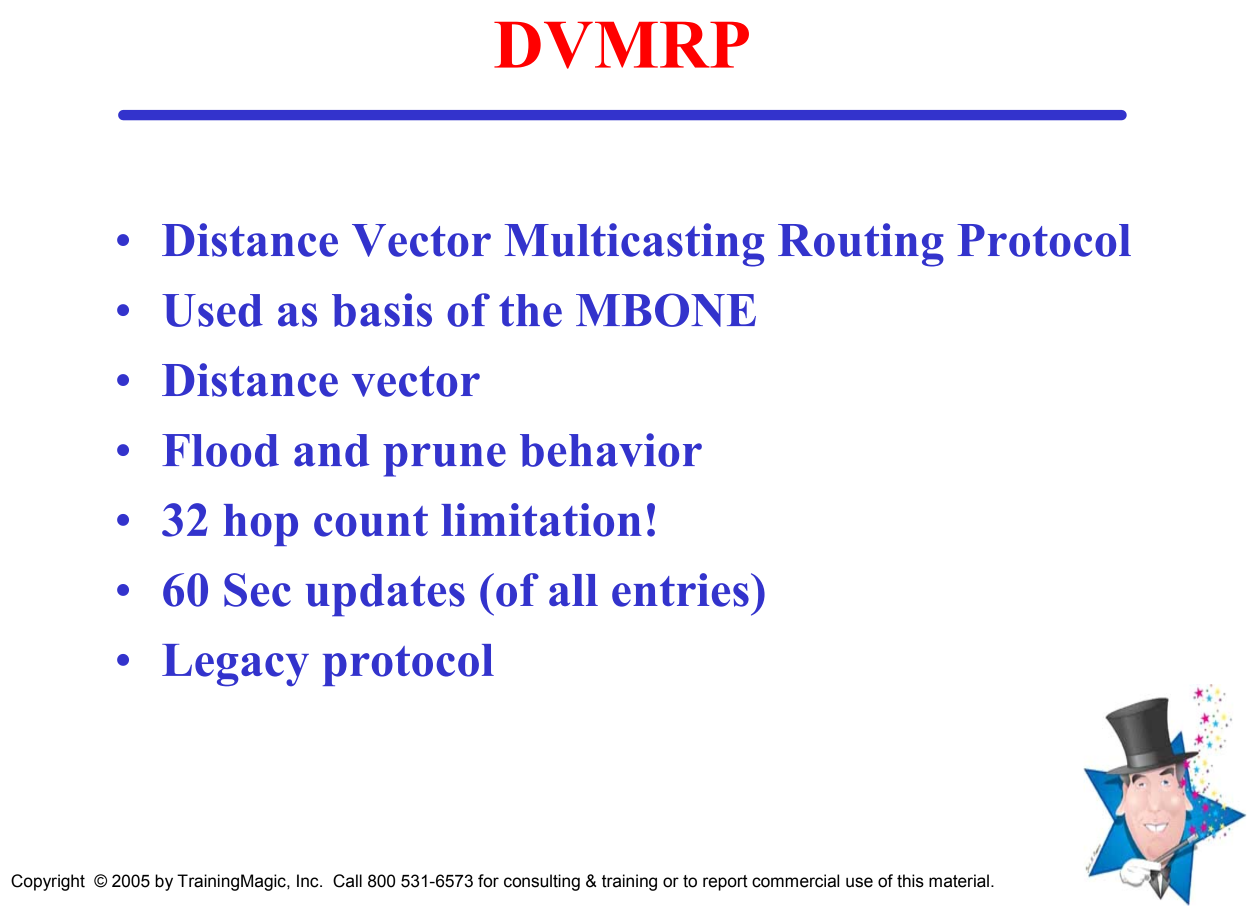


- Each router in path has a rendezvous point/group (*,G) entry

* represents a particular rendezvous point

Shared Trees (*,G)

Shared distribution trees are formed around a central router, called a rendezvous point or core, from which all traffic is distributed regardless of the location of the traffic sources. Shared distribution trees routing table entries are often represented as (*,G). The advantage of shared distribution trees is that they do not create lots of source/group state in the routers. The disadvantage is that the path delay to particular receivers may vary as a factor of the distance from the rendezvous point. This may be important for delay-sensitive applications. The rendezvous router may also be a traffic bottleneck if there are many high data rate sources.



DVMRP

- Distance Vector Multicasting Routing Protocol
- Used as basis of the MBONE
- Distance vector
- Flood and prune behavior
- 32 hop count limitation!
- 60 Sec updates (of all entries)
- Legacy protocol

DVMRP

DVMRP was the first true multicast routing protocol to see widespread use. DVMRP is similar in many ways to Routing Information Protocol, RIP, with some minor variations added to support multicast.

Although DVMRP has been deployed in the Mbone and in other intra-domain multicast networks, some rather severe scalability issues prevent it from being used in large-scale multicast environments.

Because DVMRP uses hop-count as a metric with an infinity value of 32 hops, DVMRP is ill suited for deployment as a multicast protocol for the Internet. However, because of its historical legacy, DVMRP support may become a requirement for interoperability with some customers.

MOSPF

- **Based on OSPF**
- **LSAs contain group membership**
- **Demand driven**
  - **Traffic for group is not propagated untill needed**
  - **All routers in group must maintain membership for every group and synchronize with all routers**
- **Link State database used to derive (S,G) entries as sources start talking**
- **Does not scale well**

MOSPF

Multicast extensions to OSPF (MOSPF) modifies OSPF version 2 to incorporate multicast capabilities. The extensions built into MOSPF provide routers with information on group membership by transmitting group link state advertisements (LSA). In each OSPF area, a designated router and a backup designated router are defined. The designated router transmits Internet Group Message Protocol (IGMP) messages to the network. Based on the responses, the designated router sends group LSAs to all routers in the area. On transmission of the first source group datagram, the router calculates the shortest path incorporating all group members. The packet is then forwarded based on the tree. By calculating the tree based on the link state database, flooding of the network is avoided. The forwarding information is cached until network changes require an update. In large MOSPF networks, area border routers share summary information about group membership in their areas.

Protocol Independent Multicast

- **IP routing protocol independent**
- **Uses unicast routing table to perform RPF**
- **Two modes**
  - **Dense mode**
  - **Sparse mode**
- **Simpler than DVMRP**
  - **PIM does not require computation of routing tables**

Protocol Independent Multicast

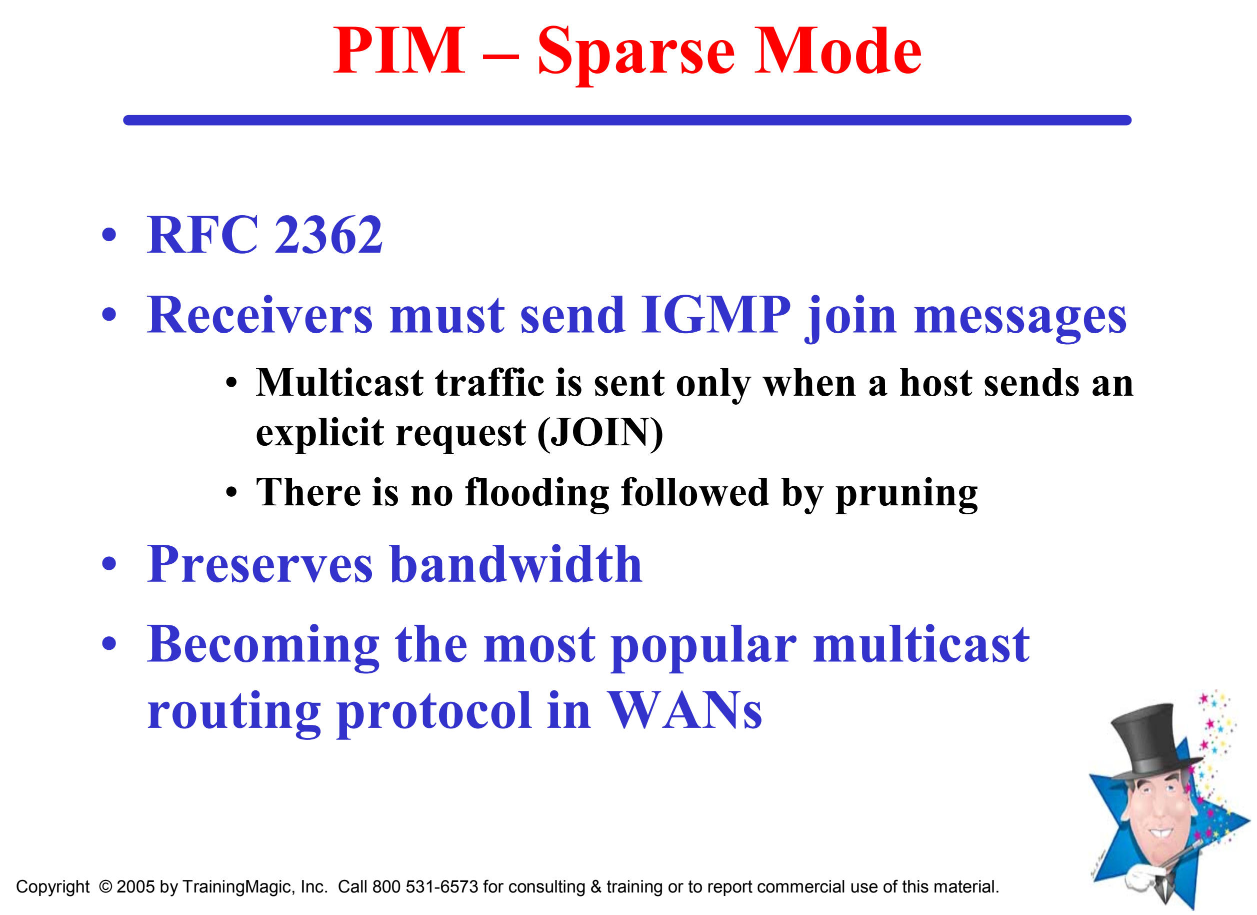
Protocol Independent Multicast, i.e. PIM, gets its name from the fact that it is IP routing protocol independent. That is, regardless of which unicast routing protocol(s) is (are) used to populate the unicast routing table, PIM uses this information to perform multicast forwarding. Although PIM is referred to a multicast routing protocol, it actually uses the existing unicast routing table to perform RPF checks instead of maintaining a separate multicast routing table.
Two implementations of PIM exist, PIM Sparse Mode and PIM Dense Mode.

PIM – Dense Mode

- **Floods all multicast traffic everywhere**
- **Router examines address**
  - **If incoming interface is used for multicast then forward packet to group**
  - **If not then delete**
- **Nonparticipating routers request to be pruned**
- **After 3 minutes repeat flood**

PIM – Dense Mode

Protocol Independent Multicast – Dense Mode operates in a similar to the broadcast/prune function of DVMRP. The network is flooded with multicast packets. If a router receives a multicast packet from the source to the group then that router checks it's routing table (remember, unlike DVMRP, PIM-DM uses the standard unicast routing table) to see if the packet came in on an interface configured for multicast. If the interface is not authorized for multicast then the packet is dropped and a "Prune (S,G) message is sent back to the source. If the interface was authorized for multicast, then the router forwards the packet to all the interfaces that have not requested to be pruned from the tree. If all the interfaces on the router have requested to be pruned, then the router will send back a prune message to the source.



PIM – Sparse Mode

- **RFC 2362**
- **Receivers must send IGMP join messages**
  - **Multicast traffic is sent only when a host sends an explicit request (JOIN)**
  - **There is no flooding followed by pruning**
- **Preserves bandwidth**
- **Becoming the most popular multicast routing protocol in WANs**

PIM – Sparse Mode

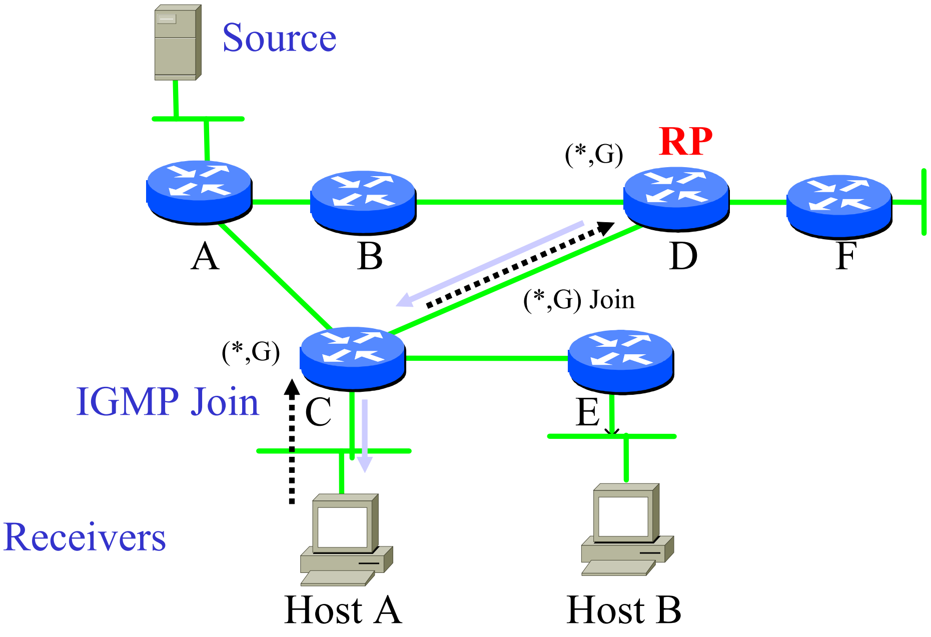
PIM-SM (defined in RFC 23630 has two key differences with existing dense-mode protocols (DVMRP, MOSPF, and PIM-DM). First, in PIM-SM protocol the routers need to explicitly announce their willingness to receiving multicast messages of multicast groups, while dense-mode protocols assumes that all routers need to receive multicast messages unless they explicitly send a prune message.

Second, PIM-SM uses the concept of "core" or "rendezvous point" (RP). Each sparse-mode domain has a set of routers acting as RPs (RP-set). Furthermore, each group in side of a domain has a single RP at any given time. In order to participate in a multicast group the router must send a join message to the RP of that group. Unless there is a reason to deny registration (policy, inadequate bandwidth, etc.) the RP will register the router as part of the multicast group.

PIM – Sparse Mode

- Support both shared and source trees
- Requires Rendezvous Point (RP)
- Each Multicast Group has a designated RP
- Receivers find sources through the RP
- Sources find receivers through the RP

Shared Tree Join



Shared Tree Join

In PIM-SM Shared trees are called the rendezvous point trees (RPT). Members send explicit joins to the central node. The result is a single tree for each multicast group, no matter how many sources there are. The only routers that know about the group are the ones that are on the tree, and data is sent only to interested receivers. With an RP, receivers have a place to join to even if no sources exist yet.

When a host wishes to join a multicast group, it sends an IGMP message to its upstream router. To accept multicast traffic for a group, the router signals the Rendezvous Point that it wishes to join the RPT by sending a PIM (*, G) Join message to its upstream PIM neighbor, in the direction of the RP. Join messages are sent multicast hop-by-hop to the address 224.0.0.13, which is the ALL-PIM-ROUTERS group. All PIM neighbors are aware of the join, but only the indicated up-stream PIM neighbor performs the join. The same message is used for both joins and prunes.

When a PIM router receives a (*, G) Join from a downstream router, it checks to see if (*, G) state exists for group G in its multicast routing table. If state already exists, then the Join message has reached the shared tree and the interface from which the message was received is entered in the group list. If no state exists, a (*, G) entry is created and the Join message is again sent towards the RP.

MSDP

- **Multicast Source Discovery Protocol**
- **RPs communicate with one another**
  - **Exchange (S,G) information**
- **Removes dependency on external RP**
- **Domains with only receivers get data without globally advertising membership**

MSDP

Multicast Source Discovery Protocol (MSDP) was designed to connect multiple PIM sparse-mode (SM) domains. MSDP allows multicast sources for a group to be known to all rendezvous point(s) (RPs) in different domains. Each PIM-SM domain uses its own RPs and need not depend on RPs in other domains. An RP runs MSDP over TCP to discover multicast sources in other domains.

MBGP

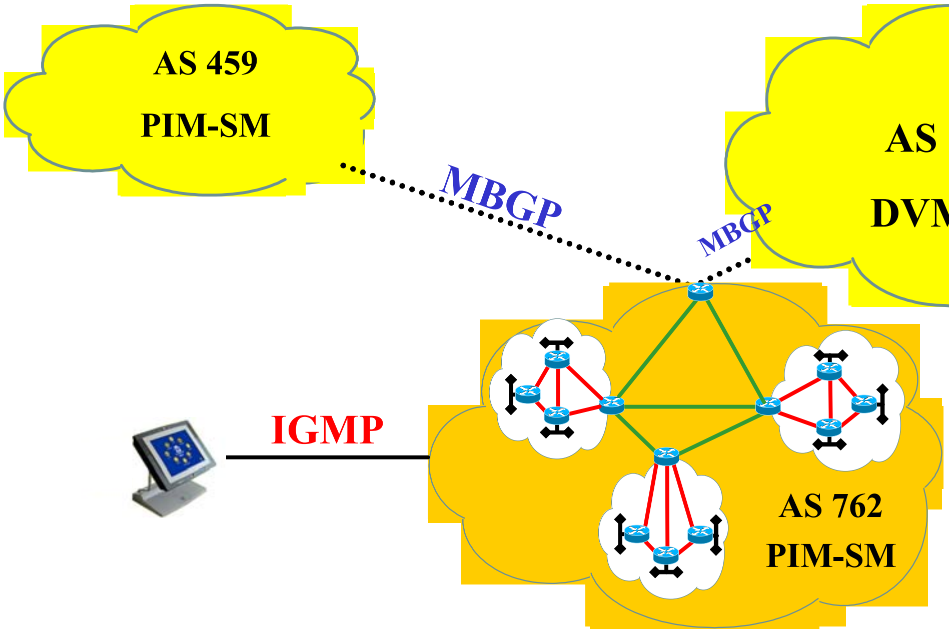
- **Multiprotocol BGP**
- **Advertises support for other protocols**
  - Allows for incongruent network interconnection
  - For multicast applications – which ASs can I talk to?
- **Carries (NLRI)**
  - Network Layer Reachability Information
  - Two sets of routes – Unicast & Multicast

MBGP

Multicast BGP, (MBGP) is an extension to BGP that allows BGP to carry multicast routing information for RPF calculations between autonomous systems
The multiprotocol BGP feature adds capabilities to BGP to enable multicast topologies to communicate within and between BGP autonomous systems. MBGP is an enhanced BGP that carries IP multicast routes. BGP carries two sets of routes, one set for unicast routing and one set for multicast routing. The routes associated with multicast routing are used by the Protocol Independent Multicast (PIM) to build data distribution trees.

Multicast BGP defines two new multiprotocol BGP attributes: MP_REACH_NLRI and MP_UNREACH_NLRI. These attributes would then be used to exchange reachability information for different address families and would be carried inside BGP update messages.

Example Multicast Network



The host reports it's multicast group memberships to the closest multicast router using The Internet Group Management Protocol.

Two different routing/forwarding protocols are used in this diagram. AS 298 is using DVMRP. The other autonomous systems are using PIM-SM. Finally, MBGP is being used to enable multicasting between AS's. It is especially useful because it exchange Multicast support information between autonomous systems that support different multicast protocols.